Collective Intelligence based Framework for Load Balancing of Web Servers

Dimple Juneja, Atul Garg

Maharishi Markandeshwar University, Mullana-Ambala, Haryana, 133207, India
atul.garg@mmumullana.org

Abstract

The paper exploits the collective intelligence referred to as ant intelligence in World Wide Web with the aim to improve the performance of online web servers by balancing the load. The central concept of this idea is that a collection of agents can individually perform relatively simple, self-centered actions, such as the selection or rejection of hyperlinks in a web page for navigation, computing the load of server and aggregate these individual actions into a common substrate. The common substrate can then be evaluated to find the best available server to perform the task. This work aims to address the challenge of distributing intelligence to World Wide Web by contributing a unique ant-based intelligent load balancing framework which is able to integrate and synthesize knowledge on a scale far beyond the capabilities of individual humans.

Keywords: *Collective Intelligence, Ant-based Frameworks, Load Balancing, World Wide Web (WWW).*

1. Introduction

Today's web is evolving beyond the traditional networks as the user is demanding more and more intelligence in the web and that too distributed throughout the web. In fact, the definition of web is now being challenged as intelligence and control distributed to entities which can improve the performance of web and ultimately increases the user satisfaction. Despite the fact that WWW's has scaled exponentially w.r.t. number of human contributors, information content and hyperlinked resources but web being intelligent is still an open challenge. Though database techniques have played a vital role in updating the websites but still it is a static hypertext repository whose dynamics such as linking of websites, are determined by explicit, externally defined design efforts. Although it has been found that link patterns gradually change over time [24], this is not necessarily a process of adaptation. The core idea of this work is to delegate intelligence to clients and servers wherein the dynamics of the web would now be controlled by intelligent entities known as ants. The ant-based web can autonomously learn from experience, synthesize and represent their world knowledge in an adaptive, distributed manner [25]. This work aims to address the question of whether we can change the present modus operandi of the WWW to accommodate certain characteristics associated with intelligent systems while maintaining its nature of a distributed, hyperlinked knowledge network [17]. The work not only focuses on finding the best available server for providing the most relevant information and minimal amount of time but also it aims to balance the load of servers so as to improve the overall performance of the web. The stated concerns have been achieved by implementing collective intelligence of ant-colony.

Collective intelligence has evolved from the simulations of biological systems such as ants where individual ants explore pathways to food sources, and leave pheromone markers to demarcate their paths. These paths can be traced back by other ants who prefer the strongest pheromone paths (determined by recency and food location), but will randomly deviate at certain positions, thereby further exploring the space of possible paths [23]. Similarly, ant intelligence has been applied to large groups of agents who can aggregate their distributed actions to perform complex tasks. For instance, ant intelligence has been applied to solve Traveling Salesman problem [22] which is a computationally demanding problem. This work is motivated from the literature [22] indicating that such systems have performed well in terms of computational efficiency.

The paper is structured as follows: Section 2 justifies the ground of proposing the framework by highlighting the works of eminent researchers who made attempts for making web more computationally effective. Section 3 presents the proposed framework and section 4 concludes by discussing the pros and cons of the proposed work.

2. Related Work

Currently, every user selects and rejects hyperlinks on the basis of individual knowledge on the relationship between hypertext pages [14]. Several researchers have put in efforts implementing algorithms which allow the WWW to integrate individual navigation patterns to optimize its global hyperlink structure, thereby aggregating the knowledge implicitly expressed by a population of its navigators. For instance, A class of such systems has been implemented as described in [15], including applications to digital library systems [16]. They have been shown to adequately map the collective knowledge of a community of users [14]. Systems which distribute documents and hyperlinks over networks of autonomous information objects have been proposed by Bollen and Nelson [17]. AntNet [1] is an agent based routing algorithm based on real ants' behavior and Ant Colony Optimization (ACO) is an optimization method where a group of artificial ants moves around a graph, which represents the instances of the problem. Each artificial ant builds a path from its source node to its destination. While an ant builds a path, it gets quantitative information about the path cost and qualitative information about the amount of traffic in the network thereby conveying the necessary information about congestion at server.

Diosteanu and Cotfas [18] proposed an agent based knowledge management application framework using a specific type of ontology that is able to facilitate semantic web service search and automatic composition which can be extended to solve more complex problems. Authors [19] in their works have utilized the Ant Miner algorithm in the field of web content classification and shows that it is more effective than C5.0. Gibbins [20] have focused their works on agent based semantic web services building ontological based web services. Bak and his co-authors [2] proposed a family of routing schemes that distribute data traffic over the whole network via bounded randomization successfully removing bottlenecks and consequently improving network performance. The proposed scheme required a simple extension to a shortest-path routing protocol.

Turning our attention to load balancing of web servers, Ali et. al [5] provided a metaheuristic optimization tool and applied collective intelligence model to several applications such as routing and load balancing. Tolksdrof [21] in his article have emphasized that "*The principles of locality and decentralization found in nature could be the key to managing Web data traffic*".

Studies reflect that that in order to balance the load on servers, the huge data is partitioned on multiple servers [13]. Many shared nothing parallel database systems use range partitioning to de-cluster a relation across the available disks for performance gain. Well-known concern in range partitioning is skew, however, as the relation evolves over time, or as workload changes, both data and execution skew pose a serious problem. In order to address the mentioned shortcoming, usage of multiple web servers was proposed. Now the concern is to select a server. The issued that are most predominantly raised are “Which path a request should follow from client to server? Which Server is free? What is the performance of CPU used by the particular server? What is the overall credibility of the Server machine? Should the reply from the server to client follow the same path?” and many more unfolded challenges are still exist.

An analytical investigation of the above literature reveals the fact that a lot of work has been done by the researchers towards applying ants in managing web traffic and hence improve the overall computational efficiency. However, it is clear at the time of listing; none of the researchers have focused towards deploying ants both at client and server end. Moreover, existing algorithms are able to perform passive load balancing while the unique contribution of our wok is active and online load balancing. The dynamics of the proposed framework is now not shaped by external design efforts but by the ability of ants to learn, integrate, synthesize and extract knowledge and hence increasing the probability of task being performed optimally.

3. The Proposed Work

The work primarily proposes to deploy two ants namely, client ant and server ant. The client ant is responsible for searching the optimal network path for the query to be processed whereas server ant is responsible for balancing the load on servers and also for finding the best available server. The server ant is also required to collaborate and hence implement collective intelligence, if the desired query cannot be completely fulfilled by itself. Every time client sends a request a client ant would be activated at client side. Ants make use of pheromone tables generated due to previous requests and thereby intelligently apply the network rules and logical processing to search the destination server. The proposed framework basically comprises of two data structures i.e. the pheromone table containing the synthesized information from ants visiting to and from the server and the client and server agent who are provided with server id and they extend the pheromone table by their individual search and existing information in pheromone table. It shall be noted that in this framework the pheromone table is to some extent shaped by the synthesized actions of client and server ant thereby aggregating their actions, while in return shaping those actions by constraining or controlling them. Collective intelligence thereby integrates the individual intelligence of each participating agent into a collective representation.

Figure 1 presents the high level view of the proposed framework. Client Ant will be activated immediately on any request generated by the client and the Server Ant will be activated when the client ant approaches a particular server with the request. Initially, the Client Ant will check the network path based on existing shortest path algorithm and later, if the path is not congested then it will check the load on server, CPU performance of server and overall credibility of server. If server is credible in all respects then it will stop at that server otherwise it will search another server and the process will go on up to three times. If all the time it has to face some problem then it will analyze all the three steps and choose the best among the three attempts. The client ant then would reach at preferred server and would adopt the goal of server ant that is responsible for analyzing the current server is capable enough to serve the request or collective efforts are required. In the later case, it would contact the peer server ants for the goal adoption and collective intelligence may result into success. The server ant would then migrate to the client exploring the pheromone table generated so far due to accumulations done by older ants. However, it may be noted that the server ant may choose to deviate from the path with maximum accumulations and may choose a random path available. Further, in case, the related data is not available on the peer server then it will send a denial message to the client and stop the process.

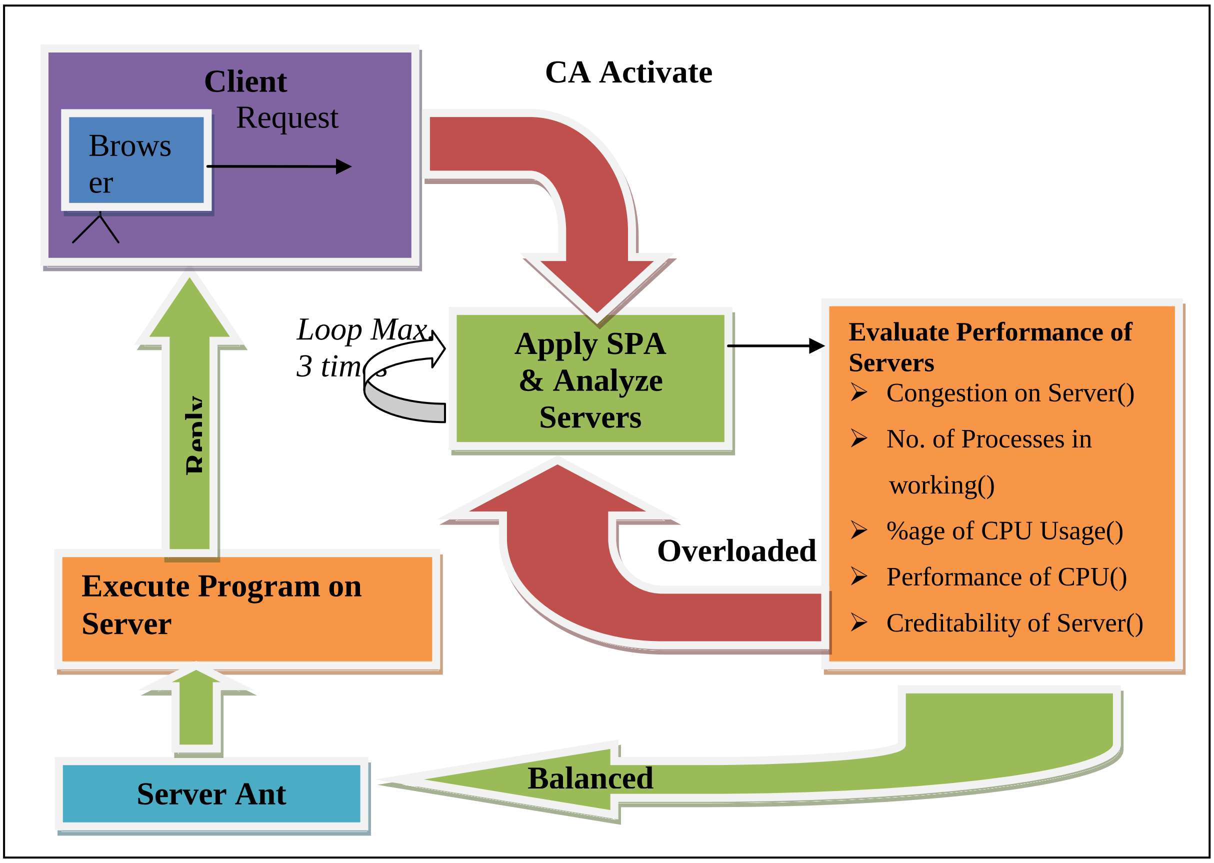

Figure 1: The Proposed Framework

Flowchart and working algorithm for the same is given in figure 2(a) and 2(b).

Working Algorithm

The client ant is activiated on receiving a request from the client. This ant applies SPA and evaluates the overall credibility of the server. In Step 2(i) the loop will move maximum three times to apply SPA if the shortest path is congested. If the path is not congested then it will move to Step 2(xiv) and check the overall credibility of the server and move upto Step 2(xix) else set a constant z=5. Value 5 is any arbitrary number. This constant is used only to check if the process moves up to Step 2 (xxi) or not. Maximum limit to apply SPA will indicate to

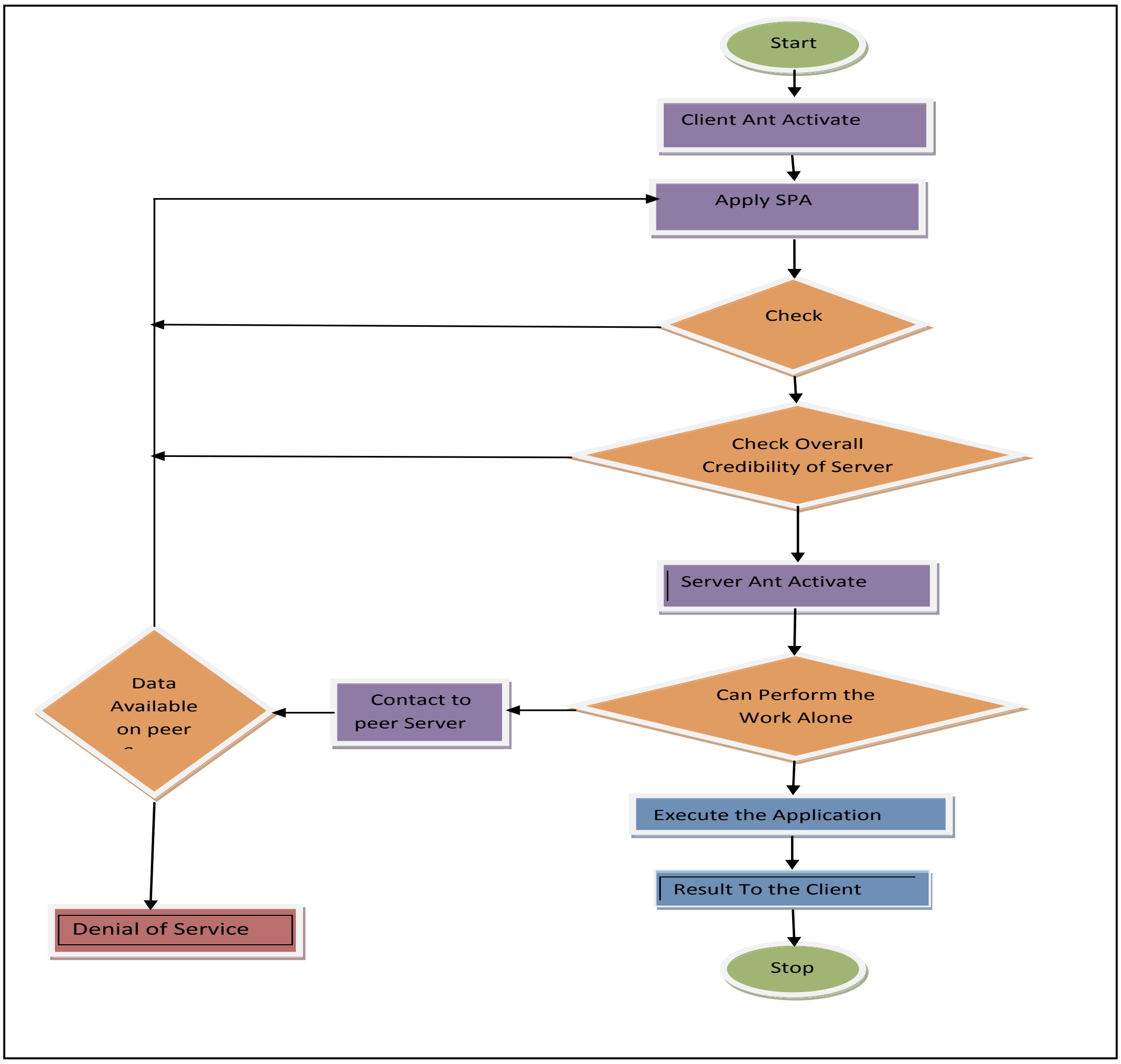

Figure 2(a): Flowchart depicting the Proposed Framework

analyze all the three servers and to choose the best one in all respects. Once the server is selected then Server Ant will activate and execute the application and reply to the client. If the server is unable to provide all the related data then the Server Ant will search the peer servers for the related data and go to Step 2(i). If the peer servers are also unable to provide the relevant data then it will send a denial of service message to the client.

4. Conclusions

The work presented in this paper is an effort to introduce the concept of collective intelligence for improving the performance of web servers. The core idea was motivated from various agent oriented applications that have demonstrated that agents can not only perform simple tasks but are also effective while solving complex and distributed tasks. It is evident from the work presented in this paper that global intelligence on the web will crucially depend on agent technologies as these are able to integrate and synthesize knowledge which is highly useful for users. The presented framework allows us to incorporate techniques and methods of analysis common in the study of collective intelligence and apply it to the quantitative study of such intelligence on the internet.

Step1. Client send a request and the CA (CLIENT ANT) Activated.
Step2. The ANT search the appropriate result

i. Loop n=1 to n=3
ii. Execute the SPA(Shortest Path algorithm)
iii. If the Path is not congested
iv. Then goto Step xiv
v. break
vi. Else if path is thickly or highly or moderately congested then repeat step i
vii. End if
viii. If n=3 then compare the congestion of three paths and select the least congested path and goto xiv
ix. Else if n=1 and z=5 then repeat step i
x. Else if n=2 and z=5 then repeat step i
xi. Else if n=3 and z=5 then goto step viii
xii. End if
xiii. End loop
xiv. If the congestion on server and number of process on server is less and percentage of CPU usage is less and performance of CPU is better and overall credibility of CPU and Server is better then
xv. If the current server can work alone
xvi. Then execute the program and send reply to the Client
xvii. Else if Check the related data on peer server and goto Step i
xviii. Else if related data is not available then send a denial of service message to the client
xix. End if
xx. Else
xxi. Set z=5 , Goto Step viii

Figure 2(b): Algorithm for the Proposed Work

References

[1] Benjamín Baran and Ruben Sosa, "AntNet Routing Algorithm for Data Networks based on Mobile Agents", Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial. No. 12 (2001), pp 75-84.

[2] Sangman Bak, Jorge A. Cob, Ernst L. Leiss, "Performance Evaluation of Load-Balanced Routing via Bounded Randomization",Clei Electronic Journal, Volume 3, Number 2, Paper 1 December 2000.

[3] S. Prasad, Y.P.Singh and C.S.Rai , "Swarm Based Intelligent routing for MANET", International Journal of Recent trends in Engineering Vol 1,No.1 ,May 2009.

[4] Hamideh Shokrani and Sam Jabbehdari, "A Survey of Ant Based Routing Algorith for Mobile Ad-Hoc Networks", International Confrence on Signal Processing Systems, 2009.

[5] Al-Dahoud Ali, Mohamed A. Belal and Moh'd Belal Al-Zoubi, "Load Balancing of Distributed Systems Based on Multiple Ant Colonies Optimization", American Journal of Applied Sciences 7 (3): 433-438, 2010.

[6] Ruud Schoonderwoerd, O Holland, J Bruten, L Rothkrantz, "Ant-based load balancing in telecommunications networks", Volume: 5, Issue: 2, Publisher: MIT Press, Pages: 169-207, available from www.hpl.hp.com.

[7] James Aweya, M. Ouellette, D.Y Montuno, B. Doray, K. Felske, "An Adaptive Load Balancing Scheme for Web servers", InternationalJournal of Network Management, 2002, pp.3-39.

[8] http://www.zenus.com/solutions/zlb/indexfull.html
[9] David Vengerov, Hamid R Berneji, Alex Vengerov, "An adaptive coordination among fuzzy reinforcement learning agents performing distributed dynamic load balancing", Proceedings of 11th IEEE International Conference on Fuzzy Systems, 2002, pp.179-184.
[10] Myung-Sup kim, Mi-Jeong Choi, James W. Hong, "Highly available and efficient load cluster management system using SNMP and Web", Proceedings of Network Operations and Management Symposium, IEEE, 2002, pp.619-632.
[11] Hemant B. More, Jie Wu, "Throughput Improvement through Dynamic Load Balance", IEEE southeast con'94, 1994, pp.339-342.
[12] Fred halsall, Mutimedia Communications, Pearson Education, 2002.
[13] D. DeWitt and J. Gray. "Parallel database systems: The future of high performance database processing", Communications of the ACM, 36(6), 1992.
[14] Bollen, J. (2001), "A cognitive model of adaptive web design and navigation. Unpublished doctoral dissertation", Vrije Universiteit Brussel, Brussels, Belgium.
[15] Bollen, J., & Heylighen, F. (1998), "A system to restructure hypertext networks into valid user models", The New Review of Hypermedia and Multimedia, 4, 189–213.
[16] Bollen, J., & Luce, R. (2002), "Evaluation of digital library impact and user communities by analysis of usage patterns", D-Lib Magazine, 8(6).
[17] Bollen, J., & Nelson, M. L. (2002), "Adaptive networks of smart objects. In J. Bollen & M. L. Nelson (Eds.)", Proceedings of workshop on distributed computing architectures for digital libraries (ICPP2002) (pp. 487–496). Vancouver, B.C., Canada: IEEE.
[18] Andreea Diosteanu, Liviu Cotfas, "Agent Based Knowledge Management Solution using Ontology, Semantic Web Services and GIS", Informatica Economică vol. 13, no. 4/2009.
[19] Nicholas Holden and Alex A. Freitas, "Web Page Classification with an Ant Colony Algorithm", Kent Academic Repository http://kar.kent.ac.uk.
[20] Nicholas Gibbins, Stephen Harris, Nigel Shadbolt, "Agent-based Semantic Web Services", Journal of Web Semantics, Available at www.elseviercomputerscience.com, Web Semantics: Science, Service and Agents on 1(2004) 141-154.
[21] Robert Tolksdorf, "Self Organizing ants find new paths to scability", http://www.perada-magazine.eu.
[22] Cormen, T. H., Leiserson, C. E., & Rivest, R. L. (1997), "Introduction to algorithms", Cambridge, Massachusetts: The MIT press.
[23] Colorni, A., Dorigo, M., & Maniezzo, V. (1992), "Distributed optimization by ant colonies", In Proceedings of the First European Conference on Artificial Life (pp. 134–142), Paris, France.
[24] Koehler, W. (2002), "Web page change and persistence–a four-year longitudinal study", Journalof the American Society for Information Science and Technology, 53(2), 162–171.
[25] Kohonen, T. (1978), "Associative memory: a system-theoretical approach", Berlin: Springer-Verlag.