

Load Balancing, Optimal Routing and Scheduling in Hyper-Local

Karan Panjwani
Sinhgad Academy of
Engineering, Pune, India.

Sahil Pathan
Sinhgad Academy of
Engineering, Pune, India.

Nitin Yadav
Sinhgad Academy of
Engineering, Pune, India

Shreyas Lokhande
Sinhgad Academy of
Engineering, Pune, India.

Bhushan Thakare
Asst. Prof. Sinhgad Academy of
Engineering, Pune, India.

ABSTRACT

Classification of different job scheduling algorithms for balancing the load on Web Servers. Static and Dynamic Scheduling Algorithms are thoroughly discussed and the strengths and weaknesses of these Algorithms. Genetic Algorithms used to get the best of the Optimal Routing and Scheduling problems.

General Terms

WWW – World Wide Web.

IWT – Initial Waiting Time.

TT – Transfer Time.

GA – Genetic Algorithm.

TSP – Travelling Salesman Problem.

SVPDP – Single Vehicle Pick-up and Delivery Problem.

SVPDPTW – Single Vehicle Pick-up and Delivery Problem with Time Windows.

Keywords

Cluster, Genetic Algorithms, Load Balancing, Nodes, Routing, Scheduling Algorithms, Traffic, Web Server.

1. INTRODUCTION

With a drastic increase in the growth of the World Wide Web (WWW), grew the increased usage of different applications and which led to more bandwidth usage and computation power [1] and the end users often experience poor response time or denial of service. Applications like E-Commerce, Multimedia, Development, Design, etc. which led to increase of network bandwidth [2]. With these applications being used on a daily basis, explosion of data, bandwidth problems, balancing the load on the web servers is become an important factor. Along with Load Balancing, Optimal Routing of the incoming inquiries from the users also play an important role. Optimal routing in Vehicle Routing Problem, Transit Routing Problem and Optimal Scheduling Problem. With some drawbacks for the optimal routing problems, accordingly Genetic Algorithms for the mentioned routing and scheduling problems are applied to overcome them.

2. TYPES OF TRAFFIC

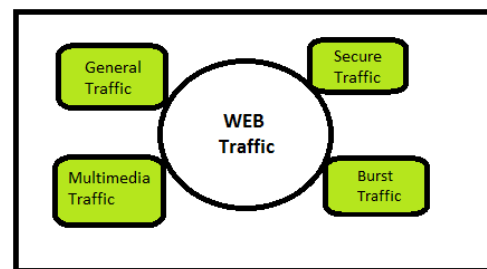


Fig 1: Web Traffic

According to Kotogiannis et.al [3], types of traffic on the web servers can be classified as-

2.1 General Traffic

The traffic that is generated on customers/users request for data. Data could be plain text documents or any static data on web sites and also the dynamic data [3].

2.2 Multimedia Traffic

Traffic related to streaming of data that could be in the format of either video or audio is Multimedia Traffic [4].

2.3 Secure Traffic

Traffic mostly generated by the E-Commerce Applications, which mostly run on SSL-TTL Protocol [4].

2.4 Burst Traffic

Traffic mainly caused due to packet which are transferred in bursts such as P2P transfers, file downloads or uploads etc. [5] [3].

3. LOAD BALANCING

Networking method distributing work between two or more processors, computers or a computer cluster, memory devices, network links, or other resources in order to channelize the resources in an efficient way and get optimized results like response time, throughputs, etc. is Load Balancing. The figure below gives an example of load balancing, distributing the workloads amongst different web servers.

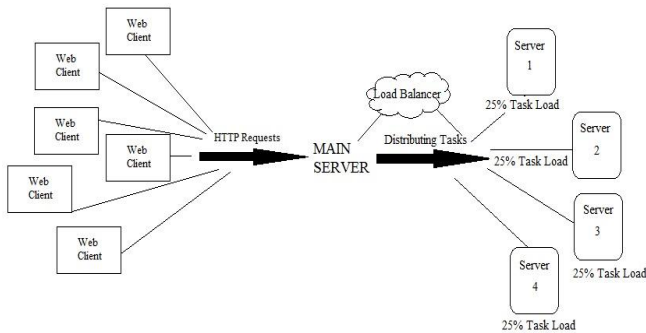


Fig 2: Load Balancing

3.1 Load Balancing Goals

- Opt a second plan when a single node or a group of nodes fail.
- Achieve maximum resource utilization.
- Minimize the response time of the system, network and avoid overloading and crashing of the system.
- Maintaining the stability of the systems connected to each other.
- Improving the performance of the overall network, so that scheduling of jobs/tasks is done in an efficient way.
- Any modifications to be done can be done easily.

3.2 Types Of Load Balancing

3.2.1 Static Load Balancing:

Load Balancing referred to as a mapping problem, task is to map a static process graph onto a fixed hardware topology in order to reduce dilation and process load differences.

3.2.2 Dynamic Load Balancing:

Mostly preferred for a distributed system to have the system load balanced evenly among all the nodes in the system/network so that the response time is very less and outcome of the system becomes efficient for the users.

3.3 Methods Of Load Balancing

3.3.1 Round Robin:

This technique is quite simple and easy to implement. With this method the incoming requests from the users are distributed among different servers sequentially, but keeping in mind that for this technique all the servers (virtual or physical) should have same resource capacities. If the condition is not met that means a low powerful server when receives a next inquiry even when it has not been able to process the current inquiry. This could lead to overloading on the weaker server.

3.3.2 Weighted Round Robin:

The drawback that the simple Round Robin had with overloading the weakest server is fulfilled by the Weighted Round Robin method. Here, while setting up the server a static weight is being assigned to the server. The administrator define capacities of the servers by weighting the servers.

For example, Server A is given a weighting of 100 and another weak server has been given a weighting of 50. This means, Server A would always receive two consecutive inquiries before server B receives its inquiry and goes on.

3.3.3 Least Connect

With increasing time the Round Robin methods does not take the number of active connections in to consideration. Hence, this method 'Least Connect' distributes the incoming inquiries on the basis of the connections that every server is currently maintaining. Among all the servers the server with the least number of active connections, automatically receives the next inquiry.

3.3.4 Weighted Least Connect

With an improvement over the Round Robin method, this method tries to combine the advantages of the weighing the servers and selecting the server with the current active connections. Using this ratio of number of connections and weightage of the server, this seems to be a quite fair distribution of inquiries. The least the ratio of the server, more the chances of it being selected.

4. THE OPTIMAL ROUTING PROBLEM

Problem is to find a path which achieves some pre-defined purpose and is efficient (i.e., it is optimal or good in some way). Two major classes of routing problems, namely the vehicle routing problem, and the transit routing problem. The sheer number of different possible routes and various different constraints representing several resource limitations make the development of such a bus system difficult.

4.1 The Vehicle Routing Problem

Problems in which the optimal closed loop paths which touch different interested points are needed to be determined. These points of interest are basically referred to as the nodes. The start node and the end node of the route are kept same and these start and end nodes are known as depot (starting and finishing points/nodes).

4.2 The Transit Routing Problem

Transit routing problems are different from the vehicle routing problems. A route is needed to be determined in which transit units (e.g. buses) will run according to some pre-defined schedule. The purpose of this technique is to get a good set of routes. The difference between the two routing problems are-

- The vehicle (e.g. bus) being routed, need not have to visit the points in demand, instead the demand points re-adjust themselves (gathering at a bus stop) to get the services.
- The demand might not be satisfied by just one route, hence these demand points can transfer from one route to another in order to reach the final destination.
- It is not a compulsion that all demand for travel be met.

According to Chakroborty and Dwivedi, an efficient route should satisfy the following:

- The route network should satisfy all of the transit demand.
- It should demand the transit demand of passengers with percentage of demand satisfied with zero transfer to be as high as possible.
- It should offer average travel time per transit as low as possible.

4.3 The Optimal Scheduling Problem

Another optimization problem related to the transit system design is scheduling of transit units (e.g. buses). Problem here is that given a set of routes, one needs to develop schedules for bus arrivals and departures at all the stops of the network. A good or efficient schedule is one which minimizes the waiting time of passengers while operating within a set of resource and service related constraints. The total waiting time of passengers have two components:

- The total Initial Waiting Time (IWT) of passengers, this is the sum of the waiting times of all the passengers at their point of origin.

The total Transfer Time (TT), this is the sum of the transfer times of all the transferring passengers.

5. GENETIC ALGORITHM'S

Given the importance of devising faster methods to obtain optimal / near-optimal solutions to the routing and scheduling problems a lot of work has been going on to develop such techniques using new tools like Tabu Search, Simulated Annealing, Ant Systems, and Genetic Algorithms. Over the last decade or so, IIT Kanpur has contributed to this effort by developing various methods using Genetic Algorithms. These methods can be found in the following articles: Chakroborty and Samanta [6], Chakroborty and Mandal [7], Chakroborty [8], Chakroborty and Dwivedi [9], Chakroborty et al. [10], Deb and Chakroborty [11], Chakroborty et al. [12], Chakroborty et al. [13], and Chakroborty et al. [14].

5.1 Ga Based Optimizer For Vehicle Routing Problems

In Chakroborty and Mandal [7] a single GA based algorithm called ROUTER was proposed for solving TSP, SVPDP, and SVPDPTW. It was shown that this algorithm was faster than similar algorithms proposed earlier (see Chatterjee et al. [15]). Further, this is the only algorithm which can handle all the three types of problems mentioned above; all other methods are problem specific and not as general as ROUTER. The greatest strengths of this algorithm are its simplicity and its speed; results show that this algorithm is about 30 times faster than an existing fast GA based algorithm (see Chatterjee et al. [15]) with only a marginal decrement in solution quality (ROUTER solutions are at the most 0.8% higher than Chatterjee et al.'s [15] solutions). [19] For solving VRP using GA, usually each individual is represented by just one chromosome, which is nothing but a chain of integers, each of them is either a customer or vehicle, depends on what the problem is meant for. So that each vehicle identifier represents a separator between different routes in the chromosome, and a string of customer identifiers a sequence of deliveries that must cover a vehicle during its route.

5.2 Ga Based Optimizer For The Transit Routing Problem

In Chakroborty and Dwivedi [9] a GA based algorithm for transit route design was developed. An adding node procedure is introduced to convert an infeasible solution to a feasible solution. Then route crossover and identical-point mutation are proposed to perform the genetic operations.

5.3 Ga Based Optimizer For The Scheduling Problem

Much amount of work done in this area having solutions to these scheduling problems with the transfer stops. The problems with the transfer stops can be deduced from the schedule of stochastic arrival times and finite bus easily. Other schedules which help in capacities have also been taken into consideration.

6. CONCLUSION

Thus from this report, load balancing on web servers are thoroughly discussed, classified and is an important factor for managing tasks among different servers. Optimal Routing and Scheduling discussed along with the Genetic Algorithms which helps in getting the best out of these optimal problems stated above.

7. REFERENCES

- [1] Jiani Guo and Laxmi Narayan Bhuyan, "Load Balancing in a Cluster-Based Web Server for Multimedia Applications", IEEE Trans. Parallel Distrib. Syst. 17, 11 (November 2006), 1321-1334.
- [2] Teixeira, M, M. Santana, M. J.Santana, R. H. C, "Analysis of Task Scheduling Algorithms in Distributed Web-server Systems", In International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS 2003), p.655{63. Montreal, Canada, Jul., 2003.
- [3] S.kontogianis, S. Valsamidis, P. Eframidis, A.karakos, "An adaptive load balancing algorithm for cluster based web systems", <http://skontog.gr/papers/duthtr-12-07.pdf>
- [4] Casalicchio, E., and Colajanni, M, "A client-aware dispatching algorithm for web clusters providing multiple services", WWW ACM (2001), 535-544.
- [5] Kant, K.; Won, Y, "Server capacity planning for web traffic workload" IEEE Transactions on Data and Knowledge Engineering, v.11, n.5, p.731 {47.
- [6] Chakroborty, P. and Samanta, S. (working paper). An asexual genetic algorithm for the vehicle routing problem with capacity constraints.
- [7] Chakroborty, P. and Mandal, A. (in press). An asexual genetic algorithm for the general single vehicle routing problem. Engineering Optimization.
- [8] Chakroborty, P. (2003). Genetic algorithms for optimal urban transit network design. Journal of Computer Aided Civil and Infrastructure Engineering, 184-200
- [9] Chakroborty, P. and Dwivedi, T. (2002). Optimal route network design for transit systems using genetic algorithms. Engineering Optimization, (1), 83-100.
- [10] Chakroborty, P., Deb, K. and Sharma, R. K. (2001). Optimal fleet size distribution and scheduling of urban transit systems using genetic algorithms. Transportation Planning and Technology, (3), 209-226
- [11] Deb, K. and Chakroborty, P. (1998) Time scheduling of transit systems with transfer considerations using genetic algorithms. Journal of Evolutionary Computation, (1), 1-24.
- [12] Chakroborty, P., Deb, K. and Srinivas, B. (1998) Network-wide optimal scheduling of urban transit

- networks using genetic algorithms. *Journal of Computer Aided Civil and Infrastructure Engineering*, 363-376.
- [13] Chakroborty, P., Deb, K., and Porwal, H. (1997). A genetic algorithm based procedure for optimal transit system scheduling. *Proceedings of Fifth International Conference on Computers in Urban Planning and Urban Management, Mumbai, India*, 330-341.
- [14] Chakroborty, P., Deb, K. and Subrahmanyam, P. S. (1995) Optimal scheduling of urban transit system using genetic algorithms. *ASCE Journal of Transportation Engineering*, (6), 544-553.
- [15] Chatterjee, S., Carrera, C. and Lynch, L. A. (1996). Genetic algorithms and traveling salesman problems. *European Journal of Operational Research*, 490-510
- [16] 2011 KEMP Technologies, Inc., loadbalancerblog.com, "Load Balancing Scheduling Methods Explained".
- [17] 2015 Liquid Web Inc., [liquid web](http://liquidweb.com) | Knowledge Base, "Understanding Load Balancing".
- [18] Joanne Suk Chun Chew and Lai Soon Lee, "A Genetic Algorithm for Urban Transit Routing Problem". *International Conference Mathematical and Computational Biology 2011. International Journal of Modern Physics: Conference Series (Vol. 9 (2012) 411–421)*.
- [19] NEO – Network and Emerging Optimization.