International Journal of Smart Sensor and Adhoc Network

Volume 2 | Issue 1 Article 9

October 2011

Techniques for Load Balancing in Wireless LAN

Krishnanjali. A. Magade
Computer Dept., D.Y.Patil College of Engineering, Akurdi, Pune, Maharashtra, India, anjali_skn@rediffmail.com

Abhijit patankar
Computer Dept., D.Y.Patil College of Engineering, Akurdi, Pune, Maharashtra, India, abhijitpatankar@yahoo.com

M. A. Potey
Computer Dept., D.Y.Patil College of Engineering, Akurdi, Pune, Maharashtra, India, mapotey@gmail.com

Techniques for Load Balancing in Wireless LAN

Krishnanjali A.Magade, Abhijit & M.A.Potey

Computer Dept., D.Y.Patil College of Engineering, Akurdi, Pune, Maharashtra, India
E-mail : anjali_skn@rediffmail.com, abhijitpatankar@yahoo.com, mapotey@gmail.com

Abstract - This suggests new strategies for balancing load in a wireless network connected in star topology. The loads are assigned to each processor using divisible load theory & Different techniques [II], [III], [IV], and [V]. Divisible load theory suggests that a load can be divided arbitrarily such that each fraction of the load can be independently assigned and computed in any processor present in the network. Wireless networks are connected in such a manner that they as assemble a distributed system most of the times, which makes load balancing an important technique to maximize the throughput from the system. A wireless sensor network generally consists of a base station (or Gateway) which communicates with other nodes present in the network. The other nodes are used for Measuring and collecting various environmental and Intelligence related data. The network that we have considered is connected with the central node being the base station and the other nodes are used for calculation of load distributed by the central node. Load balancing involves distribution of all computational and communicational activities over two or more processors, links or any other computational devices present in the network. The main thing behind this is load balancing is to reduce the execution time of the load and to make sure that all the resources present in the system are utilized optimally. The IEEE 802.11 standard does not provide any mechanism to resolve load imbalance. To reduce this deficiency, various load balancing schemes have been designed. These techniques commonly take the approach of directly controlling the user-AP association by deploying Proprietary client software or hardware. Load balancing Features in their device drivers, AP firm wares, and WLAN cards. In these solutions, APs broadcast their load levels to users via modified beacon messages and each user chooses the least-loaded AP.

Keywords *- Load balancing, WLAN, Min-Max Load balancing, Load, Access point.*

I. INTRODUCTION

Load balancing is distributing processing and communications activities evenly across a computer network so that no single device is overwhelmed. It is a technique to distribute the Workload evenly across two or more computers, network links, CPUs, hard drives, or other resources. The purposes of load balancing are to achieve optimal resource utilization, maximize throughput, minimize response time, and avoid overload. The traditional load balancing with load unit migration from one processing element to another when load is light on some processing elements and heavy on some other processing elements. It involves migration decision, i.e. which load unit(s) should be migrated and then migration of load unit to other nodes; also include [III] Divisible Load Theory (DLT). Load balancing is a technique used to distribute load on server to increase performance and speed of work designated to particular server. There are various techniques that can be used to accomplish load balancing task with the help of different types of computer hardware and software components. Application of load balancing is used by different companies that conduct huge business transactions using internet. It is normally done to ensure that all their clients are able to conduct business activities with ease and accuracy. Load balancing improves the performance of server due to distributed load and is used for busy and large networks .Without balancing load in busy networks it is very difficult to satisfy the entire request issued to server. Organization into the web services or online business normally makes use of load balancing technique and engages two web servers (or more servers accordingly). If one of the web server gets overloaded or goes off, in that case alternate server activates and access the requested load. Load balancing is done by assigning particular service time for each process in order to ensure that several requests are handled without causing traffic. In other words, specific time is assigned to each process in the server for its execution and the process no more stay in the server once service time extends. Once load balancer works actively, service time reduces for each process reduces.

Load balancing is particularly useful for parallel and distributed systems where we have to share the workload to get the maximum throughput from the System. Most distributed systems are characterized by the distribution of both physical and logical features. The architecture of a distributed system is generally modular in nature.

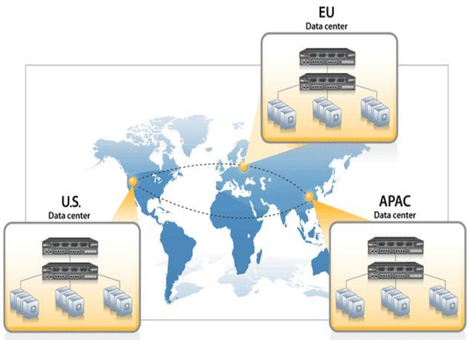

Fig. 1: Communication between different clients

Most of the distributed systems support different types and numbers of processing elements. The system resources like hardware, software, data, user software and user data are distributed across the system. An arbitrary number of system and user processes can be executed on various machines in the system. Factors to consider when selecting a machine for process execution include the availability of resources and its optimal use. In a distributed system environment, a load balancing algorithm seeks the least busy machine. At the same time, the load balancing algorithm must not overload the system. Ideally, the load balancing algorithm selects the machine for process execution based on available information about all the resources present in the system. The problem that we study is the load distribution among access points in the most client way. The standard way to connect wireless hosts (laptops, net books, smart phones, etc.)To access points is to connect to the access point with the strongest signal. This is a very good way to acquire the best bitrates possible, but it's not the optimal way to distribute the load among the available access points. The proliferation of lightweight hand-held devices with built in high-speed Wi-Fi network cards and the significant benefit of any-where any-time Internet access has spurred the deployment of wireless “hot-spot” networks [II]. It is easy to find wireless local area networks (WLANs) in classrooms, offices, airports, hotels, and malls. A key challenge for organizations that deploy WLANs is capacity management, making the best use of the available resources to derive the best return on investment while satisfying client service demands. Previous studies of public-area wireless networks have shown that client service demands are highly dynamic in terms of both time of day and location, and that client load is often distributed *unevenly* among wireless access points (APs) Clients tend to localize themselves in particular areas of the network for various reasons, such As availability of favorable network connectivity, proximity to power outlets, classrooms.

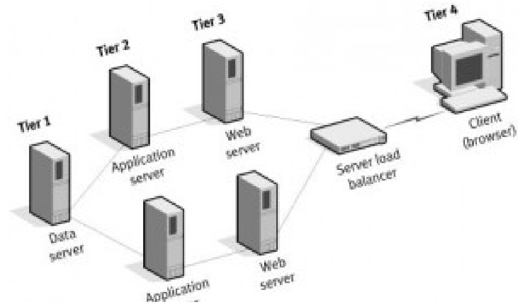

Fig. 2 : Clients connected with server

In WLANs, cell breathing can be implemented by controlling the transmission power of an AP’s beacon packets. Note that we do not change the transmission power of data packets to avoid degrading clients’ performance. More specifically, when data packets reduces, the AP may see higher data packet losses, or even adapt to a lower sending rate, both of which degrade the client’s performance. In comparison, changing the transmission power of beacon packets only affect how clients associate with APs, and does not affect the loss rate or sending rate of data packets, which matches our goal well. Finding the appropriate power assignment at APs to automatically achieve load balancing is a challenging problem. To our knowledge, the cell breathing algorithms proposed for cellular networks are based on local heuristics, and do not provide performance guarantees In this we develop power control algorithms for the following two cases: (i) APs are able to adjust Their power to any level (continuous-power assignment), and (ii) APs are able to adjust their power to only some discrete power levels (i.e., discrete-power assignment). when client demands are homogeneous (i.e., all clients have the same demand), we can always compute such a power assignment - we can set the powers of all APs in such a way that after all the clients choose their AP based on RSSI, either all the clients can be served by the APs or all the APs are fully utilized. For heterogeneous demands (i.e., clients can have different demands), we apply the same approach, and prove that it can completely satisfy at least $N - K$ clients, where N is the number of clients, and K is the number of APs (Note that K is often much smaller than N in practical scenarios). For discrete-power assignment, we develop a greedy algorithm. The high level idea of our algorithm is as follows. We start by setting the powers of all APs to the highest value, and then we choose the best power configuration resulting from iteratively decreasing the power of overloaded APs. This approach is intuitive and easy to implement. Moreover, it only requires knowledge of APs’ load, which is easy to Obtain. We show that if there exists a power assignment such that each AP has capacity to accommodate the demands assigned to it, our algorithm can find the solution in a polynomial time. In addition, we consider two extensions to the above algorithms: (1) dynamic adjustment of APs' power in response to changes in clients' load while limiting the number of hand-offs, and (2) minimizing APs' transmission power to reduce interference

II. TYPES OF LOAD BALANCING ALGORITHMS

The basic idea of a load balancing is to equalize loads at all computers by transferring loads to idle or heavily loaded computers. Load balancing algorithms can broadly be classified into three categories.

- Static algorithms
- Dynamic algorithms
- Adaptive algorithms

A. *Static Algorithms:*

In static algorithms, load balancing decisions are hard-wired in the algorithm using a priori knowledge of the system. The overhead entailed in static algorithms is almost nil.

B. *Dynamic Algorithms:*

Dynamic algorithms use system state information (the loads at nodes) to make load balancing decisions. Dynamic algorithms have t he potential to outperform the static algorithms, since they are able to exploit the short term fluctuations in the system to improve performance. But they incur overhead in the collection, storage and analysis of system state.

C. *Adaptive Algorithms:*

Adaptive algorithms are a special class of dynamic algorithms which adapt their activities by dynamically changing the parameters of the algorithm to suit the changing system state. Our results show the algorithms are effective for improving throughput. Under high load, the improvement is up to 50% for uniform client distributions, and up to an order of magnitude for non uniform distribution of clients' locations.

Another load balancing technique is *Persistent load balancing* technique. This load balancing technique does not make any use of software or hardware node and the customer is allowed to select their respective server and get services from that server. This process is very transparent and useful from client point of view as it discloses the presence of multiple servers at the backend. This technique assigns each new client with different set of server on a round robin allocation basis. This process of load balancing ensures that each new client is assigned with specific server and no server is overloaded with particular client.

III. DIVISIBLE LOAD THEORY

Divisible Load Theory (DLT) initially originated from the need of creating intelligent sensor networks, but it's most recent applications involve parallel and distributed computing. The first published research on divisible load theory appeared in a 1988 .The increasing prevalence of multiprocessor systems and data-intensive computing has created a need for efficient scheduling of computing loads, especially parallel loads that are divisible among processors and links. Divisible Load Theory involves the study of an optimal distribution of partition able loads among a number of processors and links. A load that can be arbitrarily distributed over a number of processors and communication links is called a *partition able data parallel load*. Divisible load theory offers a tractable and realistic approach to scheduling that allows integrated modeling of computation and communication in parallel and distributed computing systems. The main objective of the Divisible Load Theory is to optimally partition the processing load among a network of processors which are connected through communication links such that the entire load can be distributed and processed in the shortest possible span of time. Hence there is no precedence relations among the data involved in the computation process. DLT offers easy computation, a schematic language, and equivalent network element modeling. Since DLT does not take into account the precedence relations among data, it resumes that computation and communication loads can be arbitrarily partitioned among numerous processors and links, respectively. While it can incorporate stochastic features, but the primary model does not make statistical assumptions, which can be a drawback in a performance evaluation model. An example taken from illustrates how DLT work in a real life situation. A typical divisible load scheduling application might involve a credit card company that must process 30 million accounts each month. The company could conceivably send 300,000 records to each of 100 processors, but simply splitting the load equally among processors does not take into account different computer and communication link speeds, the scheduling policy, or the interconnection network. Divisible load theory provides the mathematical machinery to do time-optimal processing. There are so many different potential situations in which an accurate and tractable approach to divisible load scheduling would be useful.

IV. MIN-MAX LOAD BALANCING

The algorithms presented minimize the load of the congested AP, but they do not necessarily balance the load of the no congested APs. In this section consider min-max load balancing approach that not only minimizes the network congestion load but also balances the load of the no congested APs. Unfortunately, min-max load balancing is NP -hard problem and it is hard to find even an approximated solution. In this paper, 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. We present our algorithm for this problem in the limited knowledge case. Obviously, it can be used for the complete knowledge case as well. We propose an AP-centric approach to transparently balance load across different APs. The main challenge in this approach is to find appropriate transmission power for each AP such that the total client demand that APs can serve is maximized when clients use the basic associations scheme (i.e., associate with the AP with highest (RSSI). In order to formally specify the power control problem, we first introduce the following notations.

- K: The number of APs
- N: The number of clients
- Ci: The capacity of AP i
- d (i, j): The distance between AP i and client j
- α: signal attenuation factor
- Pi: AP i's power
- Di: client i's demand
- Pr (i, j): received power from AP i at client j
- Li: The total load served at AP i

Based on the notations listed above, we now formulate the power control problem as follow. Given K, N, C_i, D_i, and d (i, j), our goal is to find the transmission throughput (i.e., maximizing $\sum$i Li)ogiven I that client mj is assigned to AP i when P= min(Ci, ∑i', j) for all i′ ⍰ {1, 2, ..., K}, and L_i D_j) for all clients j that are mapped to AP i. The last equation reflects the fact that the maximum client demand the AP i can service, L_i, is bounded by its capacity and the total client demands that are assigned to it. Note that when there are multiple APs with similar RSSI, the client is randomly assigned to one of them.

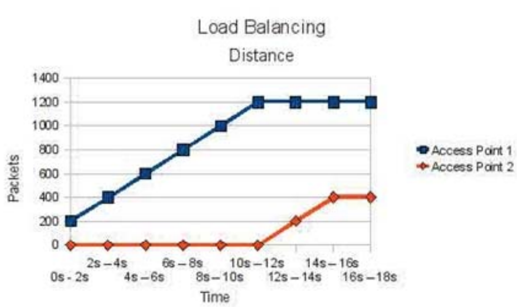

Fig. 3: Transmit to access point

V. COMPLETE KNOWLEDGE ALGORITHM

This present the complete knowledge algorithm. The user-AP association and the corresponding AP load are known a priori for all possible beacon power assignments. The algorithm starts with the maximal power state in which all APs transmit with the maximal power. This state dominates all other state and, in particular, the optimal states. The algorithm, iteratively, calculates the bottleneck set B. Based on the calculated set B, the algorithm determines whether it needs to apply another set power reduction operation or an optimal state is found. To this end, it utilizes two termination conditions. The first condition checks if B ¼ A. Recall that from it follows that reducing the transmission power of all APs does not change the user-AP association. Consequently, it cannot reduce the maximal AP load. In reality, this condition is satisfied when the load of all APs are balanced and any power reduction operation will cause some APs to be congested The second condition checks if the bottleneck set B contains an AP that transmits with the minimal transmits- power. In such case, the power level of all APs in B cannot be equally decreased and the algorithm halts. Such a case, typically, occurs when the AP load is not balanced and the algorithm attempts to reduce the maximal load by repeatedly reducing the power of the congested APs. In this section, we present power control algorithms for the cases when APs can adjust their power to any value (i.e., continuous power). The algorithms in this section require APs to estimate the received power at different clients

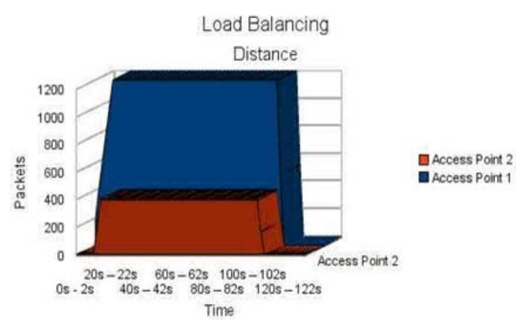

Fig. 4: Transmit with distance

Proof: Let F be the feasible (optimal) power assignment. Suppose for 1 ≤ ia ≤ h, the power of AP a in F is Pia a It is easy to see thata if ia ≤ ia, we find a power assignment in which all clients' demands are served without overloading APs (since the algorithm terminates at non-zero power only when it finds a solution in which all client demands are satisfied). Next we prove ia ≤ ia holds. Suppose by contradiction, during the while loop, there is an AP a for ia = ia + 1. Since the powers of all other APs are atleast the power in the optimal power assignment, the total demands of clients that prefer AP a can be at most the total demands assigned to an in F. This cannot be more than its capacity according to the definition of F. Therefore it is a contradiction. Reducing the transmission power of an AP affects the channel quality of all of its associated users, and this effect is not limited to those users that we intend to shift. The users who remain associated with the considered AP also experience lower channel quality and may have to communicate at a lower bit rate than before. This may result in longer transmission time of user traffic, which effectively increases the user load contributions on the AP, if the AP load is determined by considering not by the number of users but by the effective user throughput. Thus, we may end up with increasing the load of APs. In this, it presented the problem of load balancing in wireless sensor network using traditional divisible load theory. It extended an existing load balancing approach which allocates load in a wireless sensor network connected in star topology. In our proposed model we have considered the allocation time and found that the reporting time is less in case of the strategy immediate measurement and concurrent reporting than immediate measurement and simultaneous reporting. Among the existing strategies for the star topology we have found from simulations that simultaneous measurement and concurrent reporting has minimum reporting time. Among the two strategies considered for bus topology the strategy no control processor and processors with front end processor has less reporting time. Also in the existing strategies if the allocation time is considered then our proposed strategies has less reporting time.

VI. LBA

This algorithm introduces the LBA (Load Balancing Agent) to the access points which is used to create three distinct states at each access point.

- Under-loaded

The access point can accept clients that either entire the network or are roaming from neighboring access points.

- Balanced

The access point can only accept clients that enter the network.

- Over Loaded

The access point can't accept any clients that either enter the network or are roaming from neighboring access points. At this point, the access point forces the handover of current Stations (clients) to reduce its load level.

Client	Rate (Mbps) - AP1	Rate (Mbps) - AP2
Client1	50	34
Client2	44	21
Client3	11	2
Client4	24	54

Fig. 5: Results

The results about the distance algorithm show that the access point 1 has the most load according to the packets per time that there are throughout the simulation time.

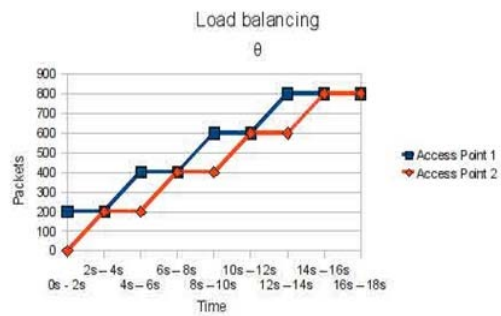

Fig. 6 : load balancing with distance

VII. CONCLUSION

We developed power control algorithms that maximize system throughput. In this section, we study how to simultaneously maximize system throughput and minimize APs' power. Power minimization is helpful to reduce interference among different APs. For ease of explanation, we consider homogeneous client demands. The same approach can be applied to split table heterogeneous client demands. To validate the effectiveness of the load balancing scheme, we evaluate the performance of the resource management policies. The performance is compared with that of two reference schemes. The first one applies a random network selection, in which an incoming service request in the overlaying area is directed to the cellular network and WLAN with equal probabilities. No vertical handoff from the cell to the overlaying WLAN is considered since this handoff is not necessary to maintain a connection. The second reference scheme uses a network selection mechanism the assignment probabilities are adapted to the traffic loads of different services. Vertical handoff between the cell and WLAN is taken into account but only limited to the boundary of the WLAN coverage. The performance of the two reference schemes is evaluated by the analytical approach .On the other hand, for the load balancing vertical handoff is also performed when a user moves within the WLAN coverage so as to relieve temporary congestion and maximize utilization. Computer simulation is used to evaluate its performance due to difficulties in analytical modeling.

ACKNOWLEDGMENT

The authors would like to thank Prof. Abhijit Patankar & Prof. M.A.Potey for his valuable suggestions.

REFERENCES

[1] 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.

[2] T. Henderson, D. Kotz, and I. Abyzov, "The Changing Usage of aMature Campus-Wide Wireless Network," Proc. ACM MobiCom, pp. 187-201, 2004.

[3] T. Togo, I. Yoshii and R. Kohno, "Dynamic Cell-Size ControlAccording 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.

[4] A. Jalali, "On Cell Breathing in CDMA Networks," Proc. IEEE Int'lConf. Comm. (ICC ' 98), pp. 985-988, 1998.

[5] 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

[6] 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.

[7] 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.

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

[9] 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.

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

[11] 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.

[12] 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.

◈◈◈