ISSN No. 0976-5697

Volume 2, No. 1, Jan-Feb 2011

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

A Load Balancing Technique for Wireless LANs

S.BalaBaskar
Department of Computer Science,
Pondicherry University, India

K.Vaitheki*
Department of Computer Science,
Pondicherry University, India
vaidehi.balaji@gmail.com

***Abstract*:** Throughput maximization in a network is one of the major key challenges in wireless LANs (WLANs) and it could be achieved when the load of access points (APs) is balanced. The AP load in WLANs is very often found to be uneven. To overcome such imbalance of load, a load balancing technique have been proposed. The load balancing techniques commonly require proprietary software or hardware at the user side for controlling the user-AP association. A new load balancing technique by controlling the size of WLAN cells (i.e., AP's coverage range), which is conceptually similar to cell breathing in cellular networks. The proposed scheme does not require any modification from the user side nor may any changes be required to the IEEE 802.11 standard [1]. It only requires the ability of dynamically changing the transmission power of the AP beacon messages. We develop a newer load balancing algorithm that will find the optimal beacon power settings which will thereon help to minimize the load of the most congested AP. The load may become shared and this may be performed with minimal delay and also maximized throughput is to be achieved.

***Keywords*:** Network throughput, Access points, WLAN, Beacon Messages, Load balancing, Throughput, Wireless LANs.

I. INTRODUCTION

In networking, load balancing is a technique to distribute workload evenly across two or more computers, network links, CPUs, hard drives, or other resources, in order to get optimal resource utilization, maximize throughput, minimize response time, and avoid overload. Using multiple components with load balancing, instead of a single component, may increase reliability through redundancy. The load balancing service is usually provided by a dedicated program or hardware device (such as a multilayer switch or a DNS server).It is commonly used to mediate internal communications in computer clusters, especially high-availability clusters. If the load is more on a server, then the secondary server takes some load while the other is still processing requests.

In WLANs, by default, a user scans all available channels and associates itself with an AP that has the strongest received signal strength indicator (RSSI), while being oblivious to the load of APs. The users are not evenly distributed, as a result some APs tend to suffer from heavy load, while their adjacent APs may carry only light load. Such load imbalance among APs is undesirable as it hampers the network from fully utilizing its capacity and providing fair services to users. The traffic load is often unevenly distributed among the access points (APs) in IEEE 802.11 wireless LANs WLANs [13]. The proposed work aims at a novel load balancing scheme that reduces the load of congested APs by forcing the users near the boundaries of congested cells to move to neighboring less congested cells. This is achieved via cell size dimensioning by controlling the transmission power of the AP beacon messages. A WLAN cell is defined as a region in which the AP beacon signal has the strongest RSSI.The approach is conceptually similar to cell breathing in cellular networks. An optimal algorithm determines min-max load balancing solutions. A WLAN is min-max load balanced, if it is impossible to reduce the load of any AP without increasing the load of other APs with equal or higher load. The proposed approach is practical since it does not require either user assistance or standard modification. Currently the IEEE 802.11 standard [20] does not provide any standard method to resolve the load imbalance. To overcome this deficiency, various load balancing schemes have been proposed by both the academia and the industry. Most of these methods commonly take the approach of directly controlling the user-AP association by deploying proprietary client software or specially-designed WLAN cards at the user computers.

Today, WLAN users frequently move between different WLANs, such as hotels, airports, shopping centers and university campuses. Different networks are managed by different organizations and likely adopt different load balancing mechanisms. It is unrealistic to require the users to have the appropriate client modules for each visiting network. The need for a new load-balancing scheme that does not require any proprietary client module nor did any modification of the standard is energized. Balachandran et al. [8] Propose to associate a user with the AP that can provide a minimal bandwidth required by the user.

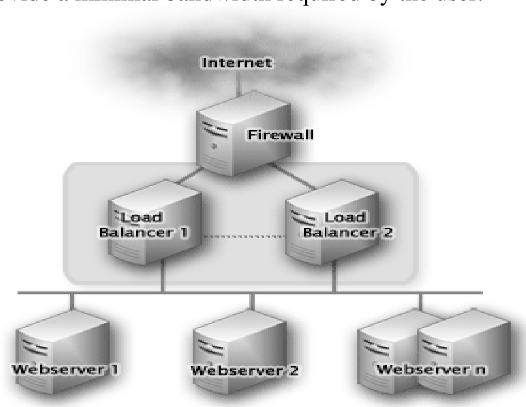

Figure 1: Load balancer

A farm of servers with the same function is the base of a load balancing cluster [17]. To distribute the user requests to several nodes, a load balancer is useful. The load balancer checks the utilization of all nodes. The node with the estimated best performance will get the next user request. This algorithm ensures the best performance available at the time is given to the users. A very important point is the quality of the load balancer. In this case, quality means the opportunity of the system to make a highly qualified forecast about which node will offer the best performance concerning an individual application. Another aspect to make the decision is to reconnect the client to an existing session. Without any kind of protection, the load balancer could be a single point of failure.

Applications of load balancing One of the most common applications of load balancing is to provide a single Internet service from multiple servers, sometimes known as a server farm. Commonly, load-balanced systems include popular web sites, large Internet Relay Chat networks, high-bandwidth File Transfer Protocol sites, Network News Transfer Protocol (NNTP) servers and Domain Name System (DNS) servers. For Internet services, the load balancer is usually a software program that is listening on the port where external clients connect to access services. The load balancer forwards requests to one of the "backend" servers, which usually replies to the load balancer. This allows the load balancer to reply to the client without the client ever knowing about the internal separation of functions. It also prevents clients from contacting backend servers directly, which may have security benefits by hiding the structure of the internal network and preventing attacks on the kernel's network stack or unrelated services running on other ports. Some load balancers provide a mechanism for doing something special in the event that all backend servers are unavailable. This might include forwarding to a backup load balancer, or displaying a message regarding the outage. An alternate method of load balancing, which does not necessarily require a dedicated software or hardware node, is called round robin DNS. In this technique, multiple IP addresses are associated with a single domain name clients themselves are expected to choose which server to connect to. Unlike the use of a dedicated load balancer, this technique exposes to clients the existence of multiple backend servers. The technique has other advantages and disadvantages, depending on the degree of control over the DNS server and the granularity of load balancing desired.

Another technique for load-balancing using DNS, which is far more intelligent than the simple "round-robin", is to delegate "www.example.com" as a sub-domain whose zone will be served out by each of the same servers that are serving the web site. This technique works particularly well where individual servers are spread around the Internet.If there exist many of such APs, the one with the strongest RSSI is selected. In [9], Velayos *et al.* introduced A distributed load balancing architecture where the load of an AP is defined as the aggregated downlink and uplink traffic through the AP. In [10], Kumar *et al.* proposed an association selection algorithm which is based on the concept of proportional fairness to balance between throughput and fairness. Most of these works heuristically determine only the association of newly arrived users.

II. CONGESTED LOAD IN WIRELESS LANS

The network throughput is maximized and the best is achieved when the load of access points (APs) is balanced.
To alleviate such imbalance of load, several a load balancing schemes have been proposed. These schemes commonly require proprietary software or hardware at the user side for controlling the user-AP association. In this paper we present a new load balancing technique by controlling the size of WLAN cells (i.e., AP's coverage range), which is conceptually similar to cell breathing in cellular networks. The proposed scheme does not require any modification to the users neither the IEEE 802.11 standard. It only requires the ability of dynamically changing the transmission power of the AP beacon messages. A set of polynomial time algorithms that find the optimal beacon power settings can be designed which minimize the load of the most congested AP. The problem of network-wide min-max load balancing is also considered. Simulation results show that the performance of the proposed method is comparable with or superior to the best existing association-based method.

A. System failure or hang due to congested load of APs in wireless LANs.
B. Client needs to wait, while server responding to other clients.
C. Traffic over the wireless LANs.
D. Server cannot able to quick response to the client
E. Signal loss.

Persistence An important issue when operating a load-balanced service is how to handle information that must be kept across the multiple requests in a user's session. If this information is stored locally on one backend server, then subsequent requests going to different backend servers would not be able to find it. This might be cached information that can be recomputed, in which case load-balancing a request to a different backend server just introduces a performance issue[16]. One solution to the session data issue is to send all requests in a user session consistently to the same backend server. This is known as "persistence" or "stickiness". A significant downside to this technique is its lack of automatic failover: if a backend server goes down, its per-session information becomes inaccessible, and any sessions depending on it are lost.

Assignment to a particular server might be based on a username, client IP address, or random assignment. Owing to DHCP, Network Address Translation, and web proxies, the client's IP address may change across requests, and so this method can be somewhat unreliable. Random assignments must be remembered by the load balancer, which creates a storage burden. If the load balancer is replaced or fails, this information can be lost, and assignments may need to be deleted after a timeout period or during periods of high load to avoid exceeding the space available for the assignment table. The random assignment method also requires that clients maintain some state, which can be a problem, for example when a web browser has disabled storage of cookies. Sophisticated load balancers use multiple persistence techniques to avoid some of the shortcomings of any one method. Another solution is to keep the per-session data in a database. The database is best used to store information less transient than per-session data. To prevent a database from becoming a single point of failure, and to improve scalability, the database is often replicated across multiple machines, and load balancing is used to spread the query load across those replicas. Microsoft's ASP.net State Server technology is an example of a session database. All servers in a web farm store their session data on State Server and any server in the farm can retrieve the data. There are more efficient approaches also. In the very common case where the client is a web browser, per-session data can be stored in the browser itself. One technique is to use a browser cookie, suitably time-stamped and encrypted. Another is URL rewriting [16]. Storing session data on the client is generally the preferred solution: then the load balancer is free to pick any backend server to handle a request. However, this method of state-data handling is not really suitable for some complex business logic scenarios, where session state payload is very big or recomputing it with every request on a server is not feasible, and URL rewriting has major security issues, since the end-user can easily alter the submitted URL and thus change session streams.

III. ADVANTAGES OF THE PROPOSED TECHNIQUE

Cell breathing is accessed mostly in the context of CDMA cellular networks. The coverage and capacity of a CDMA cell are inversely related with each other .The increase of the number of active users in a cell causes the increase of the total interference sensed at the base station. Therefore, in congested cells, users need to transmit with higher power to maintain a certain signal-to-interference ratio at the receiving base station. As the users in a congested cell increase their transmission power, they also increase their Interference to the neighboring cells since all cells use the same frequency in CDMA networks [5]. The overall network capacity may decrease as a result. The maximal transmission power of the users are bounded, the users who are far from the base station may experience poor services. To overcome these problems, the cell breathing approach was proposed. The main essence is that they reduce the size of congested cells. In existing system, when the total load exceeds a certain threshold or the bandwidth allocated to users' drops below a certain threshold. While the existing load balancing schemes achieved considerable improvement in terms of throughput and fairness, they require certain support from the client side. The beneficent part of the proposed scheme is that it does not require any proprietary client support.

A. Equal load balance of APs.
B. Overcome system failure.
C. As the server responds to another client, Client need not wait.
D. Providing high quality signals.
E. Client gets an immediate response from server.
F. The traffic over wireless LANs is highly reduced.
G. Dynamically changing the transmission power of the AP beacon messages.

IV. STRUCTURE OF THE PROPOSED SYSTEM

The problem of minimizing the load of the congested APs has been addressed in the proposed work. The Access Points with the maximal load are known as congested AP and the load is congestion load. The two polynomial time algorithms that find optimal solutions, one for the complete knowledge model and the other for the limited knowledge model could be effective. The results are intriguing, because similar load balancing problems are known to be strong NP-hard. It is particularly interesting that a polynomial time optimal algorithm exists for the limited knowledge model. Then the problem of min-max load balancing is addressed. This is a strong NP-hard problem. It has been already proved that there exists no algorithm that guarantees any coordinate wise approximation ratio, and the approximation ratio of any prefix-sum approximation algorithm is at least nlogn, where n is the number of APs. A variant of this min-max problem is solved, termed min-max priority load balancing, whose optimal solution can be calculated in polynomial time for both knowledge models [1]. Here, the AP load is defined as an ordered pair of the aggregated load contributions of its associated users and a unique AP priority.

A. Client Model

A client is an application or system that accesses a remote service on another computer system, known as a server, by way of a network. The term was first applied to devices that were not capable of running their own stand-alone programs, but could interact with remote computers via a network. These dumb terminals were clients of the time-sharing mainframe computer.

B. Server model

In computing, a server is any combination of hardware or software designed to provide services to clients. When used alone, the term typically refers to a computer which may be running a server operating system, but is commonly used to refer to any software or dedicated hardware capable of providing services.

C. Network Model

The channel quality is generally time-varying. For the ser-AP association decision, a user performs multiple samplings of the channel quality, and only the signal attenuation that results from long-term channel condition changes are utilized the proposed load model can accommodate various additive load definitions such as the number of users associated with an AP. It can also deal with the multiplicative user load contributions.

D. Cell Breathing Approach

We reduce the load of congested APs by reducing the size of the corresponding cells. Such cell dimensioning can be obtained, for instance, by reducing the transmission power of the congested APs. This forces users near the congested cells' boundaries to shift to adjacent (less congested) APs. The separation between the transmission power of the data traffic and that of the AP beacon messages. On one hand, the transmission bit rate between a user and its associated AP is determined by the quality of the data traffic channel.

Transmitting the data traffic with maximal power maximizes the AP-user SNR and the bit rate. On the other hand, each user determines its association by performing a scanning operation, in which it evaluates the quality of the beacon messages of the APs in its vicinity.

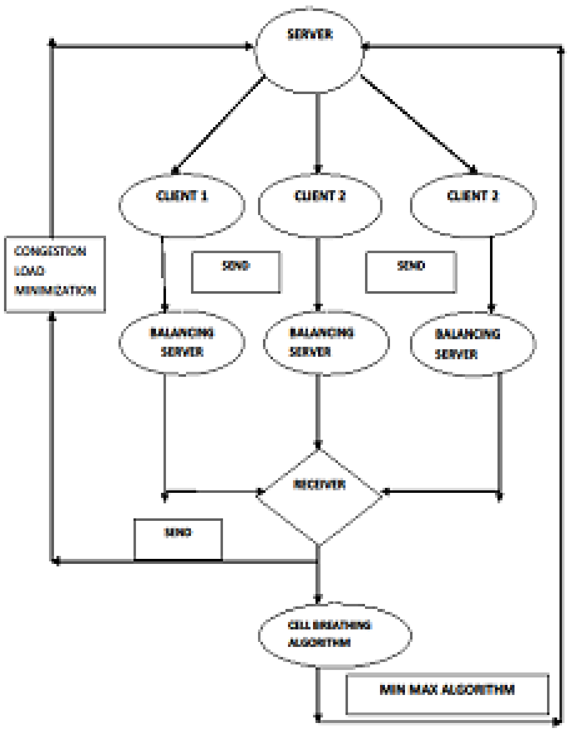

Figure 2 Data flow Diagram

E. Congestion Load Minimization

The algorithms presented the minimize the load of the congested AP, but they do not necessarily balance the load of the non-congested APs, we consider min-max load balancing approach that not only minimizes the network congestion load but also balances the load of the non-congested APs. The proposed approach can be used for obtaining various max-min fairness objectives by associating each user with appropriate load contributions. Unfortunately, min-max load balancing is NP-hard problem and it is hard to find even an approximated solution. We solve a variant of the min-max problem, termed min-max priority-load balancing problem, whose optimal solution can be found in polynomial time.

F. Complete Knowledge Model

A complete knowledge model is feasible when all users collect the RSSI information from all of the nearby APs. This model as a building block for the limited knowledge solution since most of WLANS feature is not available. A network has limited knowledge when the available information comprises only the set of users that are currently associated with each AP and the load contributions.

G. Limited knowledge Model

Unlike the complete knowledge case, in this model, the algorithm cannot calculate bottleneck set in advance.

The algorithm updates the APs beacon power settings and evaluates the AP loads. This process raises the problem of determining a "termination condition" when an optimal solution is found.

H. Min-Max Load Balancing

The min-max load balancing approach not only minimizes the network congestion load but also balances the load of the uncongested APs. Approach can be used for obtaining various max-min fairness objectives by associating each user with appropriate load contributions. We solve a variant of the min-max problem, termed min-max based on the priority-load, whose optimal solution can be found using the algorithm of congested load balancing.

The complete knowledge model and limited knowledge model minimize the congested load of access points but they do not balance the load of non congested APs. So, the min-max load balancing approach minimize load of congested APs but also balance the load of uncongested APs. This approach gives the optimal solution that found in polynomial time by using with complete and limited knowledge model. Load balancing for a parallel system is one of the most important problems which have to be solved in order to enable the efficient use of parallel computer systems. This problem can be compared to problems arising in natural work distribution processes like that of scheduling all activities (tasks) needed to construct a large building. Several objectives have to be taken into consideration: The whole work should be completed as fast as possible. The work should be distributed fairly. About the same amount of work should be assigned to every worker.

I. Implementation and Experimental Results

Figure 3 Simulation results – Three servers

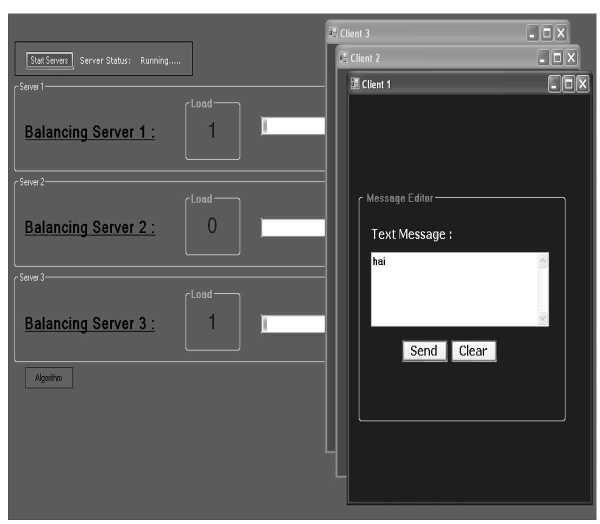

Figure 4 Sending Messages – Load Balancing in Three servers

Figure 5 Message Transferring (Load being balanced)

V. CONCLUSION AND FUTURE ENHANCEMENTS

A scheme has been proposed for optimal load balancing in IEEE 802.11 WLANs. A brief analysis of the problem is performed and the algorithm that have been proposed results in deterministic optimal solutions. The congested load algorithm minimizes the load of the congested AP(s) in the network, and the dynamic min-max algorithm produces an optimal priority based load balanced solution. These optimal solutions are obtained only with the minimal information which is readily available without any special assistance from the users or modification of the standard. Only the control on the transmission power of the AP beacon messages is assumed. The congested load balancing and the min-max algorithm that has been proposed work upon very limited knowledge models. The work of the algorithm can further be improved by increasing the load capability. The strong coverage requirements by considering the transmission power need to be considered and satisfied in the sparse access point deployments. The requirements and improvements can further be improved as the algorithm is mainly structured or modular in nature. The existing modules can be renewed by adding new modules can append improvements. The future enhancements can be made to the application, so that the functionalities of the wireless LANs are made better and more beneficial than the proposed.

VI. REFERENCES

[1] Yigal Bejerano and Seung.-Jae Han, “Cell Breathing Techniques for Load Balancing in Wireless LANs,” Proc. IEEE INFOCOM, 2006.

[2] M. Balazinska and P. Castro, “Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network,” Proc. USENIX Int’l Conf. Mobile Systems, Applications, and Services (MobiSys ’03), 2003.

[3] T. Henderson, D. Kotz, and I. Abyzov, “The Changing Usage of a Mature Campus Wide Wireless Network,” Proc. ACM MobiCom, pp. 187-201, 2004.

[4] T. Togo, I. Yoshii, and R. Kohno, “Dynamic Cell-Size Control According to Geographical Mobile Distribution in a DS/CDMA Cellular System,” Proc. IEEE Personal, Indoor, and Mobile Radio Comm. Symp. (PIMRC ’98), pp. 677-681, 1998.

[5] A. Jalali, “On Cell Breathing in CDMA Networks,” Proc. IEEE Int’l Conf. Comm. (ICC ’98), pp. 985-988, 1998.

[6] I. Papanikos and M. Logothetis, “A Study on Dynamic Load Balance for IEEE 802.11b Wireless LAN,” Proc. Int’l Conf. Comm. Control (COMCON ’01), 2001.

[7] I. Tinnirello and G. Bianchi, “A Simulation Study of Load Balancing Algorithms in Cellular Packet Networks,” Proc. ACM/ IEEE Int’l Workshop Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM ’01), pp. 73-78, 2001.

[8] A. Balachandran, P. Bahl, and G.M. Voelker, “Hot-Spot Congestion Relief and Service Guarantees in Public-Area Wireless Networks,” SIGCOMM Computing Comm. Rev., vol. 32, no. 1, pp. 59-59, 2002.

[9] H. Velayos, V. Aleo, and G. Karlsson, “Load Balancing in Overlapping Wireless LAN Cells,” Proc. IEEE Int’l Conf. Comm. (ICC ’98), 1998.

[10] A. Kumar and V. Kumar, “Optimal Association of Stations and APs in an IEEE 802.11 WAN,” Proc. Nat’l Conf. Comm., 2005.

[11] K. Premkumar and A. Kumar, “Optimal Association of Mobile Wireless Devices with a WLAN-3G Access Network,” Proc. IEEE Int’l Conf. Comm. (ICC ’06), 2006.

[12] B. Kauffmann, F. Bacilli, A. Chaintreau, K. Papagiannaki, and C. Diot, “Self Organization of Interfering 802.11 Wireless Access Networks,” INRIA Research Report RR-5649, 2005.

[13] S. Shakkottai, E. Altman, and A. Kumar, “The Case for Non- Cooperative Multihoming of Users to Access Points in IEEE 802.11 WLANs,” Proc. IEEE INFOCOM, 2006.

[14] T.-C. Tsai and C.-F. Lien, “IEEE 802.11 Hot Spot Load Balance and QoS-Maintained Seamless Roaming,” Proc. Nat’l Computer Symp. 2003.

[15] Y. Bejerano, S.-J. Han, and L.E. Li, “Fairness and Load Balancing in Wireless LANs Using Association Control,” Proc. ACM MobiCom, pp. 315-329, 2004.

[16]http://en.wikipedia.org/wiki/Load_balancing_(computing)

[17]http://www.hacc.org/uploads/pics/Load_Balancing_06.gif

[18] S.-T. Yang and A. Ephremides, “Resolving the CDMA Cell Breathing Effect and Near-Far Unfair Access Problem by Bandwidth-Space Partitioning,” Proc. IEEE Vehicular Technology Conf. (VTC ’01), pp. 1037-1041, 2001.

[19] V.V. Veeravalli and A. Sendonaris, “The Coverage-Capacity Tradeoff in Cellular CDMA Systems,” IEEE Trans. Vehicular Technology, pp. 1443-1451, Sept. 1999.

[20] IEEE Standard 802.11, .Wireless LAN Medium Access Control (MAC)and Physical Layer (PHY) specifications., 1999 Edition (ISO/IEC 8802-11:1999).