Dynamic Load Balancing Algorithm Based on Users Immigration in Wireless LANs

Sardor Q. Hojiev and Dong-Seong Kim

Abstract—This paper discusses a throughput improvement based dynamic load balancing approach for wireless LANs. The proposed scheme checks each AP's load state and balances overloaded APs by immigrating some wireless stations towards under loaded one, periodically. Each wireless stations has different throughput with different APs, respectively. Immigration can be accomplished with considering throughput value between wireless station and under loaded APs. The system overload is uniformly distributed among under loaded cells that wireless stations from overloaded cell moves to most under loaded cell first. The load balancing algorithm mostly considers throughput improvement in imbalance state of the system. Simulation results show that proposed load balancing yields significant performance gains in terms of system throughput and data loss due to overloading.

Index Terms—Dynamic load balancing algorithm, immigration, wireless local area networks.

I. INTRODUCTION

The intake in popularity of IEEE 802.11 wireless local area networks (WLANs) in recent years has been remarkable. WLANs are now the most popular technology used to provide broadband access to IP networks such as extended home networks and internet access in public locations [1], [2]. The proliferation of WLANs has resulted in an ever-increasing number of end-users with heterogeneous quality of service (QoS) requirements. In addition these users tend to gather in certain areas of the network for various reasons such as availability of favorable network connectivity, nearness to power sockets and coffee shops [3]. Such behavior leads to congestion at particular areas within the network. Such congestion creates an unbalanced load in the network and reduces overall network throughput.

A WLAN typically provides a number of Access Points (APs) that provide service to users in a particular geographical area. Users select access points based on the strongest received signal strength indicator (RSSI) [1].

As users, typically, not evenly distributed, 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. Dynamicity of workload traffic from wireless stations and heterogeneity of access points (AP) lead to network conjunction and overloading in wireless networks.

The common method to increase the performance of

WLAN is rate adaptation and configuring a new routing algorithm [4]. Another method that has been introduced to overcome congestion and an overloading in WLAN is to configure a new load balancing algorithm.

The objective of this paper is to provide efficient algorithm for solving the WLAN load balancing problem: distribute users amongst a set of APs to maximize the system throughput.

In this paper novel load balancing scheme is presented that reduces the load of congested APs by forcing the users located near the boundaries of congested cells to move to neighboring less congested cell. User's immigration is achieved via comparing throughput improvement. Amount of migration (load) to under loaded APs. To provide fairness among APs, new parameter: *migration amount* is applied. Based on migration amount of each AP, system overload can be distributed evenly.

Related Works

Several algorithms have been proposed as load balancing technique. Reference [5] proposed distributed load balancing approach (DLBA) for industrial wireless networks. To measure load of APs deadline miss ratio (DMR) is used. If DMR is over the given threshold then AP is overloaded and workstations associated to the AP should find new AP. One of the limitations of this work is workstations of overloaded AP choose new AP with better signal strength. But signal strength and capacity of AP are not related each other, so this approach is not effective. Algorithm proposed in [6] is a centralized load balancing approach. All APs are controlled by central controller. Centralized coordination requires huge processing capability as well as exchange of vast amounts of information among all users in a geographical area and the coordinating entity. Reference [7] proposed load balancing technique by controlling the size of WLAN cells, which is conceptually similar to cell breathing in cellular networks. The algorithm only controls transmission power of AP beacon massages which can decrease only number of associations but not surely workload induced on AP.

Load balancing definition is analyzed in [8]. To maximize the overall system throughput, hosts should be moved towards APs with fewer loads. A system provides a fair service if all clients have the same allocated bandwidth. Unfortunately, this fairness level can cause a significant network throughput reduction. Bandwidth allocation respects "max-min fairness" if there is no way to increase bandwidth for a specific host without decreasing bandwidth of other hosts.

In [9] Velayos *et al.* introduced a distributed load balancing architecture where the AP load is defined as the aggregated downlink and uplink traffic through the AP.

Manuscript received September 11, 2014; revised February 17, 2015. The authors are with Kumoh National Institute of Technology (KIT), South Korea (e-mail: hsqarlot88@gmail.com, dskim@kumoh.ac.kr).

Reference [10] proposed a distributed traffic-balancing routing protocol for multi-sink industrial wireless sensor networks. Each node has a gradient field that is used to decide on a neighbor node to reach a sink. The key objective of this work is to achieve traffic-balancing by detecting congested areas along the route and distributing packets along paths that have idle and under loaded nodes. This traffic balancing approach chooses path to distribute packets based on predefined gradient values. Predefining workload parameters is one of the behaviors of the static load balancing.

Reference [11] proposed a linear-programming approach for dynamic load balancing in CDMA networks. The linear program characterizes the minimum achievable base station load for a given configuration of mobiles at each time interval, and gives benchmark for the potential gains from optimizing the power assignment.

Reference [12] examines the load balancing problem, along with state-of-the-art network and wireless station based solutions. It presents experimental results using off the shelf IEEE 802.11 devices. It addresses two main issues: how to define measure load metrics and how to distribute overall traffic among available APs.

Most of the loads balancing approaches mentioned above do not give brief formulation for load balancing metrics such as: throughput, channel utilization, e.g. To overcome these issues this paper gives brief formulation for throughput optimization.

In heterogeneity system, same workstation can have different throughputs with different APs. Considering the feature of the heterogeneous system, the proposed traffic balancing approach moves workstations towards an under loaded AP with better throughput. The proposed idea is based on assumptions that, traffic overloading can eventually occur at any AP. The overloading state can be defined by comparing the load of the AP and two level thresholds. In the case of overloading, wireless stations are migrated to under loaded APs to balance state of the overloaded AP. The overloaded AP distributes its overload among under loaded APs by immigrating some wireless stations associated. In order to keep fairness the overload is evenly distributed. Before migration, allocation amount (load) for each under loaded AP is defined. Only migrations with throughput improvement are allowed.

This paper is divided into four sections. Section II describes the proposed load balancing scheme including throughput formulation and periodic decision making algorithm. Simulation setup and comparison results are given in Section III. Finally, Section IV with summary of the paper and future works concludes the paper.

II. PROPOSED LOAD BALANCING ALGORITHM

A. System Model

The proposed system model consists of APs (access points) and WSs (wireless stations) as shown in Fig. 1. APs can exchange network information through the wireless backbone. Wireless stations can be located one or more APs coverage areas. A denotes set of APs and let m denotes number of APs in the set A, i.e., m=|A|. At any given time a wireless station can be associated with one AP. U denotes

set of wireless stations within the network and let n=|U|denotes the number of wireless stations. Assume that each wireless station $u \in U$ has a weight that specifies its priority. Load of AP is aggregated load of all users connected to it. Assume there are *n* APs with load $y_1, y_2, ..., y_n$ responsibly. One or more wireless station can associate with AP at the same time with fractional association matrix $X = \{x_{a,u} | a \in A \land u \in U\}$. The load induced by user *u* on AP *a* is the time that is required of AP *a* to provide wireless station *u* a traffic volume of size $x_{a,u} \cdot \omega_u$ [5]. Consequently, we define the load, on AP *a* o be the period of time that AP *a* to provide a traffic volume of size $x_{a,u} \cdot \omega_u$ to all its associated wireless stations $u \in U$.



Fig. 1. System model.

The traffic load of AP $a \in A$ denoted by y_a is the sum of load induced by all associated wireless stations, that is,

$$y_a = \sum_{u \in U_a} y_{a,u} = \sum_{u \in U_a} \frac{x_{a,u} \cdot \omega_u}{r_{u,a}},$$
 (1)

where ω_u is workload of wireless station of u and $r_{u,a}$ is bit rate for wireless station u by AP $a \in A$.

B. Throughput Formulation

In this subsection we show the throughput measurement based on AP selection algorithm to optimize the system throughput. Throughput is the number of packets transmitted successfully in a given time period. The system throughput is aggregated throughput of all APs.

Throughput of AP *a* denoted by q_a , is the number of successfully transmitted packets in a unit time [13], that is,

$$q_a = \frac{\sum_{u \in U_a} (c_{a,u} \cdot x_{a,u})}{\sum_{u \in U_a} y_{a,u}} = \frac{c_a}{y_a},$$
(2)

where C_a is the number of successfully transmitted packets in a period of y_a time.

Throughput of the AP is determined with its load and fractional association of connected wireless stations. $C_{a,u}$ is the number of transmitted packets of wireless station $u \in U_a$ by AP *a*. Therefore, optimized throughput problem can be defined as:

$$\max \sum_{a \in A} q_a \tag{3}$$

As mentioned before an objective of the proposed algorithm is to maximize system throughput by immigrating

users (wireless stations) from hot spot areas to under loaded pools. The immigration is accomplished only based on throughput improvements of the user to be migrated into new AP's coverage area. To define throughput improvement, in this scheme we used same formulation approach proposed in [13]. Algorithm proposed in [13] makes a decision for migration based on throughput improvement that if throughput contribution of user to be immigrated with new AP is positive then the user is moved towards to the new AP. Some notations used in algorithm are given in Table I.

Assume wireless station *u* belongs to AP *k* is moving to AP *j*. Before migration system throughput is $Q_s = \sum_{a \neq j,k} q_a + q_k + q_j$, where throughput of AP *j* and *k* can be represented as:

$$q_{k} + q_{j} = \begin{cases} \frac{C_{k}}{y_{k}} + \frac{C_{j}}{y_{j}}, & \text{if } y_{j} \neq 0, \\ \frac{C_{k}}{y_{k}}, & \text{if } y_{j} = 0. \end{cases}$$
(4)

After migration, the system throughput is changed to: $Q_s^* = \sum_{a \neq j,k} q_a + q_k^* + q_j^*$, where the new throughput of the AP k and j is:

$$q_k^* + q_j^* = \frac{c_k - c_{k,u}}{y_k - c_k \cdot y_{k,u}} + \frac{c_j - c_{j,u}}{y_j - c_j \cdot y_{j,u}}$$
(5)

Difference value $D_{k,j,u}$ between $q_k^* + q_j^*$ and $q_k + q_j$ gives us throughput contribution of workstation *u* to be migrated. As mentioned before the scheme proposed in [13] allow immigrations if throughput improvement is positive.

$$D_{k,j,u} = q_k^* + q_j^* - (q_k + q_j), \qquad (6)$$

$$l_{j,u} = y_j^* - y_j,$$
 (7)

where y_j^* and $l_{j,u}$ are new traffic load of AP *j* and load contribution of new added wireless station *u*, respectively.

During the load balancing process an overloaded AP k checks the throughput contribution (improvement) $D_{k,j,u}$ of wireless stations u and traffic load at the AP j. The AP j is suitable destination for wireless stations u only with following conditions:

$$0 < l_{j,u} < (T_h - y_j), \tag{8}$$

and migration should improve throughput

$$D_{k,j,u} > 0$$
. (9)

As a contribution we add a new constraint allowing for immigrations. A newly arriving user u to AP j's cell (coverage area) will be admitted if and only if the throughput improvement and load induced at AP satisfy requirements given by (8) and (9).

C. Periodic Decision Making

The proposed load balancing algorithm periodically checks each AP's load state using two level thresholds. Higher and lower thresholds can be defined as below:

$$T_h = Y_{av} + \sigma \cdot C , \qquad (10)$$

$$T_l = Y_{av} - \sigma \cdot C \,. \tag{11}$$

where Y_{av} is a average load for each AP and it can be defined with below equation. Value σ is variance of loads of APs and *C* is a coefficient, $C \in [0,1]$.

$$Y_{av} = \frac{\sum_{a \in A} y_a}{m} \tag{12}$$

If load of AP is greater than the upper threshold T_h then the AP is overloaded, less than the lower threshold T_l is under loaded.

 $y_a > T_{h_a}$ then $AP_a \Rightarrow OverList$, (13)

$$y_a < T_h$$
, then $AP_a \Rightarrow$ UnderList. (14)

Fig. 2 illustrates *Sorting* sub process of APs. APs are sorted based on (13) and (14) into overloaded and into under loaded. Each overloaded AP distributes it over loads among under loaded APs uniformly. Load distribution is accomplished by migrating wireless stations towards under loaded APs. Before distribution each overloaded AP defines *allocation amount* for each under loaded AP. The allocation amount is illustrated in Fig. 3 and defined by following steps.

Firstly, APs in *UnderList* APs and *OverList* must be sorted based on ascending and descending order of their loads.

In this algorithm we sorted under loaded and overloaded APs in ascending and descending order, respectively:

Under List = { $AP_{un}(1), AP_{un}(2), ..., AP_{un}(j)$ } in ascending order, and Over List = { $AP_{ov}(1), AP_{ov}(2), ..., AP_{ov}(k)$ } in descending order.

Assume AP k is overloaded and its overload $L_{ov}(k)$ can be defined by

$$L_{ov}(k) = T_h - y_k \,. \tag{15}$$

At the same time $L_{ov}(k)$ overload is migration amount for AP k.

The AP k computes an under load $L_{un}(j)$ of each AP j in *UnderList* as follow:

$$L_{un}(j) = T_l - y_j$$
. (16)

The AP k sorts under loaded APs in ascending order of their under loads, and calculates *allocation amount* for each under loaded AP as shown in Fig. 3.

Newly added constraint, *allocation amount* for migration (induced by users to be migrated) insures that system overload is distributed among under loaded APs, fairly.

Sub sums of under loads $S(t)(2 \le t \le j)$ are defined as follows:

$$S(t) = \sum_{i=1}^{t} (L_{un}(t) - L_{un}(i)).$$
(17)

Next, AP k (which is overloaded) finds index t that satisfies following condition:

$$S(t) < L_{ov}(k) \le S(t+1).$$
 (18)

From the Fig. 3, *migration amount* (overload) for AP k can be defined as

$$L_{ov}(k) = S(t) + x \cdot t , \qquad (19)$$

which can be rewritten as follows:

$$x = \frac{L_{ov}(k) - S(t)}{t}.$$
 (20)

For each under loaded AP i ($i \le t$), allocation amount (load) $Alloc_k(i)$ from AP k can be defined by

$$Alloc_k(i) = x + L_{un}(t) - L_{un}(i)$$
. (21)

From (20) and (21), the allocation amount for AP i is

$$Alloc_{k}(i) = \frac{L_{ov}(k) - S(t)}{t} - L_{un}(t) - L_{un}(i) .$$
(22)

The overloaded AP k forwards $Alloc_k(i)$ amount of load to under loaded AP i for $1 \le i \le t$. As mentioned before load forwarding is accomplished by migrating WSs from overloaded AP towards under loaded APs. To improve the system throughput only migrations which meet the condition defined by (9) are allowed.

After defining allocation amounts for each under loaded AP, overloaded AP k chooses wireless station and gives *id* of the chosen under loaded AP *i*. Based on the *id* the wireless station associates with AP *i*. Decision making and migration of wireless stations are shown in Fig. 4.



Fig. 2. Flow chart of sorting APs.

TABLE I: NOTATIONS FOR PARAMETERS USED IN ALGORITHM

Parameters	Definition
Y _m	Load of AP <i>m</i>
T_h	Upper Threshold
T_l	Lower Threshold
Over List	List of Overloaded APs
Under List	List of Under loaded APs
$Alloc_k(j)$	Allocation amount for AP j
$D_{k,j,u}$	Throughput contribution of wireless station when moves from AP <i>k</i> AP <i>j</i>
$l_{j,u}$	Imposed load of wireless station at AP j
$L_{ov}(k)$	Overload of AP k



Fig. 3. Measuring immigration amount for under loaded APs.



Fig. 4. Flowchart of decision making process.

III. SIMYLATION

The proposed load balancing algorithm was simulated in OPNET Modeler 16.0 to measure and compare network performance with and without load balancing schemes. The proposed load balancing algorithm runs inside each AP in a distributed manner. Every 10 seconds the proposed scheme balances the system load. Implemented scenario is built in the area of 50×50 , Fig. 5. Initially, 70 workstations associated with 7 APs, randomly. In order to demonstrate the efficiency, the proposed traffic load balancing algorithm is compared with DLBA algorithm proposed by [5]. Table II shows simulated system setup. During the simulation we made some hot spot APs with two methods: by moving wireless stations to some APs coverage area and by increasing data arrival rate from wireless stations.

The study aimed to analyze the performance of the proposed algorithm in terms of throughput and data lost. Throughput represents the successfully transmitted bits in a given time period and data dropped is the number of lost bits in terms of bit rate (bits/sec).

TABLE II: SIMULATION PARAMETERS

Parameters	Definition
Physical Characteristics	PHY (802.11e)
Data Rate (bps)	11 Mbps
Transmit Power	0.0032 W
AP Beacon Interval (sec)	0.02
Number of APs	7
Number of Wireless stations	70
Load Balancing period (sec)	10



Fig. 5. Simulation setup.





Fig. 7. Comparison results on data loss.

Fig. 6 shows that the proposed load balancing algorithm outperformance the DLBA scheme in terms of throughput. This caused by allowing wireless station migrations with throughput improvement.

Uniformly distribution of overload leads to a reduction in packet (data) loss. As shown in Fig. 7, the proposed algorithm has better performance than the DLBA scheme.

IV. CONCLUSIONS

In this paper, a dynamic load balancing approach is proposed for WLANs. The approach lies on to keep system balanced by fairly distributing load traffic among APs. The proposed traffic load balancing approach finds imbalanced APs and moves workstations towards under loaded APs, periodically. Traffic distribution is accomplished based on immigration of workstations towards under loaded APs with better throughput improvements. Throughput improvement of workstations serves to increase the total throughput of system.

The approach introduced in this paper is restrictive in some aspects. For instance, it is assumed that all clients generate packets at the same rate. As a performance metric, only throughput is considered. As future work, the proposed model will be extended for more complicated and realistic scenarios.

REFERENCES

- E. Villegas, R. V. Ferr, and J. P. Aspas, "Load balancing in wireless LANS using 802.11k mechanisms," in *Proc. IEEE Symposium on Computers and Communications*, 2006, pp. 844–850.
- [2] D. S. Kim *et al.*, "Design and implementation of home network systems using UPnP middleware for networked appliances," *IEEE Transactions on Consumer Electronics*, vol. 48, pp. 963-972, 2002.
- [3] A. Balachandran *et al.*, "Characterizing user behavior and network performance in a public wireless LAN," in *Proc. SIGMETRICS Performance Evaluation Review*, 2002, pp. 195–205.
- [4] D. A. Nugroho and D. S. Kim, "Dynamic rate adaptation for industrial WLAN," in *Proc. ICT Conver. Conf.*, 2007, pp. 575-580.
- [5] G. P. Collotta, V. M. Salerno, and G. Scata, "Distributed load balancing approach for industrial IEEE 802.11 wireless networks," in *Proc. 18th IEEE International Conf. on Emerging Technologies and Factory Automation*, Krakow, 2012, pp. 1-7.
- [6] M. Collotta, L. L. Bello, E. Toscano, and O. Mirabella, "Dynamic load balancing techniques for flexible wireless industrial networks," in *Proc. IECON 36th Annual Conference on IEEE Industrial Electronics Society*, 2010, pp. 1329-1334.
- [7] Y. Bejerano and S. J. Han, "Cell breathing techniques for load balancing in wireless LANS," *IEEE Transactions on Mobile Computing*, vol. 8, issue 6, pp. 735-749, June 2009.
- [8] Y. Bejerano, S.-J. Han, and L. Li, "Fairness and load balancing in wireless lans using association control," *IEEE/ACM Transactions on* Networking, pp. 560–573, 2007.
- [9] V. A. H. Velayos and G. Karlsson, "Load balancing in overlapping wireless LAN cells," in *Proc. IEEE Int'l Conf. Communications*, Paris, 2004, pp. 3833-3836.
- [10] D. D. Tan, A. Prasetiadi, and D. S. Kim, "Traffic-balancing routing scheme for industrial wireless sensor networks," in *Proc. International Conf. on ICT Convergence*, 2012, pp. 595–600.
- [11] S. C. Borst, G. Hampel, I. Saniee, and P. A. Whiting, "Load balancing in cellular wireless networks," *Handbook of Optimization in Telecommunication*, Heidelberg: Springer, 2006.
- [12] K. H. C, L. H. Yen, and T. T. Yeh, "Load balancing in IEEE 802.11 networks," *IEEE Internet Computing*, pp. 57-64, 2009.
- [13] J. W. Yanxiao et al., "Throughput measurement-based access point selection for multi-rate wireless LANs," in Proc. 4th International Conf. Wireless Algorithms, Systems, and Applications, Boston, 2009, pp. 509–518.



Sardor Qurbonboyevich Hojiev was born in Olot Province, Uzbekistan, in 1988. He received the B.S. degree from Tashkent University of Information Technologies (TUIT), Tashkent, Uzbekistan, in 2010 and the M.S. degree from the Kumoh National Institute of Technology (KIT), South Korea, in 2013. He is currently a full-time researcher at Networked Systems Laboratory, KIT. His research interests include distributed systems, cloud computing, load wireless networks.

balancing, clustering, wireless networks.



Dong-Seong Kim received his Ph.D. degree in electrical and computer engineering from the Seoul National University, Seoul, Korea, in 2003. From 1994 to 2003, he worked as a full-time researcher in ERC-ACI at Seoul National University, Seoul, Korea. From March 2003 to February 2005, he worked as a postdoctoral researcher and visiting scholar at the Wireless Network Laboratory in the School of Electrical and Computer Engineering at Cornell

University, NY. He was a visiting professor with the Department of Computer Science, University of California, Davis, U.S.A. Since 2004, he has been a professor in the School of Electronic Engineering, a director of ICT-CRC and mobile research center at Kumoh National Institute of Technology, Korea. His current main research interests are industrial wireless control network, networked embedded system, and Fieldbus.