A Study of Profiled Handoff for Diffserv-Based Mobile Nodes

Muhammad Jaseemuddin[1], Junaid Ahmed Zubairi[2], and Omer Mahmoud[3]

[1]Ryerson Polytechnic University, Canada, [2]SUNY at Fredonia, USA

[3]International Islamic University, Malaysia

Abstract

There is a growing interest in providing Internet services to the mobile nodes. When mobile nodes travel from one router's service area to another, the time-sensitive applications may see degradation in service. We investigate into the effects of handoff on service quality of mobile nodes. Several experiments are conducted using various packet metering and marking schemes with or without transferring profiles to the new router. Results indicate the relative instability period following handoff, loss of packets and delay encountered by packets in profiled or un-profiled handoffs, leading to determining suitable mix of metering and marking schemes with or without context transfer.

1.0 Introduction

Best effort service has been acceptable for traditional Internet applications like web, email and file transfer but it is inadequate for new classes of applications such as audio and video streaming. These new applications demand high data throughput and low latency. Diffserv (Differentiated services) architecture [1] has been proposed as a scalable solution that can satisfy the new applications' requirements. Recent research has focussed on the provision of the Internet service in wireless network, as this is widely seen to be the next growth area for the Internet [2]. Based on this premise, researchers have identified the integration of wireless systems and Diffserv as promising research direction. A number of papers that analyze and investigate open issues in Diffserv and wireless have been published recently. One such paper [3] identifies modifications that need to be done to make Diffserv suitable for wireless networks. It is suggested to add lightweight signaling protocol to Diffserv. Diffserv architecture does not require end-to-end signaling and follows an implicit admission control mechanism. In wireless networks a simple signaling scheme is required because the mobility of the nodes creates provisioning and limitation problems.

One of the interesting areas in wireless network research is the effect of handoff on the service quality. Some authors [4] have analyzed the performance of mobile nodes with respect to the three available MIP (Mobile IP) movement detection methods, namely LCS (Lazy Cell Switching), PM (Prefix Matching) and ECS (Eager Cell Switching). In LCS, MN (mobile node) uses agent advertisement lifetime as indication of movement. In PM it can infer subnet prefixes of mobility agent in order to determine new agents. In ECS it is assumed that MNs tend to change their direction of movement very slowly. That is, if they are moving forward in one direction, it unlikely that they will stop and turn back. Hence, it is appropriate for nodes to handoff immediately upon encountering a new agent. The performance of communications involving TCP (Transport Control Protocol) over MIPv4 during handoffs is experimentally analyzed. The efficiency of MIP handoff is measured in terms of service disruption duration. The results indicate that no movement detection method can offer a MIP handoff without suffering some period of service disruption.

In a wireless access network an edge router connected to one or more base stations, called RER (Radio Edge Router), provides connectivity to a mobile node. The RER builds ***context*** for the flows communicated between the mobile and the CN (Correspondent Node). CN is a wired node that communicates with the mobile node. ***Context*** is defined as the information on the current state of a routing-related service required to re-establish the routing-related service on a new subnet without having to perform the entire protocol exchange with the mobile host from scratch [5]. A service that can potentially modify the default routing treatment of packets to and from the mobile node is a routing-related service. A Diffserv enabled access router keeps configuration and state contexts [6]. The example of state context is the estimated bandwidth computed by a meter. The parameters with which a meter is configured are known as configuration context, e.g. AVERAGE_INTERVAL, CTR and PTR for tswTCM meter.

In a Diffserv domain, an edge router performs traffic classification [1] and maintains the profile as the context. When a mobile node moves from one RER to another, the new RER lacks the context maintained by the previous RER. The context needs to be transferred to new RER to provide similar services to the mobile node. Figure 1 illustrates the situation when a mobile node moves from one RER to another.

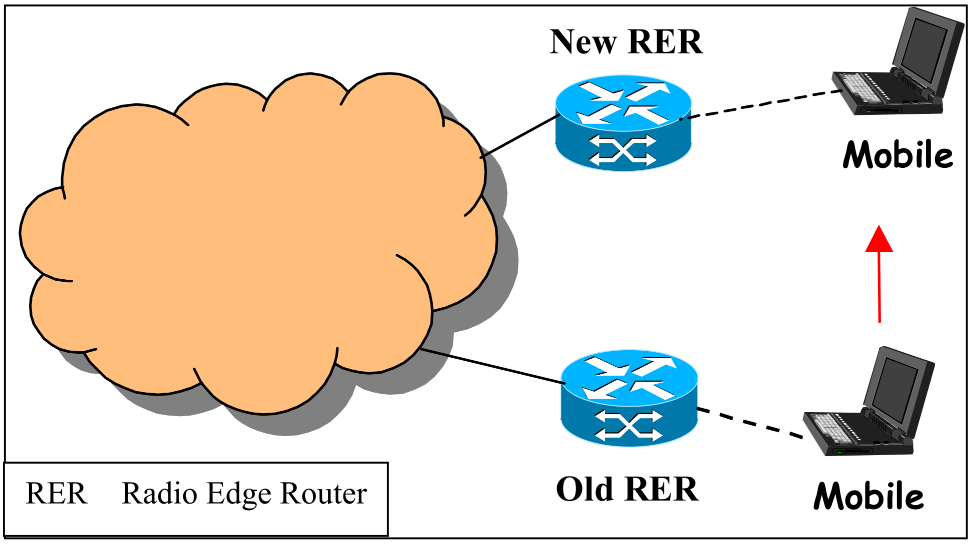

Figure 1: Mobile Node's movement from old RER to new RER

There are two possible ways to deal with the problem of transferring Diffserv context.

(1) The new RER builds the flow profile from scratch. For example, the new RER computes the average bandwidth estimate afresh.
(2) Alternatively, during handoff (transfer of mobile node from one RER to another) the flow profile may also be transferred to the new RER.

The objective of our study is to assess the benefit of context transfer during handoff. We explore the benefits of transferring context and the ***window of opportunity***, that is the time period after which the benefits diminish. The ***window of opportunity*** is defined as the time elapsed between two time instants. The first time instant is when the MN gets connected to the new RER. The second time instant is when the context has been built afresh at the new RER. At the second time instant, the utility function for the transfer of state from the old RER is reduced to zero. It shows that the state transfer is effective if it happens within a finite interval of time given by the window of opportunity. After that interval, the state built at the new RER will be good enough for correct traffic estimation and context information from old RER will not be needed. For example, in tswTCM during the AVERAGE_INTERVAL the estimated value of average bandwidth is in the neighborhood of CTR and later it reaches to the close approximation of the actual bandwidth. Hence, if the state transfer takes more than k*AVERAGE_INTERVAL then the transfer of average bandwidth estimation from the old RER is of no value. The value of integer k depends upon the traffic profile.

The performance metric used is this study is the number of packets that are colored green, yellow, and red by various marking schemes. These statistics are collected in both cases (context transfer/no transfer) in some intervals, and results are compared. We report and analyze results for Time Sliding Window Three-Color Marker (tswTCM), Single Rate Three- Color Marker (srTCM) and Two Rate Three-Color Marker (trTCM) schemes for traffic directly and indirectly affected by handoff.

The three color markers tswTCM, srTCM and trTCM can be used in conjunction with the AF PHB (Per-Hop Behavior) to create a service where a service provider can provide decreasing levels of bandwidth assurance (Green, Yellow & Red) for packets originating from customer sites. The srTCM is useful for ingress policing of a service, where only the length, not the peak rate, of the burst determines service eligibility. The trTCM is useful for ingress policing of a service, where a peak rate needs to be enforced separately from a committed rate. The tswTCM operates based on simple control theory principles of proportionally regulated feedback control. [7]

2.0 Differentiated Services

Diffserv [1] provides a scalable means of service differentiation in the Internet. No per-flow state needs to be maintained in the core routers, neither is there an explicit connection setup phase. The Diffserv architecture offers a framework within which service providers can offer each customer a range of network services, which are differentiated on the basis of performance. Diffserv offers a wide range of services through a combination of functions.

2.1 Differentiated Service Domain

A Differentiated Service Domain is a set of Diffserv nodes, which operate with a common service provisioning policy and set of PHB groups executed on each node**.** Diffserv domain boundary consists of edge routers that classify and condition the incoming traffic using TCA (Traffic Conditioning Agreement) .The traffic enters the domain through ingress and accordingly the packet is mapped to a suitable PHB from one of the PHB groups supported by the domain. Packet classification can be done according to DSCP (Diffserv Code Point) or as per MF (multiple fields) classification. Diffserv interior nodes connect to other Diffserv interior or boundary nodes within the same Diffserv domain.

The SLA (Service Level Agreement) may specify packet classification and re-marking rules and may also specify traffic profiles and actions to traffic streams which are in- or out-of-profile. The TCA (Traffic Conditioning Agreement) between the domains is derived from this SLA. Traffic conditioner, located in a Diffserv boundary node, performs metering, shaping, policing and/or re-marking to ensure that the traffic entering the Diffserv domain conforms to the rules specified in the TCA. Traffic conditioning may vary from codepoint re-adjustment to complex policing and shaping operations. Traffic streams are classified, marked, and conditioned on either end of a boundary link. A *traffic profile* is the description of the temporal properties of a traffic stream selected by a classifier. It provides rules for determining whether a particular packet is in-profile or out-of-profile. For example, a profile based on a token bucket may look like:

Codepoint = **X**, use token-bucket **R**, **B**

The above profile indicates that all packets marked with Diffserv codepoint **X** should be measured against a token bucket meter with rate **R** and burst size **B**. Thus, packets arriving to a bucket holding an insufficient number of tokens are out-of-profile packets. Traffic conditioner may contain a meter, marker, shaper and dropper. A meter is used to measure the traffic stream against a traffic profile. The state of the meter with respect to a particular packet (whether it is in- or out-of-profile) may be used to affect a marking, dropping, or shaping action. Packet markers set the DSCP of a packet to a particular codepoint, adding the marked packet to a particular PHB. Once the marker has done its job, downstream device need only provide appropriate service. Shapers delay some or all of the packets in a traffic stream in order to bring the stream into compliance with a traffic profile. A shaper usually has a finite-size buffer, and packets may be discarded if there is not sufficient buffer space to hold the delayed packets. Droppers discard some or all of the packets in a traffic stream in order to bring the stream into compliance with a traffic profile. This process is known as "policing" the stream. A dropper can be implemented as a special case of a shaper by setting the shaper buffer size to zero (or a few) packets.

A PHB is a description of the externally observable forwarding behavior of a Diffserv node applied to particular Diffserv behavior aggregate. A number of PHB's have been suggested for the Diffserv architecture [8]. Of these, two PHB's, namely AF (Assured Forwarding) and EF (Expedited Forwarding) are implemented in the routers.

AF PHB is suggested [9] for applications that require a better reliability than the best-effort service. There are four classes of service, where each AF class in each Diffserv node allocates a certain amount of forwarding resources (buffer space and bandwidth). Within each AF class, there are three drop precedences. In case of congestion, the drop precedence of a packet determines the relative importance of the packet within each class. All packets are admitted into an assured queue to avoid out of order delivery. The queue is managed by a queue management scheme called Random Early Detection RED with In and Out (RIO).

2.2 Resource Management Architecture for Diffserv

Allocating and controlling the bandwidth within a Diffserv domain is very important in order to meet the targets of an organization. One possible approach is to use the edge router to act as admission control agent. Alternatively, a bandwidth broker agent can be used for each domain.

The edge router, beside other functionality described in pervious sections, may act as an admission control agent [10,11]. In addition, it should be configured with the service level capacities available to the customer, per the SLA.

3.0 Traffic Characteristics and Marking Schemes

Network traffic can be classified into time-sensitive traffic (hard real-time traffic, soft real-time traffic) and best effort traffic [12]. Hard real-time traffic requires strict guarantees on delay and generally must be lossless (e.g. video conferencing). Soft real-time traffic also has delay bounds that need to be met, but these bounds can be slightly exceeded. Many soft real-time applications can also accept a small amount of packet loss [13]. Best-effort or data traffic has no delay requirements but short average delay is desired. Data traffic requires lossless transmission but reliable delivery is usually handled in higher layer protocols.

The four critical performance parameters of the network are

1) Throughput (data rate)

2) Delay

3) Delay variation (jitter)

4) Bit Error rate

These parameters are particularly relevant in supporting real-time applications

3.1 Throughput

"The bit rate between two communicating end-systems is the number of binary digits that network is capable of accepting and delivering per unit of time"[14]. Practical bit rate is limited to the capability of the network and the destination in accepting and processing information. From bit rate point of view, the traffic can be classified into the following two categories:

CBR (Constant Bit Rate)

Certain applications generate the traffic with a fixed rate. They are known as CBR applications. An example is digital telephony in which each conversation generates a constant bit rate equal to 64 Kbps.

VBR (Variable Bit Rate)

Some applications generate VBR streams. These applications are expected to produce traffic at a rate that varies with time (e.g. compressed video streams) VBR traffic is bursty in nature leading to difficulties in network traffic management.

The metrics to characterize the burstiness of a stream are: [14]

- ♦ MBR (Mean Bit Rate): the number of bits in the stream averaged over long period of time
- ♦ MBS (Maximum Burst Size): the maximum number of bits in the peak duration
- ♦ Burstiness Ratio: The ratio between the MBR and the MBS

3.2 Delay

Delay is the time elapsed between the emission of the first bit of a data block by the transmitting end system and its reception by the receiving end-system. It is important that the delay bounds for hard real-time applications are met. The delay experienced by the packets of interactive applications should be small.

3.3 Delay Variation (jitter)

When a stream of packets traverses the network, various packets may experience different delays due to buffering in routers. This variation in delay is called jitter. The delay variation is an important performance parameter for the real-time applications and it should be restricted to a certain threshold to avoid service failure. To overcome the delay jitter in multimedia applications, the receiving system waits for a sufficient time, called delay offset, before the play out, so that most delayed packets are given chance to arrive in time.

3.4 Bit Error Rate

Bit Error Rate (BER) defines the percentage of bits that have errors relative to the total number of bits received in a transmission expressed as ten to a negative power. For example, a transmission might have a BER of 10 to the minus 6, meaning that, out of 1,000,000 bits transmitted, one bit was in error. The BER is an indication of how often a packet or other data unit has to be retransmitted because of an error.

In order to support real-time traffic, a mechanism is needed to prioritize data. This is done by classifying traffic into service classes based on expected traffic patterns. Each service class has a data priority level and associated guarantees. Applications that need real-time guarantees first need to classify their traffic into one of the available service classes based on their expected traffic behavior before requesting a QoS (Quality of Service) guarantee from the network. Sources indicate their peak traffic rate (in bytes/sec) and their maximum burst size (in bytes). The network gives a guarantee on the peak rate and tries to minimize packet delay and packet delay variation. [15]

3.5 Marking Schemes

In order to avoid any unforeseen congestion, the users have to agree to a traffic profile before using a Diffserv domain. When the traffic enters a Diffserv domain, it may be monitored, marked and shaped at the ingress node. The purpose of marking and shaping at the ingress node is to make sure that the user is not violating the agreed profile. Some of the important parameters of agreed profile include the committed rate and allowed peak rate. Since some flows may be misbehaving and violating agreed profile, it is important to enforce the policing (metering and marking) and shaping (smoothing the bursts over time) mechanisms at the ingress node. Several algorithms have been proposed for metering and marking the user traffic streams [7,16]. The purpose of marking is to indicate whether the current packet violates the profile or not. Three color marking is considered sufficient with green indicating a good packet, yellow showing a packet that has exceeded committed profile but falls within the peak rate and red showing a violation. Colors are coded using the drop precedence of the AF (assured forwarding) class [9].

3.5.1 Single-Rate Three-Color Marker

srTCM [16] meters an IP packet stream and marks its packets green, yellow, or red. Marking is based on a Committed Information Rate (CIR) and two associated burst sizes, a Committed Burst Size (CBS) and an Excess Burst Size (EBS). A packet is marked green if it doesn't exceed the CBS, yellow if it does exceed the CBS, but not the EBS, and red otherwise. The behavior of the meter is specified in terms of its mode and two token buckets C and E; CIR is applied as token generation rate for the two token buckets. The maximum size of the token bucket C is CBS and the maximum size of the token bucket E is EBS.

The token buckets C and E are initially (at time the beginning of the traffic) full, which mean the token count Tc(0) = CBS and the token count Te(0) = EBS. Thereafter, the token counts Tc and Te are updated CIR times per second as follows,

- If Tc is less than CBS, Tc is incremented by one, else
- If Te is less then EBS, Te is incremented by one, else
- Neither Tc nor Te is incremented

When a packet of size B bytes arrives at time t, the following happens if the srTCM is configured to operate in the Color-Blind mode:

- If Tc(t)-B >= 0, the packet is green and Tc is decremented by B

  down to the minimum value of 0, else

- If Te(t)-B >= 0, the packets is yellow and Te is decremented by B

  down to the minimum value of 0, else

- The packet is red and neither Tc nor Te is decremented

3.5.2 Two-Rate Three Color Marker

The trTCM scheme [17] meters an IP packet stream and marks it green, yellow, or red. The color is coded in the Diffserv field of the packet in a PHB specific manner.

A packet is marked red if it exceeds the Peak Information Rate (PIR). Otherwise it is marked either yellow or green depending on whether it exceeds or doesn't exceed the Committed Information Rate (CIR). The trTCM is useful, for example, for ingress policing of a service, where a peak rate needs to be enforced separately from a committed rate. The behavior of the meter is specified in terms of its mode and two token buckets C and P. The difference in the metering behavior between srTCM and trTCM is that in the case of the trTCM, the token buckets operate with different rates. The token bucket labeled P is filled with PIR (Peak Information Rate) and the token bucket C is filled with CIR (Committed Information Rate). P is filled until it reaches the size PBS (Peak Burst Size)and C is filled until it hits the maximum size of CBS(Committed Burst Size).

The token buckets P and C are initially (at time 0) full, i.e., the token count Tp (0) = PBS and the token count Tc (0) = CBS. Thereafter, the token count Tp is incremented PIR times per second up to PBS and the token count Tc is incremented CIR times per second up to CBS.

When a packet of size B bytes arrives at time t, the following happens if the trTCM is configured to operate in the Color-Blind mode:

- If Tp(t)-B < 0, the packet is red, else
- If Tc(t)-B < 0, the packet is yellow and Tp is decremented by B, else
- The packet is green and both Tp and Tc are decremented by B

3.5.3 Time Sliding Window Three Color Marker

The tswTCM [7] is designed to mark packets of an IP traffic stream with red, yellow or green color. The marking is performed using the estimated average rate as compared against the Committed Target Rate (CTR) and the Peak Target Rate (PTR). The computation of estimated rate is based on a time window in order to take into account the recent behavior of the stream. Packets that confirm to CTR are marked green. Packets that exceed CTR but do not exceed PTR are marked yellow and packets contributing to the portion of the rate above PTR are marked red.

tswTCM consists of two components namely the rate estimator and a marker. The rate estimator provides an estimate of the traffic stream's arrival rate. This rate should approximate the running average bandwidth of the traffic over a specific period of time (window length). If the estimated average rate is less than or equal to the CTR, the packets of the stream are green. If the estimated average rate is greater than the CTR but less than PTR, the packet are marked yellow with probability P0 and designated green with probability (1-P0). If the estimated rate average rate is greater than PTR, packets are designated red with probability P1, designated yellow with probability P2 and designated green with probability (1-(P1+P2)). The tswTCM has been primarily designed for traffic streams that will be forwarded based on the AF PHB in core routers.

4.0 Wireless and QoS

QoS parameters for typical applications include bounds for bandwidth, packet delay, packet loss rate, and jitter. In this section, we consider major differences between wireless and wired worlds in terms of some QoS parameters [18,19]. Bandwidth is often the most obvious difference between .the wired and wireless worlds, wireless network being much slower than the wired counterpart. Wireless networks typically have considerably longer delay than the wired counterparts. The BER is often better than 1 to 10^5 even in worst case wired systems such as modems over phone lines. Wireless usage figures are often worse by at least an order of magnitude, resulting in BER of 1 to 10^4 or worse. Wireless communication is also affected by the fact that the mobile units run on batteries. If the battery goes low, packet transmission to the unit is curtailed by the RER using its power profile. [3]

4.1 Mobility

Maintaining a reservation when a mobile move between regions is a challenge because of possible blackout situations during handoff. A scheme is required to define how smooth this transition should be since it affects the QoS of an application.

4.2 Handoff

In a wireless access network, an edge router called Radio Edge Router (RER) provides connectivity to mobile nodes. Any RER has certain geographic coverage area within which a mobile node can communicate to it. This coverage area is known as a cell. Neighboring RERs overlap with each other's coverage area, thus ensuring continuity of communications when the users move from one cell to another. The procedure of maintaining a call in progress, while moving from one cell to another is called *handoff.* Handoff is caused by factors related to radio link and network management [20]. Radio link related causes reflect the quality perceived by users. Some of the major variables affecting the service quality are received signal strength (RSS) and signal-to-interference ratio (SIR). Insufficient RSS and SIR reduce the service quality. Handoff is required in the following situations [21]:

(i) When the MN approaches the cell boundary (the RSS drops below a threshold)
(ii) if SIR drops below certain level

In a less likely situation, the network may handoff a call to avoid congestion in an access point [22]. Handoff may be hard or soft. Hard handoff (HHO) is “break before make," meaning that the connection to the old access point is broken before a connection to the candidate BS is made. HHO occurs when handoff is made between disjointed radio systems, deferent frequency assignments [20]. Soft handoff (SHO) is “make before break," meaning that the connection to the old access point is not broken until a connection to the new access point is made.

In a soft handoff procedure, the MN is connected to multiple base stations for a period of time. In soft handoff algorithm, the handoff decision is based on the signal strength. If received signal strength from a new access point is higher than adding threshold *T_ADD*, it is added into the user's active set and starts the communication to the user. When the signal strength from a base station in the active set is lower than dropping threshold *T_DROP* for a period of dropping timer (*T_TDROP*) time, it is removed from the active set and loses the connection to the user [23]

4.3 Impact of Handoff on QoS

There are two key issues regarding how handoff affects the QoS of an existing connection [24]. First, when a network accepts a connection to a fixed endpoint, the requested QoS is made available to the connection during its lifetime with a high probability, in the absence of network failures. The same cannot be said for a connection to a mobile node when it changes the point of attachment to the network. The second issue is that the handoff entails a period of time during which the end-to-end connection data path is incomplete. The extent to which this disruption affects the application performance depends on the nature of the application and the period of disruption. For example, a voice application can stand short disruptions but a data application cannot tolerate such errors. Generally

- Handoff should be fast (No disruption of user service)
- Handoff should be reliable (Good post handoff quality)

- Handoff's effect on QoS should be minimal.

5.0 Experimental Setup

It is important to determine the most suitable combination of metering and context transfer schemes so that the effect of handoff on service quality experienced by the mobile nodes is minimized. This study is an effort in that direction. Here, we describe the experimental setup.

In this study, we have used NS-2 simulator on a Linux host [25]. In NS-2, the Diffserv functionality is captured in a Queue object called dsRED and policy class. The Policy class handles the creation, manipulation, and enforcement of edge router policies. Assured Forwarding is implemented using RED mechanism by enqueuing all packets for a single class into one physical queue that is made up of three virtual queues (one for each drop precedence). Different RED parameters are used for the virtual queues, causing packets from one virtual queue to be dropped more frequently than packets from another. The PHB Table handles mapping code points to precedence levels.

Figure 2 and Figure 4 depict the experimental setup. Initially the mobile node (MN) is connected through the Radio Edge Router (RER1) to the network and communicates with the correspondent node (CN), sending CBR traffic at 6Mbps. The CN is fixed node and attached to the Edge Router ER.

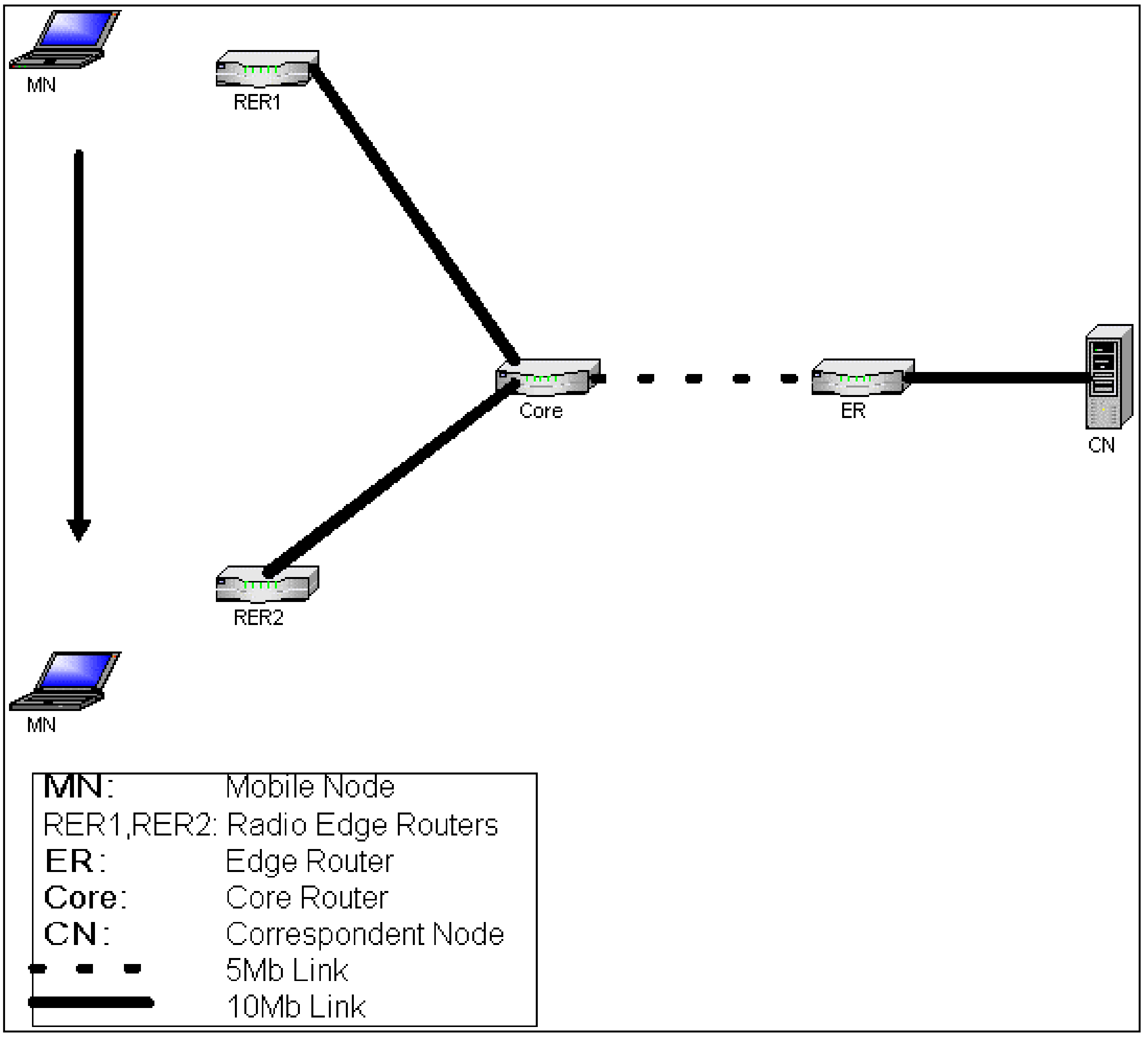

Figure 2: Experimental Setup for Handoff

CTR= 1 Mbps
PTR = 2 Mbps
Window Length (WL) = 1 second

(a) Parameters for TSW

CIR = 1Mbps CBS = 10000 bytes EBS = 30000 bytes	CIR = 1Mbps CBS = 10000 bytes PIR = 2Mbps PBS = 30000 bytes
(b) Parameters for srTCM	(c) Parameters for trTCM

Figure 3(a) Values for tswTCM (b) Values for srTCM (c) Values for trTCM

The mobile node MN remains connected to RER1 for 40 seconds and at the 40th second it moves to the coverage area of RER2. Hence, handoff takes place at the 40th second. The experiments are conducted using three different coloring schemes (srTCM, trTCM, and tswTCM). RERs implement all three coloring schemes. The set up in Figure 2 is aimed at studying the benefit of transferring the context from RER1 to RER2 .A single flow of traffic was considered from MN to CN. Using the parameters [26, 27] shown in Figure 3.

In Figure 4, a new CBR flow at 6Mbps is introduced between node S and node R through RER2. This setup is aimed at studying the impact of new flow, which is handed over from RER1 on the flows in new RER (RER2).

6.0 Results and Discussion

This section presents the results of the simulations, which capture the effect of transferring the context information on packet marking. It shows the packet distribution for green, yellow, and red packets for the case when the context information is transferred and the case when it is not transferred. When context information is not transferred during handoff, the packet marking starts from initial state. Results are presented for two phases of the study. Phase-one results show the effect of the handoff on the flow itself while Phase-two results show the effect of handoff on the other flows on the new RER.

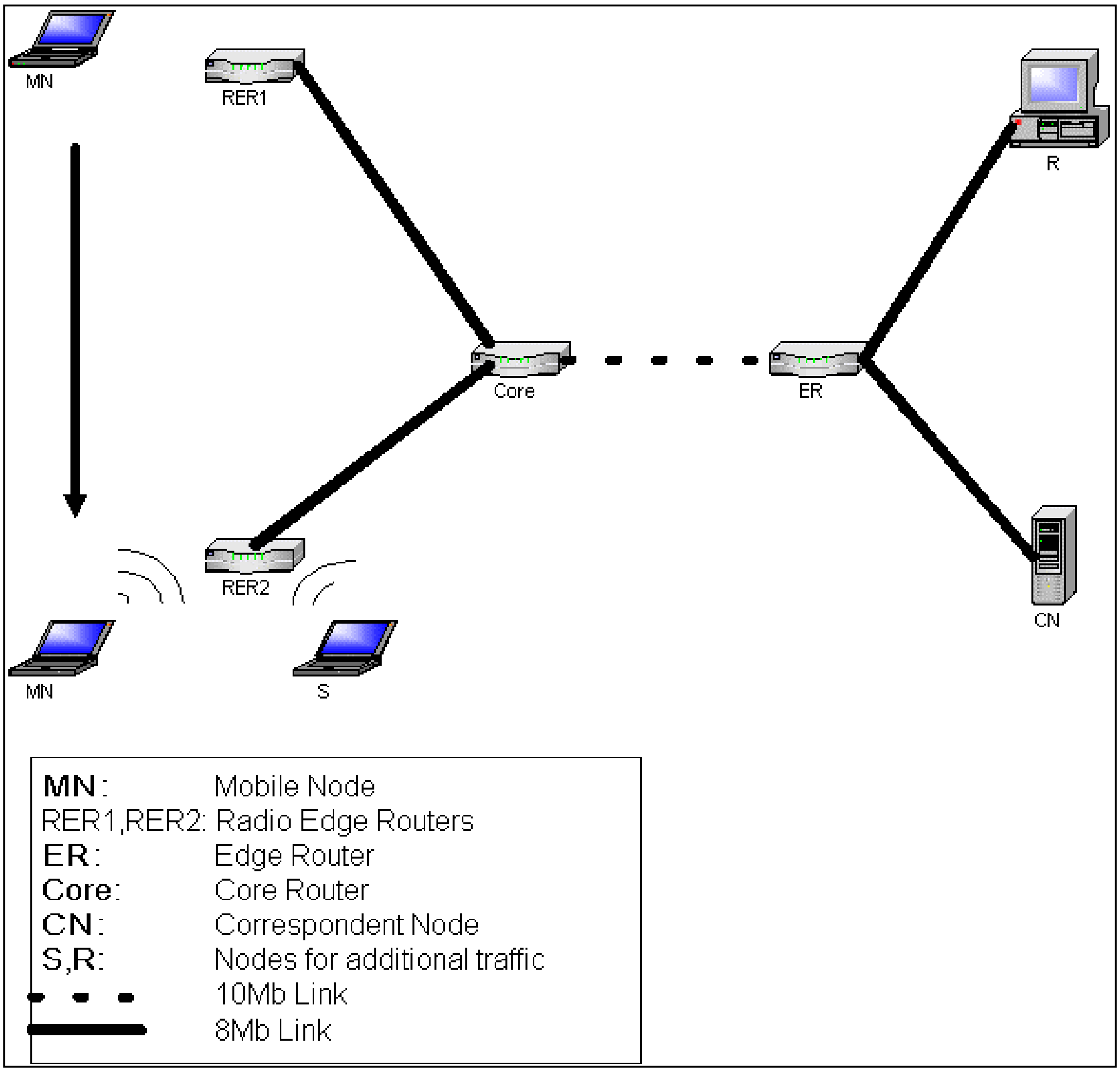

Figure 4: Second setup with new traffic into the domain

6.1 Phase One

This section presents the results obtained using the experimental setup of Figure 2, i.e. the effect of the handoff on the flow that experiences the handoff. The results were obtained by using three different marking schemes. For each marking scheme, three experiments were conducted. The first experiment represents the case when there is no handoff. The second experiment represents the case when there is handoff but without transferring the context and the third experiment includes handoff as well as transfer of context to the new RER

6.1.1 Experiments with tswTCM

A tswTCM meter estimates average packet arrival rate by including both the rate of newly arrived packet and the rate estimated within a window of history. The context information for tswTCM is thus the current average estimated rate. This section present the results obtained by running three experiments using tswTCM as described below:

1. In the first experiment packet statistics are collected for three seconds at RER1 without moving the MN to RER2, hence this is the case without handoff.

2. In the second experiment, packet statistics are collected for three seconds at RER2 without handing over the estimated average rate computed at RER1 to RER2; hence this case shows the data without context transfer. The handoff takes place at 40th second and the handoff latency is assumed to be zero. The

statistics show increase in the number of packets marked green and yellow whereas red marking was reduced substantially.

3. In the third experiment, packet statistics are collected for three seconds at RER2 with handing over the estimated average rate computed at RER1 to RER2 at 40th second (handoff time). The packet marking pattern appears to be similar to case I.

6.1.2 Analysis of tswTCM Results

Figures 5 to 7 show the packet distribution for green, yellow, and red packets that are plotted by collecting statistics at the interval of 0.5 seconds within a period of three seconds from the 40th to 43rd seconds. In these figures, abscissa is the time from 40th to 43rd seconds, while ordinate shows the packet distribution for green (G), yellow (Y), and red (R) packets. Figure 5 shows the packet distribution collected during experiment 1 (without handoff). The marking is almost stable, and it shows stability right from 40th second (starting point of the graph), because of the effect of prior estimated average rate.

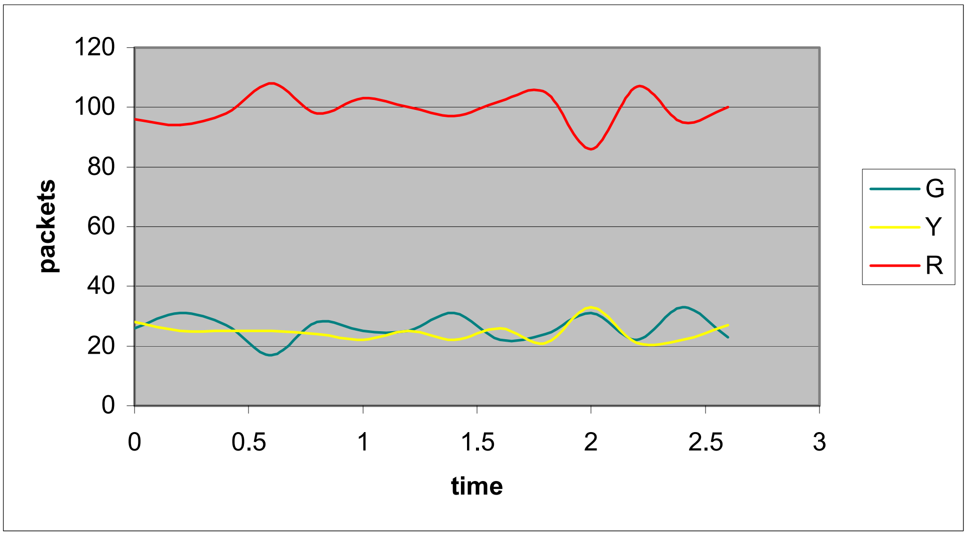

Figure 5 Marking of a flow from MN to CN atRER1 without handoff

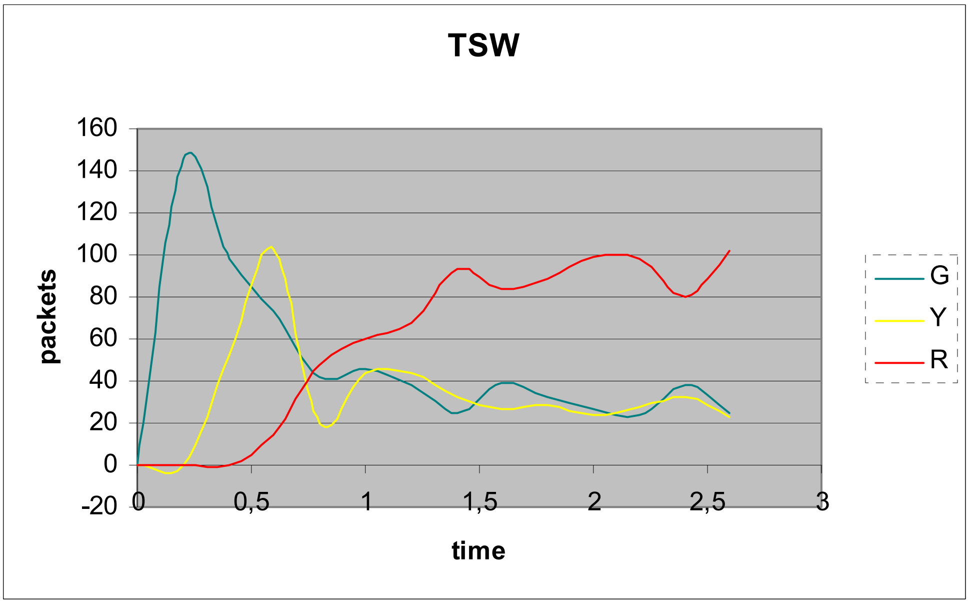

Figure 6 Marking of a flow from MN to CN at RER2 after handoff without transferring the estimated average

In Figure 6, the handoff takes place at t = 40^{th} second. In this case the estimated average rate is not transferred from RER1 to RER2 during handoff. As a result the tswTCM starts marking the incoming traffic from the initial window setting, that is using CTR as the past estimated average rate. The average rate is updated at the arrival of each packet therefore it takes a while for the average to reach a stable value that is reflective of the close approximation of the actual arrival rate. As long as the average rate is less than CTR all the packets are marked green, hence the number of the green packets is high at the beginning and then tapers off later (around one second later) to its stable value. Similar patterns are shown by the distribution of yellow and red packets as well. The graph shows that the rate estimate improves with time and consequently tswTCM marking reaches some sort of stability.

The packet distribution in Figure 7 shows that the marking reaches stability earlier than what it takes for the experiment 2 (as shown in Figure 6). This is because the estimated average rate computed at the 40th second (handoff time) is transferred from RER1 to RER2 during handoff thus calibrating the meter at RER2 to estimate the average rate closer to accuracy.

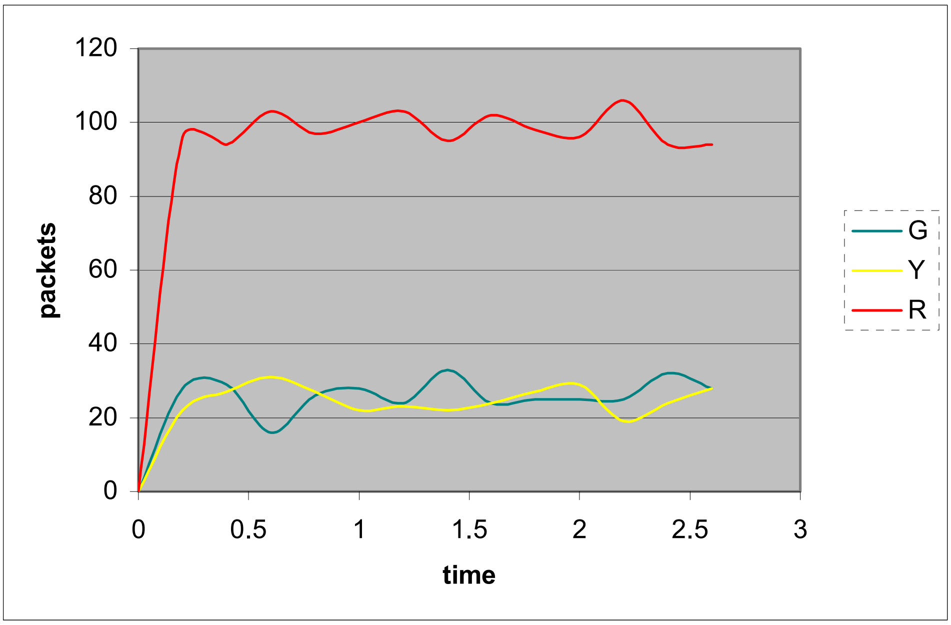

Figure 7 Marking of a flow from MN to CN at RER2 after transferring the estimated average during handoff

6.1.3 Effect of Window Length on Marking Stability for tswTCM

The graphs in Figure 8 and Figure 9 show the effect of the window length on calculating the average rate estimate for the tswTCM scheme. These graphs are obtained with experiment 2 but with window size of 0.5 second and 0.1 second respectively. It is evident from Figures 6, 8, and 9 that the instability period is proportional to the value of window size. The long window size provides more time for transferring the context, but it tends to smooth out bursts for non-CBR traffic. Hence, the initial window size cannot be increased arbitrarily large to derive a longer window of opportunity and in effect relaxing the constraint on transferring the context.

6.1.4 Experiments with srTCM

The marking of newly arrived packets in srTCM depends on the number of tokens that are available in the buckets Tc and Te. The context information for srTCM is thus the current number of tokens (token count) in the buckets Tc & Te. Here we present the results obtained by running three experiments as described below:

1. In the first experiment packet statistics are collected for three seconds at RER1 without moving the MN to RER2, hence this case is without handoff.
2. In the second experiment, packet statistics are collected for three seconds at RER2 without handing over the token count computed at RER1 to RER2; hence this case shows the data without context transfer. The handoff takes place at 40th second and the handoff latency is assumed to be zero.

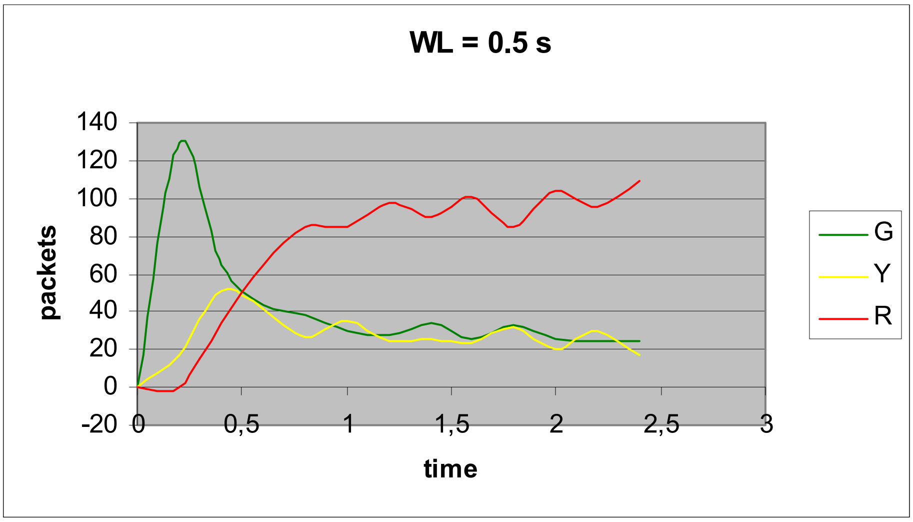

Figure 8 Marking of a flow from MN to CN at RER2 after handoff without transferring the estimated average with window size=0.5 second

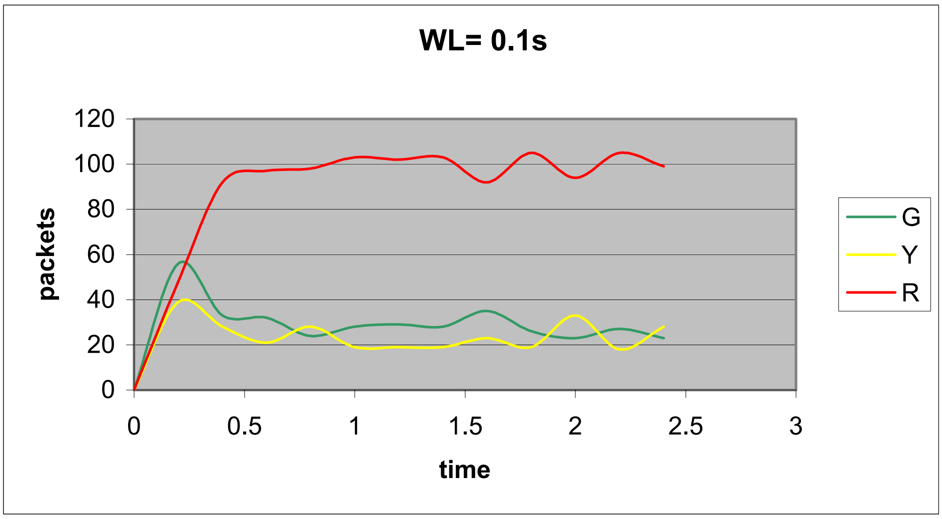

Figure 9 Marking of a flow from MN to CN at RER2 after handoff without transferring the estimated average with window size=0.1 second

3. In the third experiment, packet statistics are collected for three seconds at RER2 with handing over the token count computed at RER1 to RER2 at 40th second (handoff time), hence this case shows the data with context transfer. The handoff takes place at 40th second and the handoff latency is assumed to be zero, context is transferred in zero time.

6.1.5 Analysis of srTCM Results

Figures 10 to 12 show the packet distribution for green, yellow, and red packets that are plotted by collecting statistics at the interval of 0.5 second within a period of three seconds from the 40th to 43rd seconds. In these figures, abscissa represents time from 40 to 43 seconds, while ordinate shows the packet distribution for green (G), yellow (Y), and red (R) packets. Figure 10 shows the packet distribution collected during experiment 4 (without handoff). The marking is almost stable, and it shows stability right from 40th second (starting point of the graph).

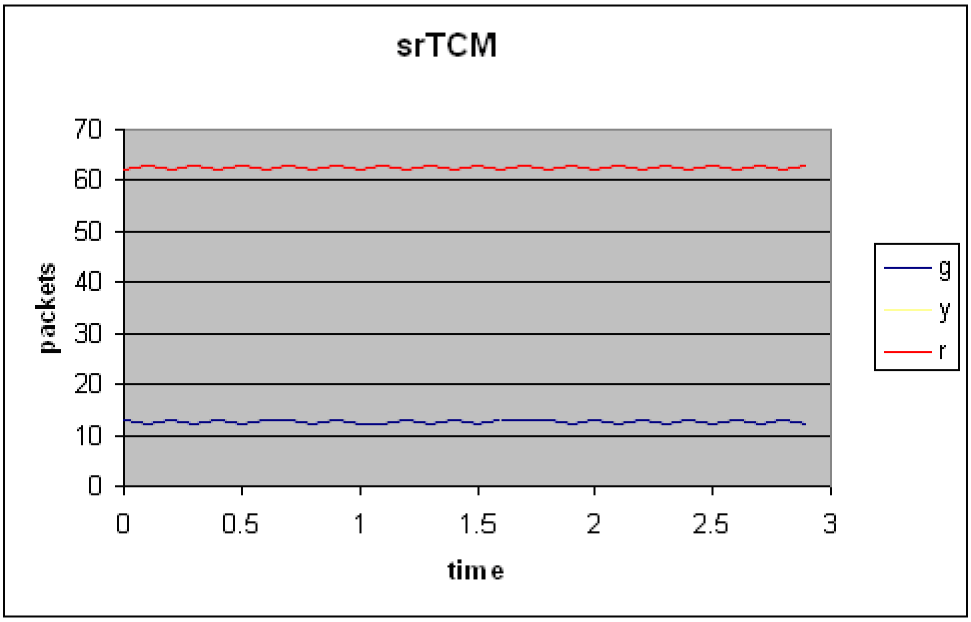

Figure 10 Marking of a flow from MN to CN at RER1 without handoff

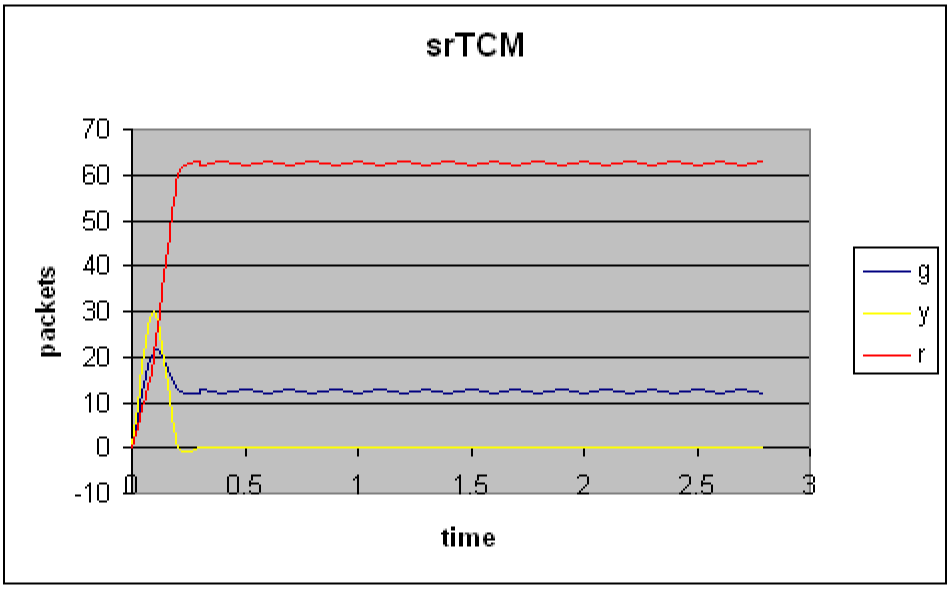

Figure 11 Marking of a flow from MN to CN at RER2 after handoff without transferring the token count

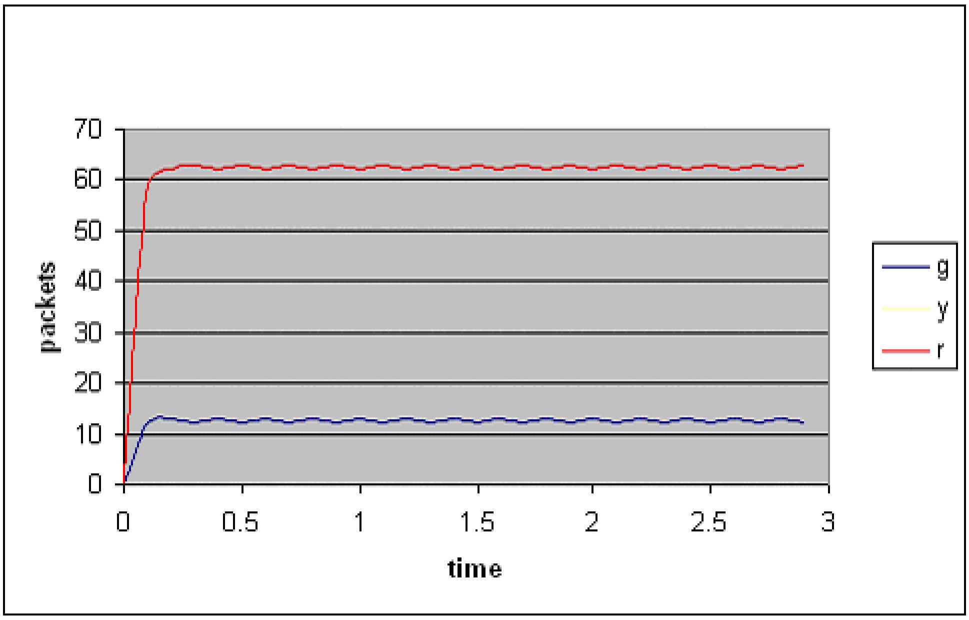

Figure 12 Marking of a flow from MN to CN at RER2 after transferring the token count during the handoff (srTCM)

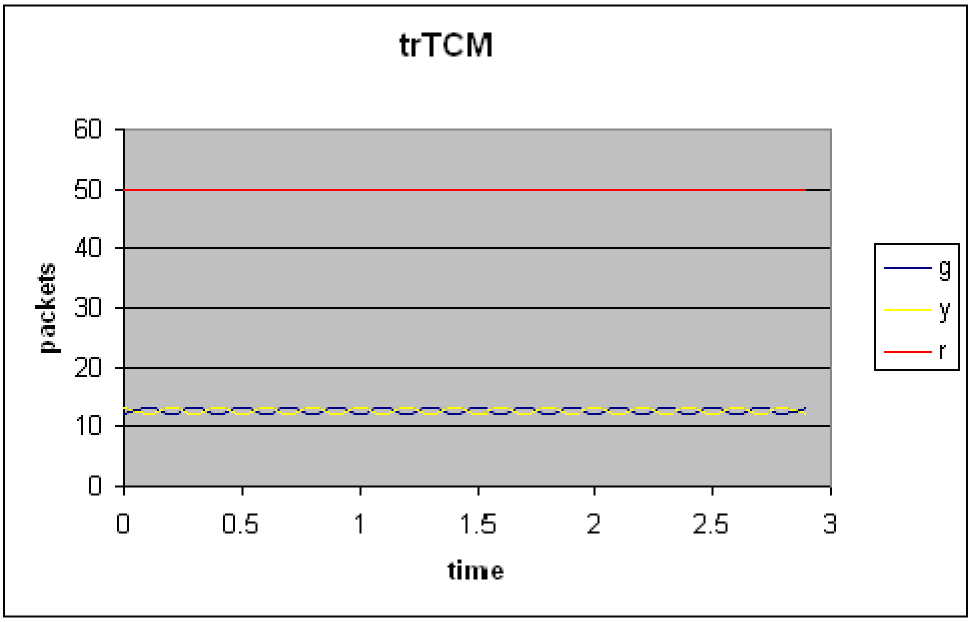

Figure 13 Marking of a flow from MN to CN atRER1 without handoff (trTCM)

In Figure 11 the handoff takes place at t = 40th second. In this case the token count of Tc and Te is not transferred from RER1 to RER2 during handoff. As a result, the srTCM starts marking the incoming traffic from the initial window setting. Tc & Te are set to the maximum value CBS and EBS. This allows a burst equal to CBS to pass as green and EBS as yellow at the beginning of the handoff as to shown in Figure 11.

The packet distribution in Figure 12 shows that the marking results are similar to the marking shown in Figure 10. This is because the token count computed at the 40th second (handoff time) is transferred from RER1 to RER2 during handoff. This calibrates the mater at RER2 to get results closer to the results in experiment 4.

6.1.6 Experiments with trTCM

The marking of a new packet in trTCM depends on the number of tokens that are available in the buckets Tc and Tp. The context information for trTCM is thus the token count in the buckets Tc& Tp. This section presents the results obtained by running three experiments as described below:

1. In the first experiment packet statistics are collected for three seconds at RER1 without moving the MN to RER2, hence this case shows the data without handoff.

2. In the second experiment, packet statistics are collected for three seconds at RER2 without handing over the token count computed at RER1 to RER2; hence this case shows the data without context transfer.

3. In the third experiment, packet statistics are collected for three seconds at RER2 with handing over the token count computed at RER1 to RER2 at 40th second (handoff time), hence this case shows the data with context transfer. The handoff takes place at 40th second and the handoff latency is assumed to be zero, context is transferred in zero time.

6.1.7 Analysis of trTCM Results

Figures 13 to 15 show the packet distribution for green, yellow, and red packets that are plotted by collecting statistics at the interval of 0.1 second within a period of three seconds from the 40th to 43rd seconds. In these figures abscissa represents time from 40th to 43rd second, while ordinate shows the packet distribution for green (G), yellow (Y), and red (R) packets. Figure 13 shows the packet distribution collected during experiment 7 (without handoff). The marking is almost stable, and it shows stability right from 40th second (starting point of the graph).

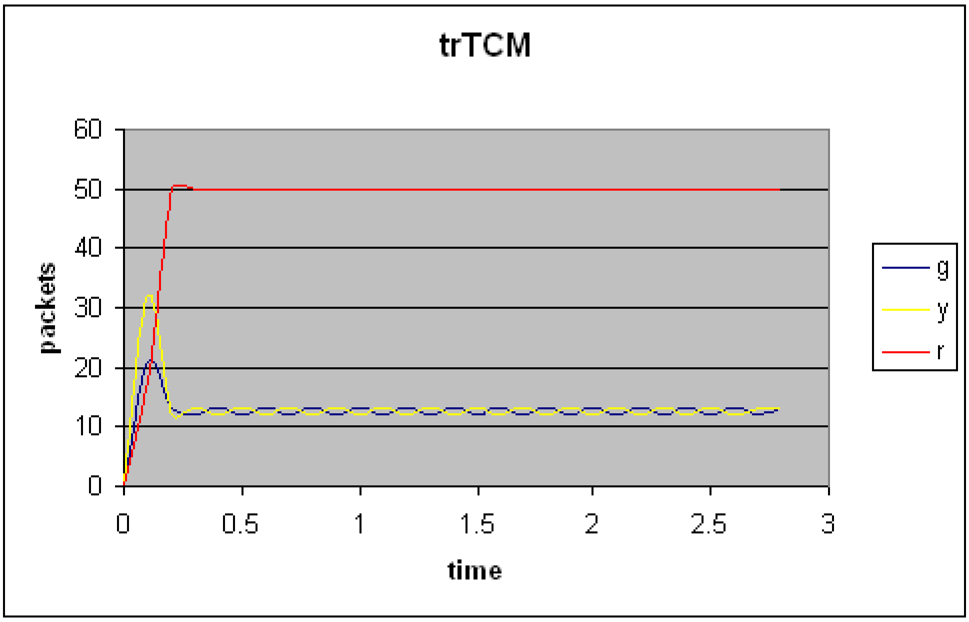

Figure 14 Marking of a flow from MN to CN at RER2 after handoff without transferring the token count (srTCM)

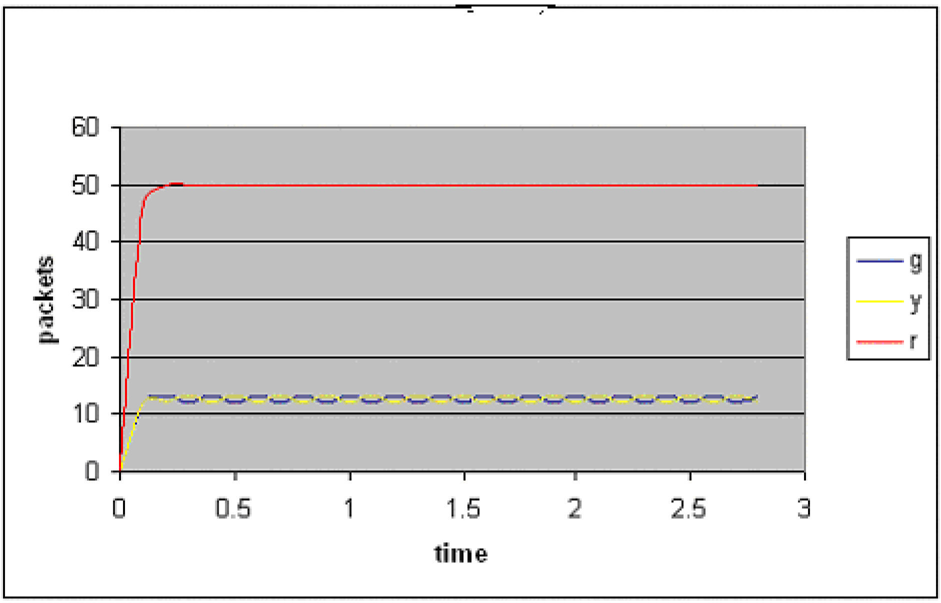

Figure 15 Marking of a flow from MN to CN at RER2 after transferring the token count during the handoff (trTCM)

In Figure 14 the handoff takes place at t = 40^{th} second. In this case the token account is not transferred from RER1 to RER2 during handoff. As a result the trTCM starts marking the incoming traffic from the initial window setting. Tc & Tp set to the maximum value CBS &PBS this allows a burst equal to CBS to pass as green and PBS as yellow at the beginning of the handoff as to shown in Figure 14.

The packet distribution in Figure 15 shows that the marking results are similar to the marking shown in Figure 13. This is because the token count of Tc &Tp computed at the 40th second (handoff time) is transferred from RER1 to RER2 during handoff.
This calibrates the mater at RER2 to get results closer to the results in experiment 7.

6.2 Losses, Delay and Jitter

Table I shows the mean delay, jitter and losses for the first case, handoff without transferring the context. Table II shows the mean delay jitter and losses for second case, handoff in which the context is transferred from old RER to the new RER. Comparing the results in Table I and II, it is obvious that tswTCM is more sensitive to context information and transferring the average rate results in decreasing the delay mean jitter and losses. This is consistent with the graphs shown in Figure 6 and 7. The packets transferred during the initial instability period experience high jitter. Since without transferring average rate after handover takes more time for marking to reach the steady state than the case where context is transferred, more packets experience high jitter causing higher mean jitter value. For srTCM and trTCM the initial instable period is short, hence the comparison shows no significant difference in jitter and losses.

Table I: The losses, delay mean and jitter for three seconds after the handoff without transferring the context

	Total	Received	Losses	Delay mean (ms)	Jitter (ms)
tswTCM	2250.	1764	486	43.6854	21.9423
srTCM	2250	1850	400	38.9405	11.9029
trTCM	2250	1849	401	38.9312	12.8883

Table II: The losses, delay mean and jitter for three seconds after the handoff when the context is transferred to the new RER

	Total	Received	Losses	Delay mean (ms)	Jitter (ms)
tswTCM	2250.	1855	395	38.9719	12.7875
srTCM	2250	1852	398	38.1372	12.708
trTCM	2250	1847	403	38.959	13.048

6.3 Phase Two

This section shows the impact of the new flow that experiences the handoff on an existing flow at the new RER. The experiments are conducted using three schemes tswTCM, srTCM & trTCM with handoffs similar to phase one. Additionally, a flow from node S to node R through RER2 is being transmitted when the handoff occurs. Figure 3 shows the experimental setup used in this section. Throughout this section "new flow" means the flow that experiences the handoff and "old flow" means the existing flow at new RER (flow from S to R through RER2).

6.3.1 Experiments with tswTCM

Handoff takes place in both experiments but the context is only transferred in the second one. The results obtained are plotted as shown in following figures. Figure 16 & 17 show the total packets, transmitted packets and dropped packets for duration of one second after the handoff. Here the context is not transferred during the handoff. Figure 17 shows the total, transmitted and dropped packets for the old flow for the duration of one second after the hand off when context is transferred to RER2 at the beginning of handoff.

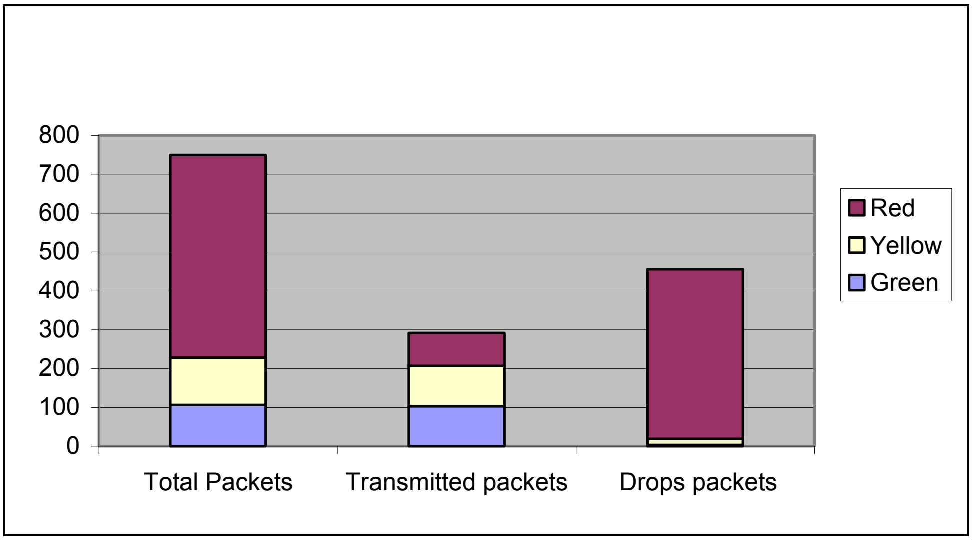

Figure 16 Packet statistic at new RER for the old flow without

Transferring the context (tswTCM)

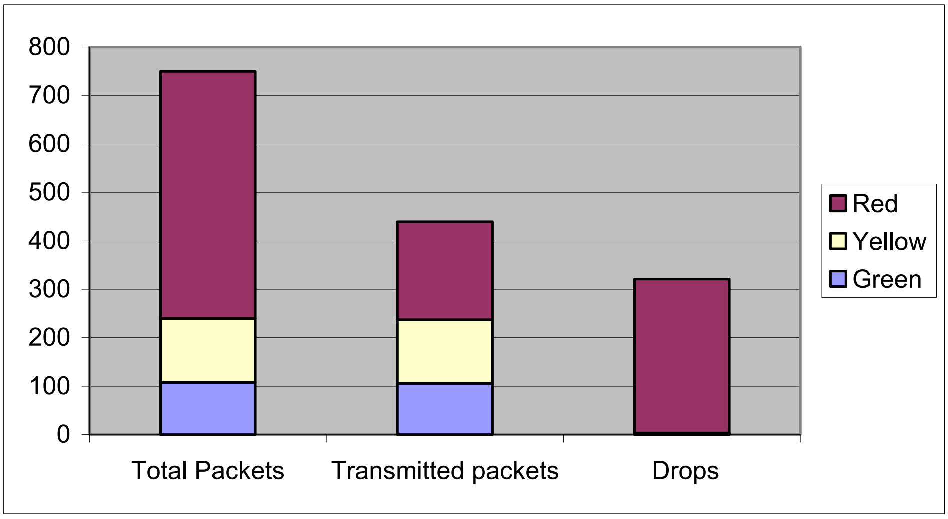

Figure 17 Packet statistic at new RER old flow with the context being transferred to

new RER (tswTCM)

As a result of instability of coloring and marking that is caused by the handoff the flows at the new RER experience loss of packets (green, yellow, and red) as shown in Figure 16. Figure 17 demonstrates that transferring the context minimizes the loss for the old flow at the new RER.

6.3.2 Experiments with srTCM

In this experiment, srTCM is implemented in RER and results are obtained running two experiments as described in pervious section. Figure 18 shows the total, transmitted and dropped packets for the old flow for the duration of one second after the handoff. Here the context is not transferred during the handoff. Figure 19 shows the packet distribution for the old traffic when the context is transmitted from the old RER to the new RER .

6.3.3 Experiments with trTCM

Figure 20 and 21 show the packet distribution obtained using trTCM at RER. The experiments were conducted twice. In Figure 20, the context is not transferred and in Figure 21, the context is transferred to the new RER.

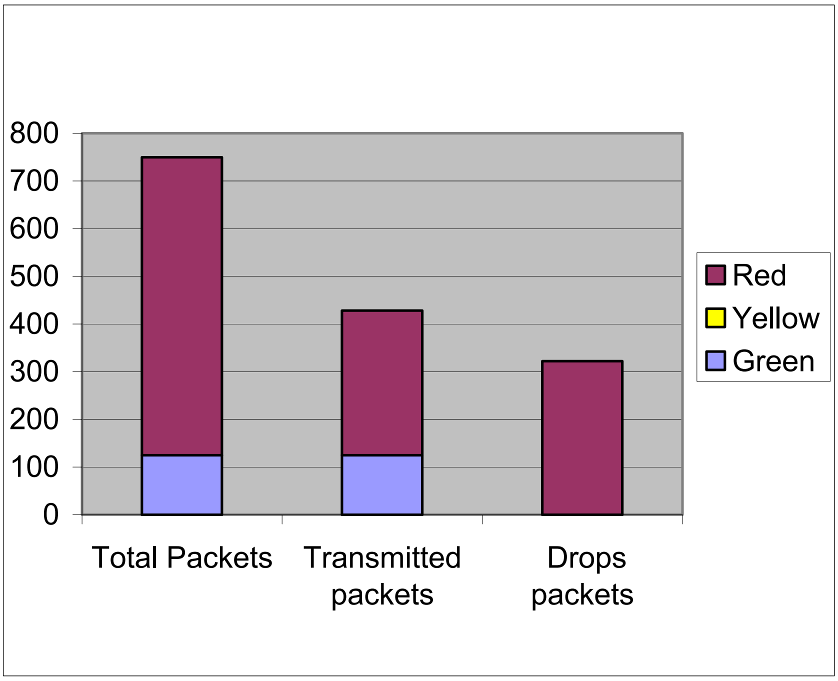

Figure18 packet statistic at new RER for old flow with out transferring the context (srTCM)

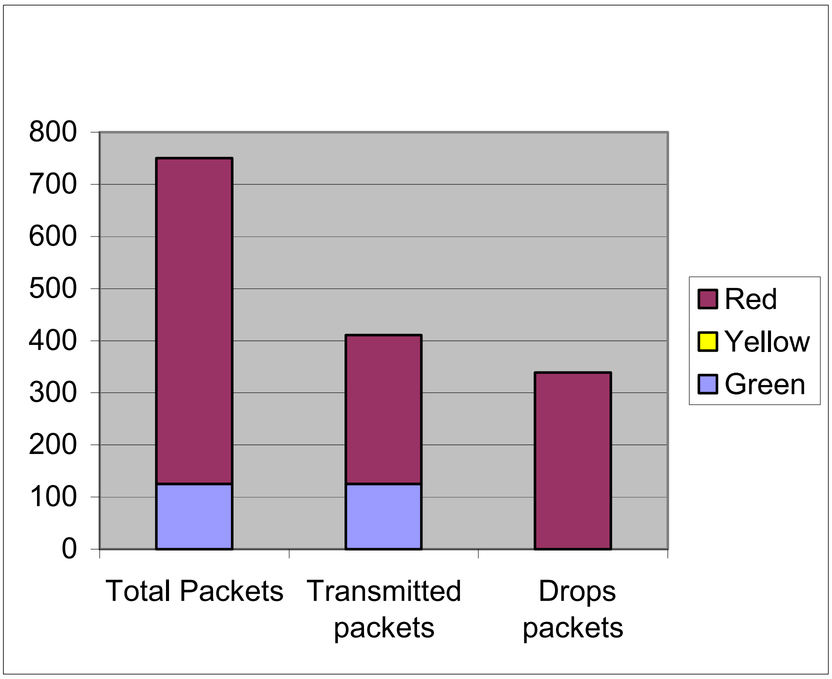

Figure 19 Packet statistic at new RER for the old flow with the context being transferred to new RER(srTCM)

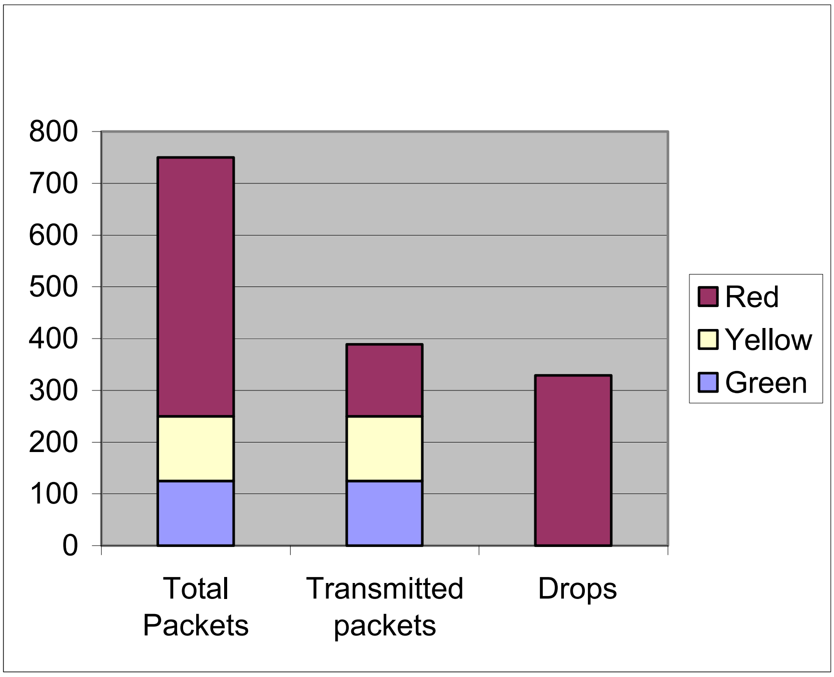

Figure 20 packet statistic at new RER for the old flow with out transferring the context (trTCM)

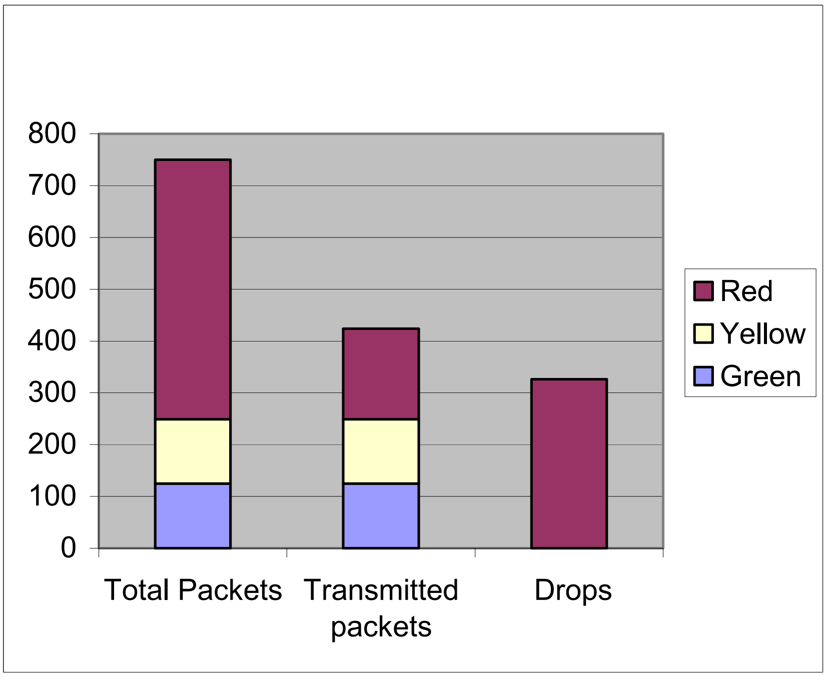

Figure 21 Packets statistic at new RER for the old flow with the context being transferred to new RER(trTCM)

The above figures show the instability caused by the handoff in marking the flow results in higher number of green and yellow nonconforming packets leading to some drops in the green and yellow packets for traffic of the same class in case of the tswTCM. For the srTCM and trTCM the effect of handoff in flow marking is minimal which means the effect on the flow at new RER is small compare to the tswTCM.

7.0 Conclusion And Future Work

There is a growing need to provide quality of service to the mobile users based on an increase in the number of time-sensitive applications and ensuring this QoS for mobile users while they are moving around. If a mobile node connects to a Diffserv domain, it is subjected to the same policing and shaping as done for the static nodes. If this node transfers over from one radio edge router to another, while maintaining connection to the same static node, the service offered to this node may go through a transitory change.

When a mobile node is connected to a RER, the RER builds some QoS context related to the communication of the mobile with correspondent nodes. When the mobile moves and changes its connectivity from one RER to another, the new RER does not have that context unless it is transferred from the old RER We have investigated the benefits of context transfer during handoff situation. Simulations are conducted using network simulator NS-2. The study is conducted considering a Diffserv domain where RERs run srTCM, trTCM and tswTCM schemes to meter and mark the packets of a flow. The results indicate that:

- Bandwidth estimation algorithm of a marker affects the marker performance. For a bursty traffic, marking shows an initial period of instability (where it absorbs bursts and marks more packets green than yellow or red) and then it reaches stability. Transferring the context information during handoff causes marking at the new RER to reach stability quickly, which results in not unusually high green packets. If the context is not transferred, then the marking at the new RER takes a while to stabilize. This is more pronounced for tswTCM where the estimated average bandwidth is transferred during handoff.
- The instability period is sensitive to the bandwidth estimation parameters that directly affect the marking. For example, in case of tswTCM the instability period is proportional to the window size setting.
- Longer marking instability shows higher delay-jitter. Hence, context transfer shows low jitter because of reducing the marking instability period. This is more pronounced for tswTCM. This shows that context transfer is important for real-time application that require strict bound on delay and no or very low jitter.
- The instability that is caused by the handoff not only affects the flow that experiences the handoff but it also affects other flows of the same class at the new RER.
- For srTCM and trTCM, although the effect of the handoff on marking the flow is small compared to tswTCM, transferring the context (token count) ensures that the flow gets the same level of marking as at the old RER.

This study has provided an opportunity to determine suitable marking schemes that maintain similar QoS across RER's while handoff occurs for the MN. It is obvious from the simulation results that if the RER's run tswTCM for marking the packets then the context information must be transferred with the handoff. tswTCM is mostly suitable for VBR streams that may be handled in the Diffserv domains using AF class. On the other hand, if the RER's are using srTCM/trTCM schemes, context transfer does not provide any major benefit and the applications will not have any major loss and/or delay of packets due to handoff. More investigation is required to analyse the relative performance impact on the traffic at the new RER caused by the flows that are handed-over to the new RER. This work is useful in designing connection admission control algorithms at the radio edge routers.

BIBLIOGRAPHY

[1] Blake S, Black D, Carlson M, Davies E, Wang Z, Weiss W. An Architecture for Differentiated Services. RFC 2475 IETF 1998.

[2] Chan E, Hong X. Analytical Model For Assured Forward Differentiated Services over Wireless Links. IEE Communications Proceedings 2001; 148(1):19

[3] Mahadevan I, Sivalingam K. Quality of Service Architectures for Wireless Networks: IntServ and Diffserv Models. IEEE ISPAN 1999; http://www.computer.org/proceedings/ispan/0231/02310420abs.htm (12/8/2001)

[4] Fikouras N, Malki K, Cvetkovic S, Smythe C. Performance of TCP and UDP During Mobile IP Handoffs in Single-Agent Subnetworks. Proc. IEEE WCNC 1999; 1258-1262.

[5] Kempf J. Problem Description: Reasons For Performing Context Transfers Between Nodes in an IP Access Network. IETF draft draft-ietf-seamoby-context-transfer-problem-stat- 03.txt, October 2001.

[6] Syed H, Jaseemuddin M. QoS (Diffserv) Context Transfer. IETF draft draft-hamid-seamoby-ct-qos-context-00.txt, June 2001.

[7] Fang W, Seddigh N. Time Sliding Window Three Color Marker. RFC2859 IETF March 2000.

[8] Brim S, Carpenter B, Faucheur F. Per Hop Behavior Identification Codes RFC2836 IETF May 2000.

[9] Heinanen J, Baker F, Weiss W, Wroclawski J. Assured Forwarding PHB Group. RFC2597 IETF 1999.

[10] Bernet Y, Ford P,Yavatkar R, Baker F, Zhang L, Speer M, Braden R,Davie B, Wroclawski J, Felstaine E. A Framework for Integrated Services Operation over Diffserv Networks RFC2998 IETF 2000.

[11] Bernet Y. White Paper: A Short Overview of QoS Mechanisms and Their Interoperation. Sep 2000. http://www.informit.com/articles (12/8/2001).

[12] Zubairi J, Elshaikh M, Mahmoud O. On Shaping and Handling VBR traffic in a Diffserv domain. ATS, Simulation Series April 2001; 33(3):10-15.

[13] Liu J. Real-Time Systems. Prentice Hall 2000.

[14] Fluckiger F. Understanding Networked Multimedia-Applications and Technology. Prentice Hall 1995.

[15] Varadarajan S. Real-Time Performance Guarantees on Switched Networks. SUNY Stony Brook RPE 1997; http://www.ecsl.cs.sunysb.edu/rpe.html (12/8/2001)

[16] Heinanen J, Guerin R. A Single Rate Three Color Marker. RFC2697 IETF 1999.

[17] Heinanen J, Guerin R. A Two Rate Three Color Marker. RFC 2698 IETF 1999

[18] Perkins C. Mobile Networking Through Mobile IP. http://www.computer.org/internet/v2n1/perkins.htm (12/8/2001)

[19] Perkins C. Mobile Networking in the Internet. Mobile Networks and Applications 1998; 3(319–334).

[20] Tripathi N. Generic Adaptive Handoff Algorithms Using Fuzzy Logic and Neural Networks. Ph.D. dissertation 1997 http://scholar.lib.vt.edu/theses/available/etd-101397-103458/ (12/8/2001)

[21] Duuren J, Kastelein P, Schoute F. Fixed and Mobile Telecommunications. Addison Wesley 1996; (201).

[22] Jiang J, Lai T. Call Admission Control vs. Bandwidth Reservation: Reducing Handoff Call Dropping Rate and Providing Bandwidth Efficiency in Mobile Networks. Proc. IEEE ICPP,2000.

[23] Yang X, Ghaheri-Niri S, Tafazolli R. Evaluation of Soft Handover Algorithms for UMTS. IEEE PIMRC 2000; (772-776)

[24] P. Prasiths. Wireless ATM An Overview. http://www.ittc.ukans.edu/~prasiths/WirelessATM/index.htm (12/8/2001)

[25] NS-2 Documentation. LBNL, UCB. http://www.isi.edu/nsnam/ns/ns-documentation.html (12/8/2001)

[26] Jaseemuddin M, Mahmoud O, Zubairi J. Effect of Context Transfer during Handoff on Flow Marking in a Diffserv Edge Router. Proc. SCI'2001; XII(87-92).

[27] Preda P, Ethridge J, Mandeep, Farhan. Network Simulator Differntiated Services Implementation Open IP. 2000 Nortel Networks.