Volume 3, Issue 12, December 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Improving handoff Latency Performance: Review

Swapnil Pawar*, Seema Ladhe
Dept Computer Engineering
MGMCET, Mumbai University, India

Abstract— Handoff is the mechanism which provides the connection to the wireless network while a movable terminal is moving across the boundaries of coverage of two wireless points of connection. The complexity of the handoff decision process has led to the examination of a number of traditional and pattern recognition handoff decision algorithms for wireless networks. Wireless local area network (WLAN) is a popular choice in Communication Based Train System (CBTC) due to the on hand WLAN equipments. WLANs were not originally designed for high speed environments with frequent handoff. Due to this frequent handoff the result which leads in communication interrupts and long latency. However, handoffs in WLAN may result in communication interruption and long latency in WLANs based CBTC networks. In this piece of writing we use scheme for CBTC networks using Stream Control Transmission Protocol (SCTP) and IEEE 802.11 advanced standard for vehicular communication to provide high availability and low latency in CBTC networks. We formulate the handoff decision problem as a Markov decision process (MDP) with the objectives of minimizing the handoff latency and maximizing the SCTP throughput.

Keywords— MDP, SCTP, SA, AP , IAPP , WAVE.

I. INTRODUCTION

Mobility is the most important feature of a wireless cellular communication system. Usually, continuous service is achieved by supporting handoff (or handover) from one cell to another. Handoff is the process of changing the channel (frequency, time slot, spreading code, or combination of them) associated with the current connection while a information transfer is in progress. It is often initiated either by crossing a cell boundary or by deterioration in quality of the signal in the current channel. Handoff is divided into two broad categories—hard and soft handoffs. They are also characterized by "break before make" and "make before break." In hard handoffs, current resources are released before new resources are used; in soft handoffs, both existing and new resources are used during the handoff process. Poorly designed handoff schemes tend to generate very heavy signaling traffic and, thereby, a dramatic decrease in quality of service (QoS). The reason why handoffs are critical in wireless cellular communication systems is that neighboring cells are always using a disjoint subset of frequency bands, so negotiations must take place between the station adapter (SA), the current serving access points (AP), and the next potential AP. Other related issues, such as decision making and priority strategies during overloading, might influence the overall performance.

A. Types of handoffs

Handoffs are broadly classified into two categories—hard and soft handoffs. Usually, the hard handoff can be further divided into two different types—intra- and inter-cell handoffs. The soft handoff can also be divided into two different types—multi-way soft handoffs and softer handoff.

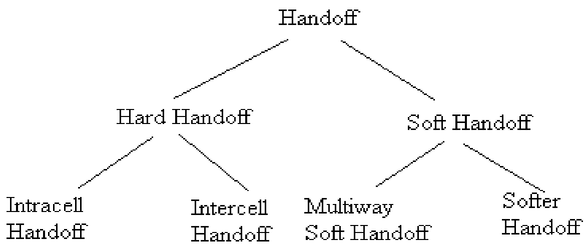

Fig 1: Types of handoffs

1. Hard Handoff

Hard handoff is one in which the channel in the source cell is released and only then the channel in the target cell is engaged. Thus the connection to the source is broken before or 'as' the connection to the target is made—for this reason such handovers are also known as break-before-make. Hard handovers are intended to be instantaneous in order to minimize the disruption to the request and reply for the particular information. It requires the least processing by the network providing service. When the station adapter is between access points, then the station adapter can switch with any of the access point, so the access points bounce the link with the station adapter back and forth. This is called ping-ponging.

1.1 Intra-cell handoff

Intra-cell handoff the source and target is one and the same cell and only the used channel is changed during the handoff. The purpose of intra-cell handoff is to change a channel, which may be interfered, or fading with a new clearer or less fading channel.

1.2 Inter-cell Handoff

In inter-cell handoff the source and the target are different cell (even if they are on the same cell site). The purpose of the inter-cell handoff is to maintain the communication link as the station adapter is moving out of the area of the one access point and entering the area of the another access point. Finally, hard handoff is permitted between members of different soft zones, but not between members of the same soft zones

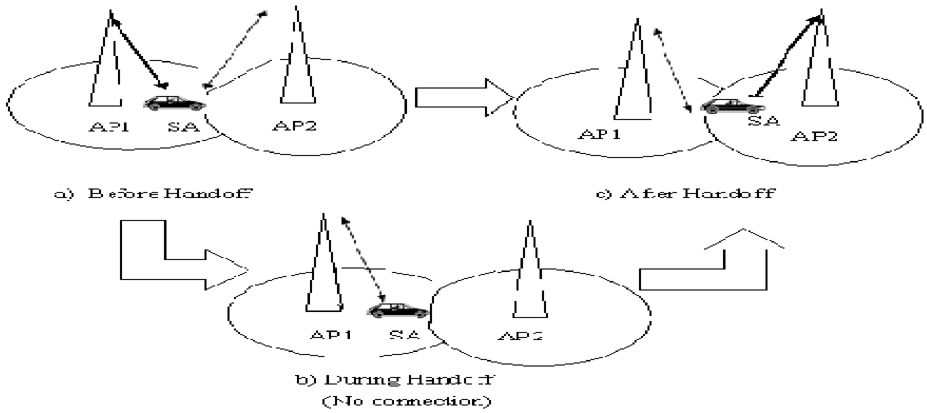

Fig 2: Mechanism for Hard Handoff

2. Soft Handoff

Soft Handoff is one in which the channel in the source cell is retained and used for a while in parallel with the channel in the target cell. In this case the connection to the target is established before the connection to the source is broken, hence this handover is called make-before-break. The interval, during which the two connections are used in parallel, may be brief or substantial. For this reason the soft handover is perceived by network engineers as a state of the call, rather than a brief event. Soft handovers may involve using connections to more than two cells: connections to three, four or more cells can be maintained by one phone at the same time. When a call is in a state of soft handover, the signal of the best of all used channels can be used for the call at a given moment or all the signals can be combined to produce a clearer copy of the signal.

2.1 Multi way Soft Handoffs

Which involves using connections to more than two cells is a multi ways handoff.

2.2 Softer Handoff

When a call is in a state of soft handoff the signal of the best of all used channels can be utilized for the call at a given moment or all the signals can be combined to produce a clear signal, this type is called softer handoff.

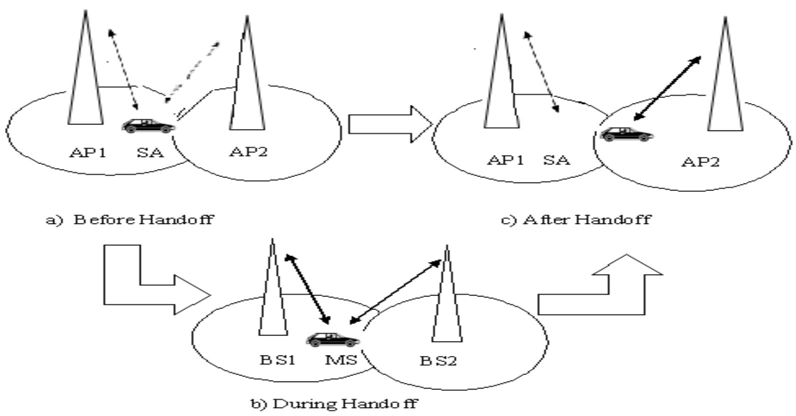

Fig 3: Mechanism for Soft Handoff

III. BACKGROUND

Wireless cellular communication system is used for safe operation of rails, vehicles. Mobility is the most important feature of a wireless cellular communication system. It is a modern successor of traditional cellular communication signaling system. In general, data between station adapter and access point are transferred bidirectional by wireless communication networks, such as global system for mobile communications and wireless local area network. Whereas in commercial wireless networks, less service availability and long latency mean less revenues or/and poor quality of services (QoSs); in wireless networks, less service availability could cause collision or even catastrophic loss of life or assets. Therefore, it is important to ensure the wireless communications are available when they are needed, and the latency is minimized in networks. Most existing WLAN-based networks are using traditional IEEE 802.11 technologies, such as 802.11a/b/g. However, IEEE 802.11a/b/g WLANs were not originally designed for high speed environments. Particularly, when a station adapter moves away from the coverage of a WLAN access point (AP) and enters the coverage of another AP along the station adapter, a handoff procedure occurs, and this handoff process may result in communication interruption and long latency. The handoff procedure can be divided into four steps, namely probing (also referred to as scanning), channel switching, authentication and association. This whole procedure may take up to several hundred milliseconds.

It is necessary to look at the handoff management at multiple layers of the protocol stack, not just at the data link layer as considered in the past. Indeed, the handoff management problem can be solved at transport layer. For example, stream control transmission protocol (SCTP), a new IETF-standardized transport layer protocol in addition to transmission control protocol (TCP) and User Datagram Protocol (UDP), can be used to solve the handoff management problem. The multi-homing, multi-stream and partial reliable data transmission features of SCTP are especially attractive for applications that have stringent performance and high reliability requirements. Compared to other handoff management approaches, transport layer schemes have many advantages, such as improved throughput and latency performance. Moreover, no third party other than the endpoints participates in handoff process, and no modification or addition of network components is required, which makes transport layer approaches attractive in WLAN-based networks. However, due to the high mobility environment, as well as the high availability and low handoff latency requirements, handoff decision policy issues are very important in designing WLAN networks, which will significantly affect the overall system performance.

III. APPROACHES USED TO MINIMIZE THE HANDOFF LATENCY PERFORMANCE

A. Sync Scan: Practical Fast Handoff for 802.11 Infrastructure Networks [1]

Wireless broadcast networks, by their very nature, provide the opportunity for user mobility. Within the range of a given wireless base station, a client may roam freely and with complete transparency to the network medium. Inexpensive 802.11-based access points (APs) provide transparent connectivity to the wired Internet at low cost and with minimal configuration overhead. However, each individual 802.11 access point (AP) has a limited range - frequently less than 100 meters indoors - and therefore large-scale deployments of access points are required to provide comprehensive coverage of a building or campus. As a client moves outside the range of one access point, it must “hand off” to another to preserve the illusion of seamless connectivity. Consequently, as a mobile 802.11 client reaches the limits of its current coverage region, it must temporarily abandon its current access point, actively probe the network to discover alternatives, and only then reconnect to the current best AP. This approach minimizes management overhead, but is slow to handoff to superior access points and worse, can produce “gaps” in connectivity of up to a second in duration. To implement this sync scan algorithm using commodity 802.11 network interfaces and show that it can improve the timing of handoff decisions (thus improving the client's signal quality) and reduce handoff delay by over an order of magnitude compared to the existing approach. The handoff process in 802.11 networks has several phases each with its own costs. First, a client must determine that it is nearing the periphery of its coverage and thus must find an alternative access point to continue. Simply, this involves detecting that packets are no longer being successfully received. However, typical commercial implementations also monitor the current signal-to- noise ratio (SNR) and will also initiate the scanning phase. Once a client has decided to attempt a handoff it must next identify the set of adjacent candidate access points. Since 802.11 does not provide a shared control channel or other means for distributing this information, the client must explicitly “scan” each channel (11 in 802.11b and 802.11g, and 8 for 802.11a indoors) for potential access points. In its simplest form, this scan can be completely passive. The client switches to a candidate channel and listens for periodic beacon packets generated by access points to announce their presence. However, the latency incurred by this approach can be quite long since the phase of beacon intervals is independent and a client must therefore wait the full interval on each channel.

802.11-based access points periodically broadcast special beacon packets to identify themselves to potential clients and to synchronize state information with currently associate clients. While access points are commonly configured to send beacons every 100ms, within this period the 802.11 standard does not restrict the absolute time at which these packets are generated. In essence, arrangement is done so that the clients can passively scan by switching channels exactly when a beacon is about to arrive. Sync Scan algorithm implies that the heart of approach is the creation of periodic schedule of beacon periods spread across channel. As all access points operating on channel 1 will broadcast beacon at time t and access point on channel 2 will do the same at time t+d ms and so on. This called as synchronization scanning or sync scan. If continuously handoff must be attempted, the cost is reduced to that of authentication and re-association. Sync scan provides the opportunity to make better handoff decisions by continuously monitoring the signal quality of multiple access points. This method has some drawback, that the accuracy of clock in access point is critical to the global synchronization of beacon timings. However, over longer time scales access point require a means of time synchronization. Multiple access point operating on the same channel will attempt to generate beacon at the same time potentially interfering with one another.

B. An Empirical Analysis of the IEEE 802.11 MAC Layer Handoff Processes [2]

The IEEE 802.11 network MAC specification allows for two operating modes namely, the ad hoc and the infrastructure mode. In the ad hoc mode, two or more wireless stations (STAs) recognize each other and establish a peer-to-peer communication without any existing infrastructure, whereas in infrastructure mode there is a fixed entity called an access point (AP) that bridges all data between the mobile stations associated to it. A Handoff occurs when a mobile station moves beyond the radio range of one AP, and enters another BSS (at the MAC layer). During the handoff, management frames are exchanged between the station (STA) and the AP. Also the APs involved may exchange certain context information (credentials) specific to the station. Consequently, there is latency involved in the handoff process during which the STA is unable to send or receive traffic. Thus the handoff is a physical layer function carried out by at least three participating entities, namely the station, a prior-AP and a posterior-AP. The AP to which the station had physical layer connectivity prior to the handoff is the prior-AP, while the AP to which the station gets connectivity after the handoff is the posterior-AP. The information transfer can be achieved by an Inter Access Point Protocol (IAPP). Looking at it another way, the handoff-latency would be strictly greater than association latency as there is an additional inter-access point communication delay involved. In this proposed scheme of handoff mainly two steps are included like discovery, Re-authentication. The handoff process starts with the first probe request message and ends with a Re-association response message from an AP. We divide the entire handoff latency into three delays which is given below.

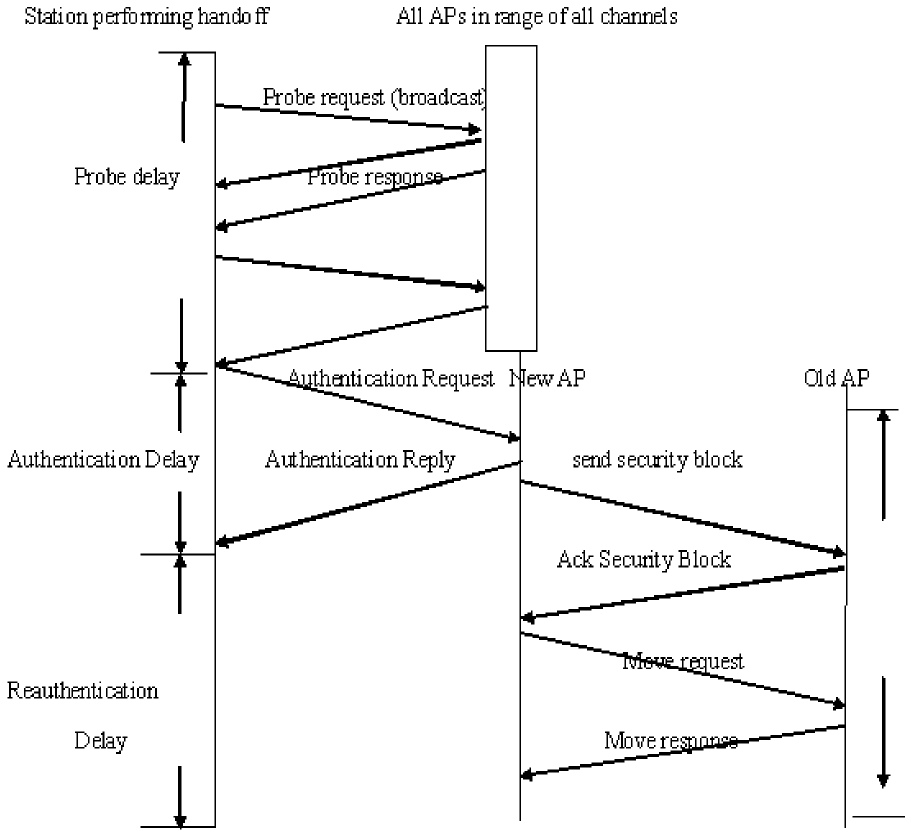

Fig4: Message exchanges of IAPP

The IAPP is a recommendation that describes an optional extension to IEEE 802.11 that provides wireless access-point communications among multi-vendor systems. The IAPP is designed for enforcing unique association throughout an Extended Service Set (ESS) and for the secure exchange of a station's security context between the current AP and the new AP during the handoff period. A problem with current 802.11 wireless equipments is that a mobile device cannot know if necessary QoS resources are available at a new access point until after a transition. Thus, it is not possible to know whether a transition will lead to satisfactory application performance.

C. Improving the latency of 802.11 hands-offs using neighbor graphs [3]

In this paper the handoff process defined for the IEEE 802.11 by using full scanning and observed scanning. In full scanning probe request is used for all channels within the range. In observed scanning the probe massage is send to some channel which is previously recorded on the list by using scan. In active scan the main problem is probe delay and probe count with respect to multiple access point operating on the same channel. In full scanning or Observed scanning algorithm. The station determines whether to stop on Min Channel Time or stay until Max Channel Time. If the medium is idle for Min Channel Time, the station can conclude to stop and probe the next channel. For neighbor graphs, overlap graph and non overlap graph considered for handoff procedure in IEEE 802.11. With consideration to this the overlap graph (OG) is an undirected graph over AP, the set of all access points in the network. An edge of an overlap graph, < APi, APj > represents an overlapping relationship between access points. APi and APj overlap if there exists a location where a mobile station can communicate to both of them with "acceptable" link quality. A Non-Overlap Graph (NOG) is a complement graph of an overlap graph, meaning that <APi,APj > is an edge in the non-overlap graph if and only if < APi,APj > is NOT an edge in the overlap graph. Two algorithms are given to define probing request in overlap and non overlap graph. The first algorithm defines neighbor graph algorithm with probe request is send to all neighboring APs on one channel. This request is kind of broadcasting the probe request on one channel for all APs. After sending the probe request, probe response is read out with Medium is idle until Min Chan Time expires else all APs on channel i have replied otherwise Max Chan Time expires. There are two factors that affect the latency, the number of channels to probe and waiting time on each probed channel. Second algorithm uses the non-overlap graph to gain additional performance improvement in probing. Once the station receives a probe response from APi, by the principle of non overlapping, it is impossible to receive a probe response from APj with acceptable link quality. Thus, there is no reason to wait for the response from APj, reducing the probe-wait time on APj's channel. This may even reduce the number of channels to probe if APj was the only AP on its channel.

The overall concept shows that the NG & NG-pruning algorithm reduces the probing latencies of full-scanning and the observed-scanning algorithms. Which provides a strong foundation for fast hand-offs in WLAN supporting VoIP and other multimedia application. This is done at the level of scanning by using probe massage only but it cannot improve the overall latency which is occurs in case of a handoff.

D. A Cross-Layer TCP Modeling Framework for MIMO Wireless Systems [4]

In this article the author wants to summarize the working of the MIMO (Multiple Input Multiple Output) technique with the TCP protocol by cross layer modeling. By applying the framework to analyze the TCP performance of two representative MIMO systems the BLAST systems and the orthogonal Space Time Block Coded (STBC).The proposed modeling framework to cross-layer analysis of MIMO wireless systems employing ARQ for both the BLAST and the orthogonal STBC systems, and study various performance metrics such as the optimal information rate that maximizes the TCP throughput, the effect of Doppler on the optimal TCP throughput, and the optimal channel coding rate that maximizes the TCP throughput under various modulation schemes.

This paper showed a simple and effective method to calculate the Gilbert model probabilities of a system employing MIMO wireless channels, taking various system parameters into account, including fading channels, space-time transmission schemes (BLAST, STBC), modulation, channel coding and ARQ. To calculate TCP throughput the new model is used is called as TCP Reno. It is shown that in Reno TCP, when the loss rate is low (as in the case of using ARQ), the loss event rate for TCP, i.e., the fraction of packets that trigger a loss indication at the TCP level, is very close to the TCP packet loss ratio. It showed that the optimal rate at which the TCP performance is maximum far from the channel capacity, and hence, from a TCP point of view, increasing the channel bandwidth efficiency is not always a good strategy. This paper also showed that increasing the ARQ retransmissions over certain number does not have a noticeable impact on the TCP throughput.

E. MDP-based handover policy in wireless relay systems [5]

Relays that receive and retransmit the signals between base stations and mobiles can be used to increase throughput extend coverage of cellular networks. Infrastructure relays do not need wired connection to network thereby offering savings in operators' backhaul costs. Mobile relays can be used to build local area networks between mobile users under the umbrella of the wide area cellular networks.

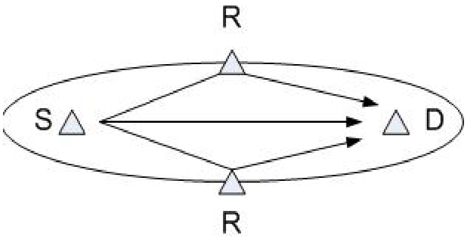

Fig 5: Relays in Wireless Communication

The relay transmission, which deploys multiple relay nodes between source node and a destination node. Due to its great potential to improve spectral efficiency and coverage area, relay transmission has been considered as a supplementary technology for the new generation wireless communication systems. The article focused study on a handover policy in wireless relay systems. At first, a relay transmission system with a single cell was introduced. The available transmission rates for both the direct transmission and relay transmission links have been analyzed. Due to the mobility of MSs, a handover decision problem is then considered, which is formulated as a Markov decision process (MDP). After that, a profit function is used to evaluate the QoS of the chosen serving node, which represents the benefit that the MS can gain by choosing a serving node. The cost function is also considered, which captures the processing and signal load incurred when the connection switch the current serving node to another. Based on the profit and cost functions, the reward function is formulated.

F. Performance Evaluation of Inter-Vehicle Communications Based on the Proposed IEEE 802.11p Physical and MAC Layers Specifications [6]

The demands of increased QoS requirements have resulted in more challenges for 802.11. Current handoff delays in 802.11 networks average in the hundreds of milliseconds. This can lead to loss of connectivity and degradation of connection quality especially for real-time voice or video applications. The delay that occurs during handoff should not exceed about 50 ms, the interval that is detectable by the human ear. Fast handoffs are thus essential for instance for 802.11-based voice and video connection. The network reconnection latency during intra-subnet handoff is solved by the existing IEEE802.11F or Inter-Access Point Protocol (IAPP). The IAPP is a recommendation that describes an optional extension to IEEE 802.11 that provides wireless access-point communications among multi-vendor systems. The IAPP is designed for enforcing unique association throughout an Extended Service Set (ESS) and for the secure exchange of a station's security context between the current AP and the new AP during the handoff period. The 802.11r working group of the IEEE is drafting a protocol that will facilitate the deployment of IP-based telephony over 802.11-enabled phones. The 802.11r standard is designed to speed handoffs between access points or cells in a wireless LAN. A problem with current 802.11 wireless equipments is that a mobile device cannot know if necessary QoS resources are available at a new access point until after a transition. Thus, it is not possible to know whether a transition will lead to satisfactory application performance.

IEEE std 802.11p-2010, also known as Wireless access in vehicular environment (WAVE), is an amendment to the IEEE Std 802.11-2007 standard that adds applications to fast changing vehicular networks. It deals essentially with the data link and physical layers of the OSI model. The focus of IEEE 802.11p lies on fast adaptation to rapid changes occurring in a highly mobile vehicular network, sacrificing identification and authentication procedures that are usually part of the IEEE 802.11 WLAN standards. For more efficient data exchange between high speed vehicles or between a vehicle and a Road Side Unit (RSU), IEEE 802.11p specifies a minimized set of parameters for the execution phase of the handoff process.

IV. MECHANISM DEFINED FOR HANDOFF LATENCY PERFORMANCE IMPROVEMENT

Handoff is the procedure providing the connection to the backbone network while a mobile terminal is moving across the boundaries of coverage of two wireless points of connection. For this purpose different protocols are used to improve the handoff latency performance in WLAN. The previous work related to the physical and MAC layer with consideration of different frequency and multiplexing concept. Afterword the same handoff latency performance improvement operation is necessary to be implemented on the transport layer to get the better result. So, we have to consider the different transport layer protocols for the same. Multimedia applications show that different transport protocols like TCP (Transmission Control Protocol) or UDP (User Datagram Protocol) are used to meet the new quality requirements. To face these new challenges, the IETF (Internet Engineering Task Force) defined a new transport protocol called Stream Control Transmission Protocol (SCTP) whose main features are multi-homing and multi-streaming. SCTP is a reliable transport protocol initially designed for signaling transport and reliable message based communication. Packets in an SS7 format need to be transported reliably and quickly. Because UDP is not connection orientated, it cannot be used for signaling. TCP is more suitable as it is connection orientated, yet it has several deficiencies. Strict byte order can cause head of line blocking. TCP is stream rather than message based and does not support multi-homing, therefore it is less reliable. It is also vulnerable to DoS attacks.

The multi-homing functionality of SCTP allows an association to be set up with multiple IP addresses at each endpoint. An association is similar to a TCP connection. This means that redundancy can be incorporated into the network so that switch over can occur between networks links to prevent lengthy interruption to data transfer. One of the IP addresses on this association can be designated the primary, the other the secondary. Data retransmission can be sent through the secondary link to better the chances of delivery. This is useful for SS7 signaling which needs to be reliable.

The multi-streaming capability of SCTP means that various sections of data can be split in terms of ordering. If packets from another stream are received out of order or delayed, another stream may be free, preventing blocking of data. This lessens the Head of Line blocking problem TCP faces. Only individually affected streams have the HOL effect, while others streams continue on passing packets up to higher layers. An example of multi-streaming can be given using a HTML page with 5 objects. TCP would create 5 separate connections for each of the objects. SCTP uses multi-streaming to send each object in a different stream, while opening only one SCTP association. Congestion control is based on TCP's rate adaptive window based scheme. It provides reliability, by detecting lost/corrupt packets and resends them. SCTP differs from TCP in that it does not have a fast recovery phase. SCTP achieves fast recovery by using a method of acknowledging packets known as SACK (Selective Acknowledgement). In Contrast to TCP, SCTP regards SACK as compulsory. This allows a more reliable reaction to multiple packets lost in a single window. This avoids the time consuming slow start stage after multiple segment losses.

The comparative study with different protocols for transport layer shown in table.

Table 1: *SCTP compared to TCP and UDP*

Feature/Service	SCTP	TCP	UDP
Allow half-closed connections	×	√	N/A
Application PDU bundling	√	√	×
Application PDU fragmentation	√	√	×

Congestion control	√	√	×
Connection-oriented	√	√	×
ECN capable	√	√	×
Flow control	√	√	×
Full duplex	√	√	√
Multi-homing	√	×	×
Multi-streaming	√	×	×
Ordered data delivery	√	√	×
Unordered data delivery	√	×	√
Reliable data transfer	√	√	×
Selective acknowledgements	√	Optional	×
Heartbeat mechanism	√	×	×

(√= yes, × = no, and N/A = not applicable)

Heartbeat Mechanism: A SCTP source should check if it is possible to reach the remote endpoint. This is done by means of the heartbeat mechanism. Alternative paths are monitored with heartbeat messages. Heartbeat messages are small messages with no user data periodically sent to the destination addresses, and immediately acknowledged by the destination. The sender of a heartbeat message should increment a respective error counter of the destination address each time a heartbeat is sent to that address and not acknowledged within the corresponding time interval (RTO, Retransmission Time Out). If this counter reaches a maximum value, the endpoint should mark this address as inactive. On the contrary, upon the receipt of a heartbeat acknowledgement, the sender of the heartbeat should clear the error counter of the destination address to which the heartbeat was sent, and mark the destination address as active.

V. CONCLUSION

Communication-based train control networks using WLANs have strict requirements for wireless communication availability and latency. The dynamic radio propagation environment and frequent handoffs can cause significant communication latency in WLAN-based CBTC. In this way, we studied the handoff management issues in CBTC networks. We presented a CBTC network based on SCTP and IEEE 802.11p WLANs to provide high communication availability and low latency in CBTC networks. The handoff decision problem was modeled as a MDP with the objectives of minimizing the handoff latency and maximizing the SCTP throughput.

REFERENCES

[1] I. Ramani and I. Ramani, “Syncscan: Practical fast handoff for 802.11 Infrastructure networks,” in *Proc. IEEE INFOCOM*, Seoul, Korea, Mar. 2005, pp. 675–684

[2] A. Mishra, M. Shin, and W. Arbaugh, “An empirical analysis of the IEEE 802.11 MAC layer handoff process,” SIGCOMM Comput. Commun. Rev., vol. 33, no. 2, pp. 93–102, Apr.2003.

[3] M. Shin, A. Mishra, and W. A. Arbaugh, “Improving the latency of 802.11 hand-offs using neighbor graphs,” in *Proc.MobiSys*, Boston, USA, Jun. 2004, pp. 70–83.

[4] A. L. Toledo and X. Wang, A Cross-Layer TCP Modeling Framework for MIMO Wireless Systems, VOL. 5, NO. 4, APRIL 2006

[5] Xiaoyu Dang, Jin-Yuan Wang2 and Zhe Cao,MDP-based handover policy in wireless relay Systems

[6] Diogo A., Igor C., Patrick Alves Performance Evaluation of Inter-Vehicle Communications Based on the Proposed IEEE 802.11p Physical and MAC Layers Specifications

[7] M. Fitzmaurice, “Use of wireless local area networks in rail and urban transit environments,” Transp. Res. Rec., J. Transp. Res. Board, vol. 1916, no. 1, pp. 42–46, 2005.

[8] R. D. Pascoe and T. N. Eichorn, “What is communication-based train control,” IEEE Veh. Technol. Mag., vol. 4, no. 4, pp. 16–21, Dec. 2009.

[9] Shaojian Fu and Mohammed Atiquzzaman ,”Performance Modeling of SCTP Multihoming” Telecommunications and Networks Research Lab School of Computer Science University of Oklahoma, Norman, OK 73019-6151, USA.

[10] Introduction to the Stream Control Transmission Protocol (SCTP): The next generation of the Transmission Control Protocol (TCP)