Learning Objectives

- Understand the problems associated with wireless or mobile networks

- How TCP handles packet loss

- Why standard TCP cannot be applied to wireless networks

- Need of specialized protocols for wireless networks

- Goals of specialized protocols for wireless networks

- I-TCP architecture and functioning

- Advantages and limitations of I-TCP

Introduction

Upto now, we have seen implication of mobility on lower layer protocols like data link and network layer. For example for data link layer we have seen that hidden and exposed terminal problems fail the application of CSMA\CD which works well in wired networks. We discussed some of the protocols specially designed for MAC layer of all layer like MACA, MAAW, DAMA, PRMA and so on. Similarly for network layer, there is mobile IP technology to support the mobility yet providing seamless connectivity but support for mobility only at lower layers is not enough. The application required mobility support at transport layer also. Mobile computing is a paradigm that provide services to the users on "anywhere" "anytime" basis. In such a scenario mobile hosts expects same services as offered to the fixed hosts hence a transport layer protocol is required for mobile domain also.

Mobile network itself is a hostile environment therefore of TCP and UDP, TCP is best choice as it is connected oriented transport layer oriented transport layer protocol that provides in order delivery of packets to the receiver. But TCP undergoes the congestion control mechanisms in the form of "Slow start" during packet loss assuming that there cannot be any other for packet loss other than congestion. But in wireless or mobile environment there are many reasons for the packet loss like higher bit error rate, effect of disconnections, improper handoff's may also lead to packet loss therefore. Replication of TCP there would cause serious drop in through put.

Therefore specialized TCP is requirement for wireless environment. In this module we will understand functioning of standard TCP and how it handles packet losses. Then we will see why we cannot replicate standard TCP in wireless networks. We will learn about need and goals of specialized protocols for transport layer in mobile computing environment. There have been several protocols proposed in this respect.

In this module we will learn about protocol I-TCP and in subsequent modules we will learn about more protocols

Regular TCP

It is a connection oriented protocol. It has been tuned to work for networks composed of wired links and stationary hosts. It is a reliable protocol. Reliability is maintained by observing acknowledging and resending lost packets. The number of packet sent is regulated by increase or decrease size of the window. TCP sender uses the cumulative acknowledgements sent by the receiver. To adapt with the problems of wired link the window size is changed. The wired data transmission is through fiber optics copper wires etc. and works without introducing transmission errors. If a software is mature enough it will not drop packets or flip bits so a packet drop cannot be due to hardware or software error. Therefore the reason for packet loss is some overload in transmission path i.e. congestion. It assumes congestion to be the main reason of packet loss.

Congestion is a condition when the buffers of the router are filled, the sum of input rates of packet is greater than output link. The router cannot forward the packets fast enough and drop the packets. It is lost. The receiver notices a gap in stream of incoming packets. It continues to acknowledge all the in-sequence packets up to the missing one. Sender notices missing acknowledgement and assumes packet loss. Therefore TCP perform following congestion control scheme. The scheme has 3 parts:

- Slow start

- Congestion Avoidance

- Fast Retransmit

Slow Start

It works as follows:

- Sender calculates congestion window for the receiver

- The transmission starts with congestion window (cwnd=1)

- Sender sends one packet; If acknowledgement received increase window size by 1 i.e. 2

- Sender now sends two packets; If acknowledgement received for two packets, window is increased by 2 one for each acknowledgement i.e. 4

- For each received acknowledge TCP increase the window exponentially only upto a threshold. Once congestion threshold is achieved, congestion phase starts.

Congestion Avoidance

- As soon as congestion threshold is received, window is increased linearly by adding 1 to the congestion window each time acknowledgement is received

- it continues to increase its window size linearly until

  - it reaches the receiver's maximum advertised window or

  - packet loss is observed due to time-out

  - duplicate acknowledgement is received for same packet

How TCP understands condition for packet loss

TCP measures how long acknowledgements take to return back and use it to determine which packets have reached the receives and provides reliability by retransmitting lost packets. For this purpose, it maintains a running average of round trip delay and an estimate of expected deviation from this average. If current delay is longer then average by more than 4 times expected deviation. TCP assumes packet loss and retransmits.

It also assumes packet loss if duplicate acknowledgements received. This is because receiver acknowledges higher in order sequence number. If it receives out-of-order packets, it generates acknowledgements for same highest in order sequence number. If it receives out-of-order sequence number and that result in duplicate acknowledgement. In this case also it assumes packet loss but not due to congestion due to error on wireless link.

Fast Retransmit

If sender receives 3 duplicate acknowledgements then it activates fast retransmit algorithm. It assumes data segment indicated by duplicate acknowledgements is lost and retransmits lost segment. No time is wasted waiting for timeout in order for retransmission. Fast retransmission is documented in RFC 2001 & RFC 2581

Illustration

- Sequence#1→Sent : ACK for 1 received
- Sequence#2→Sent : ACK for 2 received
- Sequence#3→Sent : ACK not received
- Sequence#4→Sent: received at receiver but ACK for 3 sent by receiver
- Sequence#5→Sent: received at receiver but ACK for 3 sent by receiver
- Sequence#6→Sent: received at receiver but ACK for 3 sent by receiver
  Number of duplicate acknowledgement is 3!!!
  Packet 3 is lost!!!
- Only sequence number 3 is retransmitted
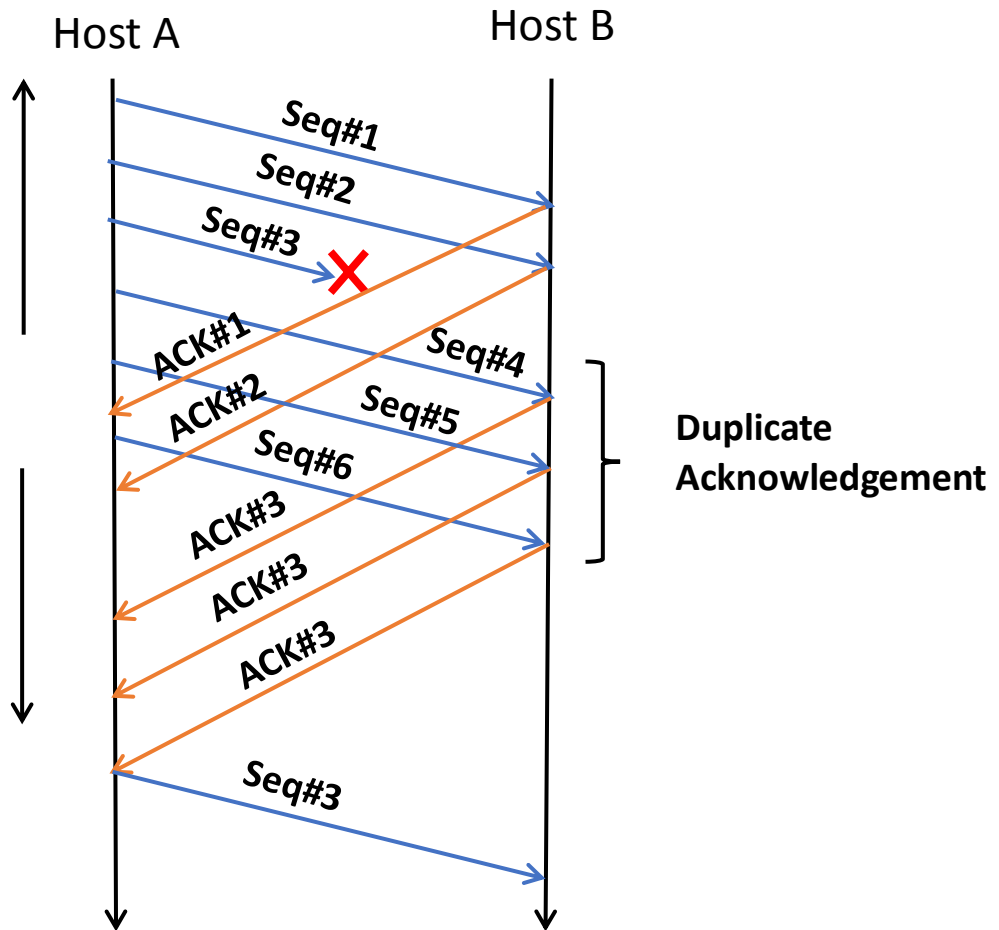
What it does in case of packet loss?

Host A   Host B

Seq#1
Seq#2
Seq#3 ✗
ACK#1   Seq#4
ACK#2   Seq#5
          Seq#6
ACK#3
ACK#3
ACK#3
Seq#3

Duplicate Acknowledgement

Figure 1 Fast retransmit

TCP reacts any packet loss by
1. Dropping threshold value into half the current window or 2 whichever is larger
2. Resetting the transmitting window size to 1 activating slow start
3. Reset retransmission timer to a back off interval that doubles with each consecutive timeout

Can TCP be applied for mobile hosts as well?

TCP was not designed for mobile host so it is not expected to perform well in wireless/mobile network when a packet is loss, regular TCP assumes that it is due to congestion. Whereas in mobile networks there can be packet losses unrelated to congestion. Some of them are:

- **High Bit Error rate:** Wireless links suffers from a high bit error rate high (can be of the order of 10^{-5}).BER causes packets to be corrupted resulting in loss of TCP data segments or acknowledgements. Missing acknowledgements triggers time out at sender causes retransmission with window size to be one which drastically reduces the throughput.

FEC at data link may handle high BER but it would be wastage of wireless bandwidth when correction is not necessary

- **Black outs:** When the mobile move from one cell to another call or service is transferred from old base station called Handoff. Hard Handover- there is brief disconnection known as black out period. Any packet sent during this period might be lost.

- **Call Blocking:** Disconnections due to call blocking; a condition when mobile station do not get any channel due to unavailability of the channels. The disconnection period can be of few seconds and can also last for 1 minute. These disconnections also result in loss of packets and acknowledgements. Since these disconnections are very lengthy the FEC scheme is ineffective cell sizes

- **Disconnection due to small cell latencies:** When the cell sizes are small to accommodate more and more users and increase the capacity results in small cell latencies and hence frequency disconnection.

- **Signal Blocking:** When radio signals are blocked due to buildings and other objects in the environment

- **Power Scarcity:** Mobile computer are battery operated hence power required is more

In such a scenario if TCP triggers congestion control procedures, it will significantly reduce the throughput. There has been lot of research in the methods to improve TCP in wireless environments. The goal of protocols:

- It is not possible to change the entire TCP because it is the base on which the entire internet is based. Therefore proposed protocol should not change standard TCP

- The protocol should not go into slow start when it is not genuine to go

- Isolate the mobility related problems from existing network protocols when fixed host and wireless station to communicate.

Indirect TCP(I-TCP)

It is a split connection protocol proposed by Bakre and Badrinath in 1995. It splits an End-to End TCP connection between fixed host and mobile host into two separate connections at an intermediator called MSR (Mobility support router).MSR can be an access point, foreign agent incase mobile IP, entry point of network like IWF in GSM, GGSN in case of GPRS.

- The link between fixed computer and AP is supported by regular TCP and between MSR and mobile node optimized TCP is used

The MSR acts as proxy between fixed computer and wireless node. It is seen as wireless node to fixed computer as fixed computer to wireless node
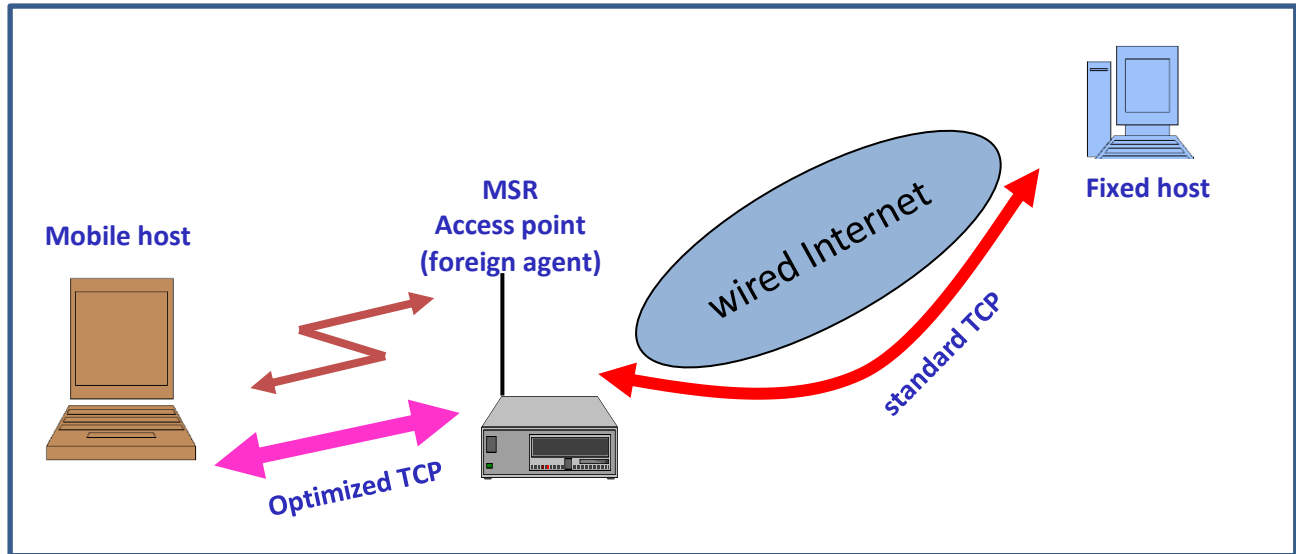

Figure 2 I-TCP

Packet goes from fixed node to wireless node

- Sender sends packet to MSR

- It acts as proxy; buffers the packets ; gives acknowledgement to the receiver

- MSR then forwards packet to mobile host

- If packet is lost on wired link, standard TCP handles it

- If packet lost on wireless link, MSR retransmits with low round trip time

Packet goes from wireless node to fixed node

- Mobile node sends the packet

- MSR acknowledges it and forwards to the fixed computer

- If packet is lost, it is retransmitted on wireless with lower round trip time

Handovers

- At the time of handover old proxy forwards the packets to the new proxy because acknowledgements are already sent for those packets

- New MSR then forwards those packets to the mobile host

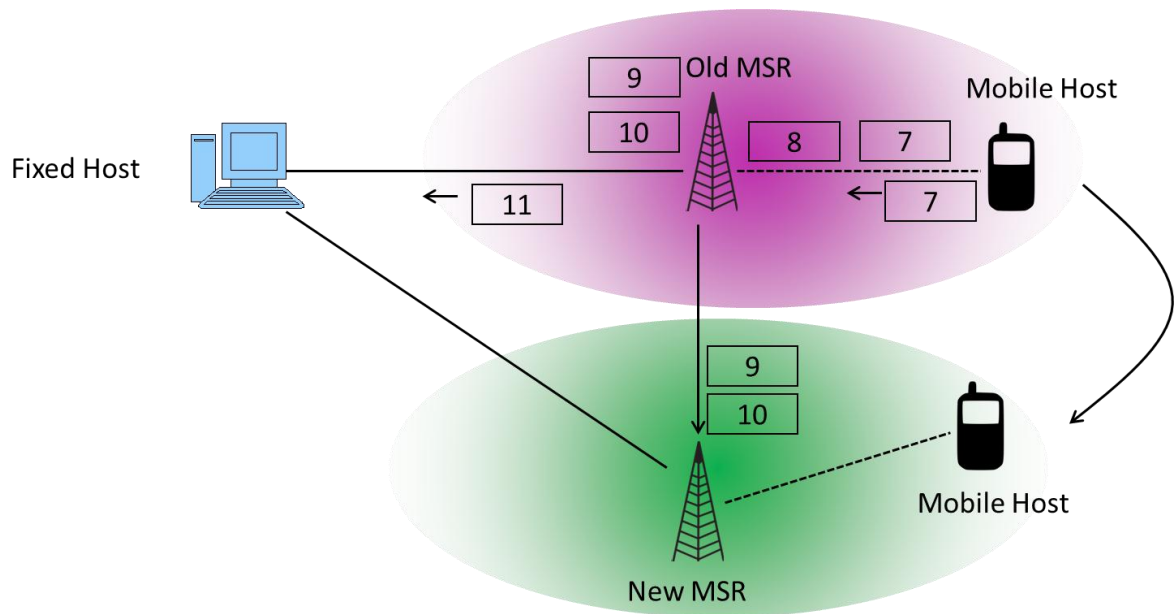- Current state of TCP (sequence numbers, address, ports)reflected by socket should also be migrated

Figure 3 Packets forwarded during handover

Software Components

The components for the software to support I-TCP

1. **MH Side:** I-TCP can be accessed as a transport protocol by application running on mobile host using special library calls. They are similar in interface and function to the socket calls made by an application using register end – to – end TCP

2. **MSR Side:** In the proposed protocol, most of the functionality for supporting I-TCP connection lies with the MSR. If acts as actual bridge connecting the wired and wireless parts of the connection. It consists of a user level UNIX process pumping data from one part of connection to the other. Hand off support for I-TCP connection is implemented in MSR kernels

Advantages

1. No need to make changes in TCP protocol used by computers in fixed network or any other device which do not need optimization

2. Transmission error on wireless links do not propagate in fixed network

3. The optimized TCP is used only on wireless link without affecting stability of internet

Disadvantages

1. Loss of end - to – end semantics packets already acknowledgements before being delivered

2. Whole scheme fails if MSR crashes

3. Increased handover tendency

4. MSR should be a trusted entity. Serious security threat if MSR is compromised