This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

A Cross-Layer Approach for an Efficient Delivery of TCP/RTP-based Multimedia Applications in Heterogeneous Wireless Networks

Tarik Taleb, *Member, IEEE*, Kenichi Kashibuchi, *Student Member, IEEE*, Alessandro Leonardi, Sergio Palazzo, *Senior Member, IEEE*, Kazuo Hashimoto, *Member, IEEE*, Nei Kato, *Senior Member, IEEE*, Yoshiaki Nemoto, *Senior Member, IEEE*

***Abstract*—Recent trends in telecommunication industry are toward the development of ubiquitous information systems where the provision of a plethora of advanced multimedia services should be possible regardless of time and space limitations. An efficient and seamless delivery of multimedia services over various types of wireless networks is still a challenging task. The underlying difficulty consists in the disparity in the bandwidth availability over each network type. Indeed, the fundamental challenge upon a handoff phenomenon in a heterogeneous wireless network consists in an efficient probing of the bandwidth availability of the new network, followed by a prompt adjustment of the data delivery rate.**

This paper presents a cross layer approach that involves five layers, namely physical, data link, application, network, and transport layers. The three former layers are used to anticipate the handoff occurrence and to locate the new point of attachment to the network. Based on their feedback, the transport layer is used then to probe the resources of the new network using low-priority dummy packets. Being the most widely used protocols for multimedia delivery, this paper addresses multimedia applications based on TCP and RTP protocols. The design of the whole cross layer architecture is discussed and enhancements to the two protocols are proposed. The performance of the enhanced TCP and RTP protocols is evaluated and compared against existing schemes through extensive simulations. The obtained results are encouraging and promising for the delivery of multimedia services in heterogeneous wireless networks.

***Index Terms*—Cross layer, heterogeneous wireless networks, next-generation wireless Internet, RTP, and TCP.**

I. Introduction

ADVANCED multimedia services are gaining momentum within the communities of both industrials and academic researchers. Indeed, along with the on-going advances in wireless technologies and the exponential growth of the mobile users' community, the provision of multimedia applications in wireless networks is likely to open a promising and strong market for service providers and operators.

Manuscript received April 13, 2007; revised September 27, 2007.

T. Taleb, K. Kashibuchi, K. Hashimoto, N. Kato, and Y. Nemoto are with the Graduate School of Information Sciences, Tohoku University, Japan (e-mails: talebtarik@ieee.org; buchiken@it.ecei.tohoku.ac.jp; kh@aiet.ecei.tohoku.ac.jp; kato@it.ecei.tohoku.ac.jp; nemoto@nemoto.ecei.tohoku.ac.jp).

A. Leonardi and S. Palazzo are with the Dipartimento di Ingeneria Informatica e delle Telecomunicazioni, University of Catania, Italy (e-mails: {aleonardi, palazzo}@diit.unict.it).

Among the protocols used for the delivery of multimedia services, the Transport Control Protocol (TCP) and the Real-time Transport Protocol (RTP), accompanied with the Real-time Transport Control Protocol (RTCP) [1], are the most notable ones. Being originally designed for wired networks, both TCP and RTP do not perform well in heterogeneous wireless networks for a number of reasons related to the protocols' syntax and semantics. Their current implementations consequently put many stringent constraints on effective multimedia streaming in wireless systems.

In wireless networks, due to users' mobility, mobile nodes freely, and sometimes frequently, change their points of attachment to the network, an operation henceforth referred to as handoff. Upon a handoff occurrence, the amount of bandwidth available at the new point-of-attachment may be different than that at the old one. This bandwidth disparity can be due to difference in traffic load in both wireless cells.

In general, when a mobile node performs handoff, two scenarios can be envisioned. If the mobile node moves from a higher bandwidth network (*e.g.,* WLAN) to a lower bandwidth network (*e.g.,* General Packet Radio Service – GPRS), and continues transmitting data without any adjustment to its sending rate, the new network will be congested and a potential number of packets will be dropped. The connection throughput will be eventually degraded. On the other hand, if the mobile node enters a higher bandwidth network, no adjustment to the sending rate of the mobile node will lead to a waste of the network bandwidth and ultimately to lower network utilization. Such a performance will obviously result in a poor Quality of Service (QoS) and will ultimately affect the credibility of the whole system.

Ideally, mobile users should be able to anticipate imminent handoff events, should be aware of the next point of attachment to the wireless network, and should get their data download rates promptly adjusted (or should themselves adjust their data sending rates) to meet the available resources of the new access point. As an attempt to realize such an ideal network, this paper proposes a cross layer architecture that involves five layers, namely physical, data link, application, network, and transport layers. The physical and data link layers monitor signal strengths and detect any impending handoff. They then advertise the event to the application layer. In its turn, the application layer refers to personal information on the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

mobile user, history on its mobility patterns, and if possible, information on the topology of the wireless network to locate the next access point (AP).

Knowing the next point of attachment, two connections are simultaneously set between the mobile node and its correspondent sender: one through the old access point and another through the new one. Assuming that the coverage areas of both APs overlap with each other, the sender continues transmitting actual data through the old connection. Meanwhile, it sends a number of low priority dummy segments through the new connection. These dummy packets are used to probe the bandwidth availability of the new network, similar in spirit to the idea presented in [2].

Application of the concept to both TCP and RTP is considered. Related issues are discussed and possible solutions are presented. Extensive simulations are conducted to evaluate the performance of the proposed modifications to TCP and RTP. The results demonstrate that the proposed concept is promising for the guarantee of QoS in wireless networks as it assures a fast handoff management, increases the system throughput, and maintains lower packet drop rates.

The remainder of this paper is structured as follows. Section II highlights the relevance of this work to the state-of-art in the context of cross layer design for wireless mobile networks. The key design philosophy and distinct features that were incorporated in the proposed cross layer architecture are described in Section III. Section IV portrays the simulation environment and reports the simulation results. Following this, the paper concludes in Section V with a summary recapping the main advantages and achievements of the proposed architecture.

II. Related Work

The traditional Open System Interconnection (OSI) layered architecture, as originally specified, did not specifically provide any interaction among its layers. A cross layer design aims at enabling such an interaction for the sake of better performance and prompt adaptation of the stack functionality in the presence of changing network conditions. Based on the involved layer, the emergence of several cross layer interactions has been highlighted in the recent literature [3], [4].

The proposed cross layer architectures and frameworks can be categorized based on the type of communication used to exchange information among layers. In [5], an architecture based on the use of the Internet Control Message Protocol (ICMP) messages is proposed. The architecture involves the physical/MAC layers, network layer, and application layer. In [6], an Interlayer Signaling Pipe (ISP) is used to propagate cross layer information through packet headers. A drawback of this technique consists in the fact that lower layers are required to read the headers of higher layers, an operation that ultimately slows down the execution of the lower layers. As a solution to this issue, the Cross Layer Signaling Shortcuts (CLASS) architecture allows direct communication between the layers [7]. Other cross layer architectures consider the addition of new component to the protocol stack. MobileMan is a notable example [8]. In [9], a cross layer manager is designed to handle events and state variables sent by the protocol layers. The state variables are used to appropriately coordinate among the link adaptation, security, QoS, and users mobility.

While the above mentioned systems are relatively generic in their design and consequently add significant complexity to the original design of the protocol stack, a number of other cross layer approaches simply uses information from different layers to optimize the protocol behavior in some circumstances [10], [11]. The RTP protocol is itself an example. Indeed, it integrates functions of both the session and presentation (and in some cases application) layers into a single protocol. By maintaining a large context related to a given multimedia session, it is possible to handle several aspects of real-time communication such as synchronization and adaptive application framing. Another example that falls in this category is the Freeze-TCP [12]. It involves the physical and link data layers as it uses their feedback to detect handoffs or to predict a temporary disconnection. If a handoff occurs, a Freeze-TCP mobile host advertises a zero window size to force the sender into frozen mode. This operation aims to avoid drops of in-flight packets. The sender restarts transmitting data only when the mobile host enters a new point of attachment. For a detailed discussion on other cross layer mechanisms related to TCP and RTP, the interested reader is referred to the related work sections of [13], [14], respectively.

III. Overview of the Proposed Cross Layer Architecture

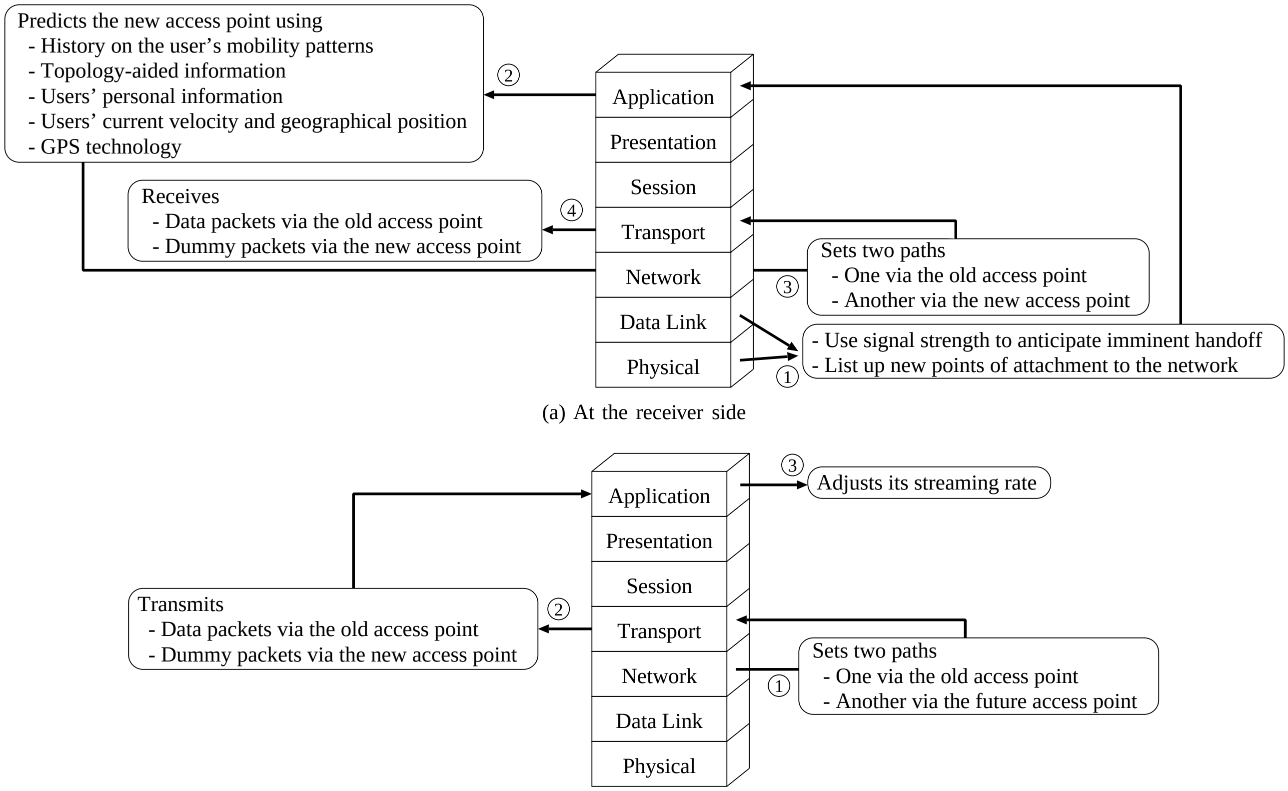
This section gives a detailed description of the proposed cross layer architecture. It first outlines the core ideas behind the architecture and its requirements. It next presents the major components of the architecture. And finally, it portrays the proposed enhancements to the working of TCP and RTP to guarantee an efficient and seamless delivery of multimedia services over wireless networks.

A. Requirements

First, it should be emphasized that this paper targets wireless networks where cells overlap with each other. The considered network is assumed to be end-to-end QoS enabled. In fact, the proposed scheme requires that all network elements along the connection path support some priority disciplines. Currently, most networks are best effort and most routers in the Internet do not apply any priority policy. However, in the near future, through the use of the Differentiated Service Model (DiffServ) [15], routers will be able to support multiple service classes. Having said that, it should be stressed out that the proposed cross layer design does not specifically require a DiffServ architecture. It simply requires a priority queuing discipline with two priority levels.

To enable mobile hosts to simultaneously access two or more different access points, mobile nodes are equipped with multiple wireless interfaces. While having multiple wireless interfaces on the same mobile device is impractical, the ongoing advances in the wireless technology have demonstrated that a single physical WLAN interface can be used to simultaneously access multiple WLANs [16]. To allow a mobile

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.



(a) At the receiver side

(b) At the sender side

Fig. 1. Envisioned cross layer design.

node to simultaneously register multiple Care-of-Addresses (CoAs), Mobile IP (MIP) simultaneous binding option [17], [18] is used. On the other hand, to keep senders always informed of these CoA registrations directly from the mobile nodes, the route optimization option [19] is used. It should be noted that the new CoA of the mobile node in the new cell should be different from the CoA used in the old cell. Finally, at the sender side, applications should be able to adjust their streaming rates by appropriately changing the quality of multimedia contents. As for the type of communication to be used in exchanging information among layers, a wide library of communication types exists as discussed in the previous section. The proposed cross layer design can consider implementation of the most adequate one taking into account the required computational load and the communication delay that may result from interactions among the layers.

B. Cross Layer Design

A cross layer optimization can be implemented either at end-devices or intermediate nodes in the network, such as access points or routers. Given the relative easiness and feasibility of the former, this paper focuses on implementing changes on mobile hosts. Fig. 1 depicts the major procedures of the proposed architecture. At the receiver side (mobile host), the physical layer of a mobile host instantly measures the radio strength or link quality. When the mobile node moves into the overlapping area of two or more wireless cells, and different signals are consequently detected by the physical and data link layers, a warning message notifying an imminent handoff event, along with a list of the new possible APs, are sent to the application layer. In case of multiple access points, the application layer refers to a set of tools to sort out the access point to which the mobile node is most likely going to be connected. Indeed the application layer may use history on the user's mobility pattern to predict the new access point. Referring to a spatial conceptual map, along with the user's personal information, its current position, and its velocity heading, the application layer can make an accurate prediction of the most probable future access point [20]. Prior knowledge on the topology of the wireless network [21] and the type of the application [22] can further increase the accuracy of the prediction. Once the next access point is determined, the sender is informed of the new base station via a new CoA binding update message from the MIP protocol. The network layer sets then two paths; one via the old access point and another via the new point of attachment. The transport layer keeps receiving data packets via the old access point and simultaneously starts receiving dummy packets via the new access point. The dummy packets are used to probe the bandwidth availability of the new network as will be explained later.

The cross layer design at the sender side is relatively simpler than that at the receiver side. First, upon receiving information on the new access point, the network layers of both the sender and receiver terminals set a new path via the new access point and at the same time maintain the old one. The transport layer of the sender terminal keeps transmitting data packets via the old path and uses dummy packets to probe the bandwidth of the new network. Once the new bandwidth is estimated, the application layer of the sender terminal should accordingly adjust its data streaming rate.

C. Enhancements to TCP

To cope with issues related to handoff management in heterogeneous wireless networks, a large body of bandwidth probing techniques has been proposed to make an estimate of the available bandwidth in the new network [23]. Most of these pioneering techniques require accurate measurements of propagation delay. Under heavy traffic load, an accurate estimate of propagation delay is usually not possible to obtain; a fact that ultimately leads to erroneous estimates in the available bandwidth. The key concept behind our proposed enhancement to TCP consists in the use of dummy packets for an efficient probing of the bandwidth availability in the new network [13]. Indeed, when the next access point (AP) is decided by the proposed cross layer architecture, as previously explained, and before reaching the middle point of the overlapping area (where the handoff usually takes place), the mobile node keeps on receiving data from the sender using the old connection through the old AP. Meanwhile, the sender sends "$rwnd$" dummy segments to the mobile node through the new AP, where $rwnd$ is the receiver window size that limits the maximum value of congestion window. The value of $rwnd$ indicates the rate at which the sender transmits dummy segments to the mobile node. The algorithm of the proposed scheme is based on the concept of using these dummy segments to probe the availability of network resources without carrying any new information to the sender. This concept was first proposed in [24] and has been used since then in several researches in the recent literature. Notable examples are TCP-Peach [25], the InterPlanetary Transport Protocol (TP-Planet) [26], and the Analytical Rate Control (ARC) [27].

Dummy segments are generated by the sender as a copy of the last transmitted data packet. They are treated as low priority segments. Accordingly they do not affect the delivery of the actual data traffic. Indeed, when a router on the connection path is congested, IP packets carrying dummy segments are discarded first. Overhead of these dummy segments in terms of bandwidth consumption should therefore not be an issue.

To distinguish dummy segments from actual data packets, dummy segments are marked using one or more of the six unused bits in their TCP headers. A simple modification of the TCP implementation is accordingly required in the end-terminals. Upon reception of a dummy segment, the mobile node can thus recognize it. In response to each dummy segment, the mobile node transmits a dummy acknowledgment (ACK) to the sender. Dummy ACK packets indicate the availability of network resources in the new cell. In response to each dummy ACK, the sender transmits, in turn, an actual data packet to the mobile node. ACKs for dummy segments are used to provide an efficient probing of the bandwidth availability in the new network. As a result, senders can adjust their sending rates to the most appropriate value within one round trip time (RTT). They either increase their transmission rates to make full utilization of the new network resources or decrease their transmission rates to avoid overloading the new network with bursty traffic.

D. Enhancements to RTP

While TCP dominates most of today's Internet traffic, RTP represents the core streaming protocol for real-time multimedia services. It does not add any delays to the transmitted data as packet retransmissions are not considered. However, it may congest the network as it does not employ any congestion control. To cope with such an issue, RTP receivers notify senders with statistics on their perceived QoS, such as cumulative of packet losses, RTP timestamp, number of packets received, and jitter. This information is periodically reported via signaling messages called Receiver Reports (RRs). Based on these RR messages, the RTP protocol assesses the network condition and accordingly controls its streaming rate. This forms the basic framework of the RTCP protocol.

One important issue that is highly missing in the design of RTCP pertains to the transmission frequency of RR messages. Indeed, the minimum time interval for transmitting two consecutive RR messages is recommended to be set to five seconds [1]. This aims to meet the 5 % fraction of the session bandwidth reserved for RTCP packets. In heterogeneous wireless networks, this policy is inefficient and may largely affect the entire system performance. As a matter of fact, in case a mobile receiver performs handoff to a lower bandwidth network without immediate transmission of a RR packet, by the time the correspondent sender gets notified of the new network conditions and starts accordingly adjusting its streaming rate, the new network may have already been overly congested and a significant number of packets may have been dropped. In case of handoff to a higher bandwidth network, no immediate adjustment of the RTP streaming rate may lead to a waste of the new network resources.

As a remedy to this issue, a RTP mobile receiver uses the above mentioned cross layer design to anticipate an imminent handoff event. It then explicitly notifies its correspondent sender with the event via newly defined RTCP packets. These packets are referred to as RTCP Handoff Notification (HN) packets throughout the remainder of this paper. While they have the same header as ordinary RTCP packets, RTCP HN packets can be distinguished by having their packet type field set to an unused value. It should be reminded that RTCP RR packets are transmitted on a periodic basis, whereas RTCP HN packets are sent only upon detecting degradation in the link quality, in other words when a handoff event is about to occur.

Upon receiving a RTCP HN packet, the RTP sender probes the available bandwidth in the new network using dummy RTP packets similar to the aforementioned enhancements proposed for TCP. These dummy RTP packets are sent through the access point of the new network at the maximum streaming rate of the multimedia data for a predefined period of time (*i.e.,* less than one second). After receiving dummy packets for the predefined period of time, the RTP receiver sends a reception quality feedback to the sender in a RTCP signaling packet. This type of packets is referred to as RTCP Handoff Report

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

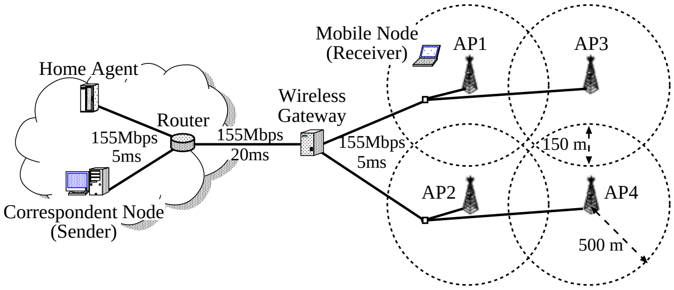


Fig. 2. Abstract configuration of the simulation network topology.

(HR) packet throughout this paper. The format of RTCP HR packets conforms to that of RTCP RR. It includes information on the reception quality measured during the reception of dummy packets. Once the optimal streaming rate of the new wireless network is known to the sender, the receiver starts receiving actual data packets via the new access point and quits its old connection with the sender by issuing a RTCP BYE packet.

IV. Performance Evaluation

Having described the details of the proposed cross layer architecture, we now direct our focus to evaluating its performance. The performance evaluation relies on computer simulation using the Network Simulator [28]. We first describe the simulation set up justifying the choices made along the way and next discuss the simulation results.

A. Simulation Setup

In the conducted simulations, particular attention is paid to the design of an accurate and realistic environment. Fig. 2 depicts the abstract configuration of the considered network. The wireless part of the network consists of a number of adjacent wireless cells. The coverage radius of each wireless cell is set to 500 meters. The distance between two adjacent access points is fixed to 850 meters. This makes the longest distance across the overlapping area between two adjacent cells equal to 150 meters. In the simulations, the actual distance across the overlapping area is varied from one meter to 150 meters. These parameters are chosen with no specific purpose in mind and do not change any of the fundamental observations about the simulation results.

The wireless domain is connected to the wired network through a single wireless gateway. The choice of a single wireless gateway serving all the APs represents a general and simple case. Indeed, considering a topology where APs are served by two different wireless gateways will simply increase the connection RTT and shall have no influence on the overall performance evaluation. To avoid packet drops due to bottlenecks at the wired network, all wired links are given similar capacities equal to 155 Mbps (*e.g.,* OC3). They have predetermined propagation delays as indicated in Fig. 2. As for the wireless links, a number of test scenarios were created by setting their capacities to different values. Their delays are minimal and are set to 0.01 ms. All links are presumed to be error-free throughout this paper. This assumption is made so as to avoid any possible confusion between throughput degradation due to packet drops and that due to wireless channel errors.

To best understand the behavior of the proposed enhancements to TCP and RTP, we consider a single handoff between two adjacent access points in the considered topology. Having prior knowledge on the position (coordinates) of each access point, a user refers to its velocity heading and its position to predict the next access point to which it will be connecting. Delay incurred by the computational load of this operation is included in the entire handoff delay. In the simulations, a mobile node receives a new network address from the new access point and issues a CoA binding update message as soon as it enters the cell overlapping area of two adjacent cells. It accordingly sets up two paths for communication through the old and new access points, respectively. As for the actual handoff, it is performed when the radio strength of mobile nodes or the wireless link quality goes down below a predefined threshold. In the simulations, handoffs are performed when a mobile node reaches the middle line of the overlapping area, which represents the most common case.

In the proposed cross layer architecture, all network elements along the connection path need to support some priority disciplines. This operation is enabled using the Weighted Random Early Detection (WRED) scheme [29]. Unless otherwise specified, the queue length of all network elements is set to 50 packets. The size of a data packet is set to 1000 bytes in TCP and 500 bytes in RTP. The maximum streaming rate in RTP-related simulations is fixed at 10 Mbps. All results are an average of multiple simulation runs.

B. Analysis: Effects of handoff on the rate control

1) TCP-based Multimedia Services: For the sake of a better understanding of the research work presented in this paper, we analyze the effects of handoff on the congestion window in case of TCP-based multimedia services. Different TCP variants are used as comparison terms. These variants include the well-known TCP NewReno [30], Freeze-TCP [12], TCP Westwood-NR which is the NewReno based version of TCP Westwood [31]. While our proposed enhancements can be implemented on any TCP variant, we consider enhancements to TCP NewReno. The reason behind the choice of the TCP NewReno among other TCP implementations underlies beneath the fact that TCP NewReno achieves faster recovery from multiple losses within the same window. It has also the potential of improving TCP's performance in the case of bursty losses. For an insightful comparison among the TCP variants, focus is on the performance of the schemes during the handoff period. The definition of the handoff period comes later.

In this analysis, we consider two scenarios where handoff occurs. In the first scenario, called H-L, the mobile node moves from high capacity (6 Mbps) cell to low capacity (1 Mbps) cell while in the second scenario, called L-H, the mobile node moves from low capacity (6 Mbps) cell to high capacity (11 Mbps) cell. We focus on the behavior of the congestion window, $cwnd$, in the four TCP versions examined.

Throughout this paper, the *handoff period* is defined as the time period during which a mobile node travels over the cell overlapped area. If we consider two time instants t_0 and t_2, with $\mathrm{t}_0 < \mathrm{t}_2$, we suppose that handoff starts at time t_0 and ends at time t_2. Handoff occurs at time t^* when the middle of the overlapping area is reached, *i.e.*, ($\mathrm{t}^* = (\mathrm{t}_0 + \mathrm{t}_2)/2$).

a) Cross-layer approach: First of all we explain the proposed cross-layer behavior during the handoff period in detail.

- $\mathrm{t}_0 \le t < \mathrm{t}_1$
  At time t_0 the mobile node enters the cell overlapping area. During the time interval $[\mathrm{t}_0, \mathrm{t}_1[$, with ($\mathrm{t}_1 < \mathrm{t}^*$), the sender sends $rwnd$ dummy segments to the receiver through the new access point, where $rwnd$ is the maximum value for the congestion window which is specified by the receiver. The receiver transmits a dummy ACK in response to each dummy packet received.
- $\mathrm{t}_0 + \mathrm{RTT} \le t < \mathrm{t}_1 + \mathrm{RTT}$
  After one round trip time (RTT) elapsed from the first dummy segment transmitted, the ACKs related to the dummy segments reach the sender. In response to each dummy ACK, the sender transmits an actual data packet to the receiver. Note that we assume ($\mathrm{t}_1 + \mathrm{RTT} < \mathrm{t}^*$), an assumption that will be later confirmed by simulation results.
- $t = \mathrm{t}_1 + \mathrm{RTT}$
  At this time, due to the mechanisms shown above, the congestion window of the connection established via the new access point results ($cwnd(t) := n_{\mathrm{ACK}}$), where n_{ACK} is the number of dummy ACKs received so far. This means that in both scenarios (H-L and L-H) the sender adjusts its sending rate to the most appropriate value within one round trip time.
- $t > \mathrm{t}_1 + \mathrm{RTT}$
  The classical TCP NewReno algorithms are used.

In Fig. 3, we show the variations of the congestion window size of the proposed cross-layer approach during the time period from 10 seconds before to 10 seconds after the handoff phase. Let AP1 and AP2 be the old and the new access points, respectively. We reproduce both H-L and L-H scenarios in Figs. 3(a) and 3(b), respectively. These plots have been obtained assuming $\mathrm{t}_0 = -4.5$ s and $\mathrm{t}_2 = 4.5$ s. Consequently, the handoff happens at $\mathrm{t}^* = 0$ s. It is observed that, in both scenarios, starting from ($t > \mathrm{t}_1 + \mathrm{RTT}$), the $cwnd$, whose value is greater than the slow start threshold, increases linearly according to the bandwidth estimated in the new cell. Moreover, we observe that when a loss is detected, the original recovery algorithms of NewReno are used.

b) Freeze-TCP: Here we describe the behavior of Freeze-TCP during the handoff.

- $t = \mathrm{t}^*$
  At this time instant, the receiver (knowing that a handoff is occurring) advertises a zero window size to the sender. This operation is performed to compel the sender into a frozen mode in order to prevent the congestion window from dropping to one.
- $t = \mathrm{t}_3$
  Let t_3 ($\mathrm{t}^* < \mathrm{t}_3 < \mathrm{t}_2$) denote the time instant when the Freeze-TCP starts retransmitting data to the new access point. The value of congestion window will not be changed from the last value, thus ($cwnd(t) := cwnd(\mathrm{t}^*)$).
- $t > \mathrm{t}_3$
  If the bandwidth of the new network, b_{new}, is greater than that of the old one, b_{old}, such as in the L-H scenario, the congestion window increases by ($\frac{1}{cwnd}$) for each ACK received, *i.e.*, ($cwnd := cwnd + \frac{1}{cwnd}$) until a congestion occurs. On the other hand, if $b_{new} < b_{old}$ (H-L scenario), the new network gets suddenly overloaded with a large number of data packets. This congests in turn the transmission queue at the bottleneck link's router and eventually results in the discard of a large number of packets. As a result, TCP decreases its $cwnd$ to one, almost immediately.

The underestimation or overestimation of the bandwidth availability in the new network in both H-L and L-H scenarios have been examined in Figs. 4(a) and 4(b), respectively. Here we show the congestion window size in packets during the handoff period. Comparing the results against those obtained in case of the cross-layer approach, we notice that Freeze-TCP exhibits a slow increase of the $cwnd$ in the L-H scenario after a loss, and a drastic reduction of the $cwnd$ in the H-L scenario as the new network gets overloaded and a number of packet drops occur.

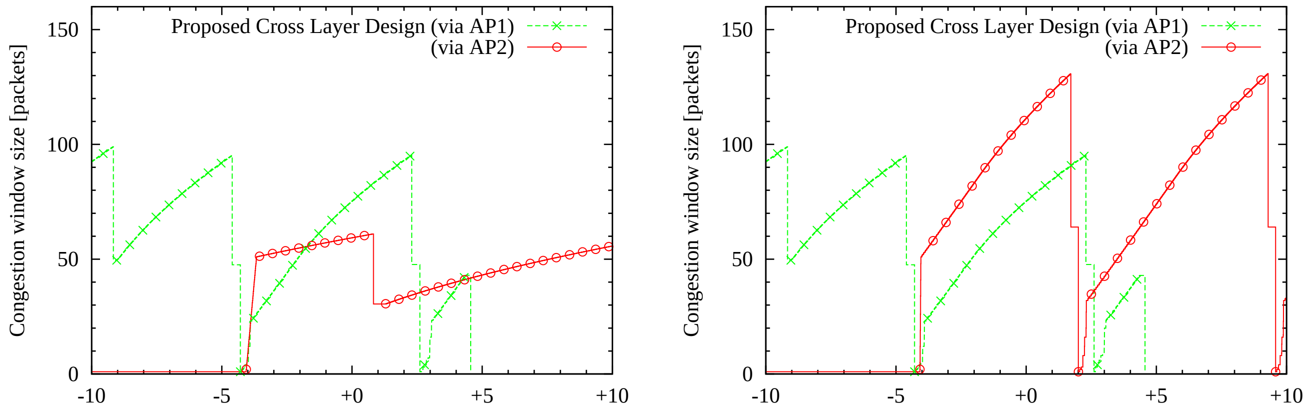
c) TCP NewReno: Here we analyze the TCP NewReno behavior in detail.

- $t = \mathrm{t}^*$
  At this time the handoff is performed. The old connection is closed and a new connection is opened via the new access point. Consequently, after a timeout, TCP NewReno sets ($cwnd := 1$).
- $t > \mathrm{t}^*$
  The sender enters the slow start phase. Upon reaching the slow start threshold ($ssthresh$), the sender switches to the Congestion Avoidance phase [30].

In Figs. 4(a) and 4(b) we observe that starting from $t \ge 0$, TCP NewReno exhibits poor performance when compared to the cross-layer approach. This is intuitively due to the slow delivery of data packets when the new connection is established.

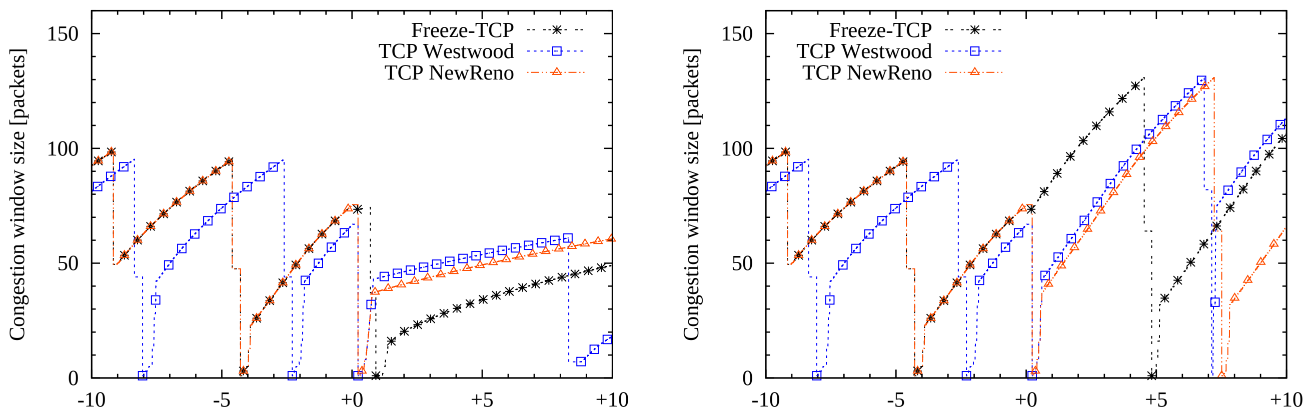
d) TCP Westwood: Finally, we examine the behavior of TCP Westwood in the two scenarios examined.

- $t = \mathrm{t}^*$
  At this time the handoff causes a timeout expiration. As a consequence, TCP Westwood sets ($cwnd := 1$) and ($ssthresh :=$ *BWE*), where *BWE* is the connection BandWidth Estimate which is defined as the rate at which data is delivered to the TCP receiver [31]. The estimate is based on the rate at which ACKs have been received and on their payload. Note that these values are obtained before the handoff happens, ($t < \mathrm{t}^*$).
- $t > \mathrm{t}^*$
  A slow start phase starts until the value of $ssthresh$ is reached. It is then followed by the Congestion Avoidance phase similar to NewReno [30]. Note that, because $ssthresh$ is set to a value calculated before the handoff



(a) H-L scenario (6 Mbps → 1 Mbps) (b) L-H scenario (6 Mbps → 11 Mbps)

Fig. 3. Variation of the congestion window size during handoff in case of the proposed cross layer design.



(a) H-L scenario (6 Mbps → 1 Mbps) (b) L-H scenario (6 Mbps → 11 Mbps)

Fig. 4. Variation of the congestion window size during handoff in case of TCP NewReno, TCP Westwood-NR, and Freeze-TCP schemes.

occurrence, TCP Westwood behaves aggressively in the H-L scenario. It therefore can result in a large number of packet drops. On the other hand, in case of the L-H scenario, TCP Westwood underestimates the new bandwidth at least at the very beginning (establishment time) of the new connection.

From Figs. 4(a) and 4(b) we observe that TCP Westwood behaves in the same way as TCP NewReno. It thus exhibits poor performance compared to that of the proposed cross-layer approach.

Compared to the other TCP versions, the proposed cross-layer approach demonstrates the best performance as it exhibits a very low packet loss rate in the H-L scenario and makes an efficient use of the new bandwidth in the L-H scenario.

2) RTP-based Multimedia Services: In RTP-related simulations, the third operation of the cross layer design at the sender side (streaming rate adjustment in Fig. 1) is performed using the Loss-Delay based Adjustment (LDA+) algorithm [32]. The reason behind this choice underlies beneath the fact that LDA+ achieves relatively good TCP-friendliness even when RTCP generates feedback messages at low frequencies. In this context, it should be noted that whilst frequent transmissions of control messages is beneficial for quick adaptation to sudden changes in network conditions, it incurs overhead in terms of bandwidth consumption. Another reason behind the choice of LDA+ consists in the fact that we assumed all wireless links to be error free in the conducted simulations. Indeed, in case of low bit error rate (BER) environments, the use of LDA+ as a rate control method suffices. However, in high BER environments, the LDA+ scheme can be substituted by more adequate schemes such as ARC [27], TCP Friendly Rate Control (TFRC) [33], or the Rate Control Scheme (RCS) [34]. As for the underlying protocol, RTP can be implemented on any network type. It can indeed work on TCP/IP, ATM, or frame relay. In the conducted simulations, User Datagram Protocol (UDP) is used as the transport protocol and IP as the network protocol.

In this analysis, we compare the transition of the streaming (sending) rate in the proposed approach with that of standard

RTP. Figs. 5(a) and 5(b) show the transitions of the streaming rate in H-L and L-H scenarios, respectively. In these plots, the handoff begins at $\mathrm{t}^* = 0\,\mathrm{s}$.

a) Cross-layer approach: Here we show the streaming rate control during the handoff.

- $t = \mathrm{t}^*$
  At this time, the handoff operation starts. The receiver transmits a RTCP HN packet to the RTP sender. In response, the sender transmits dummy RTP packets through the new access point at the maximum streaming rate of the data.
- $t = \mathrm{t}^* + \mathrm{RTT}$
  After a RTT elapsed since the transmission of the RTCP HN packet, the RTP receiver begins to receive the dummy RTP packets. Note that the sending rate of dummy packets are not reflected in Fig. 5 since they do not convey the actual data.
- $t = \mathrm{t}_4$
  Let t_4 ($\mathrm{t}_4 > \mathrm{t}^*$) denote the time instant when the RTP receiver transmits a RTCP HR packet to the sender. After receiving dummy packets for a predefined period of time, the RTP receiver sends a RTCP HR packet through the new access point.
- $t > \mathrm{t}_4$
  Upon receiving the RTCP HR packet, the RTP sender calculates the appropriate streaming rate from the information included in it, as in the case of receiving a RTCP RR packet. It then transmits normal RTP packets through the new access point at the computed rate.

Figs. 5(a) and 5(b) show that, in both scenarios, the proposed approach sends data at appropriate streaming rate according to the bandwidth estimated in the new cell after handoff.

b) Standard RTP: Here we analyze the standard RTP behavior.

- $t = \mathrm{t}^*$
  The RTP sender starts transmitting packets through the new access point, but keeps the sending rate which was adapted to the previous cell. Therefore, in L-H scenario, its rate is below the available bandwidth in the new cell. On the other hand, in H-L scenario, the network in the new cell falls into congestion and a lot of packet drops are caused.
- $t = \mathrm{t}_5$
  Let t_5 ($\mathrm{t}_5 > \mathrm{t}^*$) denote the time instant when the RTP sender receives the first RTCP RR packet since the handoff event. RTP sender calculates the streaming rate from the information included in the RTCP RR packet, and adjusts the sending rate to the computed rate.
- $t > \mathrm{t}_5$
  Even after receiving the RR packet, the new streaming rate can be inaccurate. This is due to the fact that the first RR packet includes the old information that are not valid for the new cell. Furthermore, if the streaming rate falls below the available bandwidth, it takes a long period of time to achieve the appropriate rate to the new cell because LDA+ increases it incrementally. For instance, t_5 equals to $2.039\,\mathrm{s}$ in Figs. 5(a) and 5(b).

From these figures, in both scenarios, the proposed cross-layer approach shows better performance than standard RTP as it appropriately adjusts the streaming rate immediately after handoff.

C. Simulation Results

1) TCP-based Multimedia Services: To evaluate the performance of the cross layer architecture in delivering TCP-based multimedia services, two quantifying parameters are used: average throughput and loss rate. Throughput indicates the number of bytes received by a mobile node during the handoff period. The loss rate is the ratio of the dropped packets to the aggregate sent packets during the handoff period. As dummy packets do not carry any new information, they are not considered in the computation of neither the loss rate nor the throughput.

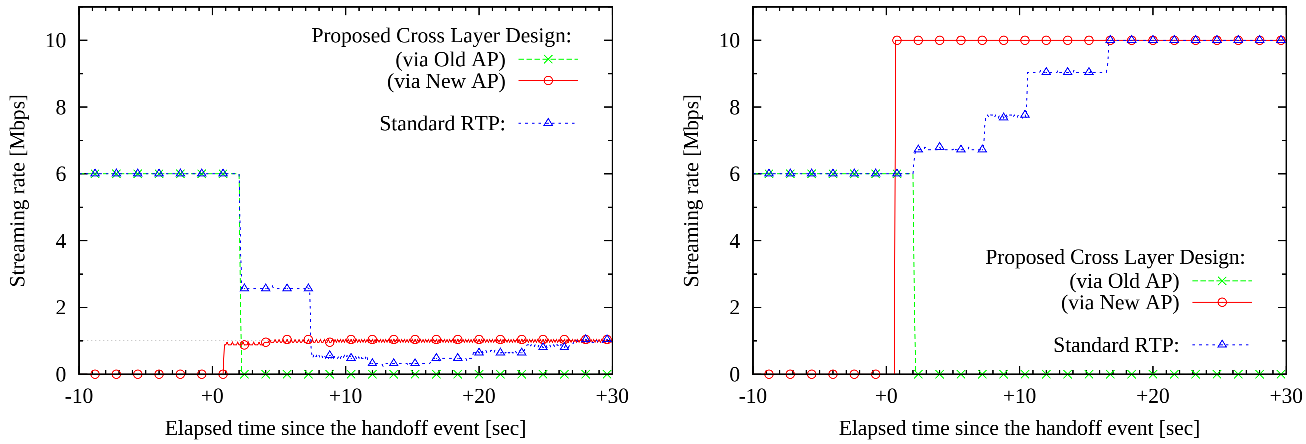
First, to investigate the robustness of the proposed cross layer design in anticipating handoff events and promptly adjusting the transmission rate to the available bandwidth in the new wireless cell, we envision a scenario where the bandwidth of the old cell is set to $6\,\mathrm{Mbps}$ and the bandwidth of the new cell is varied from $1\,\mathrm{Mbps}$ to $11\,\mathrm{Mbps}$. The moving speed of the mobile node is set to $50\,\mathrm{km/h}$.

Fig. 6(a) compares the throughput of the proposed cross layer design to that of the other three TCP variants for different disparities in the available bandwidth. The figure demonstrates that the proposed cross layer design achieves the highest throughput compared to TCP Westwood-NR, Freeze-TCP, and TCP NewReno. It shows also an abrupt increase in the throughput achieved by the proposed cross layer design. When the available bandwidth in the new network is lower (the range of negative values on the x-axis), the four simulated schemes exhibit smaller throughputs. This is simply due to the fact that the bandwidth in the new network becomes less available. On the other hand, when the new cell has a higher bandwidth (the range of positive values on the x-axis), the proposed cross layer design gains up to more than $200\,\%$ over the three other schemes. This significant gain is mainly due to the fact that dummy segments inform the sender of the extra-bandwidth, becoming available in the new cell, within a single round trip time and accordingly stimulate it to increase its sending rate.

Fig. 6(b) illustrates the performance of the four schemes in terms of packet drops. The packet loss rate is plotted as a function of the difference between the available bandwidths in the new and old networks. The results show that the proposed cross layer architecture and Freeze-TCP achieve the lowest packet drop rate. The proposed scheme outperforms further the Freeze-TCP scheme and achieves almost zero drops regardless of the available bandwidth in the new cell. The main reason beneath this performance is in the intrinsic characteristic of the proposed scheme. Indeed, the proposed cross layer design uses dummy segments to estimate the optimum rate at which the sender should send data. Accordingly, the sender avoids overloading the network with data packets that would ultimately be dropped otherwise.

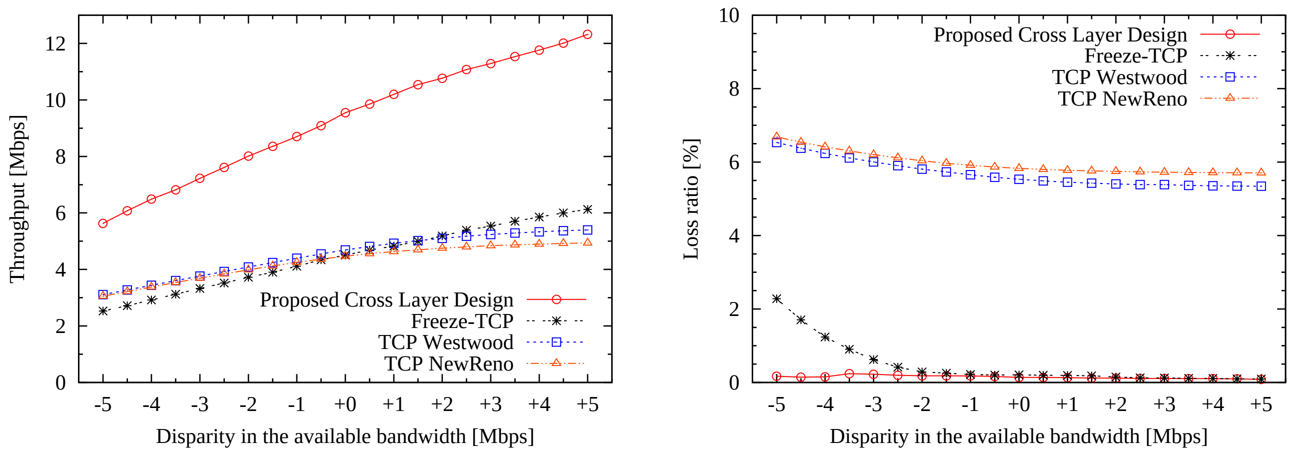
On the other hand, while TCP Westwood-NR and TCP NewReno exhibit a throughput relatively equal to that of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.



(a) H-L scenario (6 Mbps → 1 Mbps) (b) L-H scenario (6 Mbps → 11 Mbps)

Fig. 5. Transition of the streaming rate during handoff in case of the proposed cross layer design and standard RTP.



(a) Throughput (b) Packet drop rate

Fig. 6. Transmission efficiency for different disparities in the available bandwidth (Mobile node speed = 50 km/h).

proposed cross layer when the bandwidth in the new cell becomes less available ($\ll 3$ Mbps), their achieved throughput comes at the price of significant packet drops. This remark is illustrated in Fig. 6(b). Indeed, the results of the figure indicate that the two schemes experience the highest packet drop rate. This poor performance is mainly due to their bursty nature. In fact, both schemes keep transmitting data at window sizes that can not be accommodated by the new network. This leads to congestion and ultimately higher drops. In summary, since the proposed cross layer design uses dummy segments to probe the available bandwidth in the new cell, it achieves the highest throughput and maintains the lowest drop rate compared to the other three schemes.

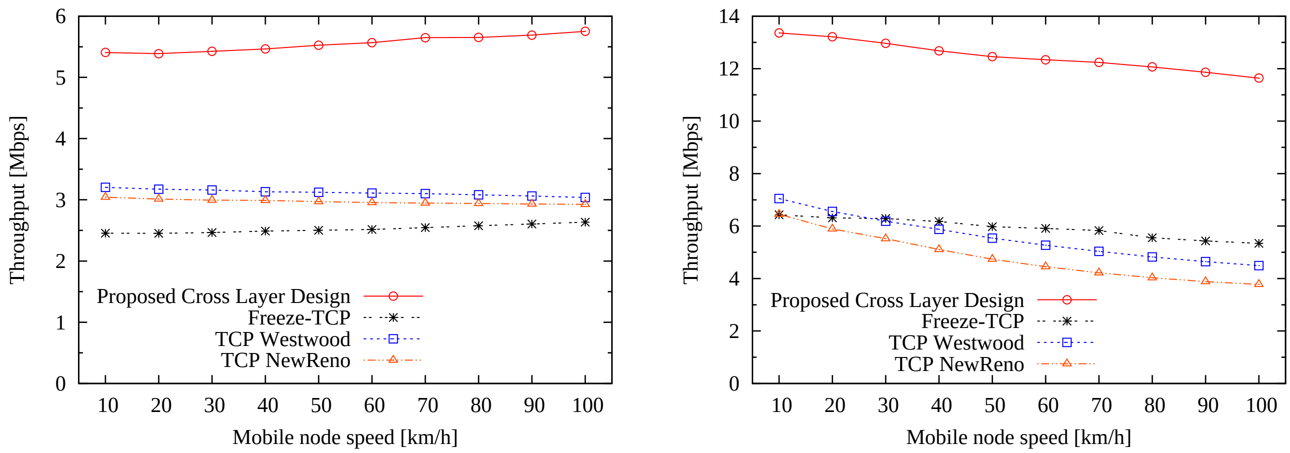
In light of the narrow surface of the cell overlapping area, the length of the handoff period becomes shorter as the mobile node speed increases. This decrease in the length of the handoff period may influence the working of the proposed cross layer design as the time required by the cross layer architecture to manage handoff and to probe for the bandwidth availability becomes shorter. To investigate such an impact, we vary the speed of the mobile node from 10 km/h to 100 km/h. We envisage two scenarios: H-L and L-H scenarios. Throughputs of the proposed cross layer design and the other three TCP variants for different mobile node speeds are graphed in Fig. 7. The figure confirms the impact of the mobile node speed on the throughputs of the four methods as their throughputs decrease with an increase in the mobile node speed. Nevertheless, it shows that the throughput of the proposed cross layer design remains the highest in both scenarios and that is for all considered speeds.

In the remainder of this section, we envision a scenario whereby a TCP connection competes for bandwidth with N TCP connections in the new cell after handoff. A roaming TCP receiver performs handoff from a 6 Mbps cell to an 11 Mbps cell. The other users remain in the same cell. As a fairness index, we use the following metric:

$$F_T = \frac{r_{\mathrm{hTCP}}}{r_{\mathrm{eTCP}}} \tag{1}$$

where r_{hTCP} and r_{eTCP} denote the throughput achieved by the roaming user via the new access point and the average

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.



(a) H-L scenario (6 Mbps $\longrightarrow$ 1 Mbps) (b) L-H scenario (6 Mbps $\longrightarrow$ 11 Mbps)

Fig. 7. Throughput variation for different mobile node speeds.

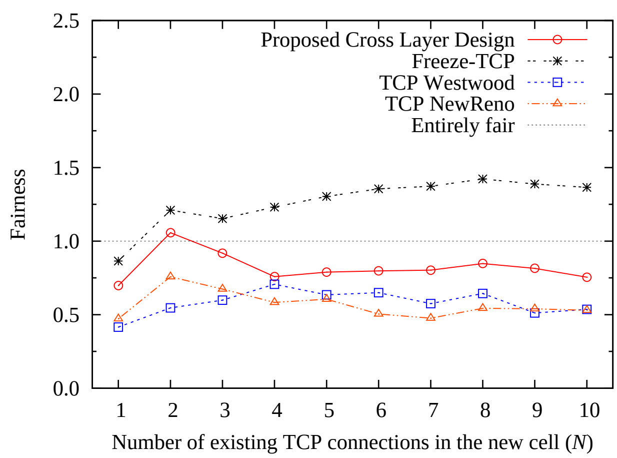


Fig. 8. Fairness to existing TCP connections after handoff.

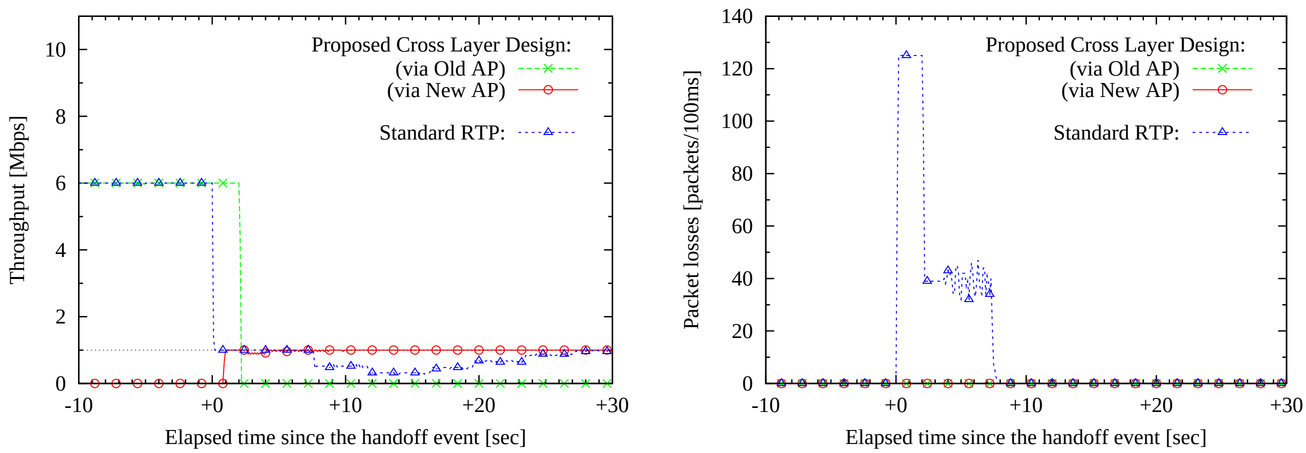
throughput of the other N TCP connections, respectively. Each throughput is measured for five seconds after the handoff occurrence time. $F_T = 1$ means that the newly-coming user and the already-existing users are evenly sharing the bandwidth. $F_T > 1$ indicates that the newly-coming user conquers a portion of the cell bandwidth higher than the old users. Fig. 8 shows that the proposed cross layer design achieves better fairness than the other TCP variants. This is due to the fact that the proposed scheme estimates the available bandwidth while taking into account the network dynamics. In contrast, Freeze-TCP affects the other traffic in the new cell as it does not adjust the window size after handoff.

2) RTP-based Multimedia Services: To highlight the efficiency of the proposed enhancements to RTP when implemented over the proposed cross layer design, we compare its performance against that of standard RTP. In case of standard RTP, we ignore both the delay that is due to the handoff operation and the in-flight packet drops that may happen during the handoff operation. The rational behind this setting is to investigate the system performance in terms of packet drops due to only delay in adjusting the streaming rate and not delay in the management of the handoff. In the performance evaluation, two metrics are used: throughput achieved by the receiver and packet losses occurred along the communication path. Here, packet losses are computed every 100 ms. They do not include dummy packet drops.

To investigate the interactions of RTP with the proposed cross layer architecture, we consider a RTP mobile receiver moving between a higher bandwidth cell (6 Mbps) and a lower bandwidth cell (1 Mbps). Fig. 9 plots the transition of the experienced packet losses and the actual throughput achieved by the mobile node when the node performs handoff to a cell with less bandwidth. The figure shows that standard RTP achieves a slightly higher throughput than the proposed cross layer design. This performance comes however at the price of significant packet drops as indicated in Fig. 9(b). This performance is attributable to delay in the adjustment of the streaming rate. Indeed, until reception of a RR packet message, the standard RTP sender keeps transmitting data at high rates that can not be accommodated by the resources of the new cell as shown in Fig. 5(a). The figure shows that even after receiving a RR packet message (2.039 seconds after the handoff event), the new streaming rate is not accurately computed and largely exceeds the available bandwidth at the new cell. This is due to the fact that the computation of the new streaming rate is based on old information that are not valid for the new cell. In case of the proposed cross layer design, when a handoff is about to occur, the sender gets notified of the event via a RTCP HN message. In response, it starts transmitting dummy packets to the receiver via the new AP. The receiver uses these dummy packets to make an accurate estimate of the bandwidth of the new network and reports it to the sender via the RTCP HR message. The sender promptly adjusts its streaming rate to the bandwidth of the new cell. This helps to avoid overloading the network with data packets and to accordingly elude packet drops as indicated in Fig. 9(b).

To ensure that the proposed cross layer design makes efficient use of the network resources when more bandwidth becomes available in the network, we consider a scenario where a mobile node roams from a lower bandwidth cell to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.



(a) Throughput (b) Packet losses

Fig. 9. Transmission efficiency in H-L scenario (6 Mbps → 1 Mbps).

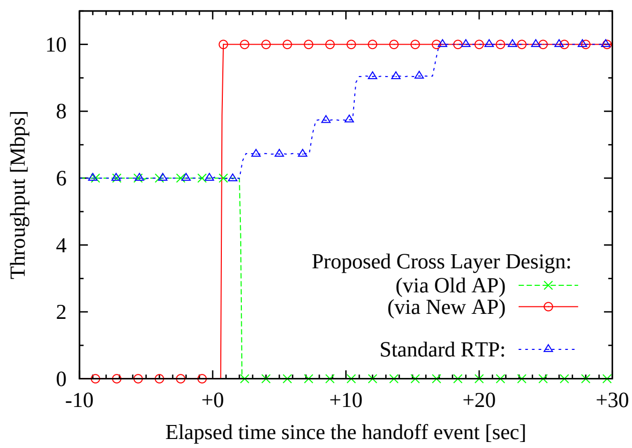


Fig. 10. Transmission efficiency in L-H scenario (6 Mbps → 11 Mbps).

a higher bandwidth cell (6 Mbps → 11 Mbps). Fig. 10 plots the transition of the mobile node's throughput 10 seconds before and 30 seconds after the handoff occurrence time. In this simulation, as no packet drops were observed neither in the proposed cross layer design nor in the standard RTP protocol, we do not graph packet losses. Fig. 10 shows that the proposed cross layer scheme achieves higher throughput compared to standard RTP immediately after the handoff event. This demonstrates the robustness of the proposed cross layer scheme to adapt to changes in the wireless network environment. It indicates also the accuracy in probing bandwidth availability using dummy packets. On the other hand, the performance of standard RTP remains limited as the sender keeps streaming data at rates far below the available bandwidth in the new network and that is for a fairly long period of time after the handoff occurrence. Moreover, the bandwidth estimation of standard RTP is inaccurate as it is based on old information provided by RTCP RR messages. The inaccuracy of the bandwidth estimation is manifested in the stair-step shape of the streaming rate graph of standard RTP (Fig. 10). In this example, the sender needed nearly 20 seconds after the handoff event till it could be able to stream data at the available bandwidth of the new network.

Finally, we evaluate the performance of the proposed cross layer design in a scenario whereby a RTP connection shares bandwidth with N TCP connections in the new cell after a handoff. In this scenario, a RTP receiver performs handoff from 6 Mbps cell to 11 Mbps cell. As a friendliness index, we use the following metric:

$$F_R = \frac{r_{\mathrm{hRTP}}}{r_{\mathrm{eTCP}}} \qquad (2)$$

where r_{hRTP} and r_{eTCP} denote the throughput of the RTP connection via the new access point after handoff and the average throughput of the N TCP connections, respectively. Each throughput is measured for five seconds after the handoff occurrence time. Fig. 11 plots the friendliness index of both schemes as a function of the total number of competing TCP connections N. The figure indicates that the proposed approach achieves TCP-friendliness faster than standard RTP. In standard RTP, RTP traffic causes a large number of packet losses in all the connections in the new cell after handoff and accordingly unfairly degrades the throughput of existing TCP traffic. In contrast, the proposed approach exhibits better friendliness as it adjusts the streaming rate and sets it to moderate rates after handoff.

D. Discussion

In the proposed cross layer approach, the bandwidth probing is based on dummy packets. Admittedly, reception of dummy segments and transmission of dummy ACKs by mobile nodes result in additional energy consumption. The proposed cross layer scheme may be thus seen as costly in terms of reducing the battery life of mobile nodes. However, the performance gains achieved by the proposed cross layer architecture in terms of both throughput and reduced packet drops are worthwhile and can be used to advocate for this additional cost. Indeed, the high throughput and low packet loss rates of the proposed cross layer design lead to significant reduction in the overall transmission time of a given data file. This intuitively

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

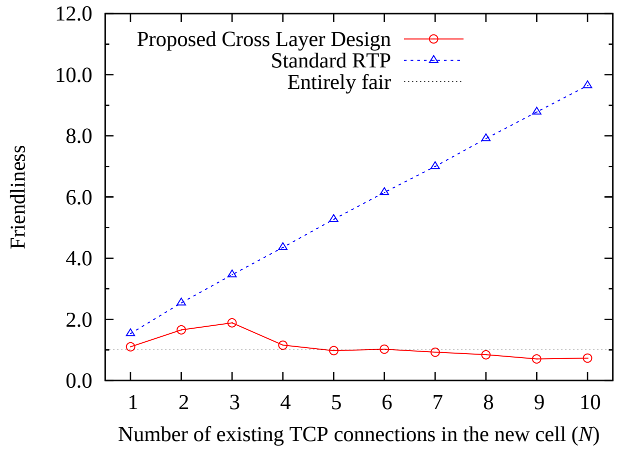


Fig. 11. Friendliness with existing TCP connections after handoff.

reduces the overall usage time of the mobile node battery and ultimately saves its energy. Moreover, apart from the rare case of mobile nodes flip-flopping over a cell overlapping area, the additional energy consumption due to dummy segments remains minimal.

To illustrate the idea at hand, we consider the following simple mathematical analysis. Let N_r and N_t denote the number of received dummy packets and the number of transmitted dummy ACKs by a mobile node. Let t_r and t_t denote the time required for receiving a dummy packet and the time required for transmitting a dummy ACK packet. Denoting by I_r and I_t the amount of the electric current required for receiving a single packet and the amount of the electric current required for transmitting a single ACK packet, the battery consumption of a mobile node due to dummy packets and ACKs can be expressed as follows:

$$B_{\text{TCP}} = N_r \cdot I_r \cdot t_r + N_t \cdot I_t \cdot t_t \tag{3}$$

As a mobile node sends back an ACK for each received dummy packet, ($N_t = N_r = N$). The above equation can be thus simplified as follows:

$$B_{\text{TCP}} = N(I_r \cdot t_r + I_t \cdot t_t) \tag{4}$$

Using the specifications of the wireless LAN CardBus adapter developed by Cisco [35], when 802.11b is in use, I_t and I_r can be set to a maximum of $539\,$mA and $327\,$mA, respectively. Additionally, the times required for receiving a data packet with a length of 1000 bytes and for transmitting an ACK packet of 32 bytes are equal to $727\,\mu s$ and $23\,\mu s$, respectively ($t_r = 727\,[\mu s]$, $t_t = 23\,[\mu s]$). Using these values, the consumed battery in case of receiving N dummy packets and transmitting N dummy ACKs is simply

$$B_{\text{TCP}} < 0.0067 \cdot N[\text{mA-min}] \tag{5}$$

Even in case of 100 dummy packets, the consumed battery is less than $0.67\,$mA-min. For a mobile phone with a battery lifetime equal to $730\,$mA-h ($43,800\,$mA-min), the consumed battery represents a negligible amount. All in all, along with on-going advances in technologies related to batteries, the use of dummy packets to probe for the bandwidth availability shall not be an issue at all for mobile users.

In a similar way, when a mobile node receives N_r dummy RTP packets, the battery consumption due to dummy packets can be calculated as[1]:

$$B_{\text{RTP}} = N_r \cdot I_r \cdot t_r \tag{6}$$

If a mobile node receives RTP packets at $10\,$Mbps for $0.5\,$sec, the consumed battery is less than $150\,$mA-min.

V. Conclusion

In this paper, we proposed a cross layer design for an efficient delivery of multimedia services in heterogeneous wireless networks. The designed cross layer architecture involves five layers. The proposed interactions between the layers are simple and practical. The key idea behind the proposed cross layer architecture is to anticipate imminent handoffs, to notify senders with these events, and to stimulate them to probe for the resources of the new wireless network. Dummy packets are used for this purpose. Two types of multimedia applications are considered, namely TCP-based and RTP-based applications. For each type, adequate enhancements are proposed.

The performance of the proposed cross layer architecture is evaluated for both TCP-based and RTP-based applications using computer simulations. Simulation results elucidate the outstanding performance of the proposed cross layer architecture in achieving high throughputs while reducing packet drops. The results also demonstrate the effectiveness of dummy packets in making accurate estimation of the available bandwidth. The resiliency of the proposed cross layer design to changing network conditions is also verified. The results are promising for streaming multimedia services in heterogeneous wireless networks where disparity in the available bandwidth is still a major issue to solve.

From the simulation results, we believe that the proposed cross layer design represents an important contribution to the field of multimedia delivery in heterogeneous wireless networks. It is all the authors' hope that the findings in this paper would stimulate further research work in the area.

References

[1] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications," RFC 3550, Jul. 2003.
[2] G. Bianchi, A. Capone, and C. Petrioli, "Throughput Analysis of End-to-End Measurement-Based Admission Control in IP," in *Proc. IEEE INFOCOM*, vol. 3, Tel Aviv, Israel, Mar. 2000, pp. 1537–1545.
[3] V. Raisinghani and S. Iyer, "Cross-Layer Design Optimizations in Wireless Protocol Stacks," *Comput. Commun.*, vol. 27, no. 8, pp. 720–724, May 2004.
[4] V. Kawadia and P. R. Kumar, "A Cautionary Perspective on Cross-Layer Design," *IEEE Wireless Commun. Mag.*, vol. 12, no. 1, pp. 3–11, Feb. 2005.
[5] P. Sudame and B. R. Badrinath, "On Providing Support for Protocol Adaptation in Mobile Wireless Networks," *Mobile Networks and Appli-cat.*, vol. 6, no. 1, pp. 43–55, Jan./Feb. 2001.
[6] G. Wu, U. Bai, and J. L. Ogielski, "Interactions between TCP and RLP in Wireless Internet," in *Proc. IEEE GLOBECOM*, vol. 1B, Rio De Janeiro, Brazil, Dec. 1999, pp. 661–666.

[1] Here, we ignore two packets (RTCP HN and HR).

[7] Q. Wang and M. A. Abu-Rgheff, "Cross-Layer Signaling for Next-Generation Wireless Systems," in *Proc. IEEE WCNC*, vol. 2, New Orleans, LA, Mar. 2003, pp. 1084–1089.
[8] M. Conti, "Cross-Layering in Mobile Ad Hoc Network Design," *IEEE Computer*, vol. 37, no. 2, pp. 48–51, Feb. 2004.
[9] G. Carneiro, J. Ruela, and M. Ricardo, "Cross-Layer Design in 4G Wireless Terminals," *IEEE Wireless Commun. Mag.*, vol. 11, no. 2, pp. 7–13, Apr. 2004.
[10] K. Chen, S. Samarth, and K. Nahrstedt, "Cross-Layer Design for Data Accessibility in Mobile Ad Hoc Networks," *IEEE Personal Commun. Mag.*, vol. 21, pp. 49–76, Apr. 2002.
[11] U. C. Kozat, I. Koutsopoulos, and L. Tassiulas, "A Framework for Cross-Layer Design of Energy-Efficient Communication with QoS Provisioning in Multi-hop Wireless Networks," in *Proc. IEEE INFOCOM*, vol. 2, Mar. 2004, pp. 1446–1456.
[12] T. Goff, J. Moronski, D. S. Phatak, and V. Gupta, "Freeze-TCP: A True End-to-End TCP Enhancement Mechanism for Mobile Environments," in *Proc. IEEE INFOCOM*, vol. 3, Tel Aviv, Israel, Mar. 2000, pp. 1537–1545.
[13] T. Taleb, K. Kashibuchi, N. Kato, and Y. Nemoto, "A Dummy Segment Based Bandwidth Probing Technique to Enhance the Performance of TCP over Heterogeneous Networks," in *Proc. IEEE WCNC*, vol. 4, New Orleans, LA, Mar. 2005, pp. 2400–2405.
[14] K. Kashibuchi, T. Taleb, A. Jamalipour, Y. Nemoto, and N. Kato, "A New Smooth Handoff Scheme for Mobile Multimedia Streaming using RTP Dummy Packets and RTCP Explicit Handoff Notification," in *Proc. IEEE WCNC*, vol. 4, Las Vegas, NV, Apr. 2006, pp. 2162–2167.
[15] A. T. Campbell, "QoS-Aware Middleware for Mobile Multimedia Communications," in *Multimedia Tools and Applications*. Kluwer Academic Publishers, Jul. 1998, vol. 7, no. 1–2, pp. 67–82.
[16] R. Chandra, Paramvir Bahl, and Pradeep Bahl, "MultiNet: Connecting to Multiple IEEE 802.11 Networks Using a Single Wireless Card," in *Proc. IEEE INFOCOM*, vol. 2, Hong Kong, China, Mar. 2004, pp. 882–893.
[17] W. Fritsche and F. Heissenhuber, "Mobile IPv6: Mobility Support for Next Generation Internet," IPv6 Forum, white paper, 2000.
[18] C. Perkins, "IP Mobility Support for IPv4," RFC 3344, Aug. 2002.
[19] R. Vadali, J. Li, Y. Wu, and G. Cao, "Agent-Based Route Optimization for Mobile IP," in *Proc. IEEE VTC fall*, vol. 4, Atlantic City, NJ, Oct. 2001, pp. 2731–2735.
[20] N. Samaan and A. Karmouch, "A Mobility Prediction Architeture based on Contextual Knowledge and Spatial Conceptual Maps," *IEEE Trans. Mobile Comput.*, vol. 4, no. 6, pp. 537–551, Nov./Dec. 2005.
[21] C. C. Tseng, L. H. Yen, H. H. Chang, and K. C. Hsu, "Topology-Aided Cross-Layer Fast Handoff Designs for IEEE 802.11/Mobile IP Environments," *IEEE Commun. Mag.*, vol. 43, no. 12, pp. 156–163, Dec. 2005.
[22] S. Khan, Y. Peng, and E. Steinbach, "Application-Driven Cross-Layer Optimization for Video Streaming over Wireless Networks," *IEEE Commun. Mag.*, vol. 44, no. 1, pp. 122–130, Jan. 2006.
[23] M. Jain and C. Dovrolis, "End-to-End Available Bandwidth: Measurement Methodology, Dynamics, and Relation with TCP Throughput," in *Proc. ACM SIGCOMM*, Pittsburgh, PA, Aug. 2002.
[24] V. N. Padmanabhan and R. Katz, "TCP Fast Start: A Technique for Speeding up Web Transfers," in *Proc. IEEE GLOBECOM, Internet Mini-Conf.*, Sydney, Australia, Nov. 1998, pp. 41–46.
[25] I. F. Akyildiz, G. Morabito, and S. Palazzo, "TCP-Peach: A New Congestion Control Scheme for Satellite IP Networks," *IEEE/ACM Trans. Netw.*, vol. 9, no. 3, pp. 307–321, Jun. 2001.
[26] O. B. Akan, J. Fang, and I. F. Akyildiz, "TP-Planet: A Reliable Transport Protocol for Interplanetary Internet," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 2, pp. 348–361, Feb. 2004.
[27] O. B. Akan and I. F. Akyildiz, "ARC: The Analyticacl Rate Control Scheme for Real-time Traffic in Wireless Networks," *IEEE/ACM Trans. Netw.*, vol. 12, no. 4, pp. 634–644, Aug. 2004.
[28] The Network Simulator - ns-2. [Online]. Available: http://www.isi.edu/nsnam/ns/
[29] Cisco IOS Quality of Service Solutions Configuration Guide. [Online]. Available: http://www.cisco.com/univercd/cc/td/doc/product/software/
[30] S. Floyd and T. Henderson, "The NewReno Modification to TCP's Fast Recovery Algorithm," RFC 2582, Apr. 1999.
[31] S. Mascolo, C. Casetti, M. Gerla, M. Y. Sanadidi, and R. Wang, "TCP Westwood: Bandwidth Estimation for Enhanced Transport over Wireless Links," in *Proc. ACM SIGMOBILE*, Rome, Italy, Jul. 2001, pp. 287–297.
[32] D. Sisalem and A. Wolisz, "LDA+ TCP-Friendly Adaptation: A Measurement and Comparison Study," in *Proc. of 10^{th} Int. Workshop on NOSSDAV*, Chapel Hill, NC, Jun. 2000, pp. 1619–1622.
[33] S. Floyd, M. Handley, J. Padhye, and J. Widmer, "Equation-based Congestion Control for Unicast Applications," in *Proc. ACM SIGCOMM*, Stockholm, Sweden, Aug. 2000, pp. 43–56.
[34] J. Tang, G. Morabito, I. F. Akyildiz, and M. Johnson, "RCS: A Rate Control Scheme for Real-time Traffic in Networks with High Bandwidth-Delay Products and High Bit Error Rates," in *Proc. IEEE INFOCOM*, vol. 1, Apr. 2001, pp. 114–122.
[35] Cisco Aironet 802.11a/b/g Wireless CardBus Adapter. [Online]. Available: http://www.cisco.com/go/wireless/

Tarik Taleb (S'04–M'05) is currently working as assistant professor with the Graduate School of Information Sciences, Tohoku University, Japan. From Oct. 2005 till Mar. 2006, he was working as research fellow with the Intelligent Cosmos Research Institute, Sendai, Japan. He received his B.E. degree in Information Engineering with distinction, M.S. and Ph.D. degrees in Computer Sciences from GSIS, Tohoku Univ., in 2001, 2003, and 2005, respectively. His research interests lie in the field of wireless networking, satellite and space communications, congestion control protocols, mobility & handoff management, on-demand media transmission, and network security. Dr. Taleb is on the editorial board of the IEEE Wireless Communications. He also serves as Secretary of the Satellite and Space Communications Technical Committee of the IEEE Communications Society (ComSoc). He has been on the technical program committee of different IEEE conferences, including Globecom, ICC, and WCNC, and chaired some of their sessions. He is a recipient of the 2007 Funai Foundation Award (Mar. 2007), the 2006 IEEE Computer Society Japan Chapter Young Author award (Dec. 2006), the Niwa Yasujirou Memorial award (Feb. 2005) and the Young Researcher's Encouragement award from the Japan chapter of the IEEE Vehicular Technology Society (VTS) (Oct. 2003). Dr. Taleb is an IEEE member.

Kenichi Kashibuchi (S'05) received his B.E. degree in Information Engineering, M.S. in Computer Sciences from Tohoku University, Sendai, Japan in 2005 and 2007, respectively. He is currently working towards his Ph.D. degree at the Graduate School of Information Sciences, Tohoku University. He has acted as reviewer for several IEEE conferences. His research interests fall in the areas of wireless networking and multimedia delivery. Mr. Kashibuchi is an IEEE student member.

Alessandro Leonardi received the Laurea Degree in Electronics Engineering and the Ph.D. in Computer and Telecommunications Engineering from the University of Catania, Italy, in 2002 and 2007, respectively. Currently he is with the Dipartimento di Ingegneria Informatica e delle Telecomunicazioni (DIIT) of the University of Catania. His current research interests include Mobile and Wireless Networks and Performance Analysis of Communication Networks. From September 2005 to February 2006 he has been a visiting scholar in the Rice Networks Group (Rice University, Houston TX) under the guidance of Prof. Edward Knightly.

Sergio Palazzo (M'92–SM'99) was born in Catania, Italy, on December 12, 1954. He received his degree in electrical engineering from the University of Catania in 1977. Until 1981, he was at ITALTEL, Milano, where he was involved in the design of operating systems for electronic exchanges. He then joined CREI, which is the center of the Politecnico di Milano for research on computer networks. Since 1987 he has been at the University of Catania, where is now a Full Professor of Telecommunications Networks. In 1994, he spent the summer at the International Computer Science Institute (ICSI), Berkeley, as a Senior Visitor. He is a recipient of the 2003 Visiting Erskine Fellowship by the University of Canterbury, Christchurch, New Zealand. Since 1992, he has been serving on the Technical Program Committee of INFOCOM, the IEEE Conference on Computer Communications. He has been the General Chair of the ACM MobiHoc 2006 Conference and currently is a member of the MobiHoc Steering Committee. In the recent past, he has been the Program Co-Chair of the 2005 International Tyrrhenian Workshop on Digital Communications, focused on "Distributed Cooperative Laboratories: Networking, Instrumentation, and Measurements", the General Vice Chair of the ACM MobiCom 2001 Conference, and the General Chair of the 2001 International Tyrrhenian Workshop on Digital Communications, focused on "Evolutionary Trends of the Internet". He currently serves the Editorial Boards of the journals IEEE/ACM Transactions on Networking, IEEE Transactions on Mobile Computing, Ad Hoc Networks. In the recent past, he also was an Editor of IEEE Wireless Communications Magazine (formerly IEEE Personal Communications Magazine), Computer Networks, and Wireless Communications and Mobile Computing. He was a Guest Editor of Special Issues in the IEEE Journal of Selected Areas in Communications ("Intelligent Techniques in High-Speed Networks"), in the IEEE Personal Communications Magazine ("Adapting to Network and Client Variability in Wireless Networks"), in the Computer Networks journal ("Broadband Satellite Systems: a Networking Perspective"), in the EURASIP Journal on Wireless Communications and Networking ("Ad Hoc Networks: Cross-Layer Issues"). He also was the recipient of the 2002 Best Editor Award for the Computer Networks journal. His current research interests include mobile systems, wireless and satellite IP networks, intelligent techniques in network control, multimedia traffic modelling, and protocols for the next generation of the Internet.

Kazuo Hashimoto (M'06) received his M.S. degree in Computer Sciences from Brown University, USA, and his Ph.D. degree in Information Sciences from Tohoku University, Sendai, Japan, in 1986 and 2001, respectively. He is currently working as short-term professor with the Graduate School of Information Sciences, Tohoku University. From 2001 till 2005, he was appointed as president & CEO of KDDI Labs USA. During this period, he was involved in R&D activities in collaboration with major research institutions and universities in the US. Under his supervision and direction, KDDI Labs USA played a key role in the mobile and Internet market and achieved a variety of collaborations between US entities and KDDI. His current research interests lie in the field of network security, network management, data mining, and multimedia information retrieval.

Nei Kato (M'03–A'04–SM'05) received his M.S. and Ph.D. degrees from the Graduate School of Information Sciences, Tohoku University, in 1988 and 1991, respectively. He has been working for Tohoku University since then and is currently a full professor at the Graduate School of Information Sciences. He has been engaged in research on computer networking, wireless mobile communications, image processing, and neural networks. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan and a senior member of IEEE. He is a recipient of the 2005 Distinguished Contributions to Satellite Communications award from IEEE Communications Society and the 2007 Funai Foundation Award. He has served on a large number of technical program and organizing committees of international conferences. From 2006, he is serving as a technical editor of IEEE Wireless Communications.

Yoshiaki Nemoto (S'72–M'73–SM'05) received his B.E., M.E., and Ph.D. degrees from Tohoku University in 1968, 1970, and 1973, respectively. He is a full professor with the Graduate School of Information Sciences, and served as director of the Information Synergy Center, Tohoku University. He has been engaged in research work on microwave networks, communication systems, computer network systems, image processing, and handwritten character recognition. He is a recipient of the 2005 Distinguished Contributions to Satellite Communications award from IEEE ComSoc society and a co-recipient of the 1982 Microwave Prize from the IEEE MTT society. He is a senior member of IEEE, a member of IEICE, and a fellow member of the Information Processing Society of Japan.