RELIABLE DOWNLINK SCHEDULING FOR WIRELESS NETWORKS WITH REAL-TIME AND NON-REAL TIME CLIENTS

A Thesis

by

ABHISHEK JAIN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	I-Hong Hou
Committee Members,	Alex Sprintson
	Radu Stoleru
Head of Department,	Chanan Singh

August 2014

Major Subject: Computer Engineering

ABSTRACT

In this thesis, we studied the problem of designing a down link scheduling policy to serve multiple types of clients from a base station in a time-varying wireless network. An ideal scheduling policy is fair among the clients, provides reliability to the clients, achieves high system throughput and prevents strategic clients from choosing incorrect means. The existing scheduling policies fail to achieve one or more of these features. The Proportional Fair scheduling policy for example, fails to provide reliability to the real time clients, while Round Robin policy provides reliability to the clients but fails to achieve high system throughput in a time-varying wireless network. Apart from these policies, there are scheduling policies which prioritize clients based on their delay requirements. Here, a client with lower priority may choose incorrect means like claiming false types of flows to obtain a better performance. A non-real time client may pretend to be a real time client if doing so, which might aid it to achieve better performance in terms of average throughput.

We proposed a new scheduling policy that is not only proportionally fair but also provides reliability to the mixture of real time and non-real time clients over a shared wireless channel. Our proposed policy aims to serve clients with different service requirements and provides best service to the clients which furnish true information about their service requirements; the client claiming false service requirements is penalized with the reduced performance.

We theoretically demonstrate the effectiveness of the algorithm by considering uniform distribution of service rates of all the clients. We then provide extensive simulation results of our scheduling policy under the fast fading Rayleigh model to show that this policy can be easily extended in wireless networks. We also show that

our policy outperforms existing policies in providing better reliability to the clients and unlike other common policies, our policy degrades the performance of a client that chooses incorrect means.

To my parents

ACKNOWLEDGEMENTS

I would like to thank my advisor Professor I-Hong Hou for his invaluable and continuous support in making this work possible. I thank him for the opportunity and freedom he gave me for working on an entirely new topic within his research group. A special thanks to Dr. Alex Sprintson and Dr. Radu Stoleru for serving on my committee.

I would like to thank my loved ones, who have supported me throughout the entire process.

NOMENCLATURE

RT	Real Time
NRT	Non-Real Time
BS	Base Station
PF	Proportional Fair
RR	Round Robin

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

1. INTRODUCTION

1.1 Problem

The rapid proliferation of Internet has led to high data service rates over wireless networks in both real and non-real time applications. Typically, the traffic over wireless networks is a mixture of real time applications such as Skype, video streaming and non-real time traffic such as emails. These real time applications have varying service requirements. These applications not only require constant expected throughput to be served with, but also prefer to have high reliabilities. In other words, they prefer to have high assurance of the throughput received rather than receiving high throughput with low assurance. For example, Skype Clients and VoIP generate packets at different time intervals and have different thresholds of throughput to be serviced with. Skype Client will prefer to get served with 95% assurance that it will receive throughput greater than 256 Kbps than being served with 1 Mbps throughput with 50% assurance at all times. Non-real time applications do not have such delay requirements and always aim at maximizing average throughput.

The problem of providing QoS to different types of wireless clients in a wireless network can be sub-divided into two parts; admission control and scheduling. In admission control, one finds the number of clients that can be served in the network provided the desired QoS for various clients are served. On the other hand, in the scheduling problem, a mechanism is implemented to allocate the clients that have been admitted using admission control over a shared wireless channel. In other words, a client is picked among the pool of clients at each time slot over a shared wireless channel. Our study is focused on the scheduling problem of a mixture of real time and non-real time clients and seeks to address the following issues:

1. The mechanism to provide different QoS to a mixture of real time and non-real time clients over a shared wireless channel.

2. Almost all the existing scheduling policies which address the above issue assume that the client will provide its true service requirements to receive its desired QoS. Therefore, it may happen that a strategic client may claim a false service requirement to achieve better performance. Our work intends to provide a mechanism to motivate these clients to tell the truth about their service requirements [1].

1.2 Motivation

The problems mentioned in 1.1 are motivated by the following considerations:

1. Since real time clients have myriad demands due to different applications they run, they try to obtain better service either by paying more or by faking their service requirements, which is unfair to other clients. To our knowledge, there is no scheduling policy which can prevent the clients from choosing such incorrect means.

2. The channel between the base station and the clients of wireless networks is time-varying due to fading and shadowing effects. Hence, the effective scheduling policy should incorporate the real characteristics of the wireless networks.

3. Finally, the scheduling policy should provide high system throughput as well as reliability to real time clients since they require constant expected throughput. Existing policies like the Round Robin provides reliability to real time clients, however, fails to provide high system throughput.

1.3 Our Contribution

In this thesis, we aim to design a new scheduling policy that considers different types of clients with varying service requirements. Our policy discourages strategic clients from claiming false service requirements by penalizing their performance. It also provides fairness to the clients as no client has to pay more to achieve better service. The policy lets the clients compete for the shared wireless channel at each time slot, however, if the real time clients fail to get serviced for their minimum number of slots, they are serviced at the end of the time interval. Consequently, our scheduling policy provides reliability in the throughput distribution of the real time clients. On the other hand, non-real time clients would achieve high throughput at the cost of unreliability in throughput distribution.

1.4 Brief Summary

We first explain a client-server system model that constitutes the characteristics of a wireless network. We have considered a cellular tower as the server and a mobile user as the client. This generic model can be applied across wide range of applications like mobile cellular networks and real time surveillance. We define the performance of a non-real time client by the long term average throughput it receives, while that of a real time client is measured using a "more reliable" relation in its throughput distribution discussed later. Under this model, we address the problem of scheduling a mixture of real and non-real time clients while providing high system throughput, such that reliability is served in the throughput distribution of the real time clients.

In this model, we have considered the channel between the base station and the client to be time-varying, which can be attributed to fading. If a client moves away from the base station, the channel quality between the base station and client decreases, since the client is mobile. To incorporate the real characteristics of a wireless network, we consider Rayleigh fading channel model [2] and assume that the service rate of each client changes over time on the scale of fast fading.

In practice, service requirements are known only to the clients and therefore strategic clients may falsify their true service requirements to receive better services. To ensure that the clients reveal their true service requirements, we design a policy where the clients announce their service requirements at the admission time. The base station then schedules the clients based on their service requirements. We propose a scheduling algorithm to serve both the real time clients with different service requirements and non-real time clients. Our proposed scheduling policy provides the best service to the clients that tell the truth about their required services, automatically penalizing clients that lie about their service. We demonstrate that our algorithm prevents the client from falsifying its service requirements by presenting the theoretical analysis while considering uniform distribution of service rates of all the clients. Next, we provide extensive simulation results of our scheduling policy for multiple real time clients and non-real time clients with varied service periods under the fast fading Rayleigh model. The results demonstrate that our policy can be easily extended to wireless networks. Finally, we compare our proposed policy against state-of-the art policies and show that the compared policies fail to provide the reliability in throughput distribution of the clients which can result in serious unfairness.

1.5 Organization

The rest of the thesis is organized as follows. Chapter II summarizes related work on the scheduling policies for a mixture of real time and non-real time clients. Chapter III describes a system model for time-varying wireless networks. In chapter IV, we describe our proposed scheduling policy that solves the addressed problems. In chapter V we theoretically analyze the effectiveness of our scheduling policy, which is to let clients report their true service requirements. Simulation results are presented in chapter VI, with the conclusion in chapter VII.

2. RELATED WORK

Existing scheduling policies like round robin (RR) provide reliability to the clients by serving them in a cyclic order, but fails to take advantage of the varying channel between the base station and client. Therefore, it fails to optimize maximum average throughput of the network. Greedy scheduling policies which focus on serving the clients with best channel conditions, fail to provide fairness to the clients with bad channel conditions. On the other hand, Proportional Fair (PF) scheduling policy [3] considers channel conditions between the server and the clients and provides fairness to the clients. The drawback with the PF policy is that it does not consider the factor for providing guaranteed service to the real time clients and therefore fails to meet the requirements for delay-sensitive networks [4]. Therefore, a new scheduling policy is desired which provides service to the clients based on their service rates, delay constraints while being fair to the clients.

The challenge of providing better service quality to both real time and non-real time clients over time-varying fading wireless channels has been an interesting area of research in recent years. Hou, Borkar, and Kumar [5] have proposed a mathematical model for providing service to the client based on its QoS, which is defined by delay constraint, delivery ratios and channel reliabilities. Hou extended his mathematical model to incorporate handling of variable bit rate applications [6]. Shakkottai [7] worked on the problem of scheduling real time clients with varied time constraints over the shared wireless channel. Earlier works [8, 9] have proposed different scheduling policies to serve the mixture of real time and non-real time clients in wireless networks. For example, the work [10] designed a scheduling policy based on the channel rate for serving delay-sensitive wireless networks. In the work [11], Wang proposed a framework to provide guaranteed services for the real time clients over the time-varying wireless networks. Li [12] proposed a mathematical model to serve different types of flow in the wireless networks. The framework prioritized the real time flows. They also designed the scheduling policy that maximized the average throughput of non real time flows and satisfied the service requirements of the real time flows. In the work [13] Patil has proposed opportunistic scheduling policy for a mixture of real time and non-real time clients in time-varying wireless networks. The work is based on the prediction of channel rate which a real time client would receive in the near future. Haci [14] proposed the scheduling policy by giving varying weights to the different types of clients. In [15], Borst analyzed the performance of a mixture of different types of clients in wireless networks using different scheduling approaches. In his work on efficient scheduler design for wireless networks [16], Uc-Rios maximized the number of clients which have their packets received before their deadlines rather than maximizing the system throughput. Based on this idea, they proposed a policy to schedule a mixture of different types of clients on the shared wireless channel. Jaramillo [17] extended their framework which provides fairness to non-real time clients, to include the QoS requirements of real time clients.

The service requirements for real time clients include delay constraints and multifarious regular inter-service times. All the above mentioned works aimed to serve clients based only on the delay requirements. Recently, in the works [18, 19, 20], Li have addressed the significance of real time clients being served regularly. The conventional scheduling policies never considered to improve the quality of the throughput received by the client. They aimed at meeting up the client's delay requirement and optimizing the system throughput. Consequently, these policies fail to provide regularities in the served flows as clients experience huge variations in the service received. Unlike the traditional policies, our work is intended to provide reliabilities in the service requirements of the real time clients.

The central assumption in all the mentioned works has been observed to be the truthfulness of the client in informing the BS about its service requirement. Since clients have varied service requirements, a strategic client may exploit the current scheduling policies and falsely report its service requirement to get better services. The works [21], [22] and [23] addressed the problem of selfish clients exploiting the 802.11 based wireless networks. In the work [1], Hou addressed this problem and proposed the policy for a mixture of real time and non-real time clients where clients are given benefits upon telling their true service requirements. Kang has studied an admission control mechanism which prevents strategic clients from choosing incorrect means in [24]. He proposed a non-monetary mechanism where clients are being admitted based on their true service requirements. The mechanism encourages selfish clients to tell the truth of their service requirements. Kavitha [25, 26] addressed the problems where the mobile clients falsify their channel states to the BS. The author designed a game based iterative policy to provide fairness among the mobile clients in the cellular network. Kavitha in the work [27] designed a scheduling policy for OFDMA system in which clients are penalized for falsifying service rates. The authors have addressed the problem where the clients falsely report their channel states to achieve better services in [28, 29]. The works [30, 31, 32, 33] have addressed the credit-based solution to prevent selfish clients from choosing not to cooperate in the wireless networks. This solution can be modified to promote the strategic clients from lying about their service requirements. In [34] Hou proposed a non-monetary mechanism in D2D networks that provides incentives to the clients for telling the truth. Compared to all the mentioned works, our work involves designing a scheduling policy which is not only fair among clients, provides reliability to real time clients, achieves high system throughput but also prevents strategic clients from choosing such incorrect means.

Current scheduling policies based on Diffserv [35] attempt to serve real time clients by providing them with higher priority. Therefore, lower priority clients may choose wrong methods to achieve better services [36, 37]. To prevent such clients from exploiting current scheduling policies, research has been done in providing an auction for the shared wireless channel [38]. Other solution to the above problem which exists is providing incentives to the lower priority clients [37], [39]. In an auction design, clients bid the shared wireless channel at regular time-intervals. The server selects the clients based on their bids and charges them. The higher priority clients pay more to achieve privileged services on the shared wireless channel. VCG auction [40] is one of the basic models on which most of the existing models are based. Myserson [41] worked out a mechanism for optimal auctions to address variety of problems based on auction design. Hou [42] proposed an auction mechanism which prevents selfish clients from lying their utility functions. In the work [43], Nuggehalli proposed a pricing scheme with an aim of achieving proportional fairness. They have assumed that the utility function of the client is known to the BS, and in the case of strategic client, if the utility function is not known, their pricing policy works on the principle of Nash equilibrium to achieve proportional fairness.

In this thesis, we aim to design a new scheduling policy that provides reliability to real time clients and discourage strategic clients to choose incorrect means. Unlike auction mechanism, we provide fairness to all the clients and no client is charged more to receive better services.

3. SYSTEM MODEL

3.1 General Model

In our model, we consider 4G cellular base station (BS) serving N clients operating cellular phones. We consider $K, (K \leq N)$ real time clients. Time is slotted and is expressed as $t \,\epsilon\, \{1, 2, 3 \ldots\}$. At each time slot BS serves only one client. The considered model is general enough to be implemented in IEEE 802.11e based wireless network [44].

3.2 Wireless Channel Model

We consider fading wireless channel and assume that the channel change over time on the scale of fast fading. The channel is considered to be constant in a time slot. We define the service rate $r_n(t)$ as the number of bits which a client n can receive at time t if it is being served. The value of $r_n(t)$ may change from time slot to time slot. The number of bits which a client n receives in time t can be represented as:

$$y_n(t) = \begin{cases} r_n(t), & \text{if client } n \text{ is scheduled,} \\ 0, & \text{otherwise} \end{cases}$$

3.3 Types of Clients

Each client informs about its type of flow to the BS at its admission time, i.e.($t = 0$). The type of flow can be either real time or non-real time traffic. At the admission time, if a client chooses real time flow, then each real time client (RT) is required to specify its *service period* to the BS, defined as the period in which throughput of that RT client is measured. Because of diverse real time applications, each RT client has different service requirements, and hence, we allow each client to have unique service period. For example, studies [45] reveal that, Skype Client generates packets at intervals that are a multiple of 16 milliseconds, while VoIP traffic generated by a Cisco IP Phone system generates packets at inter-arrival times of 8 to 15 milliseconds [46]. It is practical to measure the throughput of Skype Client in a service period of multiple of 16 milliseconds, while measure the throughput of VoIP traffic generated by a Cisco IP Phone system in a service period of 12 milliseconds.

Real time applications require high reliabilities in the throughput received. They expect a constant threshold of service to be received from the BS. For example, consider live-streaming applications like remote surgery, where the actions are performed based on the data received from the network. If the BS fails to provide high reliability in the throughput received by the respected application, the results can be fairly deteriorating. Hence, it is required by the BS to provide minimum service guarantee to RT clients in their required service periods.

We assume, that RT client n has a service period d_nN, where $d_n \geq 0 \, and n \leq K$. The service period of non-real time client (NRT) is ∞. It is inferred that, if the RT client n does not have its service period in the form of d_nN, then either the BS approximates the client's service period to the nearest integer divisible by N or if the RT client prefers not to change its service period, then it is kicked out from the pool of clients to be served by the BS. If a new RT client is added to the pool of clients being served by the BS, then the BS can prefer not to serve the new client if its service requirements are not met. As we increase the service period of a RT client, $d_nN \rightarrow \infty$, it changes to NRT client. The length of a *frame*, which we denote by T, is the least common multiple of service periods $\{d_1N \dots d_KN\}$ of all RT clients. Thus, a frame consists of T/d_nN service periods of RT_n client. All the necessary variables are defined in Table 3.1.

3.4 Performance for Different Types of Client

The performance of RT clients and NRT clients is measured differently.

Definition 1: The performance of a NRT client n is measured by its long-term *average throughput*, defined as $lim\inf_{t\to\infty}\frac{\sum_{\tau=1}^{t} y_n(\tau)}{t}$

On the other hand, RT client also wants to have high throughput and high reliability. Since, the traffic generated in a real time application is interdependent and has stringent end-to-end delay requirements, therefore it requires a constant expected throughput in its service period. Hence, it is necessary for RT clients to measure their performance regularly and differently. Consider RT client n has a service period of d_nN. Let $q_n(k)$ be the throughput obtained by RT client n in the k^{th} service period,

$$q_n(k) = \frac{\sum_{\tau=(k-1)d_nN}^{kd_nN} y_n(\tau)}{d_nN}$$

It is noted that, $q_n(k)$ is a random number obtained in k^{th} service period of RT client n. We derive the long-term CDF of throughput $[q_n(1),\ q_n(2),\ldots,q_n\lceil t/d_nN\rceil]$ for RT client n. The performance of RT client n depends on the CDF obtained. We define a relation to compare different distributions of $[q_n(k)]$ as follows:

Definition 2: Consider two distributions X and Y. X is said to be "more reliable" than Y if there exists $d > 0$ such that,

$$Prob(X \geq d) > Prob(Y \geq d)$$

and for all $c \leq d$,

$$Prob(X \geq c)\ \geq Prob(Y \geq c)$$

We denote the "more reliable" relation by $\geq_R$. As mentioned in 3.3, RT clients have high requirements for reliability; therefore it is practical for RT clients to measure their throughput in the appropriate service periods. We know that, Skype Client generates packets at inter-arrival time of 16 milliseconds. So, Skype Client can measure its throughput distribution in different service periods like that of 10 milliseconds, 20 milliseconds. and ...so on. Skype client can then compare which distribution is "more reliable" in different service periods to achieve maximum probability of receiving higher throughput. For example in Fig. 3.1, we compare the performance of client n between the two throughput CDF: A and B. It is observed that, client n under distribution A will receive throughput $> d$ with probability 0.8, while under distribution B it will receive throughput $> d$ with probability 0.75. Hence, for client n distribution A is "more reliable" than distribution B because, $P(A > c) > P(B > c)$ and $P(A > d) > P(B > d)$, for all $c \leq d$.

LEMMA 1: The $\geq_R$ relation defines a total ordering among different CDFs.

Proof: Consider a set of throughput distributions obtained for RT client n under different service period as $\{X_A, X_B \ldots X_\alpha\}$. We denote the set by X. Under total ordering, the distributions should have a binary relation that is total, transitive and antisymmetric. We prove independently all the three properties for the binary relation to be total order as follows:

- The set X holds the property of transitivity.

  Consider three different distributions X_A, X_B and X_C in set X such that, $X_A \geq_R X_B$ and $X_B \geq_R X_C$.

  Given $X_A \geq_R X_B$ implies that $\exists d_1$, such that $Prob(X_A \geq d_1) > Prob(X_B \geq d_1)$ and $\forall\, c \leq d_1$, $Prob(X_A \geq c) \geq Prob(X_B \geq c)$.

  $X_B \geq_R X_C$ implies that $\exists\, d_2$, such that $Prob(X_B \geq d_2) > Prob(X_C \geq d_2)$ and $\forall\, c \leq d_2$, $Prob(X_B \geq c) \geq Prob(X_C \geq c)$.

  Let $d = min\,(d_1, d_2)$.

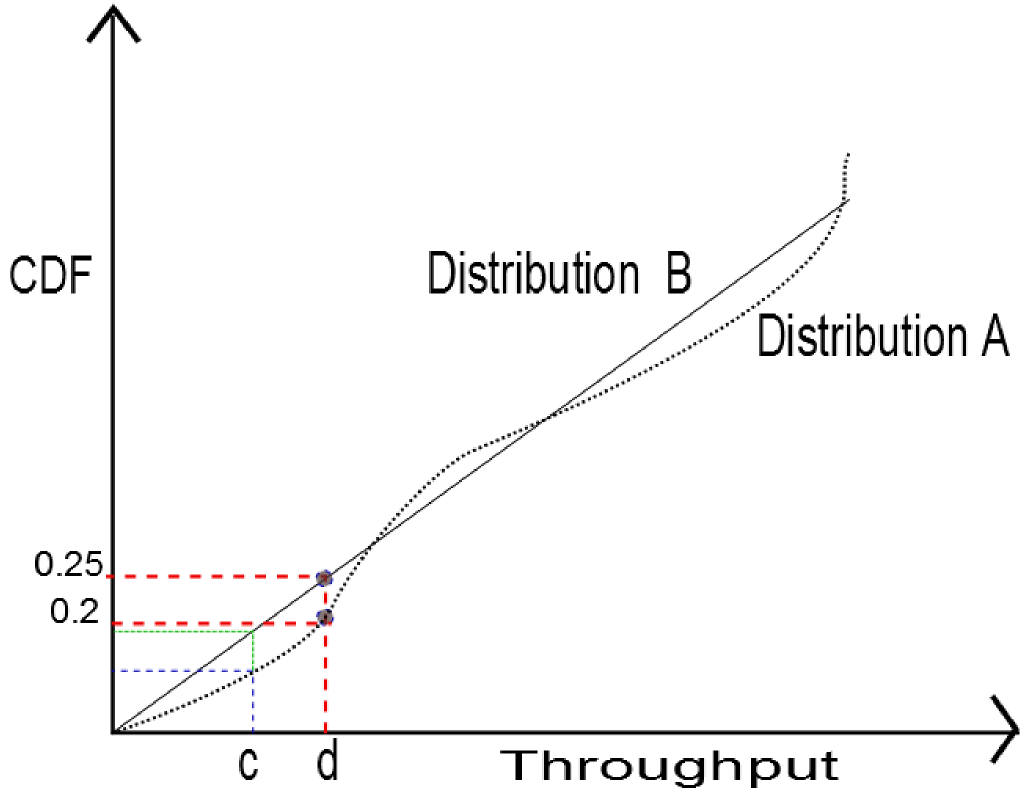

Figure 3.1: CDF Plot of Two Distributions, A and B of RT Client

If $d = d_1, \implies d < d_2,$

$$\implies Prob(X_A \geq d) > Prob(X_B \geq d) \geq Prob(X_c \geq d)$$

If $d = d_2, \implies d < d_1,$

$$\implies Prob(X_A \geq d) \geq Prob(X_B \geq d) > Prob(X_c \geq d)$$

$\forall c < d$, $c < d_1$, $c < d_2$, we have $Prob(X_A \geq c) \geq Prob(X_B \geq c) \geq Prob(X_c \geq$

c). Hence, it is observed that if $X_A \geq_R X_B$ and $X_B \geq_R X_C$, then $X_A \geq_R X_C$.

- The set X holds the property of antisymmetry.

  Consider two different distributions $X_A$and X_B in set X such that, $X_A \geq_R X_B$.

  Given $X_A \geq_R X_B$ implies that $\exists\, d_1$, such that,

$$Prob(X_A \geq d_1)\ > Prob(X_B \geq d_1)$$

  and $\forall\, c \leq d_1$

$$Prob(X_A \geq c) \geq Prob(X_B \geq c)$$

  So, assuming if X is **not** antisymmetric, then, $X_B \geq_R X_A$ exists, which implies that that $\exists\, d_2$, such that,

$$Prob(X_B \geq d_2)\ > Prob(X_A \geq d_2)$$

  and $\forall\, c \leq d_2$,

$$Prob(X_B \geq c) \geq Prob(X_A \geq c)$$

  Let $d = min\ (d_1, d_2)$

  If $d = d_1 \implies d < d_2$

$$\implies Prob(X_A \geq d)\ >\ Prob(X_B \geq d)$$

  from $X_B \geq_R X_A$, $X_A \geq_R X_B$ and $\forall\, c \leq d_1, d_2$,

$$Prob(X_B \geq c)\ \geq\ Prob(X_A \geq c) \tag{3.1}$$

and

$$Prob(X_A \geq c) \ \geq \ Prob(X_B \geq c) \tag{3.2}$$

The two equation 3.1 and 3.2 are only possible if $X_A = X_B$

If $d = d_2, \implies d < d_1$

$$\implies Prob(X_B \geq d) \ > \ Prob(X_A \geq d)$$

from $X_B \geq_R X_A$, $X_A \geq_R X_B$ and $\forall\, c \leq d_1, d_2$

$$Prob(X_B \geq c) \ \geq \ Prob(X_A \geq c) \tag{3.3}$$

and

$$Prob(X_A \geq c) \ \geq \ Prob(X_B \geq c) \tag{3.4}$$

The two equation 3.3 and 3.4 are only possible if $X_A = X_B$. Hence, the relation is antisymmetric.

- The set X holds the property of totality.

  Let $d_1 > 0$ then, either of the two relations is possible:

$$Prob\,(X_A \geq d_1) \ > \ Prob(X_B \geq d_1) \tag{3.5}$$

$$Prob\,(X_A \geq d_1) \ \leq \ Prob(X_B \geq d_1) \tag{3.6}$$

  and if $\forall\, c \leq d_1$ in equation 3.5 we have

$$Prob(X_A \geq c) \ \geq \ Prob(X_B \geq c) \tag{3.7}$$

then by definition 3.4, we say that $X_A \geq_R X_B$. But if equation 3.7 does **not** hold for $\forall\, c \leq d_1$ then we say that $X_A \ngeq_R X_B$.

By a similar argument, in equation 3.6 if $\exists\, d_2$, such that, $\forall\, d_1 \leq d_2$, we have

$$Prob(X_B \geq d_2) \;>\; Prob(X_A \geq d_2)$$

then by definition 2 we comment that $X_B \geq_R X_A$. But if equation 3.7 does **not** hold for $\forall\, d_1 \leq d_2$ in equation 3.6 then we comment that $X_B \ngeq_R X_A$.

Thus, it is observed that given 3.4, then either $X_A \geq_R X_B$ or $X_B \geq_R X_A$ holds.

Since, all the three relations, i.e. antisymmetric, transitiveness, and totality exist, therefore the $\geq_R$ follows total ordering among different CDFs..

3.5 Strategic Clients

We assume that all clients report their service periods to the BS during the admission time and the service periods of all clients remain constant at all times. The BS under some scheduling policy η, chooses to serve a client at time t based on $r_n(t)$, d_nN and clients' past history. Current scheduling policies based on DiffServ [35], prioritizes clients based on their delay requirements. In other words, RT clients are given higher priorities over NRT clients in getting served by the BS. These policies provide guarantee in meeting the delay requirements of the RT clients by serving RT clients more often than NRT clients. Also, RT clients which have very low delay requirements are serviced more frequently as compared to the RT clients which have higher delay requirements. It is observed that the drawback of these policies is, they do not analyze the packets for actual delay requirements and assume that the clients are providing their true service requirements to the BS.

Considering clients are strategic, they may report false service periods to get better performances. To be more specific, NRT client may pretend to be a RT client if doing so increases its average throughput. Similarly, a RT client may report a false service period if doing so its distribution of $[q_n(k)]$ is more reliable. We demonstrate an example, where RT client n lies its service period to achieve better performance in terms of reliability. Fig. 3.2 shows an example where a client n is considered in two different distributions A and B. In distribution B, client n has service period of $2N$ slots and in other distribution A, client n claims its service period to the BS to be N slots. In both the cases, value of N is 3. Even though client n will measure its throughput in service period $2N$, but the BS will service client n based on the reported service period, which in this example in two different distributions, is N and $2N$. The BS serves clients based on weighted round-robin scheduling policy [47]. Weights are given in increasing order of decreasing value of service periods of clients, i.e. client with service period N is given weight equal to 3, while client with service period $2N$ is given weight equal to 2. All NRT clients are given weight as 1. Therefore in case of RT client n claiming its service period to be N, client will be serviced more often in a frame and receives better performance in terms of reliability as compared to the case, when client n reports the service period to be $2N$. Hence, the policy should be chosen in such a way that it prevents the clients to tell a lie about their service periods to the BS by degrading their performance.

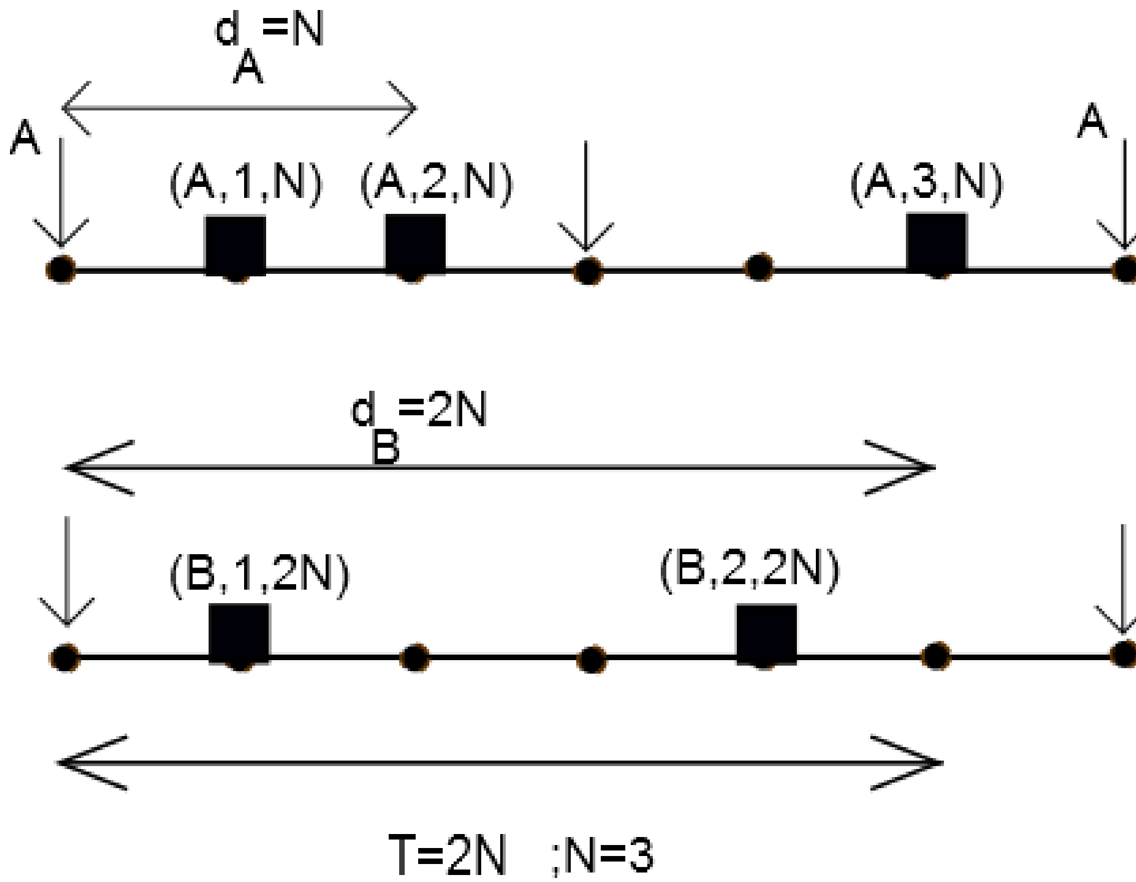

Figure 3.2: An Example of the System over a Frame of 2N Slots, where N=3 Clients, which Consists of Two Different Distributions, A and B of the Same Client. Arrows indicate Beginning of a New Period.

Table 3.1: Useful Variables

Definition	**Variables**	**Value**
Total Number of Clients	N	N
Number of Real Time Clients	K	K
A Real Time Client	RT_i	$i \in [1, K]$
A Non-Real Time Client	NRT_i	$i \in [K+1, N]$
Throughput of Real Time Client in k^{th} service period	$q_i(k)$	$i \in [1, K]$
Service Period of Real Time Clients	$[d_1 N, \ldots d_k N]$	$\text{d}_i N$ is the service period for a RT client
Frame	T	LCM of service periods
Number of Slots for a RT client in frame	S_i	L
Number of Slots for a RT client in its service period	d_i	d_i
Guaranteed slots in its service period	β_i	$i \in [1, K]$
Throughput of a client under PF policy	Thr_i	$i \in [1, N]$
Service rate of a client	$r(t)_i$	$i \in [1, N]$
Current time slot in frame	t	
Whether current slot is reserved	$\theta(X)$	$X \in [1, T]$
Next nearest service period of RT_i	τ_i	$i \in [1, K]$
Average Throughout of a client under PF Policy	$(\phi_{PF})_i$	$i \in [1, N]$

4. OUR PROPOSED SCHEDULING POLICY

4.1 Proportional Fair Scheduling

In this section, we first explain the Proportional Fair Scheduling (PF) policy that has been used in our proposed algorithm. The basic definition of proportional fair scheduling policy is that in each time slot t, it picks up the client n such that,

$$arg\, max_{n \in N} \frac{r_n(t)}{(\frac{\sum_{\tau=1}^{t} y_n^{PF}(\tau)}{t})} \tag{4.1}$$

where $(\frac{\sum_{\tau=1}^{t} y_n^{PF}(\tau)}{t})$ is the average throughput received by the client n and $r_n(t)$ is the service rate at time slot t.

It has also been found that [48], a proportionally fair scheduler maximizes the sum of logarithmic average throughput, which can be formally represented as :

$$\sum_{n \in N} log((\frac{\sum_{\tau=1}^{t} y_n^{PF}(\tau)}{t})).$$

Even though PF scheduler performs well in both fairness to the clients and system throughput, it fails to provide reliability to the clients. For example, consider a system with one RT client and all others are NRT clients. In particular, $d_1 = 1$ and $d_n = \infty$ for all $n \neq 1$. If the service rates of all the clients are in uniform distribution between $[0, 1]$, then by symmetry, the probability of each client getting served at time t by the BS is $1/N$. Therefore, the probability that the the client 1 is not served at all in k^{th} service period is $Prob(q_1(k) = 0) = (1 - \frac{1}{N})^N \rightarrow \frac{1}{e}$ as $N \rightarrow \infty$. In other words, with probability 36.7% client 1 receives 0 bits in its k^{th} service period. Since RT clients require high reliable services, therefore to overcome this downfall in PF scheduler, we need a new scheduling algorithm.

4.2 Brief Description of Our Scheduling Policy

Definition 3: We first define the long term *average throughput of client n under PF policy* as:

$$(\phi_{PF})_n(t) = \frac{\sum_{\tau=1}^{t} y_n^{PF}(\tau)}{t}$$

where $\sum_{\tau=1}^{t} y_n^{PF}(\tau)$ is the total throughput obtained by a client n in the time $[0, t]$ under PF scheduling policy η_{PF}.

We now present the brief overview of our scheduling policy η_{our}. Our policy is an extended version of PF policy with an aim to provide reliable services to RT clients. The basic idea of our scheduling policy is to provide fairness in a shared wireless channel where each client n should be served on an average $\frac{1}{N}$ of the time slots. Based on this idea, we use PF policy and impose a constraint that each RT client n is served exactly d_n times in each of its service period of $d_n N$ slots. It means that if a RT client n has already been served d_n times in its current service period, then it cannot be served even it it has the largest value of $\frac{r_n(t)}{(\phi_{PF})_n(t)}$ among all other clients. A RT client n with small value of $\frac{r_n(t)}{(\phi_{PF})_n(t)}$ may still be served, if not serving it in time slot t would make it impossible to serve d_n times in its current service period. For example, consider a system with one RT client and all other NRT clients. In particular, $d_1 = 1$ and $d_n = \infty$ for all $n \neq 1$. If a RT client has already been served one slot in its current service period, then it will not be served again. While in an other case, if a RT client is not served in the first $N - 1$ slots, then the last slot will be served to RT client irrespective of its service rate and its average throughput under PF policy.

We have explained our scheduling policy in detail in the next section.

4.3 Detailed Description of Our Scheduling Policy

We first define the definitions that are required for better understanding of the scheduling policy.

Definition 4: The *guaranteed number of time slots* is defined as the required number of time slots for RT client n to be served by the BS in its service period *of* d_nN slots. NRT clients having service periods as ∞ do not have guaranteed number of time slots which needs to be served by the BS. It is represented by $\beta_n(t)$ and is defined in client's k^{th} service period as follows:

$$\beta_n(t) = \begin{cases} \text{No. of slots that still needs to be served} & \text{at } [\,(k-1)d_nN \leq \ t \ < kd_nN], \\ d_n & \text{at } t = kd_nN \end{cases}$$

Definition 5: The term *reserve slot* is defined as the client n being strictly served at time t irrespective of $r_n(t)$ and $(\phi_{PF})_n(t)$ of client n and other $N-1$ clients.

Definition 6: The term *set of clients that can be served* is defined as the clients which are either NRT clients or RT clients which are still left with the required number of slots that still needs to be served.

The scheduling policy is divided into four steps, as depicted in Algorithm 1. We now explain our policy in greater details.

At the start of every frame, we reserve the slots to RT clients in the increasing order of their service periods (reservation is done from the end of the frame.) If the length of the frame is N slots, RT client with the smallest service period reserves its first guaranteed time slot at the N^{th} slot position, subsequently reserving its other guaranteed number of slots at $(N-1)^{th}$ slot position, and so on. For example, consider two RT clients with service periods N and $2N$, with frame length $2N$. Then, RT client with service period N would reserve its guaranteed number of slots in the frame at $(2N)^{th}$and N^{th} time slots, while RT client with service period $2N$ would reserve its guaranteed number of slots in the frame at $2N-1^{th}$, $2N-2^{th}$ time slots. Algorithm 2 explains its details.

Next, we iterate over all the slots in a frame. At each time slot, if a slot is not reserved for any RT client, we pick among the set of clients that can be served using PF scheduling policy. Otherwise, if we find that the slot is reserved to RT client n, we pick the RT client using PF policy among a pool of other RT clients who can be served and whose service periods are equal or lesser than the service period of the reserved RT client n.

In the third step, if RT client is served at the current time slot t, we decrement its count for guaranteed number of slots. We also right shift the reserved slots of the other RT clients who have service periods equal to or greater than the served RT client. Algorithm 3 explains the details.

Finally, we update the the throughput $(q_n(k))$ and reset the $\beta_n(t)$ of any RT client n, whose service period has expired, to the value equal to its guaranteed number of slots.

Algorithm 1 Our Scheduling Policy

Sort RT clients based on their service periods such that $d_1N \leq d_2N \leq \ldots \leq d_KN$
Global Variables: θ
while (true)
  $T \leftarrow$ least common multiplier of $\{d_nN \mid n \in K\}$
  RESERVATION SLOT POLICY (Algorithm 2)
  for $t = 1$ to T **do**
    M is the set of clients that can be served (if RT clients then $\beta_n(t) > 0$, or NRT)
    K is the set of RT clients that can be served ($\beta_n(t) > 0$)
    if $\theta(t) = 0$ **then**
      $Z \leftarrow arg\,max_{n \in M} \frac{r_n(t)}{(\phi_{PF})_n(t)}$
    else $\theta(t) \neq 0$ **then**
      $\tau_{\theta(t)} \leftarrow \lceil \frac{t}{d_{\theta(t)}N} \rceil * d_{\theta(t)}N$
      if $\beta_{(\theta(t))}(t)$equal to $(\tau_{\theta(t)} - t)$ **then**
        $Z \leftarrow \theta(t)$
      else find the set of RT clients $\{R \mid R \in K\}$such that $\tau_R \leq \tau_{(\theta(t))}$
        $Z \leftarrow arg\,max_{n \in R} \frac{r_n(t)}{(\phi_{PF})_n(t)}$
      end if
    end if
    if $Z \in K$ **then**
      $\beta_Z(t) \leftarrow \beta_Z(t) - 1$
      if β_Z(t) equal to 0
        $M \leftarrow M - \{Z\}$
        $K \leftarrow K - \{Z\}$
      end if
      UPDATE RESERVATION POLICY FOR RT CLIENTS (Algorithm 3)
    end if
    $Z' \leftarrow arg\,max_{n \in D} \frac{r_n(t)}{(\phi_{PF})_n(t)}$ where D is a set of all N clients.
    $Thr_{Z'}(t) \leftarrow Thr_{Z'}(t) + r_{Z'}(t)$
    for $n \in D$ **do**
      if $t \mod d_nN$ equal to 1 **then**
        $\beta_n(t) \leftarrow d_n$
        $M \leftarrow M \cup \{n\}$
        $K \leftarrow K \cup \{n\}$
      end if
      $(\phi_{PF})_n(t) = Thr_n(t)/t$
    end for
  Serve client Z at time t
  end for
end while

Algorithm 2 Reservation Policy for Multiple real time clients

```
for n = 1 to K do
    S_n ← T/d_nN
    for j = 1 to S_n do
        t ← j * d_nN
        for k = 1 to d_n do
            while θ(t) ≠ 0
                t ← t − 1
            end while
            θ(t) ← n
        end for
    end for
end for
```

Algorithm 3 Update Reservation Policy for RT Clients

```
Require: t, d_nN
τ=⌈(t/d_nN)⌉ * d_nN
for i in range (τ, t + 1) do
    θ(i) ← θ(i-1)
end for
```

5. TRUTHFUL PROPERTY

In this section, we will demonstrate that our scheduling policy discourages the strategic clients from lying their service period to the BS. For the ease of our theoretical analysis, we have assumed that the service rate of all the clients to have a uniform distribution between $[0, 1]$. In section 6 we provide extensive simulation results of our scheduling policy under the fast fading Rayleigh model to show that this policy can be extended in wireless networks with real channel characteristics.

Theorem 2: N being large and the service rate of clients being uniformly distributed between $[0, 1]$, PF policy is reduced to choosing the client with maximum $r_n(t)$ and $Prob(r_n(t) \geq 1 - \epsilon) = 1 \, \forall \, \epsilon > 0$, as $N \rightarrow \infty$.

Proof: As per PF policy the client is serviced by the BS which has $Z \leftarrow arg\, max_{n \in N} \frac{r_n(t)}{(\phi_{PF})_n(t)}$. Now, if the service rate $r_n(t)$ has a uniform distribution between $[0, 1]$, then, by symmetry at $t \rightarrow \infty$, $\phi_{PF} = (\phi_{PF})_n$ where $n \, \epsilon \, N$. Hence, PF policy picks the client which has $Z \leftarrow arg\, max_{n \, \epsilon \, N} \frac{r_n(t)}{\phi_{PF}(t)}$.

Assuming the client serviced at time slot t has a service rate of $1-\epsilon$, where $0 < \epsilon < 1$. It signifies that the serviced client has maximum service rate among all the other N-1 clients. Therefore, $Prob(r_n(t) < 1-\epsilon)$, for all $n \rightarrow Prob\{max_n(r_n(t)) < 1-\epsilon\} = (1-\epsilon)^N$. Since N is large, $(1-\epsilon)^N \rightarrow 0$. Consequently, $Prob(r_n(t) < 1-\epsilon) = 0$. Therefore, $Prob(r_n(t) \geq 1-\epsilon) = 1$. Inference is made on the probability that the served client would have service rate greater than $1-\epsilon$ would be 1.

5.1 Theoretical Analysis for Different Distributions of RT Client with Service Period N

We consider a system with one RT client and all others are NRT clients. In particular, $d_1 = 1$ and $d_n = \infty$ for all $n \neq 1$ with the length of frame to be N. As mentioned above, we have assumed that the service rate of all the clients to have a uniform distribution between $[0, 1]$. By symmetry, the probability of each client getting served at time t by the BS is $1/N$. We then study the distribution of $[q_1(k)]$ at steady state.

- **Case 1:** RT client tells the service period as N. There can be 2 events possible.
  - *(Event 1)* RT client is not serviced in the first $N-1$ time slots. Then, the last N^{th} slot will be reserved for the RT client. The probability that the RT client is not serviced at time t because it does not have the largest ratio of $\frac{r_1(t)}{(\phi_{PF})_1(t)}$ is $1-\frac{1}{N}$. Therefore, the probability that the client 1 is not served at all in the first $N-1$ time slots is $Prob = (1-\frac{1}{N})^{N-1} \to \frac{1}{e}$ as $N \to \infty$. Since, RT client will be served under uniform distribution of service rate, therefore the throughput received in its service period is uniformly between $[0, 1]$.
  - *(Event 2)* RT client is scheduled exactly once in the first $N-1$ slots. Using Theorem 2, RT client will have the service rate such that $Prob(r_1(t) > 1-\epsilon) = 1$, therefore the throughput received in its service period is $q_1(k) = 1$. The probability of this event will be $1-\frac{1}{e}$.

Hence the probability distribution of $q_1(k)$ in the k^{th} service period of RT client is:

$$q_1(k) = \begin{cases} \text{Uniformly between [0,1]} & \text{with prob.} = \frac{1}{e}, \\ 1 & \text{with prob.} = 1-\frac{1}{e}. \end{cases} \tag{5.1}$$

Therefore, CDF of $q_1(k)$ is:

$$[q_1(k) \leq c] = \begin{cases} \frac{c}{e} & c \leq 1, \\ 1 & c = 1. \end{cases}$$

The mean of $q_1(k)$ is:

$$1 - \frac{1}{e} + \frac{1}{2e} = 1 - \frac{1}{2e}.$$

- **Case 2:** RT client tells the BS its service period as ∞ while its actual service period is N. This case is similar to the case in which RT client is measuring its performance under PF scheduling policy. Therefore, the probability that the the client 1 is not served at all in N time slots is $Prob(q_1(k) = 0) = (1 - \frac{1}{N})^N \to \frac{1}{e}$ as $N \to \infty$. In other words, with probability 36.7% RT client receives 0 bits in its k^{th} service period, i.e. the throughput received is $q_1(k) = 0$. The probability that RT client is served equal to or more than once in its service period is $(1 - \frac{1}{e})$. Using Theorem 2, the throughput received by RT client is m, where m is the number of times RT client is scheduled in its service period. The probability of being serviced m times is $\binom{N}{m} \left(\frac{1}{N}\right)^m \left(1 - \frac{1}{N}\right)^{N-m} \xrightarrow{N\to\infty} \frac{1}{m!e}$.

  Hence the probability distribution of $q_1(k)$ in the k^{th} service period of RT client is:

$$q_1(k) = \begin{cases} 0 & \text{with prob. } \left(\frac{1}{e}\right), \\ m & \text{with prob. } \left(\frac{1}{m!e}\right). \end{cases} \tag{5.2}$$

  The CDF of $q_1(k)$ is:

$$Prob(q_1(k) \leq c) = \frac{1}{e} \sum_{i=0}^{c} \frac{1}{i!}$$

  The mean of $q_1(k)$ is 1.

It is observed from the equations (5.1) and (5.2), that in case RT client lies its service period, the probability of receiving zero bits in k^{th}service period is $1/e$. In other words, with probability 36.7% RT client receives 0 bits in its k^{th} service period. It is observed, when RT client lies its service period to be ∞, the theoretical CDF has a throughput mean of 1 as compared to the case when RT client tells its service period to be N, the mean is $(1 - \frac{1}{2e})$. Even though the former distribution has higher mean, yet there is no reliability guaranteed in the throughput received. Hence, the RT client will have the most reliable distribution when it tells the truth of its service period to BS which has lesser throughput but reliability is guaranteed.

In figure 5.1, we show the comparison of CDFs obtained for a RT client under different scheduling policies. It is known that CDF obtained under PF policy is similar to the case when the RT client lies its service period to be ∞. It is inferred that, PF policy provides maximum throughput with unreliable throughput distribution.Under round-robin (RR) policy, RT client will have reliable service guarantee but the throughput will not be optimized. The probability of RT client getting served in N slots will be 1 irrespective of the service rate of all the clients. The performance of RT client measured under (RR) policy will be $q_1(k)$, which is uniformly distributed between 0 and 1. The CDF obtained under RR policy for RT client will be $Prob(q_1(k) \leq c) = c$. Comparing the performance of RT client between RR and our policy, it is observed that in our policy, RT client is served with higher throughput with high reliability.

5.2 Theoretical Analysis for Different Distributions of RT Client with Service Period d_1N

Now, we consider the more generalized case where RT client has a service period of d_1N slots. We consider a system with one RT client and all others are NRT clients. In particular, $d_1 = \gamma$ and $d_n = \infty$ for all $n \neq 1$ with the length of frame to be d_1N. Let d_1N be the actual service period of RT client. As mentioned above, we have assumed that the service rate of all the clients to have a uniform distribution between $[0, 1]$. We define Irwin–Hall distribution $X(n)$ as the sum of n independent and identically distributed $[0, 1]$ random variables [49]. We then study the distribution of $[q_1(k)]$ at

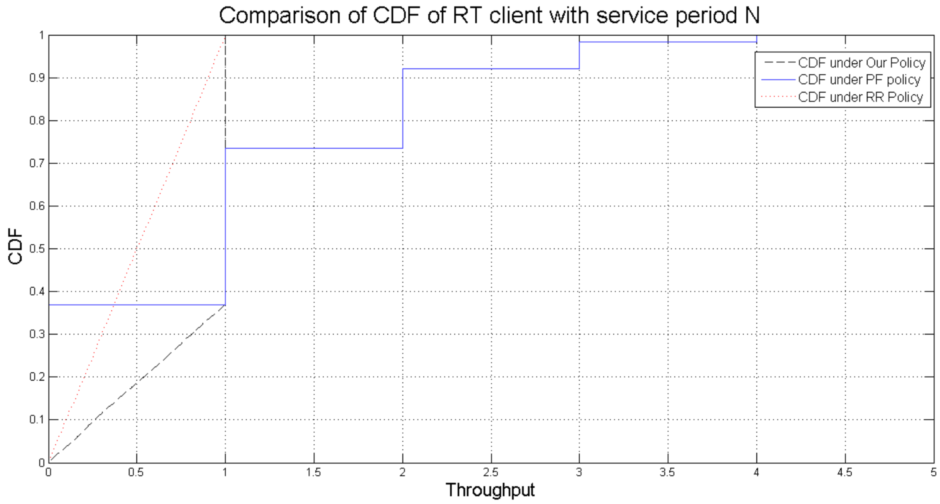

Figure 5.1: Comparison of CDFs of RT Client with Service Period N under Different Scheduling Policies

steady state of a single RT client for the three extreme cases:

1. A single RT client with service period of $d_1 N$ is telling the truth to the BS.

2. A single RT client with service period of $d_1 N$ is telling a lie to the BS claiming its service period to be N.

3. A single RT client with service period of $d_1 N$ is telling a lie to the BS of its service period to be ∞.

- **Case 1:** RT client tells the service period as $d_1 N$. There can be $d_1 + 1$ events possible.

  - *(Event 1)* RT client is not serviced in the first $d_1 N - d_1$ slots. The last d_1 slots will be reserved for the RT client. The throughput received by the RT client will be the sum of d_1 service rates generated in uniform distribution between 0 and 1. Therefore $q_1(k) = X(d_1)$. The probability of this event will be:

$$\left(1-\frac{1}{N}\right)^{d_1N-d_1} \xrightarrow{N\to\infty} \left(\frac{1}{e}\right)^{d_1}.$$

– ⋮

– *(Event j+1)* RT client is serviced exactly j times in the first $d_1N-(d_1+1-j)$ and not serviced in d_1N-d_1+j slot. Using theorem 2, RT client will have the service rate of 1 in the first j serviced slot. Then, the last d_1-j slots will be reserved for the RT client. The throughput received by the RT client in the d_1-j slots will be equal to the sum of d_1-j service rates generated in uniform distribution between 0 and 1. Hence, the throughput is $q_1(k)=j+X(d_1-j)$. The probability that the RT client is serviced exactly j times in first $d_1N-(d_1-j+1)$ slots and not serviced in d_1N-d_1+j slot will be:

$$\binom{d_1N-(d_1+1-j)}{j}\left(1-\frac{1}{N}\right)^{d_1N-(d_1+1-j)}\left(\frac{1}{N}\right)^{j}\left(1-\frac{1}{N}\right) \xrightarrow{N\to\infty} \frac{d_1^j}{j!e^{d_1}}.$$

– ⋮

– *(Event d_1+1)* RT client is serviced exactly d_1 times in d_1N-1 slots and not serviced in d_1N slot. Using theorem 2, RT client will have the service rate of 1 in all the serviced slots. Hence, the throughput received is $q_1(k)=d_1$. The probability that the RT client is serviced exactly d_1 times in d_1N-1 slots and not serviced in d_1N slot will be:

$$1-\sum \textit{Probability of all other events},$$

If N is large, then $N \to \infty$, $Prob(Event\, d_1{+}1) = 1 - \frac{1}{e^{d_1}} \left(1 + d_1 + \frac{d_1^2}{2!} \cdots \frac{d_1^{d_1-1}}{(d_1-1)!}\right)$.

Hence, the probability distribution of $q_1(k)$ of RT client is:

$$\begin{cases} X(d_1) & \text{with prob. } \left(\frac{1}{e^{d_1}}\right), \\ j + X(d_1 - j) & \text{with prob. } \left(\frac{d_1^j}{j!e^{d_1}}\right) \text{for} 1 \leq j \leq (d_1 - 1), \\ \vdots & \vdots \\ d_1 & \text{with prob. } \left[1 - \frac{1}{e^{d_1}}\left(1 + d_1 + \frac{d_1^2}{2!} \cdots \frac{d_1^{d_1-1}}{(d_1-1)!}\right)\right]. \end{cases} \tag{5.3}$$

- **Case 2:** RT client tells the BS that its service period is N slots while its actual service period is $d_1 N$. The number of its service periods served by the BS are d_1, so there are $d_1 + 1$ possible events. As we studied in Section 5.1, when the RT client had a service period of N slots, there were 2 events possible. In the first event, RT client received a throughput as a random number generated under uniform distribution between 0 and 1 with probability $1/e$. In the second event, RT client received a throughput of 1 with probability $1 - 1/e$. In this case, there are $d_1$times of service period N. Therefore, we consider j times when the RT client is serviced with event 2 and $d_1 - j$ times when the RT client is serviced with event 1.

  - *(Event 1)* RT client is serviced exactly d_1 times with event 1 when the RT client has a service period of N. Hence, the throughput is, $q_1(k) = X(d_1)$. The probability of this event will be:

$$\left(\left(1 - \frac{1}{N}\right)^{N-1}\right)^{d_1} \xrightarrow{N\to\infty} \left(\frac{1}{e}\right)^{d_1}.$$

  - $\vdots$

– *(Event j+1)* RT client is serviced exactly j times with event 2 and $d_1 - j$ times with event 1 when the RT client has a service period of N. Hence, the throughput is, $q_1(k) = j + X(d_1 - j)$. The probability of this event will be:

$$\binom{d_1}{j}\left(\binom{N}{1}\right)^j \left(\left(1-\frac{1}{N}\right)^{N-1}\right)^{d_1-j} \left(\frac{1}{N}\right)^j \left(1-\left(1-\frac{1}{N}\right)^{N-1}\right)^j,$$

$$\xrightarrow{N\to\infty} \frac{\binom{d_1}{j}}{e^{d_1-j}}\left(1-\frac{1}{e}\right)^j.$$

– $\vdots$

– *(Event d_1+1)* RT client is serviced exactly d_1 times with event 2 when the RT client has a service period of N. Hence, the throughput is, $q_1(k) = d_1$). The probability of this event will be:

$$\binom{N}{1}^{d_1}\left(\frac{1}{N}\right)^{d_1}\left(1-\left(1-\frac{1}{N}\right)^{N-1}\right)^{d_1} \xrightarrow{N\to\infty} \left(1-\frac{1}{e}\right)^{d_1}.$$

Hence, the probability distribution of $q_1(k)$ of RT client is:

$$\begin{cases} X(d_1) & \text{with prob. } \left(\frac{1}{e^{d_1}}\right), \\ j + X(d_1 - j) & \text{with prob. } \left(\frac{\binom{d_1}{j}}{e^{d_1-j}}\left(1-\frac{1}{e}\right)^j\right) \text{ for } 1 \le j \le d_1 - 1, \\ \vdots & \vdots \\ d_1 & \text{with prob. } \left(\left(1-\frac{1}{e}\right)^{d_1}\right). \end{cases} \tag{5.4}$$

- **Case 3:** RT client reports the BS that its service period is ∞ while the actual service period is $d_1 N$ slots. There are $d_1 + 1$ possible events.

  – *(Event 1)* RT client is not serviced in the service period. Here, the throughput received by RT client is $q_1(k) = 0$. The probability of this event will be:

$$\left(1-\frac{1}{N}\right)^{d_1 N} \xrightarrow{N\to\infty} \left(\frac{1}{e}\right)^{d_1}.$$

– *(Event j+1)* RT client is scheduled exactly j times in the service period. Here, the throughput received by RT client is $q_1(k) = j$. The probability of this event will be:

$$\binom{d_1 N}{j}\left(1-\frac{1}{N}\right)^{d_1 N-j}\left(\frac{1}{N}\right)^{j} \xrightarrow{N\to\infty,} \frac{(d_1)^j}{j!e^{d_1}}.$$

– *(Event d_1+1)* RT client is scheduled more than d_1 times in the service period. Here, the throughput received by RT client is $q_1(k) = d_1$. The probability of this event will be:

$$1-\sum Probability\, of\, all\, other\, d_1\, events,$$

If N is large, then $N \to \infty$, $Prob(Event\, d_1{+}1) = 1-\frac{1}{e^{d_1}}\left(1+d_1+\frac{d_1^2}{2!}\cdots\frac{d_1^{d_1-1}}{(d_1-1)!}\right)$.

Hence, the probability distribution of $q_1(k)$ of RT client for case 3 is:

$$\begin{cases} 0 & \text{with prob. } \left(\frac{1}{e^{d_1}}\right), \\ j & \text{with prob. } \left(\frac{(d_1)^j}{j!e^{d_1}}\right) \text{ for } 1 \le j \le (d_1-1), \\ \vdots & \vdots \\ \ge d_1 & \text{with prob. } \left[1-\frac{1}{e^{d_1}}(1+d_1+\frac{d_1^2}{2!}\cdots\frac{d_1^{d_1-1}}{(d_1-1)!}\right]. \end{cases} \tag{5.5}$$

Considering the equations (5.3)and (5.5), we observe that in the case, when the RT client tells BS its service period to be ∞, it receives throughput zero with probability $\frac{1}{e^{d_1}}$. Therefore, when the RT client tells a truth then its distribution is more reliable as compared to the case, when RT client tells a lie of its service period of ∞.

Our scheduling policy can be summarized for a RT client with service period d_1N and having uniform distribution of service rate between 0 and 1 under two cases:

- (Case 1) If a RT client is served under PF policy, then it will receive the service rate of 1.

- (Case 2) If a RT client is not served under PF policy but it is being served to meet up its guaranteed number of slots, then it will receive the service rate under uniform distribution between 0 and 1.

Let k times a RT client is served under PF policy. The throughput received by the RT client in its service period is:

$$q_1(k) = min(k, d_1) + X(d - min(k, d_1))$$

Now we compare the two cases, when the RT client tells the truth to the BS about its service period and when it falsify its service period to be N. Suppose the RT client is served k times in its service period under PF scheduling policy. Then, if the RT client tells the truth to the BS about its service period and is now being served under our policy then a RT client will receive a service rate of 1 for $min(k, d_1)$ slots. In the case, RT client tells a lie about its service period to be N slots, RT client will be served under our policy at a service rate of 1 with less or equal to the $min(k, d_1)$ slots in d_1 service periods of N. Obviously we know that,

$$a_1(\text{case 1}) + X(d_1 - a_1) <_{s.t.} a_2(\text{case 1}) + X(d_1 - a_2) \text{ if } a_1 < a_2. \tag{5.6}$$

where a_1 and a_2 are the number of times RT client is being served with service rate of 1 under our policy for different service periods. As mentioned above, the number of times RT client is serviced with service rate of 1 under our policy is more in the case when RT client is telling the truth to the BS than it is claiming its service period to be N. Therefore by (5.6), the throughput received by the RT client in its service period is more in the case when it is telling the truth to the BS. Thus, it is theoretically observed that when the RT client tells the truth its probability of receiving higher throughput is more or equal than the case when it claims its service period to be N. Hence, we prove that RT client achieves more reliable distribution when it tells the truth to BS about its service period. In figure 5.2, we show the comparison of all the three studied cases for RT client with service period 6N.

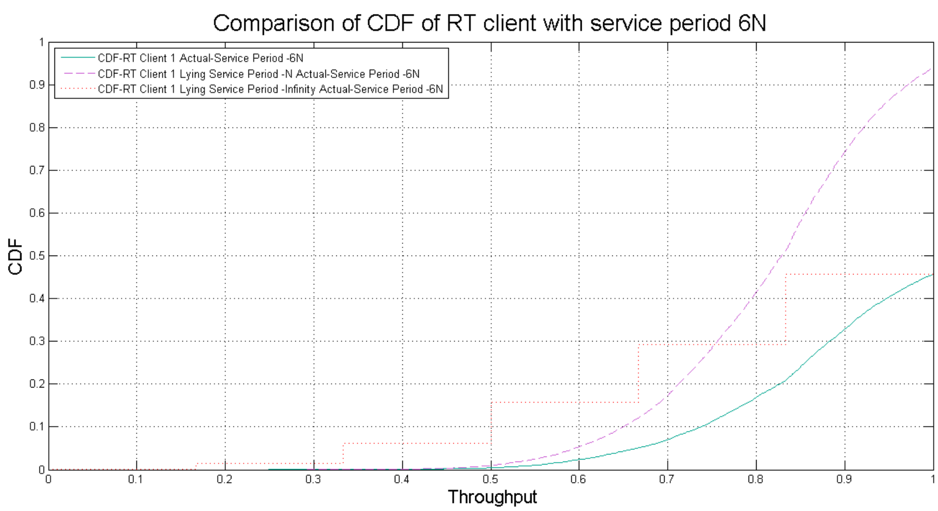

Figure 5.2: Comparison of CDFs of RT Client with Service Period 6N.

6. SIMULATION RESULTS

In this section, we present our simulation results. We have evaluated our scheduling policy considering two performance metrics,which are, reliability, as defined in Section 3 and the average system throughput. We have also compared our simulation results with the theoretical analysis explained in Section 5. Our proposed scheduling policy is compared with the traditional round-robin policy and proportional fair scheduling.

6.1 Reliability

In this section, we present the simulation results for different scenarios. We have considered two wireless channel conditions varying with time. In the first case, the independent and uniform distribution of service rate between $[0, 1]$ of all the wireless clients is considered. In the second case, we have considered the fast fading Rayleigh model under which the service rates of all the wireless clients varies with time. In both the cases, we have taken $N = 50$ wireless clients, where RT clients are $K = 20$ and NRT clients are $N - K = 30$. RT clients are given random service periods. So, there are 4 RT clients with service period of 5N, 3 RT clients with service period of 2N, 3 RT clients with service period of 3N, 2 RT clients with service period of N, 1 RT client with service period of 6N and 7 RT clients with service period of 4N. All the simulations have been performed for 1000 $frames$ in MATLAB. Results have been taken on an average for 10 runs. We have considered the three following cases:

- A RT client with service period of $6N$ is telling the truth of its service period to the BS.
- A RT client with actual service period of $6N$ is telling a lie to the BS of its

service period to be N.

- A RT client with actual service period of $6N$ is telling a lie to the BS of its service period to be ∞.

6.1.1 *Service Rate of Clients Under Uniform Distribution*

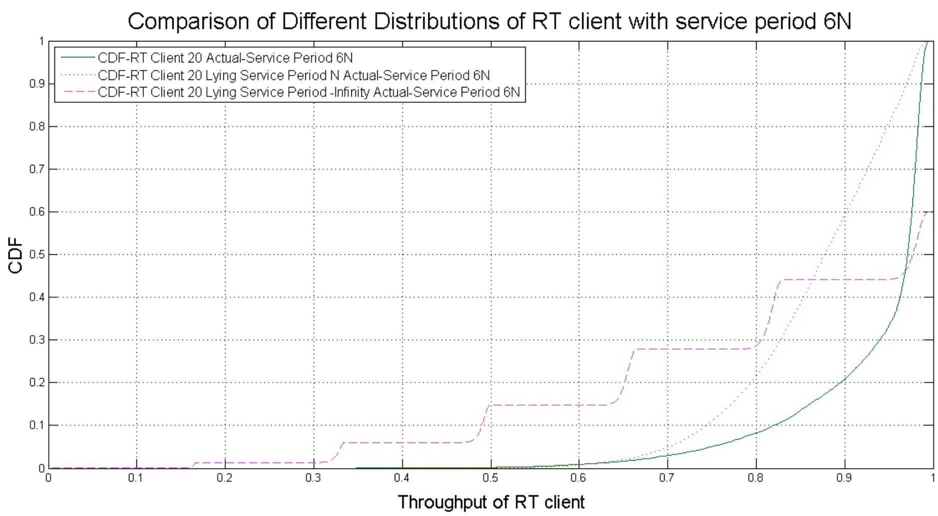

Figure 6.1: Comparison of Different Distributions of RT Client with Service Period 6N under Uniform Distribution of Service Rate

Here, the service rates of all clients are varying under independent and uniform distribution between $[0, 1]$. Fig 6.1 represents the cumulative distribution function of the throughput received for RT client with service period 6N. It is observed that, the RT client with service period 6N achieves most reliable distribution in the case, when it tells the truth to the BS. Initially, the throughput distribution remains same, for all the cases unless the RT client lies its service period to be ∞. This is because, in all the cases, probability distribution of service rate is same for the event 1 in equation 5.3 and 5.4. Also, the throughput received with 95% reliability is greater than 0.75

bits per second when the RT client reports its correct service period of 6N slots. If the RT client reports false service period of N slots its throughput received with 95% reliability is greater than 0.68 bits per second. It clearly shows that, when the RT client tells the lie it receives lesser throughput for 95% reliability. The throughput distribution is least reliable when the client tells its service period to be ∞, because there is always a probability that the RT client will receive zero throughput in its service period.

6.1.2 Service Rate of Clients Under Fast Fading Rayleigh Model

In this scenario, we have considered the fast fading Rayleigh model for the service rates of all the wireless clients at any time. The channel gain [50] for each client is derived from the following equation:

$$PL(d) = 128.1 + 37.6log10(d) + Y, \tag{6.1}$$

where $PL(d)$ is the channel gain in dB and d is distance in km. Y represents Rayleigh fast fading with a Doppler of 5 Hz. The thermal noise is randomly generated between $[3.5, 4.5] * 10^{-15}$W. As it is observed from [51] that the BS consumes 75W and 130W of power during sleep and active period respectively, we have taken 50W as standard power consumption in our macro BS. Table 6.1 summarizes the simulation parameters to run the experiment. The distance of the observed client is 1 km and is constant throughout the experiment.

Fig 6.2 represents the cumulative distribution function of the throughput received for RT client with service period 6N. It has been observed that, the results obtained in 6.1.1 are same with our proposed policy under Rayleigh fast fading channel model. The RT client with service period 6N achieves maximum reliability in the case, when it tells the truth of its service period to BS. Initially, the throughput distribution

Table 6.1: Simulation Parameters

Variables	Values
N	50
K	20
Noise	$[3.5, 4.5] * 10^{-15}$ W
Input Power	50 W
Distance	$[0.05, 2.5]$ Km
Doppler Frequency	5 Hz
Number of Periods	1000

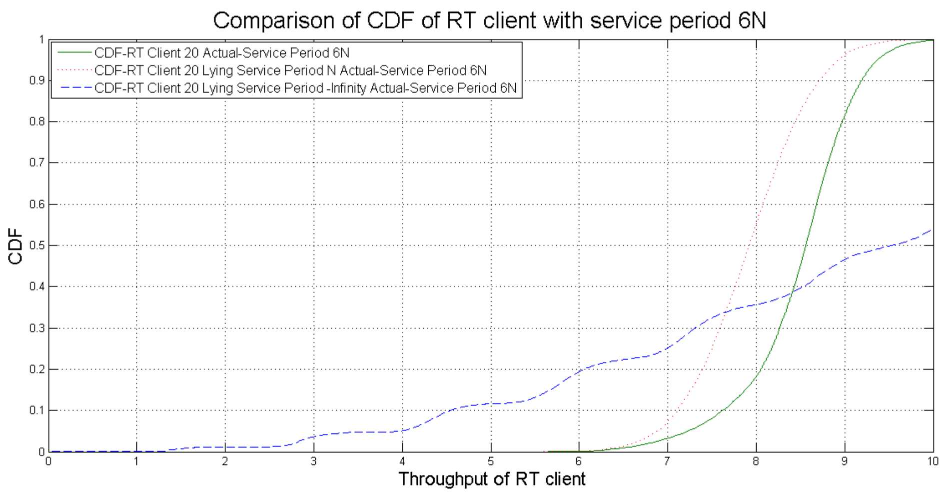

Figure 6.2: Comparison of Different Distributions of RT Client with Service Period 6N under Rayleigh Fading Model of Service Rate

remains same, for all the cases unless the RT client lies its service period to be ∞. This is because, in all the cases, probability distribution of service rate is same for the event 1 in equation (5.3) and (5.4). Also, the throughput received with 95% reliability is greater than 7.5 bits per second when the RT client reports its correct service period of 6N slots. If the RT client reports false service period of N slots its throughput received with 95% reliability is greater than 6.5 bits per second. It clearly shows that, when the RT client tells the lie it receives lesser throughput for

95% reliability. The throughput distribution is least reliable when the client tells its service period to be ∞, because there is still a probability that the RT client will receive zero throughput in its service period.

6.1.3 Reliability Comparison Among Different Policies

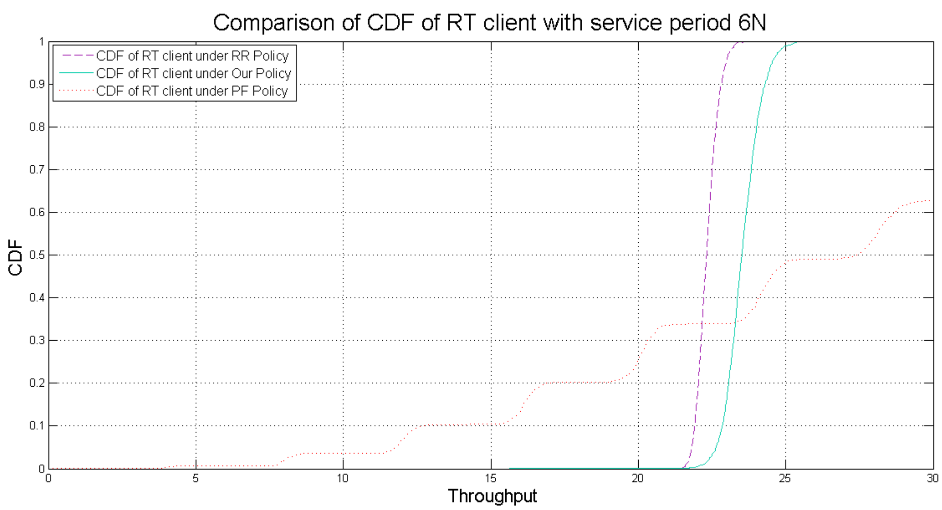

Figure 6.3: Comparison of CDF of RT Client with Service Period 6N against RR, PF and Our Proposed Policy

Fig 6.3 demonstrates the reliability comparison of the throughput received for RT client with service period 6N for different scheduling policies. The simulation has been performed when the clients have their service period under Rayleigh fast fading model. We have compared our scheduling policy with PF policy and RR policy. It is observed that PF policy serves RT client with high throughput but with no reliability. The graph demonstrates that under PF policy RT client has a probability of receiving zero bits in its service period. Hence, PF policy fails to provide reliability to RT clients which is a necessary requirement for them. The throughput distribution obtained under RR policy provides reliability to RT clients but it fails to achieve high throughput. The reason is that the clients are served in their predefined slots. As a result, the clients cannot take advantage of the shared channel when they have the largest service rates. Compared to these policies, our policy provides both reliability and high throughput to clients. It can be seen from the throughput distribution under our policy that the throughput received by the RT client with 95% reliability is greater than 23 bits per second, while RR policy provides throughput with 95% reliability is greater than 21 bits per second. Thus we can remark that, compared to the other two policies, our policy provides better reliability with high throughput to RT clients.

6.2 System Throughput

The average system throughput metric characterizes how efficient the policy is at scheduling clients that maximizes the system throughput, is defined as

$$\sum_{n=1}^{N} \sum_{t=1}^{T} y_n(t)/T$$

where $\sum_{n=1}^{N} \sum_{t=1}^{T} y_n(t)$ is the total throughput received by the system till time slot t.

Fig 6.4 and Fig 6.5 demonstrates the average throughput of the system when the clients have their service rates under uniform distribution and Rayleigh fast fading model. We have compared the average throughput of the system with the state-of-the-art policies. RR stands for Round-Robin policy, and PF stands for Proportional Fair Scheduling. Simulation was performed for $10000 * N$ times where $N = 50$. It is observed that, round robin policy performs worst as it does not consider the service rates of the clients. PF considers both the factors, i.e. service rates and fairness of

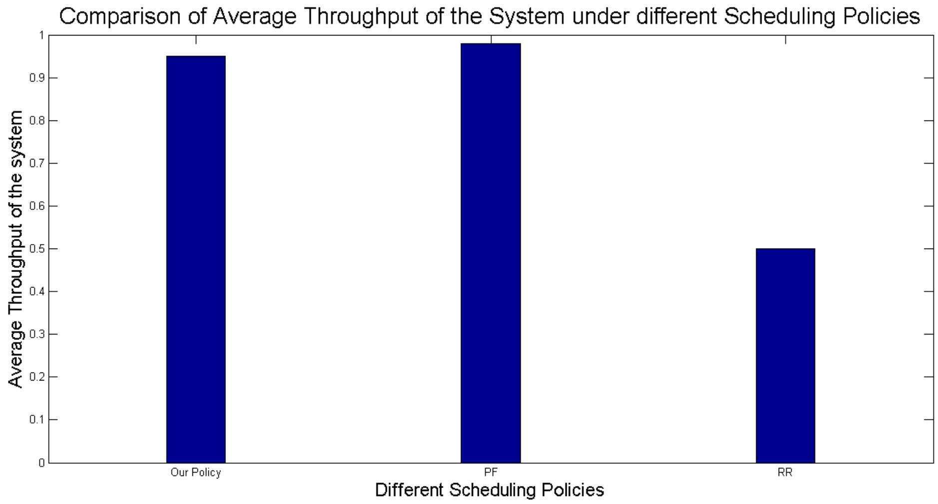

Figure 6.4: Average Throughput of the System for Different Scheduling Policies under Uniform Distribution of Service Rates

the clients and hence performs the best among all the three policies. Our policy also considers the service rate of the clients and also provides fairness to the clients. We observe a small difference in the average throughput when we compare PF with our policy. The reason is that, our policy provides guaranteed number of slots to RT clients, while PF policy does not provide consistent service. Hence, we see a small drop in the average system throughput in our policy which is justified as we are able to provide reliability to RT clients on the shared wireless channel.

6.3 Comparison of Theoretical Analysis with Simulated Results

In this experiment, we have compared the theoretical analysis with the simulation results done for the example mentioned in Section 6. Fig 6.6 depicts the comparison to verify our theoretical analysis, and to demonstrate the results for the example. In the simulated results, we have considered $N = 50$ wireless clients, where there are 49 NRT clients and 1 RT clients with $s$ervice period $6N$. All the clients have uniform distribution service rate in between $[0, 1]$. The experiment was performed

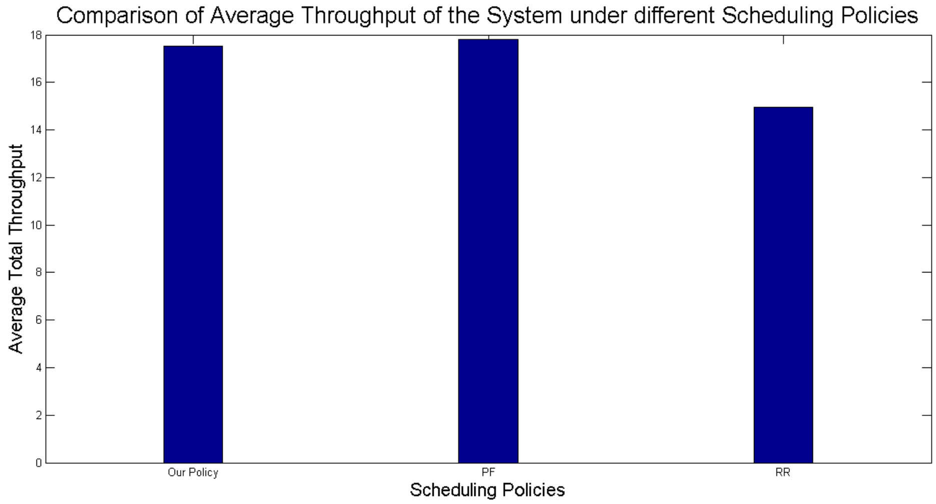

Figure 6.5: Average System Throughput of Different Scheduling Policies under Rayleigh Fast Fading Model

for 10,000 frames. It can be inferred from the figure 6.6 that our simulation results clearly match our theoretical analysis done in the example in Section 5.

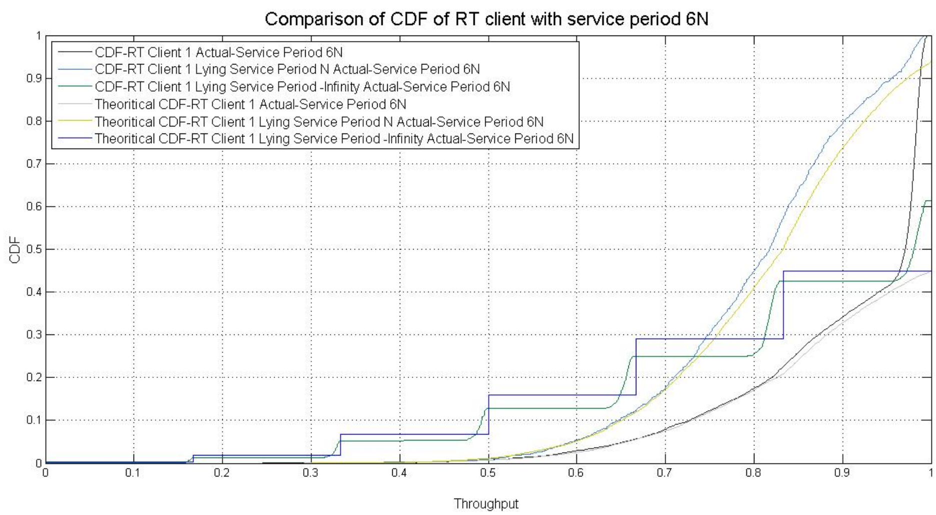

Figure 6.6: Comparison of CDF of RT Client obtained under Simulation Results with CDF of RT Client obtained under Theoretical Analysis

7. CONCLUSION

We have studied the problem of scheduling the mixture of multiple real time clients with varying service requirements and non-real time clients in a time-varying wireless network. We also explored strategic clients which can exploit current scheduling policies by choosing incorrect means to achieve better performance. We have proposed our scheduling policy that not only prevents the strategic clients from claiming false service requirements but also motivating them to choose the correct service requirements to achieve the best services. Our policy provides fairness to the clients, where no clients are given priority over the shared wireless channel. It also takes advantage of the time-varying wireless channel between the base station and the clients to serve the appropriate client. Our scheduling policy provides reliability to the real time clients and at the same time achieves high system throughput. We analyzed theoretically to prove that our policy provides the best service in the case when the client reports its true service requirement. In addition, we also performed extensive simulations under the realistic wireless channel conditions. Simulation results show that our policy outperforms other current scheduling policies in providing reliability to real time clients and in achieving high system throughput.

It is also shown that this work can easily be extended to various applications. Extending this work to other applications with more complicated and realistic settings constitutes an interesting problem for future research.

REFERENCES

[1] Jing Zhu I-Hong Hou and Rath Vannithamby. Incentive-oriented downlink scheduling for wireless networks with real-time and non-real-time flows. In *IEEE BWA Workshop, collocated with IEEE Globecom*, 2013.

[2] B. Sklar. Rayleigh fading channels in mobile digital communication systems .i. characterization. *Communications Magazine, IEEE*, 35(7):90–100, Jul 1997.

[3] K. Ramanan A. Stoyar M. Andrews, K. Kumaran and Phil Whitting. Providing Quality of Service over a Shared Wireless Link. *Communications Magazine, IEEE*, 39:150–154, February 2001.

[4] G. Barriac and J. Holtzman. Introducing delay sensitivity into proportional fair algorithm for cdma downlink scheduling. In *IEEE 7th Int. Symp. On Spread-Spectrum Tech. & Appl*, Prague, Czech Republic, September 2002.

[5] I-H. Hou, V. Borkar, and P. R. Kumar. A theory of QoS for wireless. In *Proc. of INFOCOM*, pages 486–494, 2009.

[6] I-H. Hou and P.R. Kumar. Admission control and scheduling for QoS guarantees for variable-bit-rate applications on wireless channels. In *Proc. of ACM MobiHoc*, pages 175–184, 2009.

[7] S. Shakkottai and R. Srikant. Scheduling real-time traffic with deadlines over a wireless channel. In *Proceedings of ACM Workshop on Wireless and Mobile Multimedia*, Seattle, WA, August 1999.

[8] Sanjay Shakkottai and Alexander L Stolyar. Scheduling algorithms for a mixture of real-time and non-real-time data in hdr. *Teletraffic Science and Engineering*, 4:793–804, 2001.

[9] Kapseok Chang and Youngnam Han. Qos-based adaptive scheduling for a mixed service in hdr system. In *The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2002*, volume 4, pages 1914–1918. IEEE, 2002.

[10] Khaled MF Elsayed and Ahmed KF Khattab. Channel-aware earliest deadline due fair scheduling for wireless multimedia networks. *Wireless Personal Communications*, 38(2):233–252, 2006.

[11] Xin Wang, Georgios B Giannakis, and Antonio G Marques. A unified approach to qos-guaranteed scheduling for channel-adaptive wireless networks. *Proceedings of the IEEE*, 95(12):2410–2431, 2007.

[12] Ruogu Li, Atilla Eryilmaz, Lei Ying, and Ness B Shroff. A unified approach to optimizing performance in networks serving heterogeneous flows. *IEEE/ACM Transactions on Networking*, 19(1):223–236, 2011.

[13] Shailesh Patil and Gustavo de Veciana. Managing resources and quality of service in heterogeneous wireless systems exploiting opportunism. *IEEE/ACM Transactions on Networking (TON)*, 15(5):1046–1058, 2007.

[14] H. Zhu H. Haci and J. Wang. Novel scheduling for a mixture of real-time and non-real-time traffic. In *Proceedings of IEEE Globecom*, 2012.

[15] Sem Borst and Nidhi Hegde. Integration of streaming and elastic traffic in wireless networks. In *Proceedings of 26th IEEE International Conference on Computer Communications*, pages 1884–1892. IEEE, 2007.

[16] C. Uc-Rios and D. Lara-Rodriguez. An efficient scheduler for real and non-real time services maximizing satisfied users in wireless networks. In *Proceedings of 20th International Conference on Computer Communications and Networks*, pages 1–5, 2011.

[17] Juan José Jaramillo and R Srikant. Optimal scheduling for fair resource allocation in ad hoc networks with elastic and inelastic traffic. In *Proceedings of IEEE INFOCOM, 2010*, pages 1–9. IEEE, 2010.

[18] Bin Li, Ruogu Li, and Atilla Eryilmaz. Heavy-traffic-optimal scheduling with regular service guarantees in wireless networks. In *Proceedings of the Fourteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages 79–88. ACM, 2013.

[19] Ruogu Li, Atilla Eryilmaz, and Bin Li. Throughput-optimal wireless scheduling with regulated inter-service times. In *Proceedings of IEEE INFOCOM, 2013*, pages 2616–2624. IEEE, 2013.

[20] Bin Li, Ruogu Li, and Atilla Eryilmaz. Throughput-optimal scheduling design with regular service guarantees in wireless networks. *arXiv preprint arXiv:1405.0625*, 2014.

[21] Jin Tang and Yu Cheng. Selfish misbehavior detection in 802.11 based wireless networks: An adaptive approach based on markov decision process. In *Proceedings of IEEE INFOCOM, 2013*, pages 1357–1365. IEEE, 2013.

[22] Pradeep Kyasanur and NF Vaidya. Selfish mac layer misbehavior in wireless networks. *IEEE Transactions on Mobile Computing*, 4(5):502–516, 2005.

[23] Konstantinos Pelechrinis, Guanhua Yan, Stephan Eidenbenz, and Srikanth V Krishnamurthy. Detecting selfish exploitation of carrier sensing in 802.11 networks. In *Proceedings of IEEE INFOCOM, 2009*, pages 657–665. IEEE, 2009.

[24] Xiaohan Kang, Juan José Jaramillo, and Lei Ying. A strategy-proof and non-monetary admission control mechanism for wireless access networks. In *Quality, Reliability, Security and Robustness in Heterogeneous Networks*, pages 172–187. Springer, 2012.

[25] Veeraruna Kavitha, A Eitan, Rachid El-Azouzi, and Rajesh Sundaresan. Fair scheduling in cellular systems in the presence of noncooperative mobiles. In *Proceedings of IEEE INFOCOM, 2010*, pages 1–9. IEEE, 2010.

[26] Veeraruna Kavitha, Eitan Altman, Rachid El-Azouzi, and Rajesh Sundaresan. Opportunistic scheduling in cellular systems in the presence of noncooperative mobiles. *IEEE Transactions on Information Theory*, 58(3):1757–1773, 2012.

[27] Veeraruna Kavitha, Tejas Bodas, and D Manjunath. On-demand ofdma: Control, fairness and non-cooperation. In *11th International Symposium on Modeling & Optimization in Mobile, Ad Hoc & Wireless Networks (WiOpt), 2013*, pages 224–231. IEEE, 2013.

[28] François Mériaux, Stefan Valentin, Samson Lasaulce, and Michel Kieffer. An energy-efficient power allocation game with selfish channel state reporting in cellular networks. In *6th International Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS), 2012*, pages 290–297. IEEE, 2012.

[29] Mingyi Hong and Alfredo Garcia. Mechanism design for base station association and resource allocation in downlink ofdma network. *IEEE Journal on Selected Areas in Communications*, 30(11):2238–2250, 2012.

[30] Vikram Srinivasan, Pavan Nuggehalli, Carla-Fabiana Chiasserini, and Ramesh R Rao. Cooperation in wireless ad hoc networks. In *Twenty-Second Annual Joint Conference of the IEEE Computer and Communications, INFOCOM 2003*, volume 2, pages 808–817. IEEE, 2003.

[31] Levente Buttyán and Jean-Pierre Hubaux. Stimulating cooperation in self-organizing mobile ad hoc networks. *Mobile Networks and Applications*, 8(5):579–592, 2003.

[32] Levente Buttyán and Jean-Pierre Hubaux. Enforcing service availability in mobile ad-hoc wans. In *Proceedings of the 1st ACM international symposium on Mobile ad hoc networking & computing*, pages 87–96. IEEE Press, 2000.

[33] Luzi Anderegg and Stephan Eidenbenz. Ad hoc-vcg: a truthful and cost-efficient routing protocol for mobile ad hoc networks with selfish agents. In *Proceedings of the 9th Annual International Conference on Mobile Computing and Networking*, pages 245–259. ACM, 2003.

[34] I-Hong Hou, Yu-Pin Hsu, and Alex Sprintson. Truthful and non-monetary mechanism for direct data exchange. In *Allerton*, pages 406–412, 2013.

[35] Y. Bernet, P. Ford, R. Yavatkar, F. Baker, L. Zhang, M. Speer, R. Braden, B. Davie, J. Wroclawski, and E. Felstaine. A framework for integrated services operation over diffserv networks, 2000.

[36] Suong H Nguyen, Hai L Vu, and Lachlan LH Andrew. Service differentiation without prioritization in ieee 802.11 wlans. *IEEE Transactions on Mobile Computing*, 12(10):2076–2090, 2013.

[37] Pavan Nuggehalli, Mahasweta Sarkar, Kishor Kulkarni, and Ramesh R Rao. A game-theoretic analysis of qos in wireless mac. In *Proceedings of IEEE INFOCOM 2008*. IEEE, 2008.

[38] I-Hong Hou and PR Kumar. A survey of recent results on real-time wireless networking. *Proceedings of Real-Time Wireless for Industrial Applications*, 2011.

[39] Jennifer Price, Pavan Nuggehalli, and Tara Javidi. Incentive compatible mac-layer qos design. In *5th IEEE Consumer Communications and Networking Conference, 2008*, pages 586–590. IEEE, 2008.

[40] William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, 16(1):8–37, 1961.

[41] Roger B Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73, 1981.

[42] I-H. Hou and P.R. Kumar. Utility maximization for delay constrained qos in wireless. In *Proc. of IEEE INFOCOM*, 2010.

[43] Pavan Nuggehalli, Jennifer Price, and Tara Javidi. Pricing and qos in wireless random access networks. In *Proceedings of IEEE GLOBECOM 2008*, pages 1–5. IEEE, 2008.

[44] Stefan Mangold, Sunghyun Choi, Peter May, Ole Klein, Guido Hiertz, and Lothar Stibor. Ieee 802.11 e wireless lan for quality of service. In *Proc. European Wireless*, volume 2, pages 32–39, 2002.

[45] A. Heyde P. Branch and G. Armitage. Rapid identification of skype traffic. In *ACM NOSSDAV 2009*, pages 47–56, Williamsburg, Virginia, USA, June 2009.

[46] L. H. Do and P. Branch. Real time voip traffic classication, 2009. Techni-cal Report 090914AIR].[S. 1.]: CAIA.

[47] E. Hossain L. Le and A. Alfa. Service differentiation in multirate wireless networks with weighted round-robin scheduling and arq-based error control. *IEEE Transactions on Communications*, 54:208–215, Feb. 2010.

[48] David Tse. Multiuser diversity in wireless networks. In *Wireless Communications Seminar, Standford University*, 2001.

[49] Philip. Hall. The distribution of means for samples of size n drawn from a population in which the variate takes values between 0 and 1, all such values being equally probable. *Biometrika*, 19(3/4):240–245, 1927.

[50] ZI. Yang Z. Niu S. Zhou, JIE. Gong and P. Yang. Green mobile access network with dynamic base station energy saving. In *Proc. of ACM Mobicom*, 2009.

[51] C. Desset I. Godor P. Skillermark M. Olsson M. Imran D. Sabella M. Gonzalez O. Blume G. Auer, V. Giannini and A. Fehske. How much energy is needed to run a wireless network? *IEEE Wireless Comm.*, 18(5):40–49, Oct. 2011.