IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

A Review of Use of TCP over Wireless Cellular Networks

Vaibhav Vaidya[1], Prof. Amutha Jeyakumar[2]

Electronics Engineering, Veermata Jijabai Technological Institute, Matunga, India[1, 2]

Abstract: This paper is a review of the issues faced when TCP is used over wireless cellular networks further to that proposed solutions to these issues. The problem is first characterized by the TCP operation followed by characteristics of wireless channels that affect performance. Different Wireless generations are compared as they determine nature and severity of problems when TCP is used. Further to this paper provides proposals made to resolve the problem. Performance degradation is considered based on examples from previous studies.

Keywords: Standard TCP, mobile wireless network, TCP performance, end-to-end, 3DA, M-TCP, Freeze TCP.

1. INTRODUCTION

The development of mobile computing growing exponentially. Nowadays all the mobile computing moves around data network which basically relies on past work and standards used for wired networks, including the TCP/IP protocol suite. The definition of Mobile IP found in [8] was the first move in development of the Wireless Internet, but many problems are yet to be solved. One most significant problem that has emerged when transmission control protocol is used over wireless cellular network is degraded performance. Since TCP was developed originally for wired communication, its use in wireless communication network results in unforeseen problems due to uncertainty in wireless links at radio link level. High bit errors and frequent handoffs in the radio link are interpreted by TCP as congestion, and the actions taken to mitigate congestion result in poor or degraded performance. This problem will be further characterized in Section II, while Section III deals with the proposed solutions made to resolve the problem.

2. PROBLEM ANALYSIS

A. TCP operation and response to lost segment

TCP operation and response to lost segment Transmission Control Protocol is one of transport level protocol that takes data from upper layer, divides it into small chunks, and adds a TCP header creating a TCP segment. The TCP segment is further encapsulated with IP header to form packets, and then exchanged with peers. TCP protocol is connection oriented protocol. The overall operation of protocol can be described in terms of how TCP prepares, negotiates, establishes, manages and terminates connections.TCP operations may be divided into three phases, connection establishment, data transfer, connection termination. Connections establishment requires a 3-way handshake process between peers.

Once connection is established TCP enters into data transfer phase, post complete data transfer between peers, the connection termination closes established virtual connection and releases all allocated resources.[9] TCPs response to lost segments is designed for congestion in wired networks, but in wireless networks, it causes poor performance. Some characteristics of TCP and its response to congestion are explained here causing poor performance and later exploring possible solutions. Below is a summary of TCPs response with more details. When TCP transmits a segment, TCP at sender end starts timer which tracks how long it takes for an acknowledgment forthat segment to return. Time taken by segment to transmit and receive back its acknowledgment is called as round-trip time (RTT). Retransmission timeout(RTO) timer at sender, for a TCP connection is determined by these RTT over that link. These RTT measurements are collected and RTO is determined as a sum of smoothed RTT. RTO measurement is very crucial and directly effects on utilization of resources. If the RTOis set to very high value, retransmission will be delayed by long time, this will result in slow recovery of network. If the RTO is set to too short, unnecessary retransmissions will occur and effective throughput will be decreased.[9].

A sent TCP segment is assumed to be lost if no acknowledgement (ACK) is received for the segment within retransmission timeout (RTO)or if multiple duplicate ACKs were arrive for the segment sent prior to the one that was lost. When segment loss is determined by RTO event, TCP initiates an exponential backoff and enters in slow start and congestion avoidance mode. In exponential backoff RTO timer value is doubled every time on expiration of RTO after packet has been re-transmitted. Along with this packet transmission rate is also reduced so that congestion can be avoided. This is known as slow start, congestion window is the number of packets that can be transmitted over network without causing any congestion. In slow start congestion window is set to value one, with each successful acknowledgment this window is exponentially increased. Once this value reaches its threshold, which is half of its value when segment loss was determined previously, TCP enters into

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

congestion avoidance and increase congestion window linearly instead of exponentially. While considering transmission over a wireless channels, it is important to note that multiple lost packets due to lossy channels will cause TCP to enter in congestion control and reducing slow start threshold repeatedly. In this situation multiple invoke of congestion avoidance mode will cause packet transmission rate to grow very slowly. This can lead to degradations in throughput.[4] If sender encounters with duplicate ACK's of transmitted packet packet loss can be determined, in this case TCP beings in fast retransmit and fast recovery mode. In this situation TCP dose not wait for RTO event to occur and starts retransmission of segments that are not acknowledge. With fast retransmit time is not lost waiting for RTO event to occur. Further to this fast recovery algorithm is invoked skipping slow start phase. This will avoid excessive decrease in transmission rate. In fastrecovery congestion window is not set to one instead it is set to half of current value and congestion avoidance mechanism is invoked. Fast retransmit and fast recovery is later introduced as an alternative to conventional congestion control mechanism.

B. Lossy Channels in Cellular Networks

The wireless media and wired media has very different characteristics. Wireless medium is generally affected by 3 factor which are path loss, surrounding noise and the sharing of the radio spectrum[5]. Path loss and ambient noise causes higher bit error rates (BER) in wireless link when correlated with a wired link. As an example, typical values for bit error rates are 10^{-6} or worse over wireless paths, while a error rates of 10^{-12} or better for wired path typically fiber links[1]. Addition to this mobility in wireless network causes additional degradation. Mobility in wireless network causes different types of fading due to which radio channel status changes intermittently, this increase bit error rate. In cellular approach coverage area is thus divided into smaller areas as per requirement and resources are shared between them further causes losses. Handoff between one cell to another cell is required when mobile user crosses boundary. When handoff occur mobile terminal must get synchronized with set of other resources. This handoff causes additional delay and data losses if connectivity is got lost during handover. So as compared wireless links are highly unreliable and causes frequent connection and disconnection.

C. Different Cellular Networks

Wireless networks are bounded by different bandwidth and cell coverage areas based on type of wireless network. Characteristics of wireless network determines round trip delay based on latency, also handoff frequency is different for different networks. These characteristic directly affects TCP performance over wireless link. Below are described different cellular wireless networks and its characteristics that affect TCP performance.The focus of most studies of TCP over wireless links is for current mobile cellular network environments. Mobile cell coverage radii are usually on the order of hundreds of meters to few kilometers, with bit rates in the range of 2 to 100 Mbps. For example, the Global System for Mobile Communication (GSM) has a bit rate of 9.6kbps and RTTs are in hundreds of milliseconds. While in case of Long Term Evolution (LTE) network, bit rate is upto 150mbps and the RTT on a LTE network averaged 50 ms[3]. Given the traits of these various types of cellular wireless systems, it is important to note how TCP performance will vary over the wireless links[2]. Link delay or latency in network directly affects TCP performance. In case of longer delay, TCP must wait for longer time for data transmission increasing transmission window. In case of higher speed links TCP takes more time to reach its original peak post packet loss. One more factor that affects performance is radio cell coverage areas. Smaller the coverage area more frequent handoffs, which reduces overall throughput. Cellular networks are generally deployed in hierarchical structures, meaning different radio coverage overlaps each other in order to increase capacity. For example, a LTE microcell located within a LTE macrocell should in the future allow for seamless handoff between the two cells. These constitute additional handoff within single system which cause additional problems resulting in degraded performance.

D. Measured Performance

The comprehensive study primarily aimed at measuring TCP's performance in cellular wireless networks was performed. The main focus of this study was on the effect of crossing cell boundaries and it makes measurements for mobile networks. Based on analysis of three different handoff situations it is found that throughput decreases significantly with handoff when compared without handoffs. Also, different adverse effects including pauses in communication, packet loss, and slow recovery are measured. Some details of observation found in are given below. The reduction in throughput found in [1] depends on the nature of the handoff. While handoffs between overlapping cells, the mobile stay connected and never loses connectivity, and the throughput reduces by only few ten percent. This loss in throughput is due to signalling procedures to change forwarding path. These encapsulation and forwarding delay causes short pauses after handoff. When cell boundaries are non-overlapping, the degradation in throughput is more. In such case, packets are lost while routing tables are being updated. Also, waiting periods for very high RTOs causes long pauses in transmission, which is determined to be the main cause of the decrease in throughput. For a handoff which is occurring at the instant the mobile crosses the cell border, throughput reduces by few 30 percent. Whenever there is a lag time during handoff and the mobile loses connectivity for some period of time, the throu ghput is measured to reduce by 40 percent. Other measurements taken in [1] involve the long pauses occurring during handoff. For TCP without fast retransmission and recovery, causes disconnection in communication upto 800ms after a handoff from non-overlapping cells. By implementing fast retransmissions reduces disconnection

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

period upto 200-300ms. Due to exponential backoff of the RTO, the pauses in transmission grow exponentially with increasing lag time during handoff. In worst cases, disconnections are measured to last several seconds after post handoff completion which is not desirable.

3. PROPOSED ARCHITECTURE

Many solutions have been put forth for improving TCP performance over wireless links. This section describes a number of these proposed solutions, though the possibilities described here are certainly not exhaustive. As the problem is caused by poor interaction between the link and transport layers, most solutions are suggested for either one of these two layers. Link layer solutions work with the aim of improving link characteristics or hiding non-congestion-related losses from the transport layer. Transport layer solutions instead attempt to adapt the TCP protocol to make it aware and respond appropriately to losses that are not related to congestion.[3]

A. End to End Solutions

In end to end design solutions, end to end semantics between mobile host and fixed host is maintained. Fig 1 shows E-2-E TCP stream between mobile host and fixed host irrespective of intermediate path and nodes.

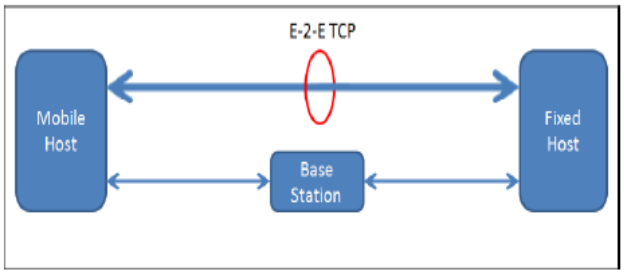

Fig. 1. End to End Connection approach

Below are the few end to end connection approach.

1) 3 duplicate ack approach

The approach proposed in, 3-duplicate acknowledgements approach, requires network layer to provide information about ongoing mobility to the TCP layer at MH. At the time of reconnection, the MH sends three duplicate acknowledgements to FH. These acknowledgements cause the FH to enter the fast recovery phase and restart transmission. This scheme reduces the idle period after an MH is re connected, otherwise TCP at the FH would have waited for an RTO event to our before restarting transmission.[1]

2) Snoop protocol

In this approach, Base station routing code is modified by adding a module, called snoop, that monitors every packet that passes through the connection in either direction. The snoop module maintains a cache of TCP packets sent from the FH (fixed host) that haven't yet been acknowledged by MH[10]. When a new packet arrives from the FH, snoop module add it to its cache and passes the packet on to the routing ode which performs the normal routing functions. The snoop module also keeps track of all the acknowledgment sent from the mobile host. When a packet loss is detected(either by the arrival of a duplicate acknowledgment or by a local timeout), it retransmit the lost packet to the MH if it has the packet cached. Thus, the base station(snoop) suppresses the packet loss by not propagating duplicate acknowledgments towards FH. These mechanisms together improve the performance of the connection in both directions, without sacrificing any of the end-to-end semantics of TCP, modifying host TCP code in the fixed network or relinking existing applications. This combination of improved performance, preserved protocol semantics and full compatibility with existing applications is the main contribution of our work.[10] The main drawback of the split connection approach is that the semantics of TCP acknowledgments are violated. In contrast, the snoop protocol maintains the end-to-end semantics of the TCP connection between the fixed and mobile hosts by not generating any artificial acknowledgments. Though end to end semantics are maintained but when IP level encryption is used snoop module approach fails. Handoffs in this approach require the transfer of a significant amount of state.

3) Freeze TCP

The main idea of this approach is that when MH sense that it going to disconnect from the network it advertised its window to zero which cause FH sender to go into persist mode. And when mobile host connect back to network, it send 3-duplicate acknowledgment of last received byte. Thus forcing FH-sender to take fast retransmit and fast recovery action and start sending data immediately after MH connect back to network. In case of impending handoff or disconnection, it can advertise a zero window size, to force the sender into the ZWP mode and prevent it from dropping its congestion window. This approach does not deal with high bit error rate of wireless channel. The approach requires the network layer to give an indication of impending disconnection.To implement this scheme, only the clients TCP code needs to change and there is no need for an intermediary[6].This makes Freeze TCP approach to interwork with different networks as network dependency is not required. Sense period for impending disconnection should not be longer or shorter than RTT led to worse average performance in freeze TCP. Freeze-TCP mechanism to enhance TCP throughput in the presence of frequent disconnections (and reconnections) which characterize mobile environments. It is a true end-to-end signaling scheme and does not require any intermediaries to participate in the flow control. It does not require any changes to TCP code on the sending side. Changes to TCP code are confined entirely to the receiver side and are easy to implement.

4) Selective acknowledgement

One modification to TCP that would improve performance over wireless links is the use of Selective

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

Acknowledgement (SACK). The use of cumulative ACKs in traditional TCP result in poor performance when multiple packets are lost during one transmission window. The cumulative ACKs do not provide information quickly enough to allow for fast recovery. The SACK proposal[11] would modify ACKs to contain segment sequence numbers. This could allow for lost segments to be quickly identified and resent within a single RTT. In [3]it is noted that SACK implementation is particularly useful in bursty error channels. Performance improvements due to SACK are shown to be significant, though not as great as improvements provided by other link-layer solutions[3]. Explicit loss notification (ELN) is another solution which provides explicit notification to sender to distinguish between losses due to congestion and those due to errors on the wireless link. This is achieved by marking cumulative ACKs to identify a loss on the wireless link that is not related to congestion. Thus, the sender can retransmit segments without initiating congestion-control mechanisms that would unnecessarily reduce throughput. A drawback of ELN is that it may be difficult to identify which packet losses are due to errors on the wireless link.

B. Split connection solutions

In split connection approach TCP stream between mobile host and fixed host is splitted into 2 separate TCP stream. Fig 2 shows 2 separate TCP streams one between mobile host and base station and other between base station and fixed host. Below are the few split connection approach proposed.

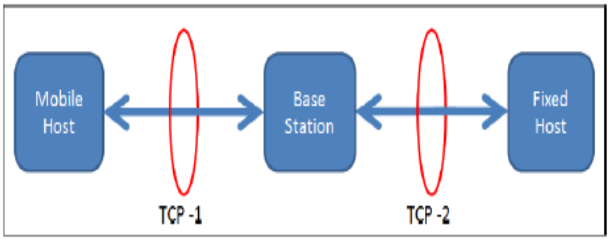

Fig. 2. Split Connection approach

1) I-TCP

In I-TCP a transport layer connection between an FH(Fixed host) and MH(Mobile host) is splitted in two separate connection one between FH to BS, another between BS to MH. Normal TCP run on connection between FH to BS and specialized protocol over BS to MH connection can be used for flow and congestion control tuned for wireless condition. When a packet is received on FH to BS connection it is acknowledged independent of second connection between BS to MH and vice-versa. When a MH moves out of old BS region and enter new BS region, the whole state consisting of two socket per connection is transferred to new BS. It may increase time involve in handoff. This scheme does not preserve semantics of TCP that FH gets the acknowledgement only after peer TCP at MH has got the data. It separates the flow control and congestion control functionality on the wireless link from that on the fixed network. This separation is desirable because of the vastly different characteristics of the two kinds of links the fixed links which are becoming faster and more reliable whereas the wireless links which are still very slow and are extremely vulnerable to noise and loss of signal due to fading[12]. A separate transport protocol can have event reporting functionality.

2) M-TCP

It assume 3-tier architecture for network[4]. Many MH can be connected to MSS(Mobile Support Station) and many MSS are connected to SH(Supervisory host) and SHs are interface to fixed networks. MSS have minimum capability just to support MH for communication and SH are responsible for bandwidth management and mobility management. It also assume that transport layer see low bit error rate over wireless link. And M-TCP approach mainly take are of frequent handoff related problem. It uses the split connection approach i.e. every TCP connection is split in two at SH. The TCP sender on fixed network uses unmodified TCP to send data to the SH while the SH uses modified version of TCP called M-TCP for delivering data to MH. The TCP client at the SH (SH-TCP) receives segments transmitted by the sender and it passes these segments to the M-TCP client for delivery to MH. ACKs received by M-TCP at the SH are forwarded to TCP client for delivery to the TCP sender.

4. CONCLUSION

This report is a review of the issue of TCP performance over wireless links. The problem is characterized by describing baisc TCP operation, TCPs response to lost packets and the fundamentals of wireless links. The effects on different types of cellular wireless networks are noted, and measured performance results from past studies are given. Some proposed solutions and their effectiveness are then explored and compared. It can be observed that it is desirable for an approach to require modification only at MH compare to requiring support from BS also. It enable the approach to be applicable even when traffic is encrypted or acknowledgement pass through different path. When these different approaches are compared based on IP level Encryption support, interpretability and scalability only 3DA and freeze tcp approaches fall into this category. 3DA approach focuses on reducing idle time period after reconnection but same time has the degrading side effect on throughput. It casues tcp to enter in fast recovery phase post receving 3 duplicate acknowledgment. Whereas Freeze TCP approach requires lower layer to predict disconnection, also accurate prediction is required based on RTT otherwise will lead to degraded performance. All these factors lead us to propose new approach which require modification at mobile host only, does not require prediction of future disconnections, does not reduce the congestion window of the sender after mobile host get reconnected while it reduces the idle time successfully. It should also focus on MH to FH data transfer.

DOI 10.17148/IJARCCE.2016.5710

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2016

REFERENCES

[1] R.Caceres and L.Iftode, Improving the Performance of Reliable Transport Protocols in Mobile Computing Environments, IEEE Journal on Selected Areas in CommunicationsJune 1995, vol.13 no.5, pp.850-857.
[2] G.Xylomenos, G.C.Polyzos, P.Mahonen, and M.Saaranen, TCP Performance Issues over Wireless Links, IEEE Communications Magazine, April 2001.
[3] H.Balakrishnan, V.N.Padmanabhan, S.Seshan, and R.H.Katz, A Comparison of Mechanisms for Improving TCP Performance over Wireless Links, IEEE/ACM Transactions on NetworkingDec 1997, vol.5 no.6, pp.756- 769.
[4] W.R.Stevens, A Comparison of Mechanisms for Improving TCP Performance over Wireless LinksTCP/IP Illustrated, Volume 1: The Protocols, Addison-Wesley Longman, Reading, Massachusetts, 1994.
[5] L.Ahlin and J.Zander, Principles of Wireless Communications, Studentlitteratur, Lund 1997.
[6] Tom Goff, James Moronski, D. S. Phatak Freeze-TCP: A true end-toend TCP enhancement mechanism for mobile environments, Electrical Engineering Department State University of New York, Binghamton, NY 13902-6000in Proceedings of ACMSIGCOMM96, Palo Alto, CA, Aug 1996, pp. 256 269.
[7] Chunlei Liu, Raj Jain Approaches of Wireless TCP Enhancement and A New Proposal Based on Congestion Coherence, Department of Computer and Information Science Ohio State University, Columbus
[8] C.E.Perkins, ed., IP Mobility Support, IETF RFC 2002,1996
[9] www.ietf.org/rfc/rfc793.txt RFC 793,
[10] Hari Balakrishnan, Srinivasan Seshan, and Randy H. Katz, Improving Reliable Transport and Handoff Performance in Cellular Wireless Networks, In ACM Wireless Networks Journal December 1995.
[11] M.Mathis, J.Mahdavi, S.Floyd, and A.Romanow, Selective acknowledgement options, IETF RFC 2018 1996. 5
[12] Ajay Bakre, B.R. Badrinath I-TCP: Indirect TCP for Mobile Hosts, Tech Rep., Reuters university, May 1995, http://www.cs.rutgers.edu/badri/journal/ontents11.html.
[13] K. Brown and S. Singh M-TCP: TCP for Mobile Cellular Networks, ACM Computer Communications Review, vol27, no.5, 1997.