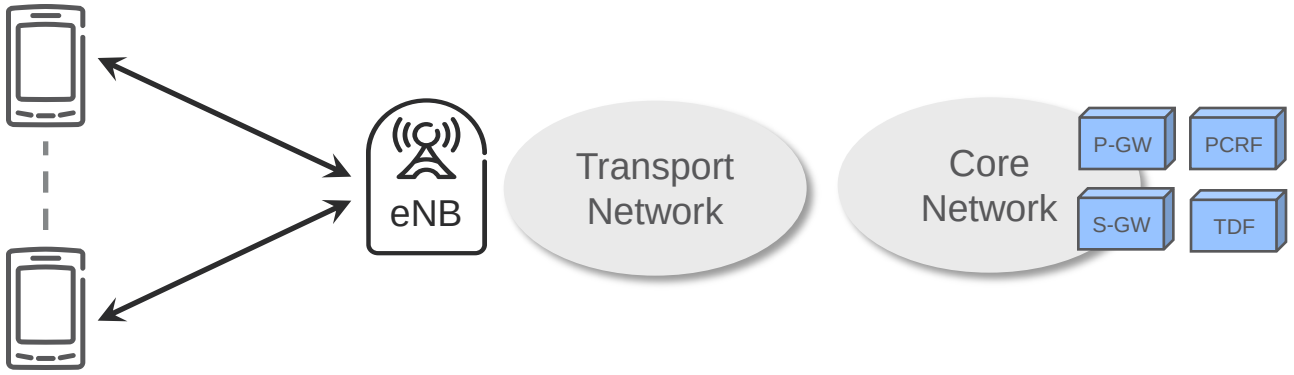
An open loop congestion management scheme in mobile networks

András Zahemszky (Ericsson Research, Sweden)
Joint work with György Miklós (Ericsson Research, Hungary)

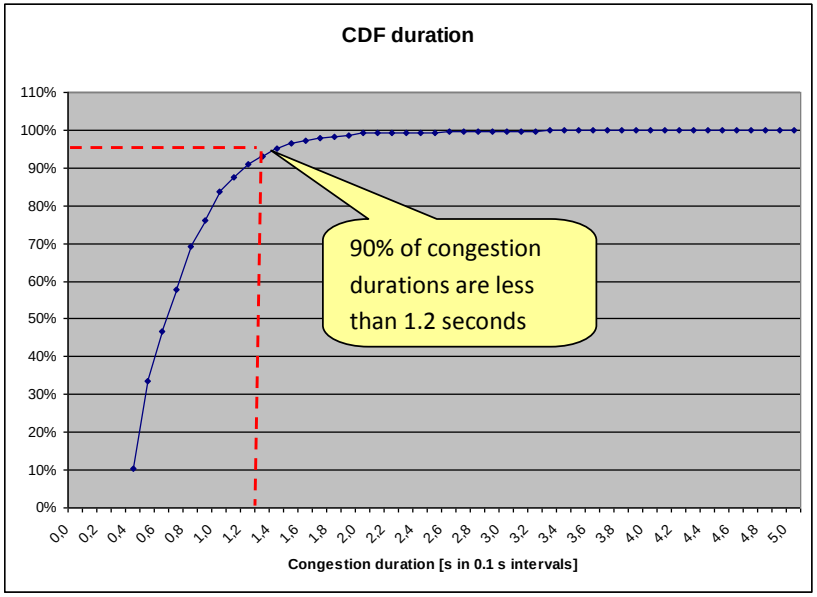
COST ACROSS Workshop, 14 February, 2014

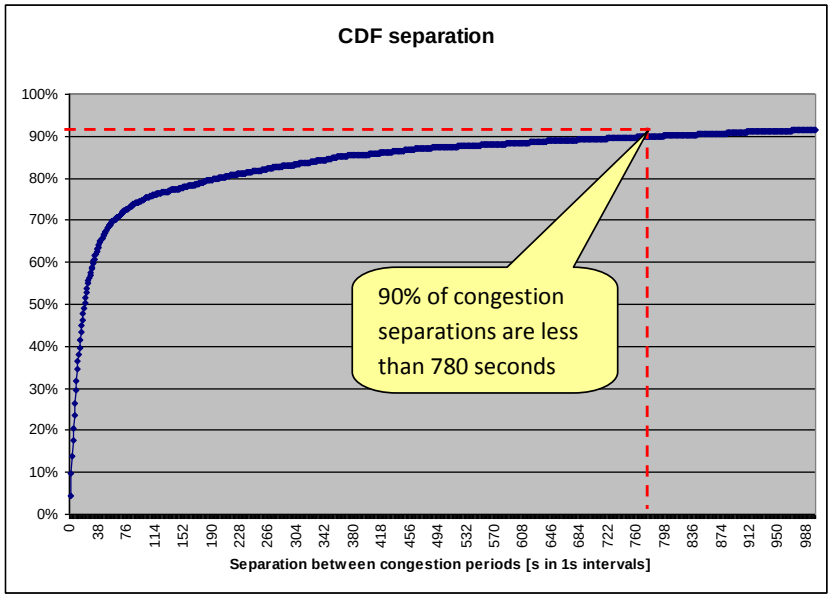
Congestion management in mobile networks

- Congestion: the demand for (RAN) resources exceeds the available (RAN) capacity for a period of time
- Congestion may lead to decreased user experience
- Scope:
  - Make use of available resources efficiently in congestion
  - Handling of user plane traffic when RAN congestion occurs based on e.g.
    - Subscription of the user
    - Type of content
    - Application



Congestion characteristics in mobile networks





- Measured on a live network
- Congestion periods are short
- Congestion periods are relatively frequent

- Source of the figures: 3GPP TR 22.805

Quality of Service in mobile networks

- Based on bearers (association between the PDN GW and the UE)
- QCI (QoS Class Identifier) value determines required treatment (delays, guaranteed bitrate, priority.. etc)
- 9 standardized QCI values

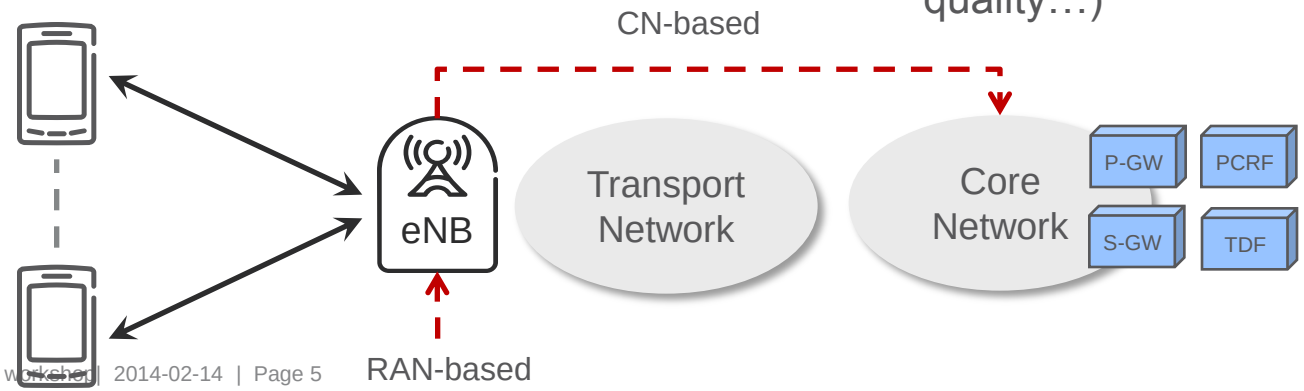
QCI	Resource type	Priority	Packet delay budget	Packet error loss rate
1	GBR	2	100 ms	10^{-2}
2	GBR	4	150 ms	10^{-3}
3	GBR	3	50 ms	10^{-3}
4	GBR	5	300 ms	10^{-6}
5	Non-GBR	1	100 ms	10^{-6}
6	Non-GBR	6	300 ms	10^{-6}
7	Non-GBR	7	100 ms	10^{-3}
8	Non-GBR	8	300 ms	10^{-6}
9	Non-GBR	9	300 ms	10^{-6}

From 3GPP TS 23.203

CN-based vs RAN-based congestion management

- **Core Network based (closed loop)**
  - Congestion information sent from RAN to CN
  - CN takes specific actions based on congestion feedback
- Challenges
  - Inherent delay in the feedback loop
  - Under-utilization might occur
  - Prone to potentially high signalling
- Advantages
  - Centralized management

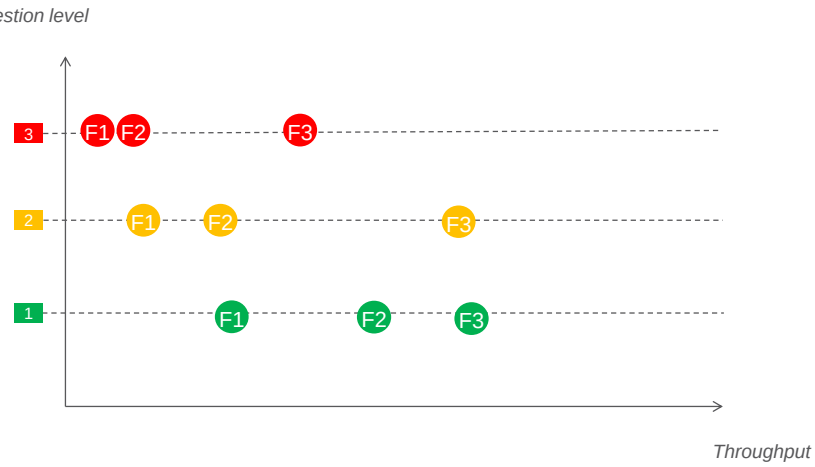
- **Radio Access Network Based (open loop)**
  - Core Network marks the packets to indicate preferred treatment
  - Congestion mitigation is done in the RAN
- Challenges
  - Packet marking
- Advantages
  - Congestion handled where it occurs
    - Fast reaction
    - Radio scheduler has the full picture (demands, channel quality…)



FQI – A RAN-based (open loop) congestion management scheme

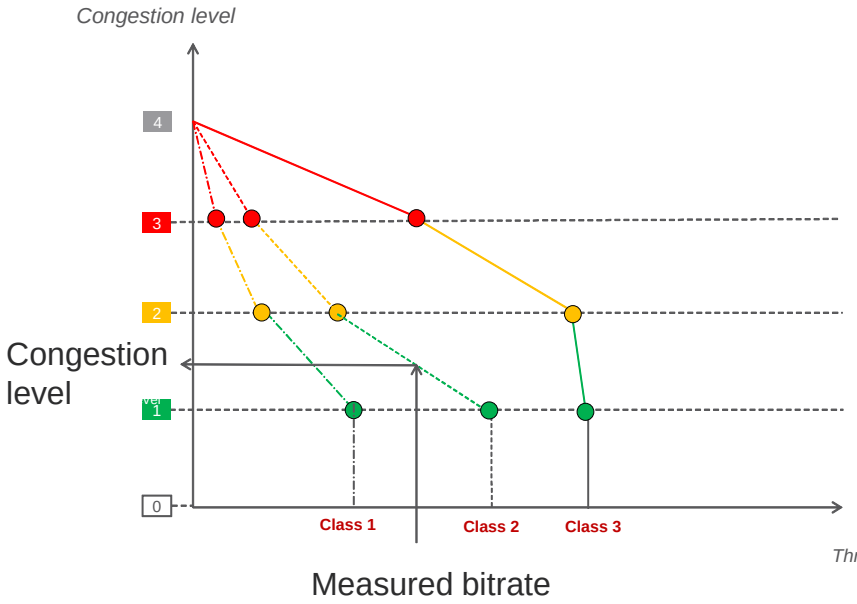
- Downlink data packets are marked with **FQI – Flow QoS Index**
- FQI points to a specific **congestion handling behaviour**
  - E.g. bitrate in a **given congestion level** (e.g. x kbps in “low” congestion, y kbps in “high” congestion)
- QCI and FQI jointly define the treatment of the flow in congestion situations

Throughput targets for flows F1, F2 and F3 in congestion levels 1,2 and 3:



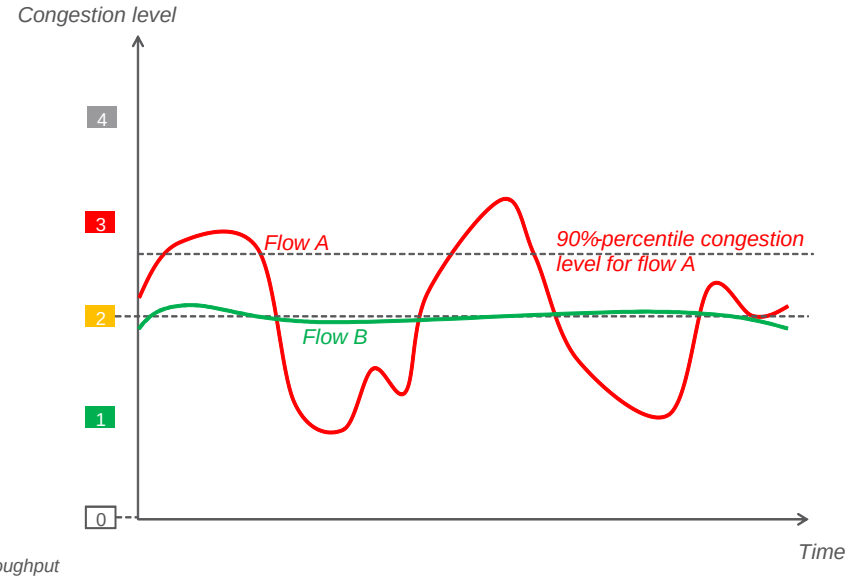
Evaluation - Metrics

Realized bitrate -> congestion level



- Bitrate is translated to congestion level based on the target charts

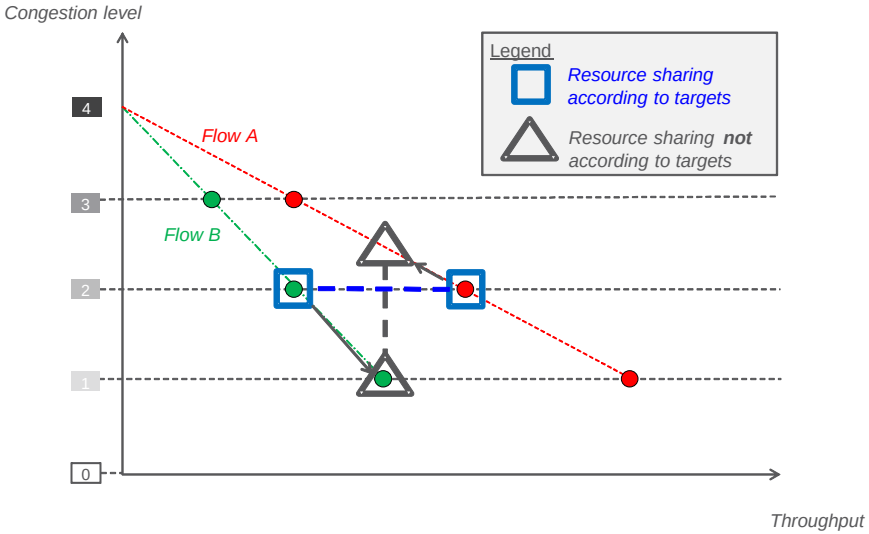
Average vs. 90% congestion level



- Flow A and Flow B has the same average congestion level over time
- 90% percentile congestion level better captures the congestion experience of the user

Evaluation - Metrics

Congestion level for different users

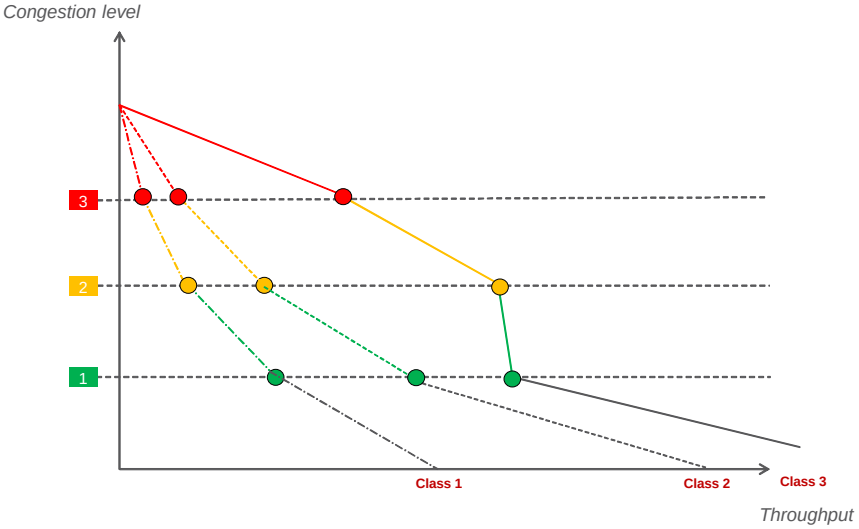


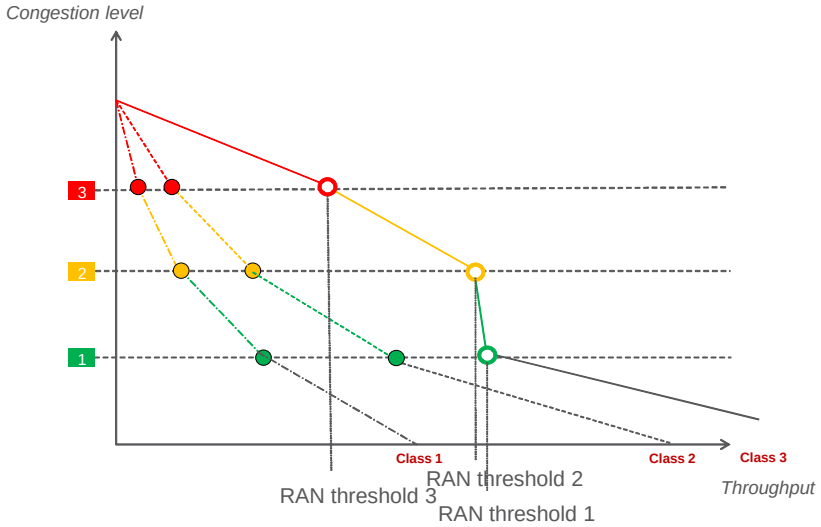
- Average congestion level for situation □ > average congestion level for situation △
- 90% congestion level weighted by the amount of traffic better captures the overall goodness of the resource distribution

- Congestion levels vary over time and between users
- Proposed metric
  - Measure the bitrate for each flow for 1 second
  - Determine the congestion level corresponding to the measured bitrate
  - Calculate the 90% percentile of all the samples, weighted by their bitrates

FQI vs a CN-based scheme

- Two schemes has to work with the same operator targets
- We are interested in how the schemes can follow operator targets

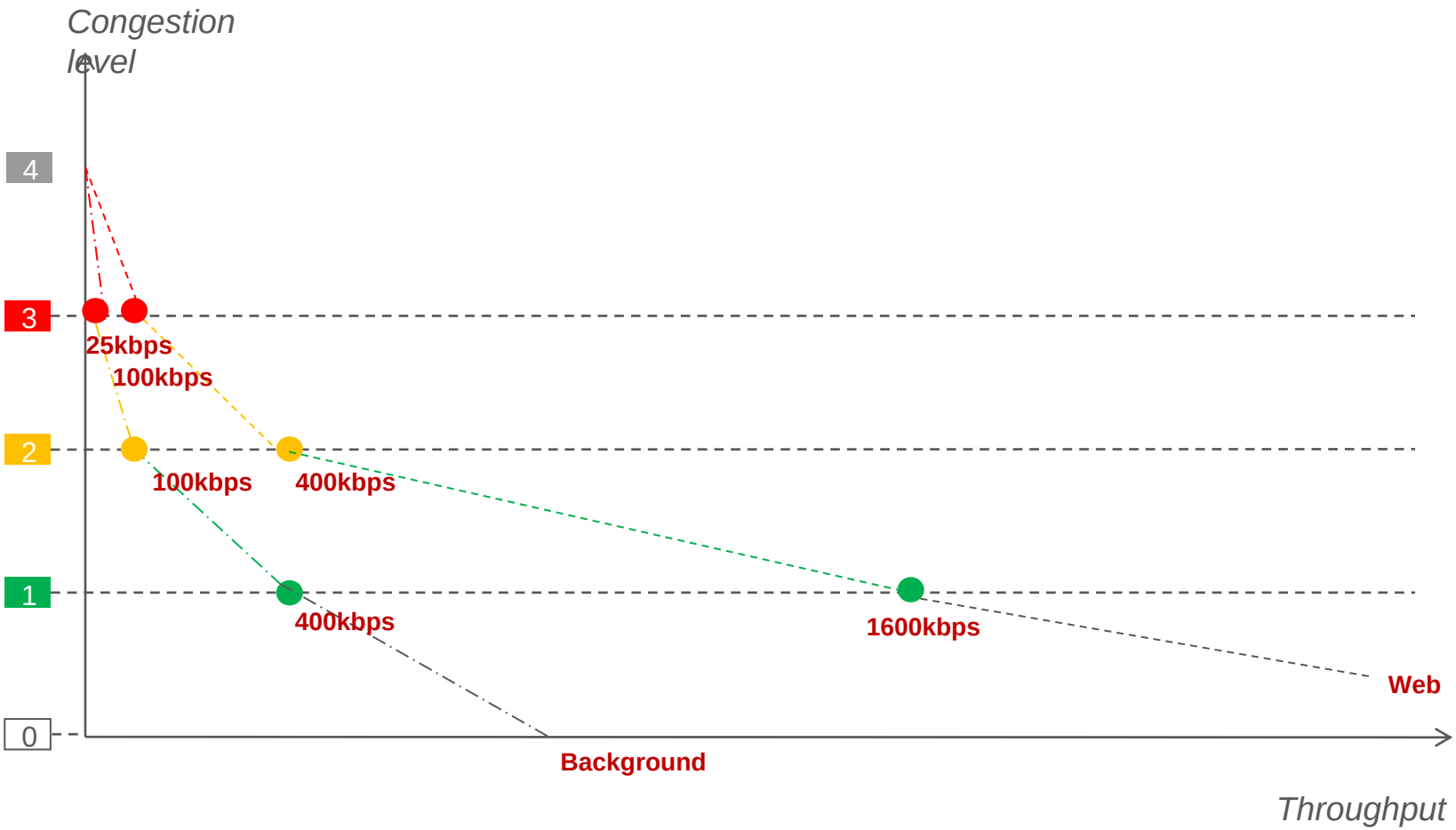




- FQI:
  - Allocate resources to fulfill the lowest congestion level possible
  - Distribute excess resources proportionally

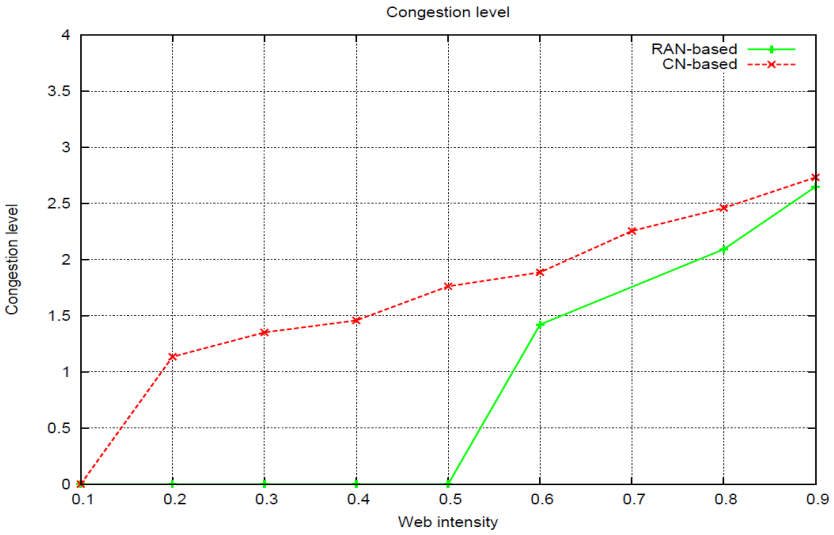
- CN-based scheme:
  - Report the achieved throughput in RAN per flow periodically (e.g. 60 seconds)
  - Use the report to determine overall congestion level
  - Shape the flows in the CN accordingly; all but most important class is shaped
  - RAN distributes resources equally

Example scenario - 1



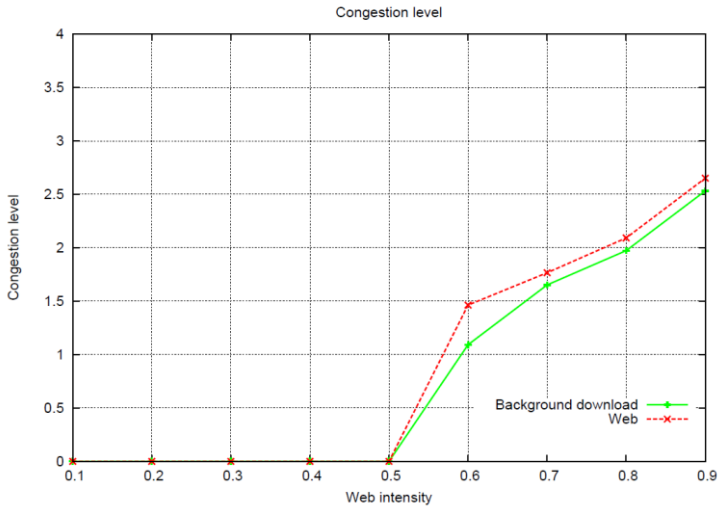
- Web and background FTP downloads
- Web should get significantly higher throughput in all situations

Results – web and background

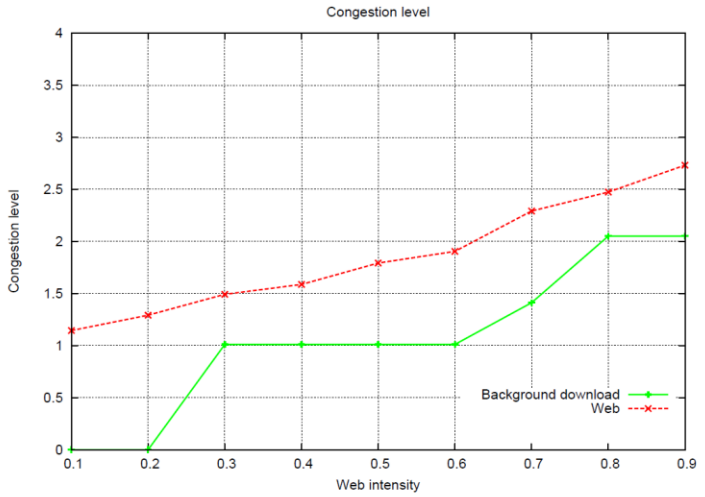


- Main parameters: 20 Mbps cell, 10 background users, web users (news portal) arriving by Poisson
- 90% congestion level is lower with RAN-based (FQI) solution
- FQI has similar congestion levels for both type of flows -> targets succesfully followed
- CN-based approach shows divergent congestion levels for different type of flows

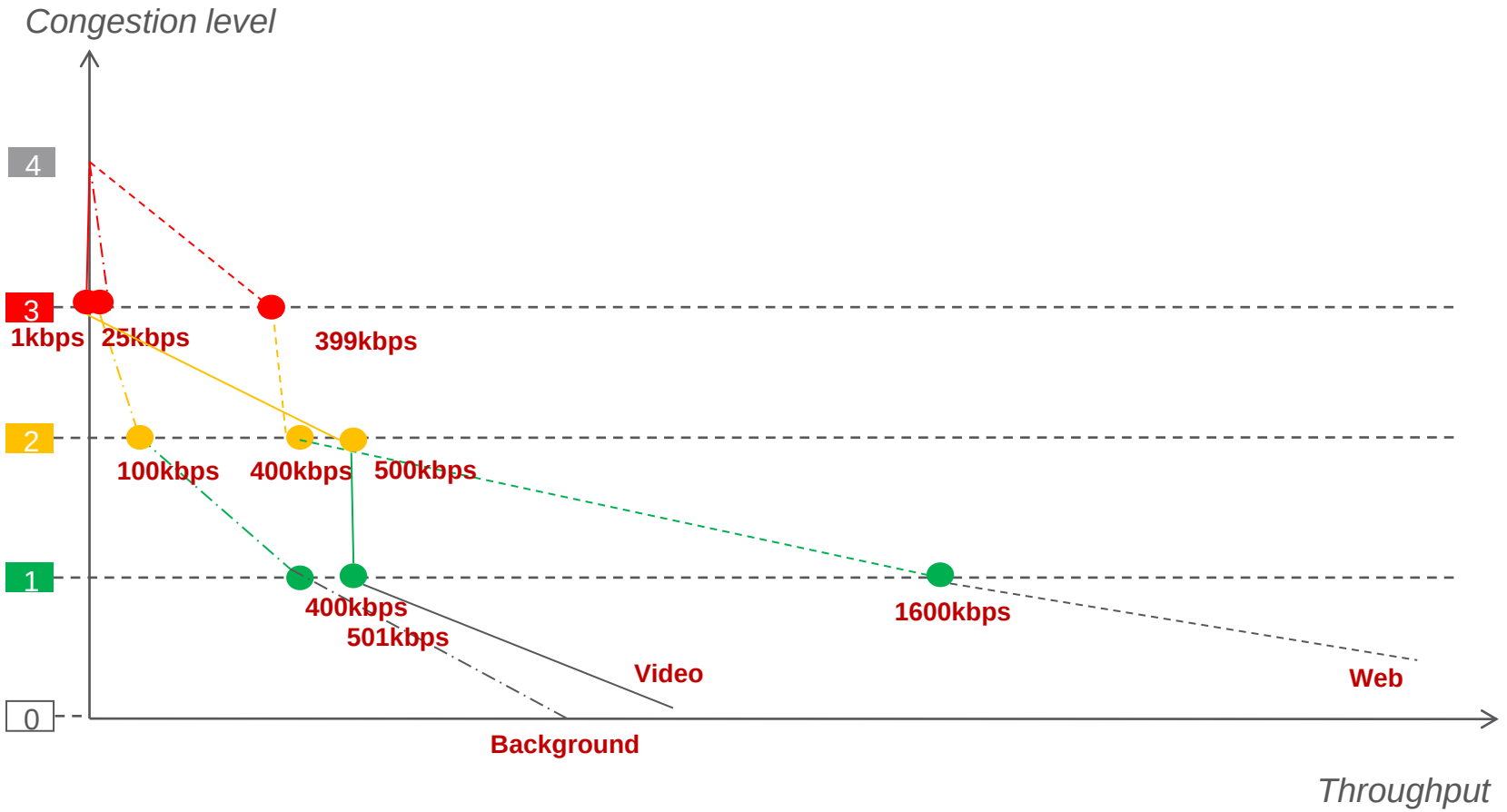
RAN-based (FQI)



CN-based

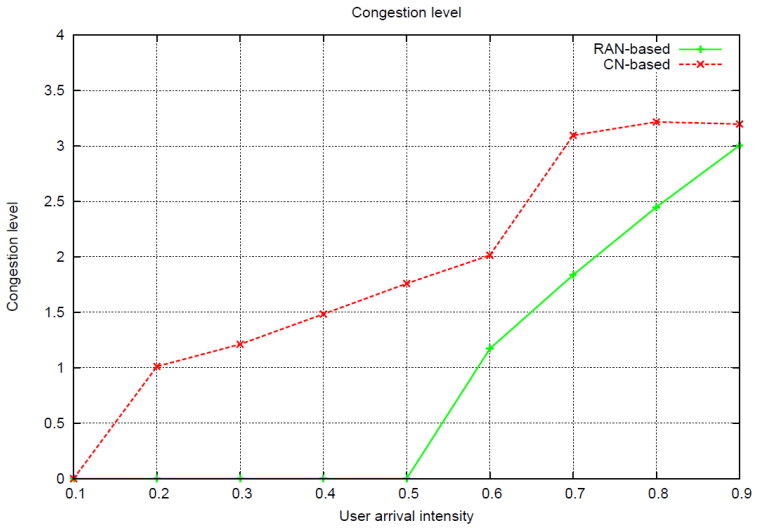


Example scenario - 2



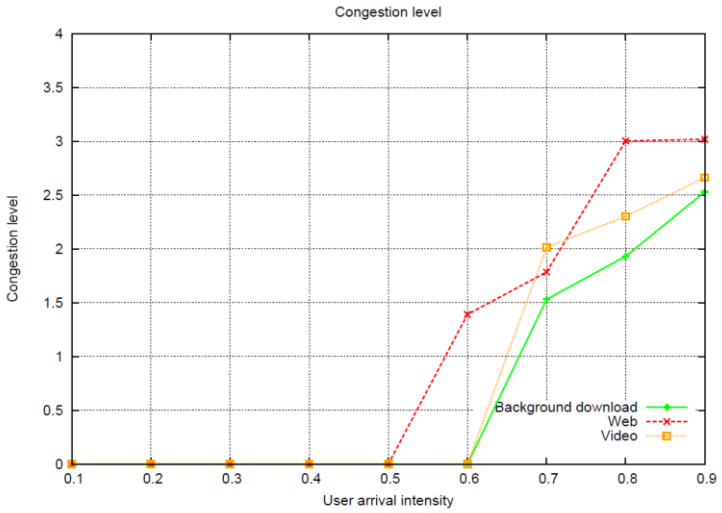
- Web, video and background FTP downloads
- At high congestion video is blocked; at medium congestion web is shaped to protect video flows; at low congestion video is shaped as it does not need higher throughput for continuous playout

Results – web, video and background

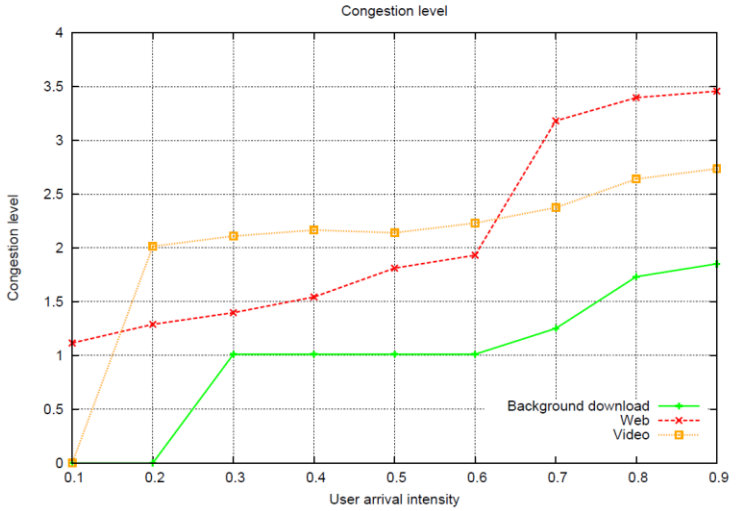


- Main parameters: 20 Mbps cell, 10 background users, 50% web users (news portal) and 50% video users arriving by Poisson; 400 kbit/s video in 12 ~200KB chunks
- 90% congestion level is lower with RAN-based (FQI) solution
- FQI has similar congestion levels for both type of flows -> targets succesfully followed
- CN-based approach shows divergent congestion levels for different type of flows

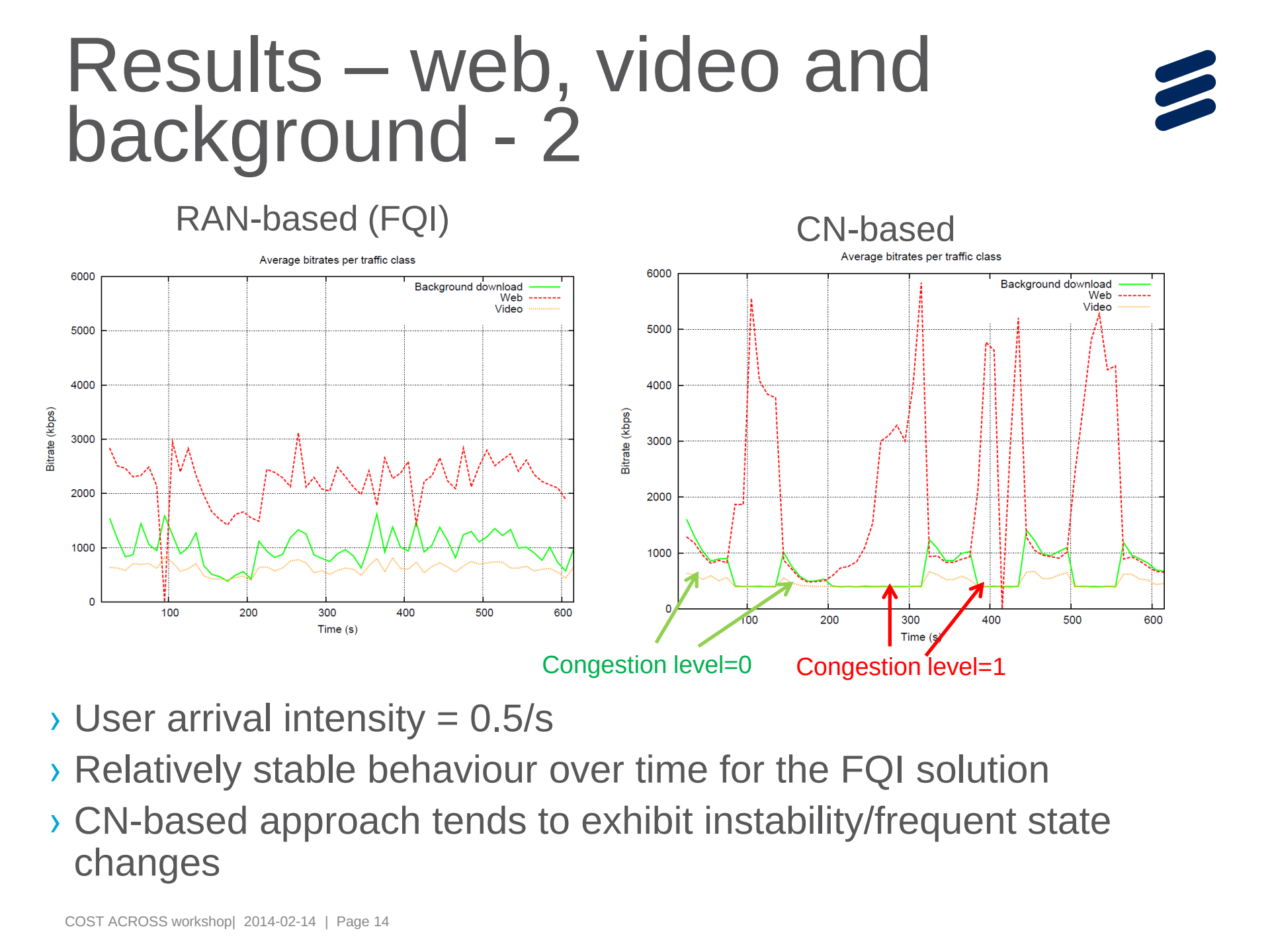
RAN-based (FQI)



CN-based



Results – web, video and background - 2



RAN-based (FQI)

CN-based

- User arrival intensity = 0.5/s
- Relatively stable behaviour over time for the FQI solution
- CN-based approach tends to exhibit instability/frequent state changes

Summary & conclusions

- FQI scheme extends QoS model and offers flexibility in defining desired behaviour in different congestion levels
- We provided a methodology how to evaluate congestion management schemes in mobile networks
- We illustrated the difference between the FQI solution and a CN-based approach with packet-level simulations
  - FQI based solution keeps the system in lower congestion level -> better user experience during congestion
  - FQI based solution can follow a wide range of operator targets

ERICSSON