The impact of association on the capacity of WLANs

Thomas Bonald, Ali Ibrahim, James Roberts

HAL Id: hal-01244788
https://hal.science/hal-01244788

Submitted on 16 Dec 2015

The Impact of Association
on the Capacity of WLANs

Thomas Bonald, Ali Ibrahim, James Roberts

Orange Labs

38-40 rue general Leclerc, Issy-les-Moulineaux, France

{thomas.bonald,ali.ibrahim,james.roberts}@orange-ftgroup.com

Abstract—This paper contributes to the definition of an association policy for a multi-channel, multiple AP WLAN that depends on both physical rate and realised throughput. We show that existing proposals are inefficient when several APs share the same channel and define a new policy that is demonstrated to be close to optimal. Policies are compared through their traffic capacity defined as the network stability limit. We determine this capacity analytically using the fluid limit method for some simple network configurations deriving insight into the structure of the optimal policy. The quasi-optimality of our proposal is verified by simulation of a more complex network configuration.

I. INTRODUCTION

We consider a network where users are able to connect to the Internet via one of a set of WLAN access points (AP). It is well-known that the way user stations choose the AP with which they associate has a significant impact on realized performance. In particular, the usual approach where users associate with the AP having the largest Received Signal Strength Indication (RSSI), offering therefore the highest physical rate, can lead to performance degradation due to traffic imbalance.

Several authors have advocated that users should instead associate with the AP offering greatest throughput given current occupancy and propose techniques to estimate this metric [2], [9], [15], [16]. It has been observed, however, that this locally optimal policy can prove globally inefficient since the new user may unduly reduce the throughput of the users that are already active. This occurs because the IEEE 802.11 MAC protocol ensures all users associated with an AP realize the same throughput. If one user has a low physical transmission rate (because it is far from the AP), it will occupy the AP for a disproportionate amount of time reducing the throughput of the other users.

Throughput degradation is taken explicitly into account in the association metric proposed in [1]. This metric is a weighted sum of the throughput provided to the new user and the average lost transmission time of users that are already associated with the AP in question.

In [3] the authors propose a hybrid dynamic association approach applicable also to mesh networks. Users first choose to associate with the AP offering maximum throughput but can later change to the AP with maximum signal strength if a measure of available bandwidth drops below a certain threshold.

The proposal in [11] is to choose the AP which maximizes a particular linear combination of rate and throughput. This is a relatively simple criterion derived empirically from the results of an investigation into the nature of an optimal policy performed using dynamic programming [10].

The three approaches are in fact very similar in principle and all provide significant performance improvement compared to using either signal strength or throughput metrics alone. In the present work we build on the association policy proposed in [11].

Our objective is to maximize the capacity region of the set of APs under a dynamic traffic model. Users arrive at a point within the network coverage area, associate with an access point, realize a download and leave. The arrival process is Poisson and, for simplicity of presentation, the size of the download has an exponential distribution. Like [11] we ignore upstream traffic assuming this is of secondary importance for user - AP association. Network capacity is defined by the limiting load beyond which the number of active users would grow indefinitely as the arrival rate in some region is greater than the rate at which downloads complete.

Unlike [11], we assume several APs can share a common frequency channel, limiting scope for load balancing and leading to non-optimality of the simple rate and throughput policy. We propose therefore the following association scheme: for each frequency, users elect the AP offering the highest RSSI; they then choose one among this subset of APs by applying the metric from [11].

We analytically derive stability conditions for the considered association policies for some simple network configurations using the well-known fluid limit approach and drawing on results of our previous analysis of multi-cell WLANs [6], [7]. The analysis of these simple networks provides insight into why and how rate and throughput metrics must be combined to realize an efficient association policy. We are able to characterize the optimal policy for a particular class of networks and explain why our simple practical approach turns out to be quasi-optimal.

II. MODELLING ASSOCIATION POLICIES

We introduce the considered multi-cell WLAN configuration and define the evaluated association policies. The traffic model and the notion of network capacity through which we compare the policies are described.

A. Access points

We consider a set of N access points (AP) indexed by i. These APs use F independent frequency channels. We denote by f_i the frequency channel of AP i. A dynamic number of users compete for available bandwidth. Specifically, each user associates with an AP before downloading data, and remains associated with this AP during the whole data transfer. We are interested in the impact of the association policy on capacity, measured in terms of maximum traffic intensity.

The region covered by the N access points is divided into an arbitrary set of M sub-regions such that any two users belonging to the same sub-region experience the same radio conditions. We refer to such a sub-region as a class and index classes by j. Any class-j user associated with AP i receives data at rate R_{ij} when served[1]. The network state is characterized by the numbers x_{ij} of class-j users associated with AP i, for all $i = 1, \ldots, N$ and $j = 1, \ldots, M$. We denote the corresponding vector by x. Let $x_i = \sum_j x_{ij}$ be the number of users associated with AP i in state x. We say that AP i is active if $x_i > 0$.

B. Resource sharing

We assume that all APs using the same frequency channel interfere with each other so that any two such APs cannot transmit simultaneously. This is representative of dense networks. Networks with multiple collision domains will be considered in future work. We also assume that all users use a common packet size. Since all APs sharing the same frequency channel are equally likely to transmit during idle periods of the medium, the throughput of each active AP using frequency channel f, expressed in packets per time unit, is given by:

$$\frac{1}{\sum_{i:f_i=f,x_i>0} \tau_i},\tag{1}$$

where τ_i is the average transmission time of a packet by AP i when active.

Now assume all users associated with the same AP get the same throughput. This is typically the case for TCP flows if the AP is the only bottleneck, or in the presence of a round-robin scheduler at the AP. The average transmission time of a packet by AP i is then given by:

$$\tau_i = \sum_j \frac{P_{ij}}{R_{ij}},\tag{2}$$

where P_{ij} is the probability that AP i serves a class-j user when active:

$$P_{ij} = \frac{x_{ij}}{x_i}.$$

Using (1) and (2), we deduce the throughput of any user associated with AP i:

$$T_i = \frac{1}{\sum_{k:f_k=f,x_k>0} \sum_j \frac{P_{kj}}{R_{kj}}} \times \frac{1}{x_i}.\tag{3}$$

C. Association policies

We consider the association policies based on the following criteria:

- **Rate:** An incoming user associates with the AP that provides the highest peak rate, i.e., a class-j user associates with AP $i^\star = \arg\max_i R_{ij}$. Equivalently, each user chooses the AP that offers the best signal-to-noise ratio. This is typically the policy implemented in current systems.
- **Throughput:** An incoming user associates with the AP that provides the highest throughput, i.e., a class-j user associates with AP $i^\star = \arg\max_i T_i$. The throughput received from each AP is assumed to be estimated over a short time period preceding the actual data transfer.
- **Rate-throughput:** An incoming user associates with the AP that provides the highest linear combination of rate and throughput, i.e., a class-j user associates with AP $i^\star = \arg\max_i(\alpha R_{ij} + \beta T_i)$, where α and β are fixed parameters. This is the scheme introduced in [10], [11]. The scheme actually only depends on the ratio $\gamma = \alpha/\beta$.
- **Rate plus rate-throughput:** For each frequency channel f, users choose the AP that offers the highest peak rate, i.e., AP $I(f) = \arg\max_{i:f_i=f} R_{ij}$ for a class-j user. An incoming user then selects the frequency channel that provides the highest linear combination of rate and throughput, i.e., a class-j user selects frequency channel $f^\star = \arg\max_f(\alpha R_{I(f)j} + \beta T_{I(f)})$, where α and β are fixed parameters. Thus a class-j user associates with AP $i^\star = I(f^\star)$.

In the following, these policies are referred to as R, T, RT and R2T, respectively. Note that the proposed R2T policy is equivalent to the R policy if all APs share the same frequency channel and to the RT policy if all APs use different frequency channels.

D. Traffic characteristics

Users are assumed to become active at random, according to a Poisson process of intensity λ. The file size distribution is exponential[2] with parameter μ. We refer to the traffic intensity as the quantity $\rho = \lambda/\mu$ (in bit/s). An incoming user is of class j with probability p_j, with $\sum_j p_j = 1$. Class-j traffic intensity is denoted by $\rho_j = \rho p_j$.

The evolution of the network state x defines a Markov process with the following transition rates:

$$
\begin{aligned}
x_{ij} \rightarrow x_{ij} + 1 &\quad : \quad \lambda p_j 1\{i^\star = i\} \\
x_{ij} \rightarrow x_{ij} - 1 &\quad : \quad \mu T_i x_{ij}
\end{aligned}
$$

where $i^\star$ depends on the association policy, as described above. It is the stability behaviour of this Markov process that we use to characterize network capacity.

---

[1]This is a net throughput that is assumed to include all overheads: MAC headers, MAC acknowledgment, backoff timers, etc.

[2]This assumption simplifies the description of the stochastic process but is not critical for the stability issues addressed in the paper.

E. Network capacity

We define network capacity as the maximum traffic intensity ρ such that the above Markov process is ergodic. This is the maximum traffic intensity sustainable by the network without any overflow at any AP.

We refer to the capacity region as the set of vectors of traffic intensities $(\rho_1, \ldots, \rho_M)$ such that the Markov process is ergodic. This will be used to analyse the impact of traffic distribution $(p_1, \ldots, p_M)$ on capacity.

III. CAPACITY OF SINGLE CLASS NETWORKS

In this section, we provide analytical results in the case of two APs with a single user class, as shown in Figure 1. This simple scenario turns out to provide very useful insight into the efficiency of the considered association policies. Specifically, we successively consider the case of a common frequency channel and the case of two different frequency channels and show that no association policy except R2T yields nearly maximum capacity in both cases. For convenience, we drop the index of class from all notations. We let $R_1 < R_2$ by convention.
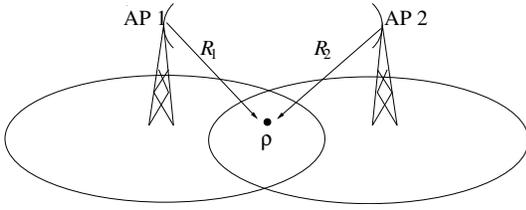

Fig. 1. Two access points with a single user class.

A. A common frequency channel

Assume that both APs use the same frequency channel. In view of (3), the total throughput of each active AP is given by:

$$x_1 T_1 = \begin{cases} R_1 & \text{if } x_2 = 0, \\ R_0 & \text{otherwise,} \end{cases} \qquad x_2 T_2 = \begin{cases} R_2 & \text{if } x_1 = 0, \\ R_0 & \text{otherwise,} \end{cases}$$

where R_0 is the common throughput of each AP when both are active:

$$R_0 = \frac{1}{1/R_1 + 1/R_2}.$$

We denote by $\bar{R} = 2R_0$ the total network throughput when both APs are active, which corresponds to the harmonic mean rate. We have the following results for the R, T and RT schemes:

- **Rate:** All users associate with AP 2. The stability condition is $\rho < R_2$ so that capacity is maximal, equal to R_2.
- **Throughput:** An incoming user associates with AP1 if and only if $x_2 \geq x_1 > 0$ or $x_1 = 0$ and $x_2 \geq R_2/R_1$. The stability condition is given by Proposition 1 below. Capacity is equal to $\bar{R}$, which is less than R_2; the use of AP 1 (which is idle under the R association policy) reduces capacity.

Proposition 1: Under the T association policy, the stability condition is $\rho < \bar{R}$.

This and all other propositions of this and the next section are proved in the appendix.

- **Rate-throughput:** An incoming user associates with AP1 if and only if:

$$x_1 > 0, \quad \frac{1}{x_1 + 1} \geq \frac{1}{x_2 + 1} + \frac{R_2 - R_1}{\gamma R_0}$$

$$\text{or} \quad x_1 = 0, \quad x_2 \geq \frac{\gamma R_2}{\gamma R_0 + (R_1 - R_2)} - 1.$$

The stability condition is given by Proposition 2. Capacity grows from $\bar{R}$ to R_2 as γ increases from 0 to ∞.

Proposition 2: Let $K = \lceil \gamma R_0/(R_2 - R_1) - 1 \rceil$. Under the RT association policy, the stability condition is given by:

$$\left(\frac{\rho}{R_0}\right)^{K+1} < \frac{R_2}{R_0} + \frac{\rho}{R_0} + \left(\frac{\rho}{R_0}\right)^2 + \ldots + \left(\frac{\rho}{R_0}\right)^K.$$
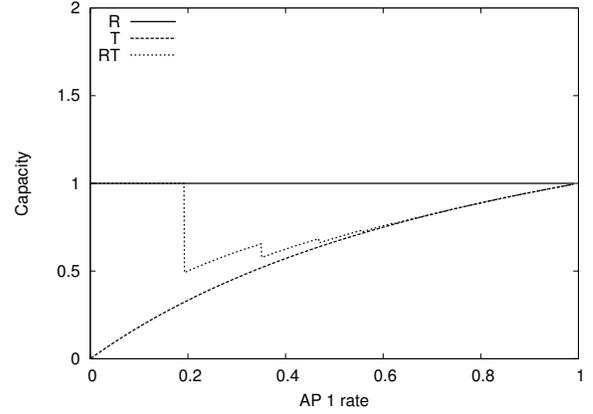

Fig. 2. Capacity with respect to AP-1 rate R_1 for a single frequency channel when $R_2 = 1$.

The above results are illustrated by Figure 2 for $R_2 = 1$ and $\gamma = 5$, the value recommended in [10]. Note that RT is suboptimal whenever R_1 is higher than a certain value around 0.2, beyond which K is positive. The R2T scheme, which is equivalent to R in this case, is optimal.

B. Two channels

If the APs use different frequency channels, they become independent and have total throughputs when active of R_1 and R_2, respectively. We have the following:

- **Rate:** All users associate with AP 2. The stability condition is $\rho < R_2$ so that capacity is equal to R_2, which is suboptimal since only one frequency channel is used.
- **Throughput:** An incoming user associates with AP1 if and only if

$$\frac{R_1}{x_1 + 1} > \frac{R_2}{x_2 + 1}.$$

The stability condition is given by Proposition 3. Capacity is maximal and equal to $R_1 + R_2$.

Proposition 3: Under the T association policy, the stability condition is $\rho < R_1 + R_2$.

- **Rate-throughput:** An incoming user associates with AP1 if and only if:

$$\frac{\gamma R_1}{x_1 + 1} - \frac{\gamma R_2}{x_2 + 1} \geq R_2 - R_1.$$

The stability condition is given by Proposition 4. Capacity decreases from $R_1 + R_2$ to R_2 as γ increases from 0 to ∞.

Proposition 4: Let $L = \lceil \gamma R_1/(R_2 - R_1) - 1 \rceil$. Under the RT association policy, the stability condition is:

$$\left(\frac{\rho}{R_1}\right)^{L+1} < \frac{R_2}{R_1}\left(1 + \frac{\rho}{R_1} + \ldots + \left(\frac{\rho}{R_1}\right)^L\right).$$
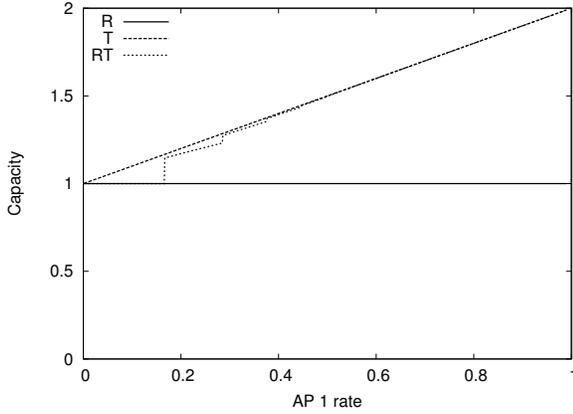

Fig. 3. Capacity with respect to AP 1 rate R_1 for two frequency channels when $R_2 = 1$.

The above results are illustrated by Figure 3 for $R_2 = 1$. The RT scheme, which coincides with R2T in this case, is quasi-optimal whenever R_1 is greater than a certain value around 0.3, beyond which L is larger than 1.

IV. CAPACITY OF MULTI-CLASS NETWORKS

We now consider the case of two classes, as illustrated by Figure 4. By convention, we assume that class-i users have a higher transmission from AP i, for $i = 1, 2$, i.e., $R_{11} > R_{21}$ and $R_{22} > R_{12}$. The objective is to analyse the impact of heterogeneous radio conditions on the efficiency of the considered association policies. A similar scenario is considered in [10] where it is assumed that both APs use different frequency channels.
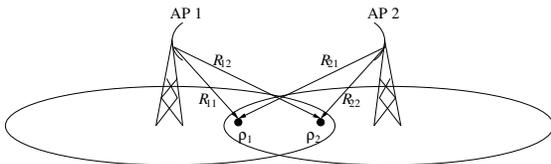

Fig. 4. Two access points with two user classes.

A. A common frequency channel

If both APs use the same frequency channel, their common throughput R_0 when active satisfies:

$$\frac{1}{R_0} = \frac{x_{11}}{x_1 R_{11}} + \frac{x_{12}}{x_1 R_{12}} + \frac{x_{21}}{x_2 R_{21}} + \frac{x_{22}}{x_2 R_{22}}.$$

We have the following results for the R, T and RT schemes:

- **Rate:** Since $R_{11} > R_{21}$ and $R_{22} > R_{12}$, class-1 users associate with AP 1 and class-2 users associate with AP 2. The stability condition is given by $\rho_1/R_{11} + \rho_2/R_{22} < 1$.
- **Throughput:** Assume that $R_{11} > R_{12}$, $R_{22} > R_{21}$ and:

$$\frac{1}{R_{21}} \leq \frac{1}{R_{11}} + \frac{2}{R_{22}}, \quad \frac{1}{R_{12}} \leq \frac{2}{R_{11}} + \frac{1}{R_{22}}. \quad (4)$$

In the symmetric case $R_{11} = R_{22}$, $R_{12} = R_{21}$ for instance, this implies that $R_{21}/R_{11} = R_{12}/R_{22} \geq 1/3$. The stability condition is given by the following result:

Proposition 5: Under the T association policy and assumption (4), the stability condition is given by:

$$\frac{\rho_1}{2}\left(\frac{1}{R_{11}} + \frac{1}{R_{21}}\right) + \frac{\rho_2}{2}\left(\frac{1}{R_{12}} + \frac{1}{R_{22}}\right) +$$
$$\frac{\rho_1\rho_2}{3(\rho_1 + \rho_2)}\left(\frac{1}{R_{11}} + \frac{1}{R_{22}} - \frac{1}{R_{21}} - \frac{1}{R_{12}}\right) < 1.$$

- **Rate-throughput:** The capacity region of the RT policy cannot be stated explicitly as in the previous examples. However, the stability region can be computed numerically using the fluid limit approach described in the appendix. First observe that, in the limit, the RT association policy is equivalent to R when all fluid volumes are positive. The network is therefore stable if

$$\rho_1, \rho_2 < \frac{1}{R_{11}} + \frac{1}{R_{22}},$$

and unstable if

$$\rho_1, \rho_2 > \frac{1}{R_{11}} + \frac{1}{R_{22}}.$$

Now assume

$$\rho_2 < \frac{1}{R_{11}} + \frac{1}{R_{22}} < \rho_1.$$

The fluid volume at AP 2 empties after some finite time. While the fluid volume at AP 1 is positive, the throughput of any incoming class-2 user is null at AP 1; since $R_{12} < R_{22}$, all class-2 users associate with AP 2. The stability condition follows on computing the fraction of class-1 users that associate with AP 1 (which gives the arrival rate at AP 1) and the fraction of time AP 2 is idle (which gives the service rate at AP 1). These two quantities can be evaluated numerically by determining the stationary distribution of the user population at AP 2 that forms a 2-dimensional Markov process.

Figure 5 shows the capacity regions of the R, T and RT association policies for $R_{12} = R_{21} = 1$ and $R_{11} = R_{22} = 2$. Again, the R2T scheme is equivalent to R and is optimal. The capacity of RT is greater than that of T and smaller than that
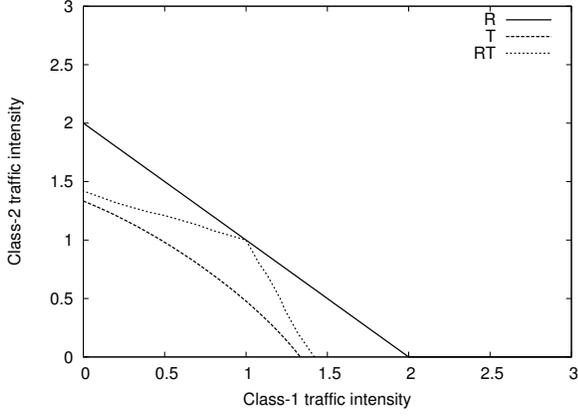
Fig. 5. Capacity region for a single frequency channel.

of R. It coincides with R if the network is symmetric. The reason why T and RT perform badly when both APs use the same channel is that some users needlessly associate with APs that offer lower physical transmission rates.

B. Two channels

In the case of two channels, the stability region cannot be stated explicitly except for the R association policy where it is given by:

$$\frac{\rho_1}{R_{11}} < 1, \quad \frac{\rho_2}{R_{22}} < 1.$$

For the T and RT policies, we can evaluate capacity numerically for given rates R_{ij} using the fluid limit approach, as for a single channel.
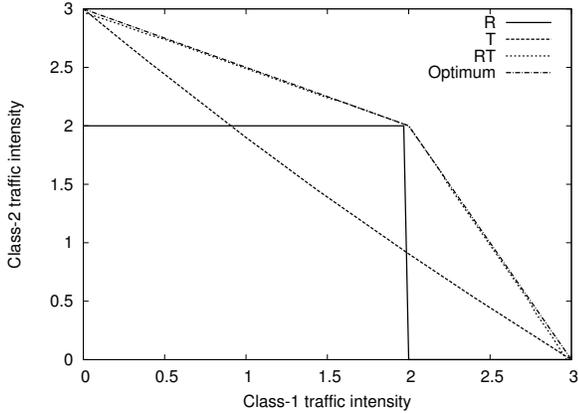

Fig. 6. Capacity region for two frequency channels (right).

Figure 6 shows the capacity regions of R, T and RT as well as the optimal capacity region, derived from the results of Section V, when $R_{12} = R_{21} = 1$ and $R_{11} = R_{22} = 2$. Both R and T are quite far from optimal; T is not better than R unless traffic is highly asymmetric. R2T is equivalent to RT and is quasi-optimal, in agreement with the results of [10]. We discuss the quasi-optimality of the R2T policiy in the next section. In the present configuration, it can be shown

that the difference between RT and the optimal policy decays exponentially with increasing γ.

V. QUASI-OPTIMALITY OF R2T

Results of the previous two sections show that R2T is quasi-optimal in the particular network configurations considered. In this section we characterize the optimal allocation and explain in relation to a further more general configuration why R2T performs so well.

A. Optimal association policy

To determine the optimal stability region of a general multi-channel single collision domain, for each direction of the traffic intensity vector p, we have to solve the following optimisation problem.

$$\begin{cases} \text{Minimise} & \max_{1 \leq f \leq F} \left(\sum_{1 \leq i \leq N, f_i = f} \sum_{1 \leq j \leq M} \frac{\alpha_{ij}\rho_j}{R_{ij}} \right), \\ \text{subject to:} \\ & 0 \leq \alpha_{ij} \leq 1, \quad 1 \leq i \leq N, \quad 1 \leq j \leq M, \\ & \sum_{1 \leq i \leq N} \alpha_{ij} = 1, \quad 1 \leq j \leq M. \end{cases}$$ (5)

where α_{ij} designates the fraction of class-j traffic that associates with AP i. Here we assume the users associate with AP i with probability α_{ij} in any occupancy state. It is clear on considering the corresponding fluid limit that this yields the same capacity as any state dependent policy realizing the same traffic split.

We consider the particular case $F = 2$ with just one AP for each channel. Extension to more than two APs is straightforward.

We show below that the optimal policy consists in starting with an R association and then shifting traffic classes from one AP to the other to achieve optimal load balancing. The crucial question is which classes should be moved. Note that shifting traffic tends to consume more network resources since available rate is necessarily smaller. To preserve resources we should therefore preferentially move the classes for which the relative rate difference is smallest. Proposition 6 confirms this insight.

Proposition 6: Consider a two AP network and assume the APs use different frequency channels. The optimal association rule can be derived as follows:

1) Arrange the traffic classes in decreasing order of R_{1j}/R_{2j}.
2) Temporarily associate classes using R.
3) Balance AP load as follows:
   - If the load of AP 1 is greater than that of AP 2, then give a fraction of AP 1 traffic to AP 2 starting with the classes j that have the lower R_{1j}/R_{2j}.
   - If the load of AP 2 is greater than that of AP 1, then give a fraction of AP 2 traffic to AP 1 starting with the classes j that have the higher R_{1j}/R_{2j}.

Proof: For this special case, problem (5) can be expressed as follows.

$$
\begin{cases}
\text{Minimise} & \sum_{1 \le j \le M} \frac{\alpha_{1j}\rho_j}{R_{1j}}, \\
\text{subject to:} & \\
& \sum_{1 \le j \le M} \frac{\alpha_{1j}\rho_j}{R_{1j}} = \sum_{1 \le j \le M} \frac{\alpha_{2j}\rho_j}{R_{2j}}, \\
& 0 \le \alpha_{ij} \le 1, \quad 1 \le i \le 2, \quad 1 \le j \le M, \\
& \alpha_{1j} + \alpha_{2j} = 1, \quad 1 \le j \le M.
\end{cases} \tag{6}
$$

It is sufficient to show that applying Proposition 6 yields a local minimum of system (6) since the objective function is convex. Ignoring trivial exceptions, the solution of Proposition 6 has the following form. There exists a certain class l with $R_{1l}/R_{2l} > 1$ whose traffic is (strictly) split between the APs and is such that:

$$
\begin{cases}
\alpha_{1j} = 1 & \text{for} \quad 1 \le j < l, \\
\alpha_{1j} = 0 & \text{for} \quad l < j \le M.
\end{cases} \tag{7}
$$

Here we have assumed that, after the R association, AP 1 has the higher load. Consider the $M - 1$ independent variables $\alpha_{1j}, j \ne l$. For a small variation $d\alpha_{1j}$ we have

$$
d\alpha_{1j} + d\alpha_{2j} = 0, \quad d\alpha_{1l} + d\alpha_{2l} = 0,
$$
$$
\frac{\rho_j}{R_{1j}}d\alpha_{1j} + \frac{\rho_l}{R_{1l}}d\alpha_{1l} = \frac{\rho_j}{R_{2j}}d\alpha_{2j} + \frac{\rho_l}{R_{2l}}d\alpha_{2l}.
$$

After some manipulations we deduce

$$
\frac{d}{d\alpha_{1j}}\left(\sum_{1 \le j \le M} \frac{\rho_j \alpha_{1j}}{R_{1j}} \right) = \frac{R_{2j}R_{1l} - R_{1j}R_{2l}}{R_{1j}R_{2j}(R_{1l} + R_{2l})}\rho_j. \tag{8}
$$

For $1 \le j < l$, the right hand side of (8) is negative implying that decreasing α_{1j} below 1 increases the load $\sum_{1 \le j \le M} \rho_j \alpha_{1j}/R_{1j}$. For $l < j \le M$, the right hand side of (8) is positive implying that increasing α_{1j} above 0 also increases the load. We conclude that the α_{ij} of Equation (7) are indeed a local optimum. ∎

B. Two access points

We now consider a particular network where users and APs are situated on the line segment [0,1]. User positions are uniformly distributed on the segment and the transmission rate depends on their distance d from the AP in question. The rate is 1 while $d < d_0$ and equal to $\log_2(1 + d_0/d)$ otherwise. We set d_0 to 0.1.

The two APs are located at 0 and D, respectively, where $D \in [0, 1]$ denotes the inter-AP distance.

1) A common frequency channel: Simulation results for the single channel case are plotted in Figure 7. They confirm those obtained in Section IV-A. The capacity regions satisfy T $\subset$ RT $\subset$ R = R2T which is optimal. At $D = 0.1$, the APs are close so that users practically see a single AP. This explains why all policies than have similar performance. At $D = 1$, the network is symmetric and RT is nearly optimal while T has the worst performance, as expected from Figure 5.
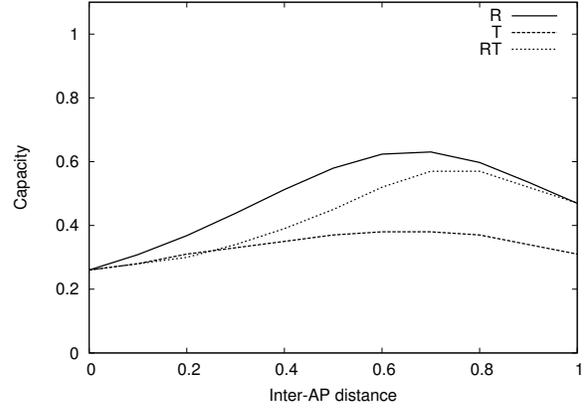

Fig. 7. Capacity with respect to inter-AP distance for a single frequency channel.
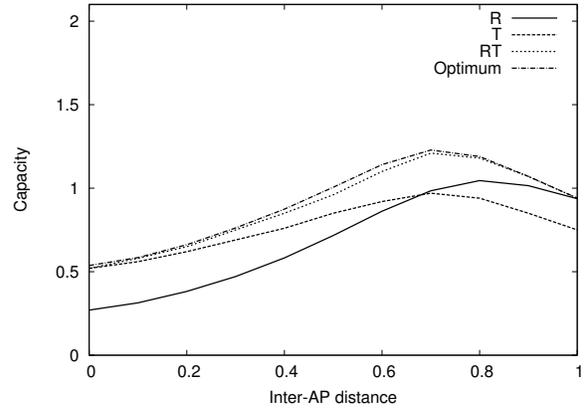

Fig. 8. Capacity with respect to inter-AP distance for two frequency channels.

2) Two channels: Results for two channels are shown in Figure 8. The optimal capacity is determined using Proposition 6. The figure confirms the results of Section IV-B. RT is quasi-optimal while T and R can be significantly worse. R2T coincides with RT and is therefore also quasi-optimal. T outperforms R as network asymmetry increases. At $D = 0.1$, practically all traffic is given to AP 2 under R while AP 1 is almost idle. Under T, on the other hand, traffic is split equally between the APs and capacity is close to th optimum. For $D = 1$ the network is symmetric and R is optimal while T performs poorly, as expected from the results of Figure 6.

The optimal association determined by Proposition 6 is in fact somewhat counter-intuitive in this example. The fact that R2T is quasi-optimal derives from the fact that it mimics this counter-intuitive behaviour. In Figure 9, we plot the ratio $R_1(d)/R_2(|d - D|)$ of physical rates as a function of user position d with D set to 0.5. Surprisingly, this ratio is not monotonic so that the optimal strategy does not exhibit a "threshold distance" with users below associating with AP 1 and users above with AP 2. Instead, the optimal strategy divides the segment $[0, 1]$ into 3 sub-segments as shown in the figure. The two outer segments are assigned to AP 1 while the inner one is given to AP 2.

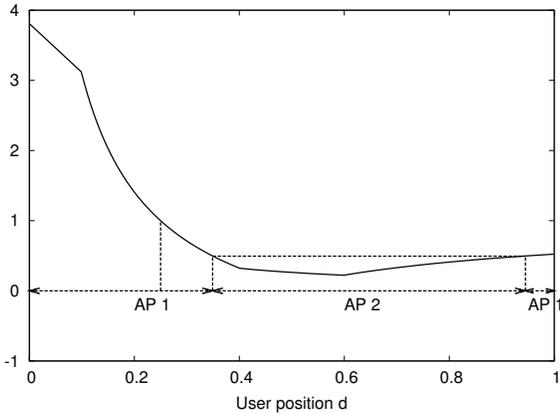To understand why R2T (and RT) mimic this behaviour,

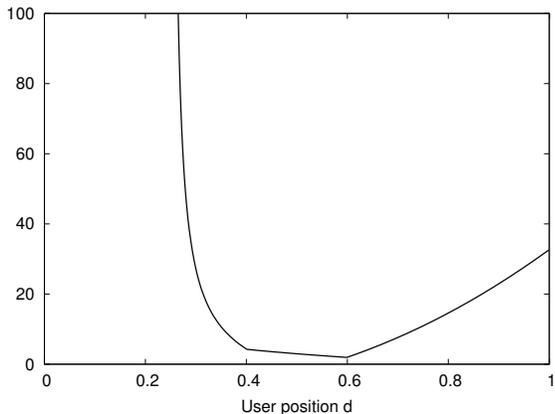Fig. 9. Ratio of physical rates $R_1(d)/R_2(|d - D|)$ for $D = 0.5$.

Fig. 10. The threshold $\gamma/((R_2(|d - D|) - R_1(d)) - 1/R_1(d)$.

consider how users associate under the fluid limit in this strategy. While both APs are saturated, R2T acts like R and AP 2 would receive more traffic than AP 1. AP 1 eventually empties while AP 2 remains saturated. In this regime, a fraction of the users closer to AP 2 will now associate with AP 1. It can be shown by studying the behaviour of the stochastic process that the probability a user at distance d associates with AP 1 is proportional to the quantity $\gamma/((R_2(|d-D|) - R_1(d)) - 1/R_1(d)$. This function, as shown in Figure 10, has the same u-shape as Figure 9 implying a similar division of the segment into three regions with AP 1 receiving both close and distant users.

C. Random networks

In all previous network scenarios, R2T was either identical to R (single channel scenarios) or to RT (multiple channels scenarios). In this final section, we present a network scenario where R2T is different from both R and RT.

We again consider the line segment with the same propagation model but this time assume a certain number of APs are placed randomly on [0,1]. Starting from the left, the APs are allocated one of F channels in a cyclic order. For each network scenario, we consider only a single random instance and compute the maximal traffic density achieved by each policy. The results are shown in Tables I and II. Results

confirm that R2T indeed outperforms the three other schemes, in particular when $F = 3$ when both R and RT are suboptimal.

N	R	T	RT	R2T
3	0.69	0.42	0.52	0.69
6	0.83	0.47	0.67	0.83
9	0.99	0.53	0.98	0.99
12	0.99	0.55	0.98	0.99

TABLE I
CAPACITY WITH RESPECT TO THE NUMBER OF APS FOR $F = 1$.

N	R	T	RT	R2T
3	1.03	1.40	1.68	1.68
6	1.34	1.44	1.81	2.35
9	2.12	1.54	2.89	2.99
12	2.13	1.54	2.86	2.98

TABLE II
CAPACITY WITH RESPECT TO THE NUMBER OF APS FOR $F = 3$.

VI. CONCLUSION

In this paper we have studied the performance of some association policies and evaluated their impact on network capacity. We have found that performance depends strongly on the assignment of frequency channels to APs, a property that has largely been ignored in previous work. Our analytical results provide insight into why both physical rate and load dependent throughput must be taken into account in determining the AP with which a user should associate. They also explain why simply combining rate and throughput in a single metric is not always an adequate solution. Our proposal is that users first determine for each channel the AP with the highest rate and then apply the rate-throughput metric to choose one AP from this set.

The present work has a number of limitations that we intend to address in future work. Using network capacity as a criterion for comparison does not necessarily take proper account of how well policies perform at normal loads. In particular, the parameter γ is not critical for capacity (large values being marginally advantageous) but may have a significant impact on realized throughput when demand is somewhat below the stability limit. Our analysis is confined to networks where all APs assigned the same channel constitute a single domain of collision. Future work will remove this restriction. We also intend to evaluate the interest of dynamic policies like those proposed in [4], [5], [12] where users may change their AP during the service time. Finally, it is necessary to examine the impact of upstream traffic on association, especially when this is significant in volume due, for instance, to the use of peer-to-peer applications.

REFERENCES

[1] M. Abusubaih, J. Gross, S. Wiethoelter, and A. Wolisz. On access point selection in IEEE 802.11 wireless local area networks. In *proceedings of the 31st IEEE Conference on Local Computer Networks*, 2006.
[2] G. Athanasiou, T. Koraki, O. Ercetin, and L. Tassiulas. Dynamic cross-layer association in 802.11-based mesh networks. In *INFOCOM*, 2007.

[3] G. Athanasiou, T. Korakis, O. Ercetin, and L. Tassiulas. A cross-layer framework for association control in wireless mesh networks. *IEEE Transactions on Mobile Computing*, 8(1):65–80, January 2009.

[4] A. Balachandran, P. Bahl, and G. M. Voelker. Hot-spot congestion relief in public-area wireless networks. In *proceedings of the fourth IEEE Workshop on Computing Systems and Applications*, 2002.

[5] Y. Bejerano, S. J. Han, and L. Li. Fairness and load balancing in wireless LANs using association control. *IEEE/ACM Transactions on Networking*, 15(3):560–573, June 2007.

[6] T. Bonald, A. Ibrahim, and J. Roberts. Traffic capacity of multi-cell wlans. In *SIGMETRICS*, 2008.

[7] T. Bonald, A. Ibrahim, and J. Roberts. Enhanced spatial reuse in multi-cell wlans. In *INFOCOM Mini Conference*, 2009.

[8] J. Dai. On positive harris recurrence of multiclass queueing networks: a unified approach via fluid limit models. *The Annals of Applied Probability*, 5(1):49–77, 1995.

[9] Y. Fukuda, T. Abe, and Y. Oie. Decentralized access point selection architecture for wireless LANs. In *Wireless Telecommunications Symposium*, 2004.

[10] G. S. Kasbekar, J. Kuri, and P. Nuggehalli. Online association policies in IEEE 802.11 WLANs. In *WiOpt*, 2006.

[11] G. S. Kasbekar, P. Nuggehalli, and J. Kuri. Online client-AP association in WLANs. In *RAWNET*, 2006.

[12] B. Kauffmann, F. Baccelli, A. Chaintreau, V. Mhatre, K. Papagiannaki, and C. Diot. Measurement-based self organization of interfering 802.11 wireless access networks. In *INFOCOM*, 2007.

[13] Leonard Kleinrock. *Queueing Systems Volume 1: Theory*. 1975.

[14] S. P. Meyn. Transience of multiclass queueing networks and their fluid models. *The Annals of Applied Probability*, 5(4):946–957, 1995.

[15] Karthikeyan Sundaresan and Konstantina Papagiannaki. The need for cross-layer information in access point selection algorithms. In *IMC*, 2006.

[16] S. Vasudevan, K. Papagiannaki, C. Diot, J. Kurose, and D. Towsley. Facilitating access point selection in ieee 802.11 wireless networks. In *IMC*, 2005.

APPENDIX

We use the fluid limit method of [8], [14] to determine the capacity regions. The basic idea consists in transforming the complex stochastic queueing system into a simpler deterministic dynamical system called the fluid limit model. The fluid limit is said to be stable if the fluid volumes of all user classes, represented by some vector $\bar{x}(t)$, reach 0 after a finite time, independently of the initial fluid volumes. It is shown in [8] that stability of the fluid limit is a sufficient condition for stability of the stochastic system. The major difficulty of this method applied to multidimensional processes resides in the need to account for the non-deterministic behaviour of components whose fluid volume disappears before that of other components.

In this paper we study a particular type of queueing system where the arrival and service rates are state dependent. The corresponding fluid limit model is thus also state dependent. Determining the capacity region using the fluid model method proceeds three steps. In the first step we show, using conservation law equations and certain properties of service rate functions, that the fluid volume vector $\bar{x}(t)$ converges to a certain direction in the state space. In the second step we determine the fluid limit in this particular direction using the underlying stochastic system. The positive components of $\bar{x}(t)$ are considered to have an infinite user population while the null components have a finite stochastically varying population. The fluid limit is computed by taking the average of arrival and service rates over this stochastic system. In the

third step we determine the stability conditions of the fluid limit model in the investigated direction.

A. Proof of Proposition 1

The fluid model satisfies:

$$\frac{d\bar{x}_1}{dt} = a_1 - R_0, \qquad \frac{d\bar{x}_2}{dt} = a_2 - R_0$$

if $\bar{x}_1 > 0$ and $\bar{x}_2 > 0$, with

$$(a_1, a_2) = \begin{cases} (\rho, 0) & \text{if } \bar{x}_1 < \bar{x}_2, \\ (0, \rho) & \text{if } \bar{x}_2 < \bar{x}_1, \end{cases}$$

and $a_1 + a_2 = \rho$ in all cases since no traffic is lost. We deduce that $\bar{x}_1 = \bar{x}_2$ after some finite time. Since

$$\frac{d\bar{x}_1}{dt} - \frac{d\bar{x}_2}{dt} = a_1 - a_2,$$

this implies $a_1 = a_2$. Finally, we obtain $a_1 = a_2 = \rho/2$ and the stability condition $\rho/2 < R_0$.

B. Proof of Proposition 2

Again, the fluid limit dynamics are given by:

$$\frac{d\bar{x}_1}{dt} = a_1 - R_0, \qquad \frac{d\bar{x}_2}{dt} = a_2 - R_0,$$

if $\bar{x}_1 > 0$ and $\bar{x}_2 > 0$, with $a_1 + a_2 = \rho$. If $\bar{x}_1 > 0$, there is an infinite user population at AP 1 so that $T_1 = 0$. Since $R_1 < R_2$, all users associate with AP 2 under the considered RT policy. Now if $\bar{x}_1 = 0$ and $\bar{x}_2 > 0$, there is an infinite user population at AP 2 so that $T_2 = 0$. An incoming user associates with AP 1 if:

$$\alpha \frac{R_0}{x_1 + 1} + \beta R_1 > \beta R_2,$$

, i.e., if $x_1 < K$. Thus the number of users associated with AP 1 behaves like an $M/M/1/K$ queue of load $\varrho = \rho/R_0$. Blocked users associate with AP 2. In view of [13], the arrival rate at AP 2 is given by:

$$a_2 = \rho \times \frac{\varrho^K}{1 + \varrho + \ldots + \varrho^K}.$$

Moreover, AP 1 is idle a fraction of time equal to:

$$p = \frac{1}{1 + \varrho + \ldots + \varrho^K},$$

in which case the service rate of AP 2 is equal to R_2. We deduce the stability condition:

$$a_2 < pR_2 + (1 - p)R_0,$$

which corresponds to that stated in the proposition.

C. Proof of Proposition 3

The fluid model satisfies:

$$\frac{d\bar{x}_1}{dt} = a_1 - R_1, \quad \frac{d\bar{x}_2}{dt} = a_2 - R_2$$

if $\bar{x}_1 > 0$ and $\bar{x}_2 > 0$, with

$$(a_1, a_2) = \begin{cases} (\rho, 0) & \text{if } \bar{x}_1/R_1 < \bar{x}_2/R_2, \\ (0, \rho) & \text{if } \bar{x}_2/R_2 < \bar{x}_1/R_1, \end{cases}$$

and $a_1 + a_2 = \rho$ in all cases since no traffic is lost. Hence $\bar{x}_1/R_1 = \bar{x}_2/R_2$ after some finite time. Since

$$\frac{d\bar{x}_1}{dt}\frac{1}{R_1} - \frac{d\bar{x}_2}{dt}\frac{1}{R_2} = \frac{a_1}{R_1} - \frac{a_2}{R_2},$$

we obtain $a_1/a_2 = R_1/R_2$. Thus

$$a_1 = \rho \frac{R_1}{R_1 + R_2}, \quad a_2 = \rho \frac{R_2}{R_1 + R_2},$$

and the stability condition is $\rho < R_1 + R_2$.

D. Proof of Proposition 4

The proof is similar to that of Proposition 2 (case of a single channel). The equations satisfied by the fluid model are given by:

$$\frac{d\bar{x}_1}{dt} = a_1 - R_1, \quad \frac{d\bar{x}_2}{dt} = a_2 - R_2,$$

if $\bar{x}_1 > 0$ and $\bar{x}_2 > 0$, with $a_1 + a_2 = \rho$. If $\bar{x}_1 > 0$, all users associate with AP 2. Now if $\bar{x}_1 = 0$ and $\bar{x}_2 > 0$, an incoming user associates with AP 1 if $x_1 < L$. Thus the number of users associated with AP 1 behaves like an $M/M/1/L$ queue of load $\varrho = \rho/R_1$. The arrival rate at AP 2 is given by:

$$a_2 = \rho \times \frac{\varrho^L}{1 + \varrho + \ldots + \varrho^L}.$$

We deduce the stability condition $a_2 < R_2$, which is that stated in the proposition.

E. Proof of Proposition 5

Let $\bar{R}_0$ be the common throughput of each AP in the fluid model:

$$\frac{1}{\bar{R}_0} = \frac{\bar{x}_{11}}{\bar{x}_1 R_{11}} + \frac{\bar{x}_{12}}{\bar{x}_1 R_{12}} + \frac{\bar{x}_{21}}{\bar{x}_2 R_{21}} + \frac{\bar{x}_{22}}{\bar{x}_2 R_{22}},$$

with $\bar{x}_1 = \bar{x}_{11} + \bar{x}_{12}$ and $\bar{x}_2 = \bar{x}_{21} + \bar{x}_{22}$. The equations of the fluid limit are given by:

$$\frac{d\bar{x}_{ij}}{dt} = a_{ij} - \bar{R}_0, \quad i, j = 1, 2,$$

when $\bar{x}_1 > 0$, $\bar{x}_2 > 0$, with

$$\begin{cases} a_{11} = \rho_1, \ a_{12} = \rho_2, & \text{if } \bar{x}_1 < \bar{x}_2, \\ a_{21} = \rho_1, \ a_{22} = \rho_2, & \text{if } \bar{x}_2 < \bar{x}_1, \end{cases}$$

and $a_{11} + a_{21} = \rho_1$, $a_{12} + a_{22} = \rho_2$ since no traffic is lost. We deduce that $\bar{x}_1 = \bar{x}_2$ after some finite time.

It remains to calculate the arrival rates $a_{11}, a_{21}, a_{12}, a_{22}$ when $\bar{x}_1 = \bar{x}_2$, that is when the difference $x_1 - x_2$ is finite in

the underlying stochastic network. A class-1 user associates with AP 1 if and only if:

$$\frac{R_0(x_{11} + 1, x_{12}, x_{21}, x_{22})}{x_1 + 1} \geq \frac{R_0(x_{11}, x_{12}, x_{21} + 1, x_{22})}{x_2 + 1},$$

where the dependency of R_0 on state x is made explicit. This can be written as:

$$\frac{x_2 + 1}{x_1}\left(\frac{x_{11}}{R_{11}} + \frac{x_{12}}{R_{12}}\right) + \frac{x_{21} + 1}{R_{21}} + \frac{x_{22}}{R_{22}} \geq$$
$$\frac{x_{11} + 1}{R_{11}} + \frac{x_{12}}{R_{12}} + \frac{x_1 + 1}{x_2}\left(\frac{x_{21}}{R_{21}} + \frac{x_{22}}{R_{22}}\right).$$

If $x_1 = x_2$, this inequality is equivalent to:

$$\frac{1}{x_1}\left(\frac{x_{11}}{R_{11}} + \frac{x_{12}}{R_{12}}\right) + \frac{1}{R_{21}} \geq \frac{1}{R_{11}} + \frac{1}{x_2}\left(\frac{x_{21}}{R_{21}} + \frac{x_{22}}{R_{22}}\right),$$

which is satisfied since $R_{12} < R_{11}$ and $R_{21} < R_{22}$. Now if $x_1 < x_2$, we get:

$$\frac{x_2 + 1}{x_1}\left(\frac{x_{11}}{R_{11}} + \frac{x_{12}}{R_{12}}\right) + \frac{x_{21} + 1}{R_{21}} + \frac{x_{22}}{R_{22}} \geq$$
$$\frac{x_{11}}{R_{11}} + \frac{x_{12}}{R_{12}} + \frac{x_{21} + 1}{R_{21}} + \frac{x_{22}}{R_{22}} \geq$$
$$\frac{x_{11} + 1}{R_{11}} + \frac{x_{12}}{R_{12}} + \frac{x_{21}}{R_{21}} + \frac{x_{22}}{R_{22}} \geq$$
$$\frac{x_{11} + 1}{R_{11}} + \frac{x_{12}}{R_{12}} + \frac{x_1 + 1}{x_2}\left(\frac{x_{21}}{R_{21}} + \frac{x_{22}}{R_{22}}\right),$$

where the second inequality results from the fact that $R_{21} < R_{11}$. Finally, if $x_1 > x_2$, we get:

$$\frac{x_2 + 1}{x_1}\left(\frac{x_{11}}{R_{11}} + \frac{x_{12}}{R_{12}}\right) + \frac{x_{21} + 1}{R_{21}} + \frac{x_{22}}{R_{22}} \leq$$
$$\frac{x_{11}}{R_{11}} + \frac{x_{12}}{R_{12}} + \frac{x_{21} + 1}{R_{21}} + \frac{x_{22}}{R_{22}} \leq$$
$$\frac{x_{11} + 1}{R_{11}} + \frac{x_{12}}{R_{12}} + \frac{x_1 + 1}{x_2}\left(\frac{x_{21}}{R_{21}} + \frac{x_{22}}{R_{22}}\right),$$

where the second inequality, which is equivalent to

$$\frac{1}{R_{21}} \leq \frac{1}{R_{11}} + (x_1 - x_2 + 1)\left(\frac{x_{21}}{x_2 R_{21}} + \frac{x_{22}}{x_2 R_{22}}\right),$$

follows from (4) and $R_{21} < R_{22}$. We conclude that class-1 users associate with AP 1 if and only if $x_1 \leq x_2$. We verify that this property also holds if $x_1 = 0$ or $x_2 = 0$. Similarly, class-2 users associate with AP 2 if and only if $x_2 \leq x_1$.

Now consider the difference $\Delta = x_1 - x_2$. This defines a continuous time Markov process on $\mathbb{Z}$. If $\Delta < 0$, all users associate with AP 1. Thus each arrival increases Δ by 1; a departure from AP 2 increases Δ by 1 while a departure from AP 1 decreases Δ by 1. We deduce that the transition rates of Δ from n to $n - 1$ are equal to $\bar{R}_0$ and from n to $n + 1$ to $\rho_1 + \rho_2 + \bar{R}_0$, for any state $n < 0$. Similarly, the transition rates of Δ from n to $n + 1$ are equal to $\bar{R}_0$ and from n to $n - 1$ to $\rho_1 + \rho_2 + \bar{R}_0$, for any state $n > 0$. Finally, if $\Delta = 0$, then class-1 (resp. class-2) users associate with AP 1 (resp. AP 2). The transition rates of Δ from state 0 to state -1 are

then equal to $\rho_2 + \bar{R}_0$ and from state 0 to state 1 to $\rho_1 + \bar{R}_0$. The stationary distribution of Δ yields:

$$a_{11} = \rho_1 \mathbb{P}[\Delta \le 0] = \frac{\rho_1(\rho_1 + 2\rho_2 + \bar{R}_0)}{2(\rho_1 + \rho_2 + \bar{R}_0)},$$

$$a_{12} = \rho_2 \mathbb{P}[\Delta < 0] = \frac{\rho_2(\rho_2 + \bar{R}_0)}{2(\rho_1 + \rho_2 + \bar{R}_0)},$$

$$a_{21} = \rho_1 \mathbb{P}[\Delta > 0] = \frac{\rho_1(\rho_1 + \bar{R}_0)}{2(\rho_1 + \rho_2 + \bar{R}_0)},$$

$$a_{22} = \rho_2 \mathbb{P}[\Delta \ge 0] = \frac{\rho_2(2\rho_1 + \rho_2 + \bar{R}_0)}{2(\rho_1 + \rho_2 + \bar{R}_0)}.$$

This characterizes the fluid model and completes the first part of the proof.

Now let:

$$\bar{\xi}_{11} = \frac{\bar{x}_{11}}{\bar{x}_1}, \quad \bar{\xi}_{22} = \frac{\bar{x}_{22}}{\bar{x}_2}.$$

The direction of the fluid volume vector is completely determined by $\bar{\xi}_{11}$ and $\bar{\xi}_{22}$. The derivative of $\bar{\xi}_{11}$ is given by:

$$\frac{d\bar{\xi}_{11}}{dt} = \frac{\rho_1(\rho_1 + 2\rho_2 + \bar{R}_0)\bar{\xi}_{12} - \rho_2(\rho_2 + \bar{R}_0)\bar{\xi}_{11}}{2(\bar{x}_{11} + \bar{x}_{12})(\rho_1 + \rho_2 + \bar{R}_0)}.$$

It has the same sign as the following second degree polynomial in $\bar{\xi}_{11}$:

$$\rho_1(1 - \bar{\xi}_{11}) - \rho_2\bar{\xi}_{11} + \left(\rho_1(\rho_1 + 2\rho_2)(1 - \bar{\xi}_{11}) - \rho_2^2\bar{\xi}_{11}\right) \times$$

$$\left(\bar{\xi}_{11}\left(\frac{1}{R_{11}} - \frac{1}{R_{12}}\right) + \frac{1}{R_{12}} + \frac{1}{R_{21}} + \bar{\xi}_{22}\left(\frac{1}{R_{22}} - \frac{1}{R_{21}}\right)\right).$$

The limits of this polynomial when $\bar{\xi}_{11}$ tends to 0 and 1 are respectively:

$$\rho_1 + \rho_1(\rho_1 + 2\rho_2)(\frac{1}{R_{12}} + \frac{1}{R_{21}} + \bar{\xi}_{22}(\frac{1}{R_{22}} - \frac{1}{R_{21}}))$$

and

$$-\rho_2 - \rho_2^2(\frac{1}{R_{11}} + \frac{1}{R_{21}} + \bar{\xi}_{22}(\frac{1}{R_{22}} - \frac{1}{R_{21}})),$$

which are respectively positive and negative. Since a second degree polynomial cannot have more than two roots, we deduce that for each $\bar{\xi}_{22}$, there exists a unique solution $\bar{\xi}_{11}^*$ to the equation $d\bar{\xi}_{11}/dt = 0$ such that $d\bar{\xi}_{11}/dt > 0$ for all $\bar{\xi}_{11} < \bar{\xi}_{11}^*$ and $d\bar{\xi}_{11}/dt < 0$ for all $\bar{\xi}_{11} > \bar{\xi}_{11}^*$. Similar results exist also for $\bar{\xi}_{22}$. We deduce that the system $d\bar{\xi}_{11}/dt = 0$ and $d\bar{\xi}_{22}/dt = 0$ has a unique solution and that $\bar{\xi}_{11}$ and $\bar{\xi}_{22}$ converge to that unique solution after some finite time. Therefore the service rate $\bar{R}_0$, and thus the arrival rates a_{ij}, can now be considered as constant.

The network is stable if $a_{11} + a_{12} < \bar{R}_0$ and $a_{21} + a_{22} < \bar{R}_0$. Since $a_{11} + a_{12} = a_{21} + a_{22} = (\rho_1 + \rho_2)/2$, the stability condition is simply given by $(\rho_1 + \rho_2)/2 < \bar{R}_0$. At the boundary of the stability region, we have $(\rho_1 + \rho_2)/2 = \bar{R}_0$ so that:

$$a_{11} = \frac{\rho_1(3\rho_1 + 5\rho_2)}{6(\rho_1 + \rho_2)}, \quad a_{12} = \frac{\rho_2(\rho_1 + 3\rho_2)}{6(\rho_1 + \rho_2)},$$

$$a_{21} = \frac{\rho_1(3\rho_1 + \rho_2)}{6(\rho_1 + \rho_2)}, \quad a_{22} = \frac{\rho_2(5\rho_1 + 3\rho_2)}{6(\rho_1 + \rho_2)}.$$

Using the fact that at each AP, the fluid volumes of each class are proportional to the arrival rates, we get:

$$\frac{1}{\bar{R}_0} = \frac{2}{\rho_1 + \rho_2}\left(\frac{a_{11}}{R_{11}} + \frac{a_{12}}{R_{12}} + \frac{a_{21}}{R_{21}} + \frac{a_{22}}{R_{22}},\right)$$

We deduce the stability condition:

$$\frac{a_{11}}{R_{11}} + \frac{a_{12}}{R_{12}} + \frac{a_{21}}{R_{21}} + \frac{a_{22}}{R_{22}} < 1,$$

which is equivalent to that given in the proposition.