Price-based Congestion-Control in Wi-Fi Hot Spots

Roberto Battiti, Marco Conti, Enrico Gregori, Mikalai Sabel

Price-based Congestion-Control in Wi-Fi Hot Spots

Roberto Battiti(*), Marco Conti(**), Enrico Gregori(**), Mikalai Sabel(*)

(*) University of Trento,
Dip. Informatica e Telecomunicazioni
Trento, Italy
{battiti, msabel}@dit.unitn.it

(**) Consiglio Nazionale delle Ricerche
IIT Institute
Pisa, Italy
{marco.conti, enrico.gregori}@iit.cnr.it

ABSTRACT. Wireless networks are now proliferating due to the success of the IEEE 802.11b protocol, also known as "Wi-Fi" (Wireless Fidelity). A Wi-Fi network is characterized by a set of base stations (also called access points) placed throughout the environment and connected to the traditional wired LANs. This technology allows nomadic users a broadband access to the Internet if they are in the transmission range of an access point. A new business model, named Wi-Fi Hot Spots, is now emerging to exploit the potentialities of this technology. A hot spot is a "critical" business area, e.g., airports, stations, hotels, where users can have wireless access by subscribing a contract with the hot spot operator, or with a wireless Internet service provider (WISP). Due to the random access nature of the Wi-Fi technology, if the number of users connected to the same access point increases, the QoS experienced may quickly degrade. This generates complains from the users that, as a consequence, may change their WISP. In order to be competitive, a Wi-Fi hot spot operator needs simple and effective mechanisms to control the congestion therefore guaranteeing the QoS, and (at the same time) maximizing his/her revenues. In this paper we present and evaluate a price-based policy for the access control in a Wi-Fi hot spot. Our policy, named *Price-based Congestion Control (PCC)*, controls the hot spot traffic by dynamically determining the access *cost* as a function of the current load in the hot spot. We develop a theoretical framework to compute for any load condition the access cost to maintain the hot spot in its optimal operating point, for any load condition. The effectiveness and robustness of the PCC policy has been evaluated by simulating a Wi-Fi hot spot. Both in saturated and not-saturated conditions the PCC policy provides a better channel utilization than the legacy Wi-Fi policy.

1. Introduction

Wi-Fi broadband Internet access is mushrooming and competing with legacy wireless (cellular) networks to support nomadic and data-centric applications. Telecom operators and service providers are gradually changing their marketing strategies and complementing their traditional offer with Wi-Fi. To reach an efficient use of the scarce bandwidth resources, market mechanisms appropriate for the Wi-Fi technology have been proposed ([IEEE802], [IEEE11b]).

A well-designed market should encourage spectrum efficiency and innovation. The two driving forces characterizing the Wi-Fi evolution are: the low cost-barrier for a service provider to enter in the market (no expensive infrastructure is required to start with), and the emerging tendency (mainly in USA) to deregulate the spectrum environment to create a secondary wireless market [Cro02], secondary with respect to the primary cellular market.

In this work we propose a price-based policy for access control in a Wi-Fi hot spot. We consider a WLAN Operator and we investigate policies to control the traffic generated by its hot-spot users in order to maximize the revenues. The idea is to control the hot spot traffic by implementing a congestion-control pricing policy. At any given time instant, the access cost depends on the current load in the hot spot. When congestion increases, the WLAN Operator increases the access cost until some users give up transmitting. The objective is to maintain the quality of service, and to maximize the revenues.

In this paper, we assume a per-packet cost model, i.e., the user will be charged for the number of successfully transmitted and received packets. By increasing or decreasing the per-packet cost the WLAN Operator will control the number of active users and hence the congestion in the hot-spot. The aim is to drive the hot spot to operate in a status that maximizes the operator revenues and maintains the system in the most efficient state, i.e., the maximum aggregated throughput. Specifically, we identify as the *optimal operating point* of the system, the status corresponding to the maximum aggregated throughput. The rational for this choice is based on the following observations:

- when the offered load is below this point, the WLAN Operator can stimulate additional traffic, thus increasing the revenues without congesting the system;
- when the offered load is higher than this point, to maintain the same revenues of the optimal operating point, the WLAN Operator must increase the costs and thus the users have to pay more for a worse quality of service. This policy, of course, cannot be acceptable by the users that can hence migrate to another WLAN Operator implementing a more fair policy and therefore providing a better quality at a lower cost.

This research is partially supported by the Province of Trento (Italy) in the framework of the WILMA (Wireless Internet and Location Management) project (http://www.wilmaproject.org/)

For example, a simple cost model which drives the system towards the optimal operating point can be the following. The cost model has only two values: a low and constant cost, say C_0, whenever the load is below the maximum aggregated throughput of the system, and a very high value, say $C_{infinite}$ whenever the load exceeds the maximum aggregated throughput. The flat access cost stimulates the users to access the system until it is not overloaded, while the "infinite" cost $C_{infinite}$ forces the users to stop transmitting when the load exceeds the offered load. Obviously, this policy needs to be refined to avoid sharp fluctuations in the offered load as all users will give up when the $C_{infinite}$ is announced. Simple mechanisms to smooth these fluctuations can be introduced. For example gradually increasing the costs, or the cost increase can be announced to subsets of randomly selected users.

In this paper we discuss a *Price-based Congestion Control (PCC)* policy that can be implemented in an 802.11 based WLAN hot-spot. To define the PCC policy we exploit some results derived in [CG02]. The cited work presents a distributed mechanism for controlling the congestion in a IEEE 802.11 network to guarantee that the system operates below the optimal operating point. Furthermore, the proposed mechanism asymptotically (with respect to the number of active stations) drives the system close to the optimal operating point. The main drawback of that approach is the requirement that all the network stations use a network interface card obtained by modifying the Wi-Fi cards. Obviously, this constraint is not acceptable in a WLAN hot spot in which the WLAN operator cannot force the use of modified Wi-Fi cards. In this paper, we exploit some of the ideas presented in [CG02] to construct a "centralized" congestion control policy that only requires additional hardware/software in the Access Point while the users adopt standard Wi-Fi cards.

The paper is organized as follows. In Section 2, we present the main Wi-Fi characteristics and summarize previous results about Wi-Fi modeling that are relevant for this paper. In Section 3, we introduce the theoretical framework of the Price-based Congestion Control policy. The performance of the PCC policy are investigated, via simulation, in Section 4.

2. Wi-Fi

Wi-Fi cards implement the IEEE 802.11b standard. The 802.11b standard extends the 802.11 standard by introducing a higher-speed Direct Sequence Spread Spectrum (DSSS) Physical Layer in the 2.4 GHz frequency band. Specifically, 802.11b enables transmissions at 5.5 Mbps and 11 Mbps, in addition to 1 Mbps and 2 Mbps supported by 802.11. On top of the DSSS physical layer, the IEEE 802.11b standard implements a MAC layer that offers two different types of service: a contention- based and contention-free service. Hereafter, we focus on the contention- based service only. This service is implemented by the *Distributed Coordination Function* (DCF). According to the DCF, before transmitting a data frame, a station must sense the channel to determine whether another station is transmitting. If the medium is found to be idle for an interval longer than the *Distributed InterFrame Space* (*DIFS*), the station continues with its transmission. If the medium is busy, the transmission is deferred until the end of the ongoing transmission, and then the backoff procedure is activated. A random interval, (the *backoff time*) is selected and used to initialize the *backoff timer*. The backoff timer is decreased for as long as the channel is sensed as idle, stopped when a transmission is detected on the channel, and reactivated when the channel is sensed as idle again for more than a DIFS. The station is enabled to transmit its frame when the backoff timer reaches zero.

To guarantee fair access to the shared medium, a station that has just transmitted a packet and has another packet ready for transmission must perform the backoff procedure before initiating the second transmission.

The backoff time is slotted: its value is an integer number of slots uniformly chosen in the interval (0, *CW*-1). *CW* is defined as the Backoff Window, also referred to as *Contention Window*. At the first transmission attempt *CW*=*CWmin*, and it is doubled at each retransmission up to *CWmax*. In the 802.11b standard *CWmin* and *CWmax* are 32 and 1024, respectively. Obviously, it may happen that two or more stations start transmitting simultaneously and a collision occurs. In the CSMA/CA scheme, stations cannot detect a collision by hearing their own transmissions (as in the CSMA/CD protocol used in wired LANs). Therefore, an immediate positive acknowledgement scheme is employed to ascertain the successful reception of a frame. In detail, upon reception of a data frame, the destination station initiates the transmission of an acknowledgement frame (ACK) after a time interval called *Short InterFrame Space* (SIFS). The SIFS is shorter than the DIFS in order to give priority to the receiving station over other possible stations waiting for transmission. If the ACK is not received by the source station, the data frame is assumed to have been lost, and a retransmission is scheduled. The ACK is not transmitted if the received packet is corrupted. A Cyclic Redundancy Check (*CRC*) algorithm is used for error detection. After an erroneous frame is detected (due to collisions or transmission errors), a station must remain idle for at least an *Extended InterFrame Space* (EIFS) interval before it reactivates the backoff algorithm. Specifically, the EIFS shall be used by the DCF whenever the physical layer has indicated to the MAC that a frame transmission was begun that did not result in the correct reception of a complete MAC frame with a correct CRC value. Reception of an error-free frame during the EIFS re-synchronizes the station to the actual busy/idle state of the medium, so the EIFS is terminated and normal medium access (using DIFS and, if necessary, backoff) continues following reception of that frame. A MAC data frame contains the control information (hereafter denoted as MAC_{hdr}), a variable length *data payload*, which contains the upper layers data information, and the CRC field. The MAC_{hdr} contains the MAC addresses in addition to other control information.

The values for the protocol parameters are summarized in Table 1, where τ is the air propagation time

Table 1. IEEE 802.11b parameter values.

$Slot\_Time$ t_{slot}	τ	MAC_{hdr}	CRC	*Bit Rate(Mbps)*
20 μsec	1 μsec	240 bits (2.4 t_{slot})	32 bits (0.32 t_{slot})	1, 2, 5.5, 11
DIFS	*SIFS*	*ACK*	CW_{MIN}	CW_{MAX}
50 μsec	10 μsec	112 bits	32 t_{slot}	1024 t_{slot}

2.1 Wi-Fi Modeling

The behavior of a IEEE 802.11 network can be closely approximated by an equivalent *p-persistent IEEE 802.11* protocol [CCG00]. A *p-persistent IEEE 802.11* protocol differs from the standard protocol only in the selection of the backoff interval. Instead of the binary exponential backoff used in the standard, the backoff interval of the *p-persistent IEEE 802.11* protocol is sampled from a geometric distribution with parameter p. In that paper it is also shown that the *p*-persistent IEEE 802.11 protocol closely approximates the standard protocol (at least from the protocol capacity standpoint) if the average backoff interval is the same. This is obtained when the following equation holds:

$$p = 1/\left(E[B]+1\right) \tag{1}$$

where $E[B]$ is the average backoff interval.

Due to its memory-less backoff algorithm, the *p*-persistent IEEE 802.11 protocol is suitable for analytical studies. By exploiting the similarity of this protocol with the standard one, the analytical results derived from the *p*-persistent model can be used to infer the behavior of the standard protocol.

In this section, we assume a system with a variable number, *M*, of active stations accessing the channel according to the *p*-persistent IEEE 802.11 protocol. The stations transmit messages whose length (expressed as number of 802.11 time slots) is a random variable *L* with average *l*.

The behavior of the 802.11 channel access scheme can be described as a sequence of *virtual transmission times*. A virtual transmission time is the interval between two successful transmissions and, as shown in Figure 1, it includes a successful transmission and may include several collision intervals and idle periods.[1]

In the following we use the following notation:

- $Idle\_p$ is the number of consecutive empty slots;
- $Coll$ is the time the channel is busy due to a collision given that a transmission attempt occurs, also referred to as *collision interval* (see Figure 1). Obviously, $Coll$ is equal to zero if the transmission attempt is successful, otherwise it is equal to the collision length;
- t_{slot} is the time duration of a time slot;
- $E[\ ]$ denotes the average operator, i.e., given a random variable X, $E[X]$ is its average;

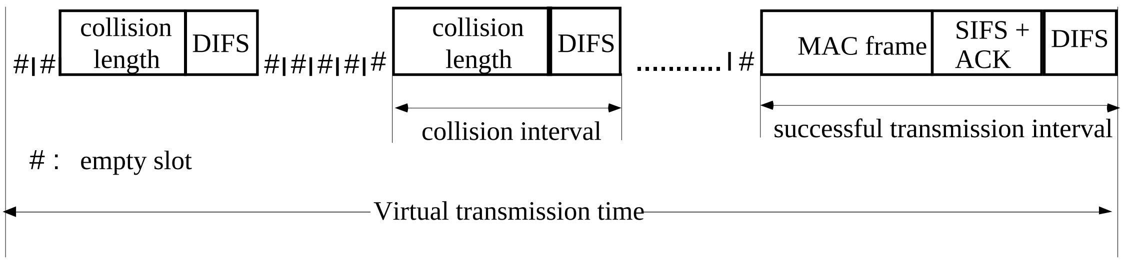

Figure 1: Structure of a virtual transmission time

By exploiting the analytical tractability of the *p-persistent IEEE 802.11* the following property, relevant for defining our congestion pricing policy, was proved ([CG02], [BCG02]):

Lemma 1. *For* $M >> 1$ *and* $l > 1$ *the optimal operating point, corresponds to the system status that satisfies the following equation:*

$$E[Idle\_p] \cdot t_{slot} = E[Coll] \tag{2}$$

[1] It is worth noting that, hereafter, the channel idle period that follows a collision (at least a DIFS), and the SIFS, ACK and DIFS intervals that follow a successful transmission are counted as overheads associated to a collision or to a successful transmission.

Closed form expressions for the $E[Idle\_p]$ and $E[Coll]$ in a *p-persistent IEEE 802.11* with *M* active stations are derived (by exploiting classical probabilistic arguments) in [CCG00a]. The formulas indicate that the optimal operating point is a function of the number of active stations (*M*), the length of transmitted packets, and the average congestion window (i.e., the *p* value). Therefore, by fixing the *p* value, and the length of transmitted packets, it is possible to determine the number of stations that must be active (in a hot spot) in order to maintain the system in the optimal operating point. Therefore, Equation (2) provides an effective tool that can be used by the Access Point to control the congestion status of the network. It is worth noting that the perspective of the PCC policy is complementary to the policies developed in the previous works ([CG02], [BCG02]). In those papers *p* is a variable whose value is dynamically set, while *M* cannot be modified. On the contrary, in this paper the *p* value (i.e., the average backoff window size) cannot be controlled, while the Access Point can try to modify the value of *M* (by increasing or decreasing the transmission cost). In detail, for each network condition, the Access Point tries to drive the *M* value to the optimal operating point, say M_{opt}, that guarantees the satisfaction of Equation (2).

3. The Price-based Congestion Control (PCC) policy

To simplify the presentation, we assume that the PCC policy takes its decisions (i.e., the new price to be communicated to the hot spot users) at the end of each virtual transmission time. However, it is also possible to implement the same policy taking as a reference other events occurring on the channel, or assuming a periodic re-computation (and announcement) of the optimal price. As shown in the Section 4, the periodic announcement is the most robust solution for PCC.

The PCC policy operates in two steps. In the first step, it identifies the percentage increase/decrease in the number of active stations in the controlled hot spot to drive the system to the optimal operating point. In the second step, the policy identifies the price level to achieve the desired increase/decrease in the number of active stations. Before presenting the PCC policy in detail, we introduce some assumptions.

i) After each successful transmission the AP decides the transmission-cost *c* to be applied to the future transmissions. The time it takes to the AP to compute the new cost and to communicate it to the clients is assumed to be negligible.

ii) To model the users behavior, we introduce a probability function $P_{give\_up}(c)$. Specifically, $P_{give\_up}(c)$ represents the probability that a user does not accept the *c* transmission-cost announced by the AP, and leaves the network for a random time. It is reasonable to assume that $P_{give\_up}(c)$ is monotone not decreasing, i.e., when the cost increases the probability that a user gives up increases or remain the same. The AP estimates the $P_{give\_up}(c)$ from the system history, by observing the users' reaction to changes in the price level.[2]

iii) We model the arrival of new customers (new clients plus the clients that make a new attempt after a give-up period) in the hot spot through a Poisson process, with rate λ. λ is estimated by applying a moving window estimator to the recent history of the system.

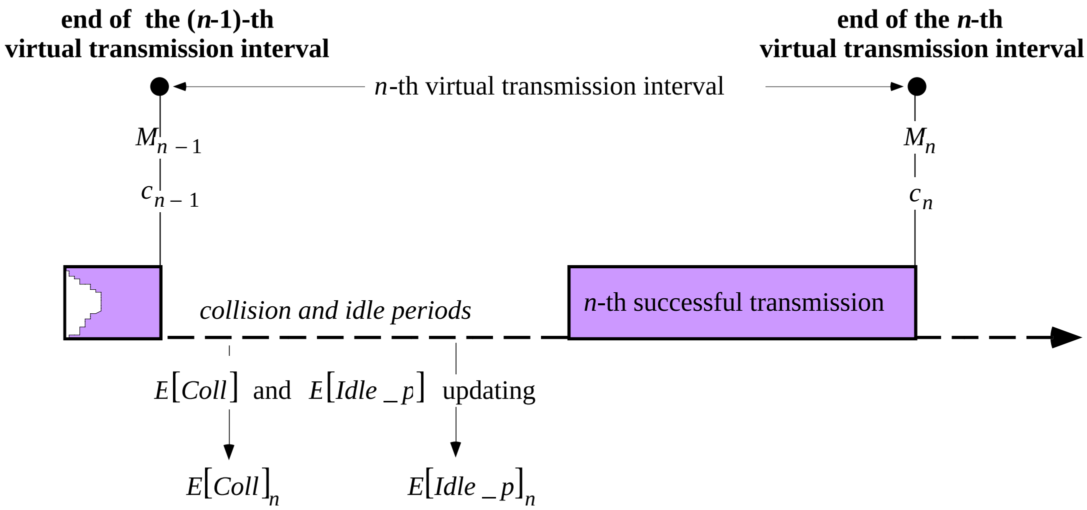

Figure 2: PCC policy behavior

Step 1. Percentage increase/decrease in the number of active stations

The behavior of the PCC policy is shown in Figure 2. As mentioned before, PCC operates at the end of each virtual transmission time. As shown in Figure 1, this time interval (from the channel status standpoint) consists of one or more

[2] $P_{give\_up}(c)$ can be computed offline by exploiting the statistics collected on a daily (or longer) basis.

idle periods, and one or more busy periods. An idle period is made up of consecutive idle slots, while the busy period is an interval in which the channel is busy due to either a collision or a successful transmission.

To explain the behavior of the PCC policy, we focus on the end of the *n*-th virtual transmission interval. In addition, let's assume that, after computing $E[Idle\_p]_n$ and $E[Coll]_n$, Equation (2) does not hold. Hence, the aim of PCC is to estimate the desirable number of stations that should be active in the future in order to equate the collision and idle costs. Specifically, we denote with M_n the number of active stations that will guarantee (in the future) the balance between the idle-period length, and the collision cost, thus satisfying Equation (2).

By assuming that M_n stations will be active in the future, by applying the Taylor formula to the closed formulas, we can express $E[Idle\_p]_{n+1}$ and $E[Coll]_{n+1}$ as follows[3] ([BCG02], [CG02]):

$$E[Idle\_p]_{n+1} \approx \frac{1-p\cdot M_n}{p\cdot M_n} \quad , \quad E[Coll]_{n+1} \approx l\cdot\frac{p\cdot M_n}{2} \qquad (3)$$

where *l* is the average collision length (in time units) given that two stations collide.

By exploiting (3), PCC increases or decreases the *M* value to have $E[Idle\_p]_{n+1} = E[Coll]_{n+1}$. The following lemma directly defines M_n.

Lemma 2. *In a network adopting the p-persistent 802.11 access protocol, in which the message lengths {Li}, normalized to* t_{slot}, *are a sequence of i.i.d. random variables, the M value corresponding to the optimal operating point is:*

$$M_{opt} \cong \frac{\sqrt{1+2l}}{l\cdot p} \qquad (4)$$

Proof. Equation (4) is obtained by equating the expressions of $E[Idle\_p]_{n+1}$ and $E[Coll]_{n+1}$ defined by (3).

Lemma 2 provides to the AP the expression of the desirable *M* value. However, to exploit this expression the AP must have an estimate of the average backoff window size used in the WLAN (see Equation (1)).[4] This may be difficult to determine. To avoid this problem, in the following we present a method that computes the percentage increase/decrease of *M* without requiring any knowledge of *p*. To this end, we express M_n as a function of an unknown quantity *x*, such that

$$M_n = M_{n-1}(1+x) \qquad (5)$$

where *x* denotes the percentage increase/decrease in the M_{n-1} value.

By using Equations (5), (3) and (2), we obtain Lemma 3. This Lemma provides a closed formula to estimate the *x* value starting from quantities that are easy to estimate.

Lemma 3. *By denoting with* $E[Idle\_p]_n$ *and* $E[Coll]_n$ *the estimates of* $E[Idle\_p]$ *and* $E[Coll]$ *that the Access Point has got at the end of the n-th virtual transmission interval, the percentage increase/decrease in the M value to achieve the optimal operating point is:*

$$x = \frac{-\left(t_{slot}+2\cdot E[Coll]_n\right)+\sqrt{\left(t_{slot}\right)^2+4\cdot E[Coll]_n\cdot t_{slot}\left(1+E[Idle\_p]_n\right)}}{2E[Coll]_n} \qquad (6)$$

Proof. By substituting (5) in the expressions of $E[Idle\_p]_{n+1}$ and $E[Coll]_{n+1}$ defined by (3), and by assuming $E[Idle\_p]_{n+1} = E[Coll]_{n+1}$, after some algebraic manipulations, we obtain:

$$x = \frac{-\left(t_{slot}+2\cdot E[Coll]_n\right)\pm\sqrt{\left(t_{slot}\right)^2+4\cdot E[Coll]_n\cdot t_{slot}\left(1+E[Idle\_p]_n\right)}}{2E[Coll]_n} \qquad (7)$$

Equation (7) provides two possible solutions to the problem, but only the solution given by (6) is correct for the PCC policy. This can be verified by focusing on the case in which $E[Idle\_p]_n$ is equal to $E[Coll]_n$. In this case only by applying Equation (6) we obtain the correct value *x*=0, as the system is already in the optimal operating point.

Once the desirable percentage increase/decrease in the *M* value, the PCC computes the cost level that must be applied in the future to stimulate/discourage the users to access the system in order to approach the desired *M* value.

---

[3] The new *M* value will of course influence the future average backoff window (and hence the *p* value), hereafter we assume that the impact of the *M* change on *p* is negligible. The accuracy of this assumption increases with the increase of the *M* value.

[4] The AP can approximate this value with the estimate of the average backoff window size it has adopted in the past.

It is worth noting that Equation (6) completely defines the percentage increase/decrease in the *M* value in terms of quantities that the Access Point estimates by measurements on the network. These measurements may introduce fluctuations. To avoid harmful fluctuations, smoothing factors are normally introduced. Specifically, $E[Idle\_p]_n$, and $E[Coll]_n$ can be computed as:

$$E[Idle\_p]_n = \alpha \cdot E[Idle\_p]_{n-1} + (1-\alpha) \cdot E[Idle\_p(n)]$$
$$E[Coll]_n = \alpha \cdot E[Coll]_{n-1} + (1-\alpha) \cdot E[Coll(n)] \tag{8}$$

where α is a smoothing factor and $E[Idle\_p(n)]$ and $E[Coll(n)]$ are the average length of the idle periods and collision costs during the n-th virtual transmission time, respectively.

The use of a smoothing factor is widespread in the network protocols to obtain reliable estimates from the network estimates by avoiding harmful fluctuations, e.g. RTT estimation in TCP. Previous work has shown that $\alpha = 0.9$ is a good compromise between accuracy and promptness [CCG00b]. For this reason we assume $\alpha = 0.9$ as the default value. A similar approach can be used to avoid cost fluctuation. In this case the smoothing strategy must be modified to avoid frequent (small) modification in the transmission cost.

Step 2. The cost level that stimulate/discourage new users

In the Step 1, the AP has identified the percentage increase/decrease of *M* (i.e., *x*). In this second step, the *x* value is used to determine the transmission cost to apply for the future transmissions in order to have $M_n = M_{n-1}(1+x)$ active stations. The basis of this analysis is given by:

$$\lambda \cdot E[t_v]_n \cdot (1 - P_{give\_up}(c)) + M_{n-1} \cdot (1 - P_{give\_up}(c)) = M_n \tag{9}$$

where $E[t_v]_n$ denotes the average length of the virtual transmission time and it is estimated by applying a smoothing factor, α, see also Equation (8).

The rationale behind Formula (9) is the following. The first term in the left hand side (l.h.s.) of (9) represents the expected number of users that accept to start a new session if a transmission cost equal to *c* is applied. The second term in the l.h.s of (9) represents the number of active users that accept to continue their transmissions given that they will pay a cost per-packet equal to *c* for future transmissions.

With some algebraic manipulation of Equation (9) we finally obtain:

$$P_{give\_up}(c) = \max\left\{0, \frac{\lambda \cdot E[t_v]_n - M_{n-1} \cdot x}{\lambda \cdot E[t_v]_n + M_{n-1}}\right\} \quad . \tag{10}$$

Equation (10) identifies the new *c* value, say c_{new}, to achieve the optimal operating point. To use Equation (10) the AP requires the knowledge of M_{n-1}, i.e., the current number of active stations. In principle the AP can estimate this number by observing the MAC addresses of the stations with an active session. This is a conservative estimate because it may happen that some stations are connected to the Access Point but are not currently transmitting any data. However, we can avoid this problem, and at the same time further simplifying Equation (10), by noting that the intervals between successive updating of the price are quite small (in the order of a second). Hence, the number of new stations[5] that connects to the access point during a measurement interval is expected to be low. Therefore, we can avoid the complexity in Equation (10) by neglecting in Equation (9) the impact of stations that will wake up in the future, and simply concentrating on optimizing the system for the current number of active stations. The impact of stations that will wake up in the next future will be taken into account in the successive price computations phases. According to this simplifying assumption, Equation (9) reduces to:

$$M_{n-1} \cdot (1 - P_{give\_up}(c)) = M_n \tag{11}$$

and by substituting (5) in (9) we obtain:

$$P_{give\_up}(c) = -x \tag{12}$$

Equation (12) provides a simple way to identify the new price without requiring any explicit knowledge of the number of the active stations. Unfortunately, (big) errors can be introduced if $P_{give\_up}(c)$ is not a continuous function in the range [0,1]. For example in the case in which all users have the same threshold, say $\bar{c}$, $P_{give\_up}(c)$ is:

$$P_{give\_up}(c) = \begin{cases} 0 & c < \bar{c} \\ 1 & c \geq \bar{c} \end{cases} \tag{13}$$

In this case, given a value *x* (0<*x*<1), any choice of the *c* value introduces a big approximation with negative effects on the overall system behavior. If the AP announces a price *c*, such that $c < \bar{c}$, the system remains congested (as no station gives

[5] This number includes also the stations that reconnect after a give-up period.

up) and the QoS remains low. On the other hand, by announcing a price c, such that $c \geq \bar{c}$, all stations give up and the system utilization, and hence the provider revenue, drops to zero.

The problem pointed out in the previous example occurs whenever $P_{give\_up}(c)$ is defined only for a (small) subset of the price values. Hence given a value x it may not exist a c value that satisfies $P_{give\_up}(c) = -x$ but we have two possible approximations, c_{low} and c_{up}, such that: $P_{give\_up}(c_{low}) = -x_{low}$, $P_{give\_up}(c_{up}) = -x_{up}$ and $x_{low} < x < x_{up}$.

To eliminate (or at least to reduce this error) we propose to use selective announcements to a randomly selected number of users. Specifically, the algorithm operates as follows:

- the AP selects the price c_{up};
- for each station that is currently connected, the AP announces the new price according the following probability distribution:

$$\Pr\{c = c_{low}\} = 1 - \frac{x}{x_{up}} \quad , \quad \Pr\{c = c_{up}\} = \frac{x}{x_{up}}$$

For example, in the 0-1 threshold case (see Equation ()), $x_{up} = 1$, and if we wish to obtain a $x = 0.40$ reduction in the number of active stations, the AP announces the high cost, $c_{up} = \bar{c}$, to $x/x_{up} = 0.40$ of the stations.

This policy based on selective announcements to randomly selected stations may be questionable in a real Wi-Fi hot spot. However, the focus of this work is to provide some theoretical foundations to define price-based congestion control policies in a Wi-Fi environment. The discussion of how such a policy can be implemented in a real environment is for further studies.

In the following sections we show the effectiveness of PCC policy by simulating the performance of a Wi-Fi network with or without the PCC policy.

4. PCC Performance Analysis

In this section we will extensively validate the capacity of the PCC policy to drive a Wi-Fi network close to the optimal operating point. We investigate the performance of a Wi-Fi hot spot in which there are *M* stations. A station can be either active, or in the sleep state. An active station accesses the channel whenever it has a frame to transmit. A station enters the sleep state when the AP announces a transmission price that is too high for the station to accept. Specifically, in our experiments we randomly assigned to each station a threshold for the maximum price it accepts to pay. When the price exceeds its threshold a station enters in the sleep state, and all packets in its queue are discarded. A station remains in the sleep state for an exponentially distributed random time. The average “sleep time” is the same for all stations and it is equal to 1 sec. When the station wakes up it starts to fill its queue again.

In our study, the maximum price accepted by a station is sampled from a truncated (to zero) normal distribution with average 50 and a standard deviation equal to 10, if not mentioned otherwise.

We perform two types of experiments: *saturated* and *not-saturated*. In saturated mode each station always has a non-empty queue, and therefore it is always willing to transmit. In not-saturated mode, each station generates new packets to transmit according to a Poisson distribution. The rate of the Poisson distribution is selected such that a station's bandwidth requirement is only a small fraction of the channel capacity. A large number of stations is therefore required to saturate the channel. Hence, in the not-saturated model, each station alternates between ON (i.e., with packets to transmit), and OFF (i.e., no packets to transmit) states.

The performance indices computed via simulation, have been estimated with the independent-replication technique. The deviation among the estimates achieved in the different runs are generally very small, and hence the error-bars are often not visible.

As said at the beginning of Section 3, the PCC policy may operate at different time instants. Hereafter, we analyze two different possibilities: updating at the virtual transmission time boundaries (denoted as Tv), and periodic updating (denoted as Tp). The periodic updates occur every 102.4 msec, i.e., a typical value for Wi-Fi beacon signal frequency.

4.1 Long packets

The main target of this performance study is to investigate the relationship between the channel utilization and the network congestion. In our experiments, several congestion levels are simulated by running a set of experiments with different *M* values.

As explained in Section 2, the MAC data frame contains in addition to the MAC_{hdr}, a variable length *data payload*, hereafter referred to as packet. The results presented in this section are obtained with packets sizes sampled from a normal distribution with an average packet size of 1500 bytes (value typical for wired networks). The minimum packet size is 0, while the maximum size of 2130 bytes is below the limit of the Wi-Fi standard, i.e., 2312 bytes.

The effectiveness of the proposed PCC policy is shown in Figures 3 and 4. These figures show the channel utilization levels achieved by adopting the PCC system, and compare this index with the utilization levels achieved with the legacy Wi-Fi approach (*no pricing*). The results show that the PCC policy significantly increases the Wi-Fi channel utilization as the network congestion increases. Simulative runs performed in our not-saturated scenario show that for networks with less than M=15 stations the price mechanism is never active, and the performance with or without the PCC policy are similar. After this congestion level, the PCC mechanism becomes more and more effective. In this case there are not significant differences between the Tv and the Tp policies. On the other hand, results under saturated conditions show interesting and relevant differences between the Tv and the Tp policies. Specifically, we observe that in few "unlucky" runs the PCC policy with the Tv updating enters in a deadlock state. In detail, for small M values, it may happen that during a run all the stations give up due the price announced by the access point. In this case, a virtual transmission interval never ends (there are no successful transmissions), and hence no new price announcement will occur. This is the reason of the very wide error-bars that appear in Figure 4 associated to Tv plots.

The deadlock problem cannot occur with the Tp updating policy, and hence in this case the error-bars are always very small.

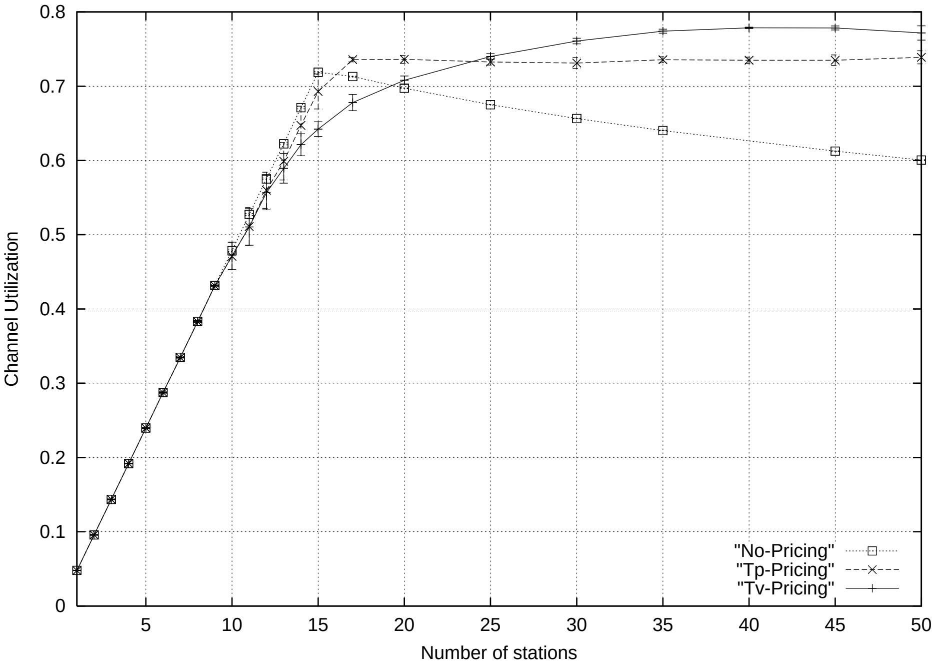

Figure 3: not saturated (long packets)

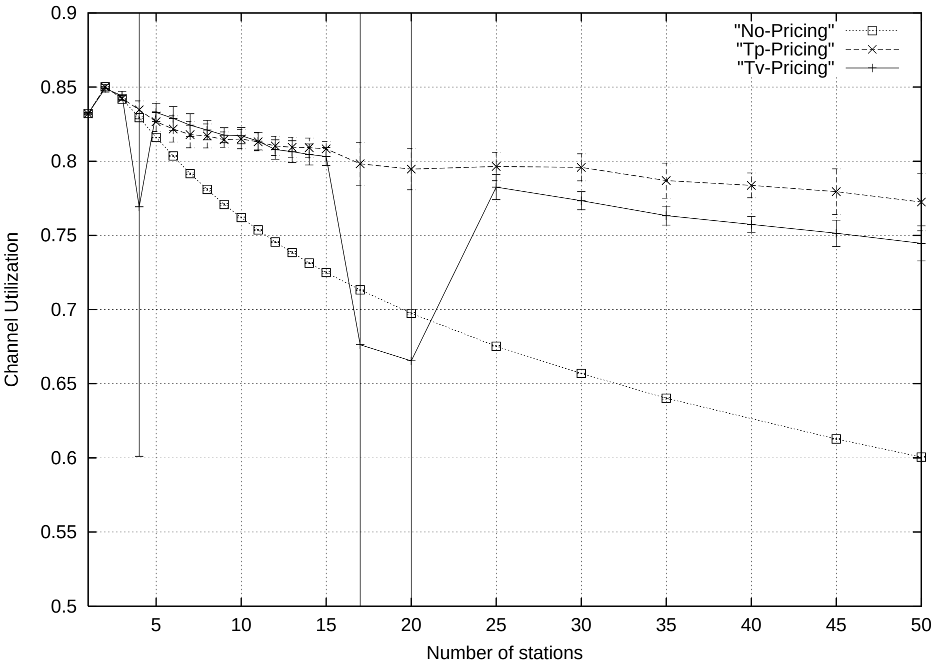

Figure 4: saturated (long packets)

4.2 Mixed-packets

In the previous section we assumed a traffic made of long messages only. As shown by TCP-traffic studies [Ste 94], on a byte count basis, 90% of the traffic is made up of maximum size packets while the remaining 10% consists of very short packets. This means that, on a packet-count basis about 50% of the packets are long messages (up to 1500 bytes), and the remaining 50% are short packets 40-byte long. Hence, in a more realistic environment the message length distribution should be bimodal. For this reason, hereafter we evaluate our system by assuming a mixed distribution of the packet lengths. In detail, we assume 50% long messages and 50% short messages. Short messages have a 2-slot constant length (about 40 bytes of payload). The length of long messages is sampled from the same normal distribution used in Section 4.1.

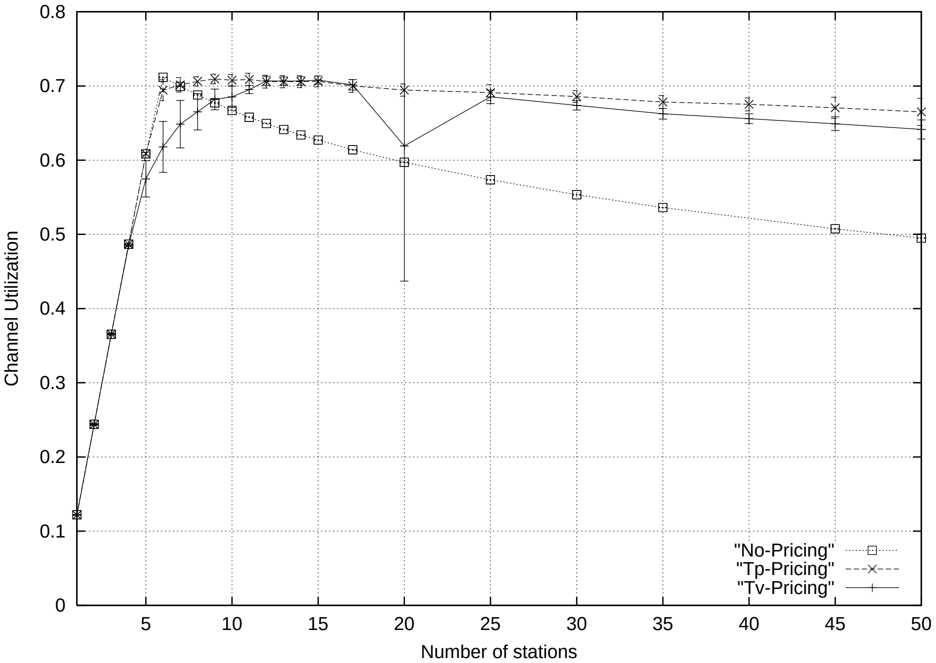

Figure 5: not-saturated scenario (mixed traffic)

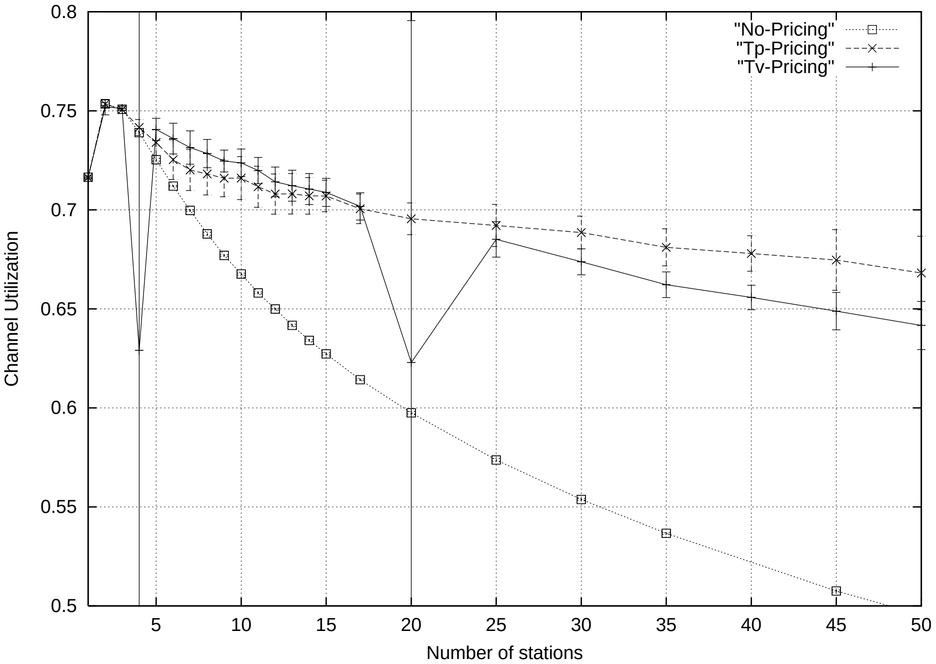

Figure 6: Saturated scenario (mixed traffic)

From a qualitative standpoint, Figures 5 and 6 confirm the behavior observed with long packets, only:

i) the PCC policy is effective when the network congestion increases;
ii) the Tv and Tp updating policies have the same asymptotic (with respect to M) behavior;
iii) the Tp policy is always deadlock free, while the Tv policy may produce deadlocks both in the saturated, and in the not-saturated conditions. These events are highlighted by wide error-bars on the Tv plots and correspond to deadlocks occurred in some simulative runs.

5. Conclusions

In this paper we proposed and evaluated a *Price-based Congestion Control (PCC)* policy that can be applied to dynamically control the network contention level in a Wi-Fi hot spot. The basic idea is to control the traffic inside the hot spot by determining the access cost as a function of the current load in the hot spot. By increasing/decreasing the per-packet transmission cost, the hot spot operator encourages/discourages her/his customers. In the paper we developed a theoretical framework to compute the access cost to maintain the hot spot in its optimal operating point, for any load condition. Via simulation, we evaluated the performance of a Wi-Fi hot spot with or without the PCC policy. The performance analysis was carried out both in saturated and not-saturated conditions. Results obtained show that the PCC policy is effective for all the network and traffic configurations analyzed. In detail, we considered two implementations of PCC depending on the event that triggers the notification of the new price to the users: periodic and virtual-transmission-time updating. The results show that the periodic updating policy is robust while the virtual-transmission-time updating may generate deadlock conditions.

Several assumptions used in our implementation of the PCC policy may be considered unrealistic (e.g., detailed knowledge of the channel status in terms of idle and collision periods) or too complex (e.g., the per-packet transmission cost) for a real setting. It is worth pointing out that the main scope of this paper is to demonstrate the feasibility of a price-based policy for congestion control in a Wi-Fi hot spot. In this sense, our results constitute a theoretical foundation for price-based congestion control in Wi-Fi networks. The definition of a realistic implementation of a PCC policy is beyond the scope of this paper and it represents the next step of this research activity.

References

[BCG00b] R. Bruno , Conti M., E. Gregori, "Optimal Capacity of p-persistent CSMA Protocols", IEEE Comm. Letters (to appear).

[CCG00a] F. Cali' , Conti M., E. Gregori, "Dynamic Tuning of the IEEE 802.11 Protocol to Achieve a Theoretical Throughput Limit", IEEE/ACM Transactions on Networking, Volume 8 , No. 6 (Dec. 2000), pp. 785 - 799.

[CCG00b] F. Cali' , Conti M., E. Gregori, "Dynamic IEEE 802.11: design, modeling and performance evaluation", IEEE Journal on Selected Areas in Communications, 18(9), September 2000. pp. 1774-1786.

[CG02] M.Conti, E. Gregori, “Optimization of bandwidth and energy consumption in Wireless Local Area Networks” in *Performance Evaluation of Complex Systems: Techniques and Tools* (Performance 2002 Tutorial Lectures), LNCS 2459, M. Calzarossa, S. Tucci (Editors), pp. 435-462.

[Cro02] J. Crowcroft, R. Gibbens, S. Hailes, “BOURSE - Broadband Organisation of Unregulated Radio Systems through Economics", http://www.cl.cam.ac.uk/~jac22/out/bourse.pdf

[IEEE802] Official Homepage of the IEEE 802.11 Working Group, http://grouper.ieee.org/groups/802/11/.

[IEEE11b] IEEE standard 802.11b-1999, “Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Higher-Speed Physical Layer Extension, in the 2GHz Band.

[Ste94] W.R.Stevens. TCP/IP Illustrated, Volume 1: The Protocols, Addison-Wesley, Reading, MA, 1994.