Passive Monitoring of RTT spikes

Jorma Kilpi

VTT Information Technology

P.O.Box 1202, 02044 VTT, FINLAND

Email: Jorma.Kilpi@vtt.fi

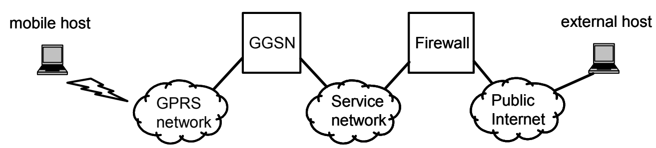
(Joint work with Pasi Lassila form HUT)

5th November 2004

Problem formulation:

- Sudden increase in the RTT of a TCP/IP connection is called *an RTT spike*.
  - Such a phenomenom may exist specially in mobile/wireless environments.
  - Spikes are *not congestion related* but, instead, *are unpredictable for the TCP*.
  - In mobile connections *mobility management* and *mobile routing* are possible sources of such spikes.
  - In wireless connections *the distance to the base station* is one possible source of spikes.
- In the worst case the RTT estimation algorithm of TCP gets confused, spurious timeouts and unnecessary retransmissions cause loss of goodput for TCP. (A *Strict* RTT spike.)
- Even without a retransmission the goodput of TCP gets worse. (A [non-strict] RTT spike.)
- What is the statistical significance of this phenomenom and how to monitor it passively? Statistical inference and modeling?

Available measurements: (In GSM/GPRS, EDGE/GPRS)



- All down- and upstream traffic of a TCP flow goes through the same GGSN.
- Time difference of a data segment and the corresponding ACK, measured at GGSN is called 'half' RTT.

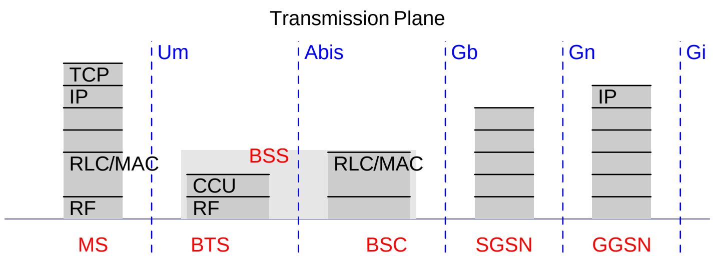
- There are four different 'half' RTT cases:

End host	**Mobile**	**Internet**
Client	RTT_{MC}	RTT_{IC}
Server	RTT_{MS}	RTT_{IS}

- We concentrate for RTT_{MC}.

Approach/methodology used: (For statistical inference and modeling purposes)

- We use *spectroscopy (Radon transform)* to find out link characteristics like the number of uplink *Packet Data Channels (PDCHs)* and *channel codings* used in the uplink.



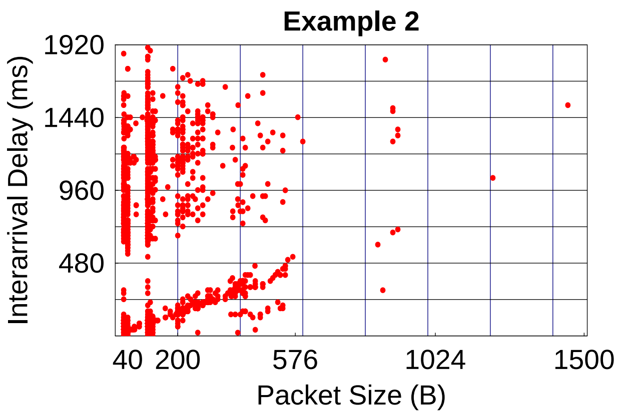
- Given packet sizes B_i and packet interarrival times d_i, measured at the GGSN, we analyze empirical probabilities

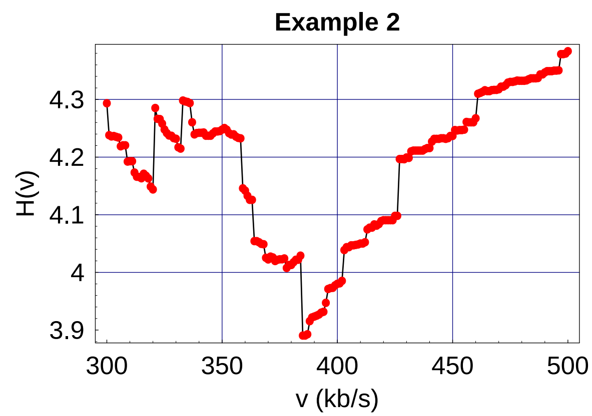
$$p(B,d) = \frac{1}{n-1} \sum_{i=1}^{n-1} \mathbf{1}_{\{(B,d)=(B_{i+1},d_i)\}}.$$

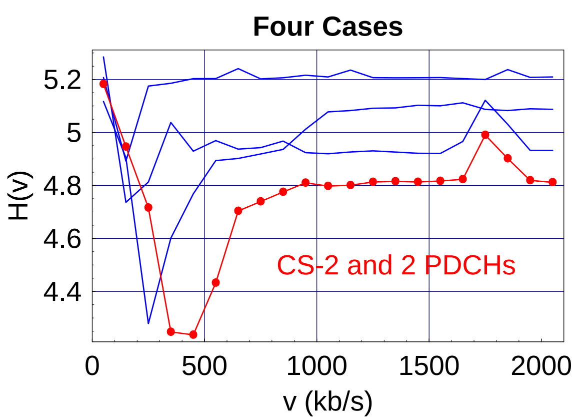
- The Radon transform in our case is defined as

$$p_R(r,v) = \sum_B p\left(B, r + \frac{18.5+20}{N_{PDCH}} N_{block}(B,CS) + \lceil (8 \times B)/v \rceil\right)$$

Preliminary results: (How the Radon transform method works)

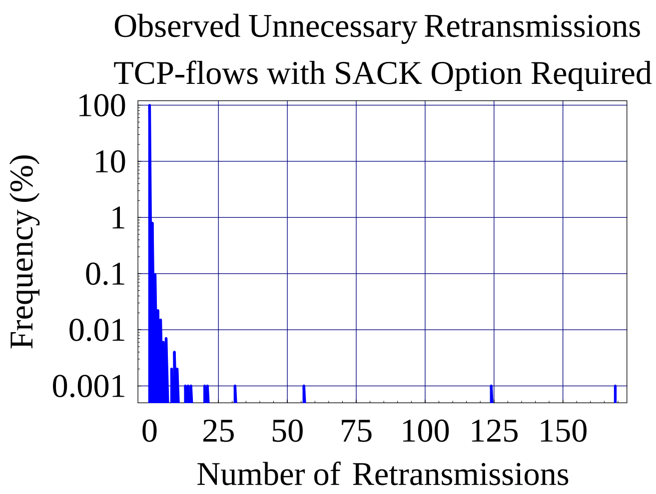


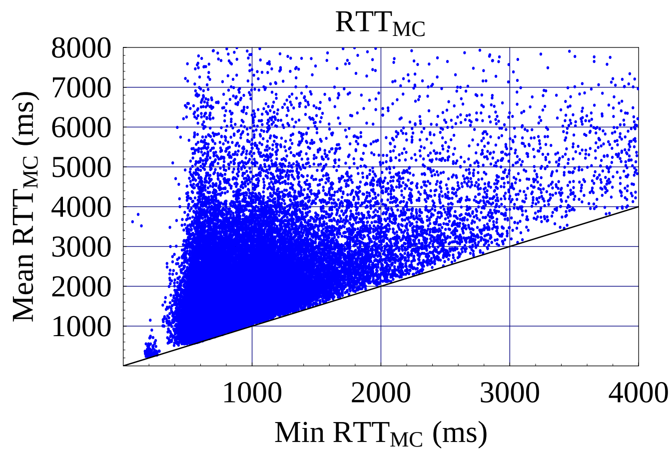


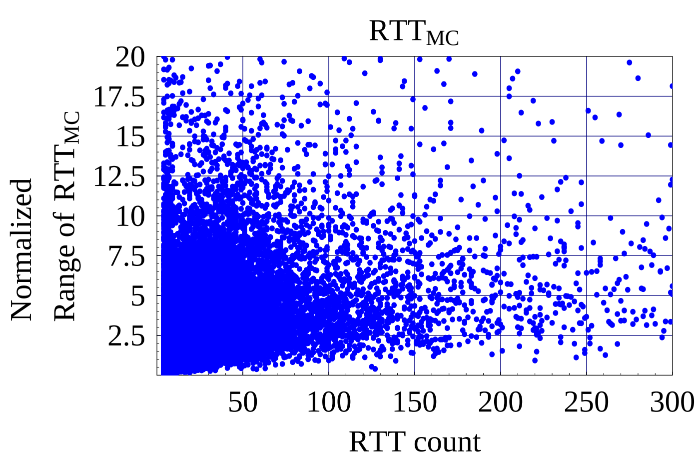


- This analysis is based *only* on time stamps and packet sizes.
- Minimum RTT_{MC}?
- According to link characteristics we try to group TCP flows into *as homogeneous classes as possible*.

Validation: (That RTT spikes can be observed passively)







- It seems that at least 1% of TCP flows suffer from strict RTT spikes.
- There is some evidence that non-strict RTT spikes are much more common.
- An RTT spike for an EDGE user is not a spike for a GSM user!

Next steps: (Work in progress and still in early phases)

- Practical issues:
  - How to get high-quality mobile TCP data. (Subscribers' privacy, operator's business secrets, organization, technical issues *et cetera*).
- Theoretical issues:
  - Statistically useful definition of an RTT spike? (Unnecessary retransmission may identify a strict RTT spike but non-strict spikes are also of importance.)
  - Homogeneous classes according to the link characteristics would be required for more fruitful statistical inference!
  - Robustness and minimum requirements of the Radon transform method?