Mobile Transport Layer

- TCP review
- Motivation
- TCP mechanisms
- Indirect TCP
- Snooping TCP
- Mobile TCP
- Fast retransmit/recovery
- Transmission freezing
- Selective retransmission
- Transaction oriented TCP
Motivation

TCP protocol designed for
- fixed, wired networks
  - low delay and delay jitter
  - low packet loss rate
- packet loss is caused by router congestion

TCP congestion control mechanism
- packet loss in fixed networks typically due to (temporary) overload situations
- router have to discard packets as soon as the buffers are full
- TCP recognizes congestion only indirectly via missing acknowledgements (timeout)
- retransmissions unwise, they would only contribute to the congestion and make it even worse
=> TCP applies slow-start algorithm as reaction

Wireless link scenario
- high frame error rate requires link layer ARQ
- link layer ARQ and buffering results in high delay and jitter
- persistency of link layer ARQ is usually configured for a trade-off between delay/jitter and residual packet loss rate
TCP Review

TCP slow-start algorithm – timeouts

- sender calculates a congestion window for a receiver
- start with a congestion window size equal to one segment
- exponential increase of the congestion window up to the congestion threshold (ssthresh), then linear increase
- missing acknowledgement (timeout) causes
  - reduction of the congestion threshold (ssthresh) to one half of the current congestion window (cwnd)
  - congestion window (cwnd) starts again with one segment

TCP fast retransmit/fast recovery – dupacks

- TCP sends an acknowledgement only after receiving a packet
- if a sender receives several acknowledgements for the same packet (dupack), this is due to a gap in received packets at the receiver
- however, the receiver got all packets up to the gap and is actually receiving packets
- therefore, packet loss is not due to congestion, continue with current congestion window (do not use slow-start)
Influence of wireless transmission on TCP

TCP assumes congestion if packets are dropped
- typically wrong in wireless networks, here we often have packet loss due to transmission errors
- furthermore, mobility itself can cause packet loss, if e.g. a mobile node roams from one access point (e.g. foreign agent in Mobile IP) to another while there are still packets in transit to the wrong access point and forwarding is not possible

The performance of an unchanged TCP degrades severely
- however, TCP cannot be changed fundamentally due to the large base of installations in the fixed network, TCP for mobility has to remain compatible
- TCP on server does not know whether peers are mobile or not
- the basic TCP mechanisms keep the whole Internet together
Indirect TCP – Principle

Indirect TCP (I-TCP) segments the connection

- no changes to the TCP protocol for hosts connected to the wired Internet, millions of computers use (variants of) this protocol
- optimized TCP protocol for mobile hosts
- splitting of the TCP connection at, e.g., the foreign agent into 2 TCP connections, no real end-to-end connection any longer
- hosts in the fixed part of the net do not notice the characteristics of the wireless part

![Diagram of Indirect TCP](image-url)
Indirect TCP – Socket and state migration due to handover

A handover between access points requires the migration of the TCP sockets and the TCP state!
Indirect TCP – Discussion

Advantages

- no changes in the fixed network necessary, no changes for the hosts (TCP protocol) necessary, all current optimizations to TCP still work
- transmission errors on the wireless link do not propagate into the fixed network
- simple to control, mobile TCP is used only for one hop between, e.g., a foreign agent and mobile host
- therefore, a very fast retransmission of packets is possible, the short delay on the mobile hop is known

Disadvantages

- loss of end-to-end semantics, an acknowledgement to a sender does now not any longer mean that a receiver really got a packet, e.g. wireless link may drop or foreign agent might crash
- higher latency possible due to buffering of data within the foreign agent and forwarding to a new foreign agent
- access point needs to be involved in security mechanisms (e.g. IPsec)
Snooping TCP – Principle

„Transparent“ extension of TCP within the foreign agent

- buffering of packets sent to the mobile host
- lost packets on the wireless link (both directions!) will be retransmitted immediately by the mobile host or foreign agent, respectively (so called “local” retransmission)
- the foreign agent therefore “snoops” the packet flow and recognizes acknowledgements in both directions, it also filters ACKs
- changes of TCP only within the foreign agent
Snooping TCP

Data transfer to the mobile host
- FA buffers data until it receives ACK of the MH, FA detects packet loss via duplicated ACKs or timeout
- fast retransmission possible, transparent for the fixed network

Data transfer from the mobile host
- FA detects packet loss on the wireless link via sequence numbers, FA answers directly with a NACK to the MH
- MH can now retransmit data with only a very short delay

Integration of the link layer
- link layer often has similar mechanisms to those of TCP

Problems
- snooping TCP does not isolate the wireless link as good as I-TCP
- snooping might be useless depending on encryption schemes, e.g. does not work with IPsec due to encryption of IP payload (including TCP segment number)
Mobile TCP

Special handling of lengthy and/or frequent disconnections

M-TCP splits as I-TCP does

- unmodified TCP fixed network to supervisory host (SH)
- optimized TCP between SH and MH (no slow start)

Supervisory host (SH)

- no caching, no retransmission (different from Indirect-TCP)
- monitors all packets, if disconnection detected
  - set sender window size to 0
  - sender automatically goes into persistent mode
- old or new SH reopens the window (set to old size)

Advantages

- maintains semantics, supports disconnection, no buffer forwarding

Disadvantages

- loss on wireless link propagated into fixed network (no buffering)
- adapted TCP on wireless link
Forced fast retransmit/fast recovery

Change of foreign agent often results in packet loss
- TCP reacts with slow-start although there is no congestion

Forced fast retransmit
- as soon as the mobile host has registered with a new foreign agent (Mobile IP), the MH sends duplicated acknowledgements on purpose
- this forces the fast retransmit mode at the communication partners (instead of slow start)
- additionally, the TCP on the MH is forced to continue sending with the actual window size and not to go into slow-start after registration

Advantage
- simple changes result in significant higher performance

Disadvantage
- focus on problems due to (fast) handover, not on temporarily poor wireless link quality
- mix of Mobile IP and TCP, no transparent approach
Transmission/timeout freezing

Mobile hosts can be disconnected for a longer time
- no packet exchange possible, e.g.,
  - discontinued communication in a tunnel
  - disconnection due to overloaded cells
  - preemption by higher priority traffic (scheduling)
- TCP disconnects after time-out completely

TCP freezing
- PHY/MAC layer is often able to detect interruption in advance
- PHY/MAC can inform TCP layer of upcoming loss of connection
- TCP stops sending, but does now not assume a congested link
- PHY/MAC layer signals again if reconnected

Advantage
- scheme is independent of data

Disadvantage
- TCP on mobile host has to be changed
- mechanism depends on lower layers
Selective retransmission

TCP acknowledgements are often cumulative

- ACK \( n \) acknowledges correct and in-sequence receipt of packets up to \( n \)
- if single packets are missing quite often a whole packet sequence beginning at the gap has to be retransmitted (go-back-n), thus wasting bandwidth

Selective retransmission as one solution

- RFC2018 allows for acknowledgements of single packets, not only acknowledgements of in-sequence packet streams without gaps
- sender can now retransmit only the missing packets
- mechanism is supported by newer TCP implementations

Advantage

- much higher efficiency

Disadvantage

- more complex software in a receiver, more buffers needed at the receiver
Transaction-oriented TCP

TCP phases
- connection setup, data transmission, connection release
- Using 3-way-handshake needs 3 packets for setup and release, respectively
- Thus, even short messages need a minimum of 7 packets!

Transaction oriented TCP
- RFC1644, T-TCP, describes a TCP version to avoid this overhead
- Connection setup, data transfer and connection release can be combined
- Thus, only 2 or 3 packets are needed

Advantage
- Efficiency for TCP short connections

Disadvantage
- Requires changed TCP on all hosts
Comparison of different approaches for a “mobile” TCP

<table>
<thead>
<tr>
<th>Approach</th>
<th>Mechanism</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indirect TCP</td>
<td>splits TCP connection into two connections</td>
<td>isolation of wireless link, simple</td>
<td>loss of TCP semantics, higher latency at handover</td>
</tr>
<tr>
<td>Snooping TCP</td>
<td>“snoops” data and acknowledgements, local retransmission</td>
<td>transparent for end-to-end connection, MAC integration possible</td>
<td>problematic with encryption, bad isolation of wireless link</td>
</tr>
<tr>
<td>M-TCP</td>
<td>splits TCP connection, chokes sender via window size</td>
<td>Maintains end-to-end semantics, handles long term and frequent disconnections</td>
<td>Bad isolation of wireless link, processing overhead due to bandwidth management</td>
</tr>
<tr>
<td>Fast retransmit/</td>
<td>avoids slow-start after roaming</td>
<td>simple and efficient</td>
<td>mixed layers, not transparent</td>
</tr>
<tr>
<td>fast recovery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transmission/time-out</td>
<td>freezes TCP state at disconnect, resumes after reconnection</td>
<td>independent of content or encryption, works for longer interrupts</td>
<td>changes in TCP required, MAC dependant</td>
</tr>
<tr>
<td>freezing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selective retransmission</td>
<td>retransmit only lost data</td>
<td>very efficient</td>
<td>slightly more complex receiver software, more buffer needed</td>
</tr>
<tr>
<td>Transaction oriented TCP</td>
<td>combine connection setup/release and data transmission</td>
<td>Efficient for certain applications</td>
<td>changes in TCP required, not transparent</td>
</tr>
</tbody>
</table>
References

Jochen Schiller: Mobile Communications (German and English), Addison-Wesley, 2000 (chapter 10 provides an overview on different approaches)


Selected papers on TCP over wireless:

