9. QoS and QoE Optimization

Dr. David Soldani

(david.soldani@nokia.com, tel. +358.50.3633527)

S-38.3215 Special Course on Networking Technology for Ph.D. students at TKK

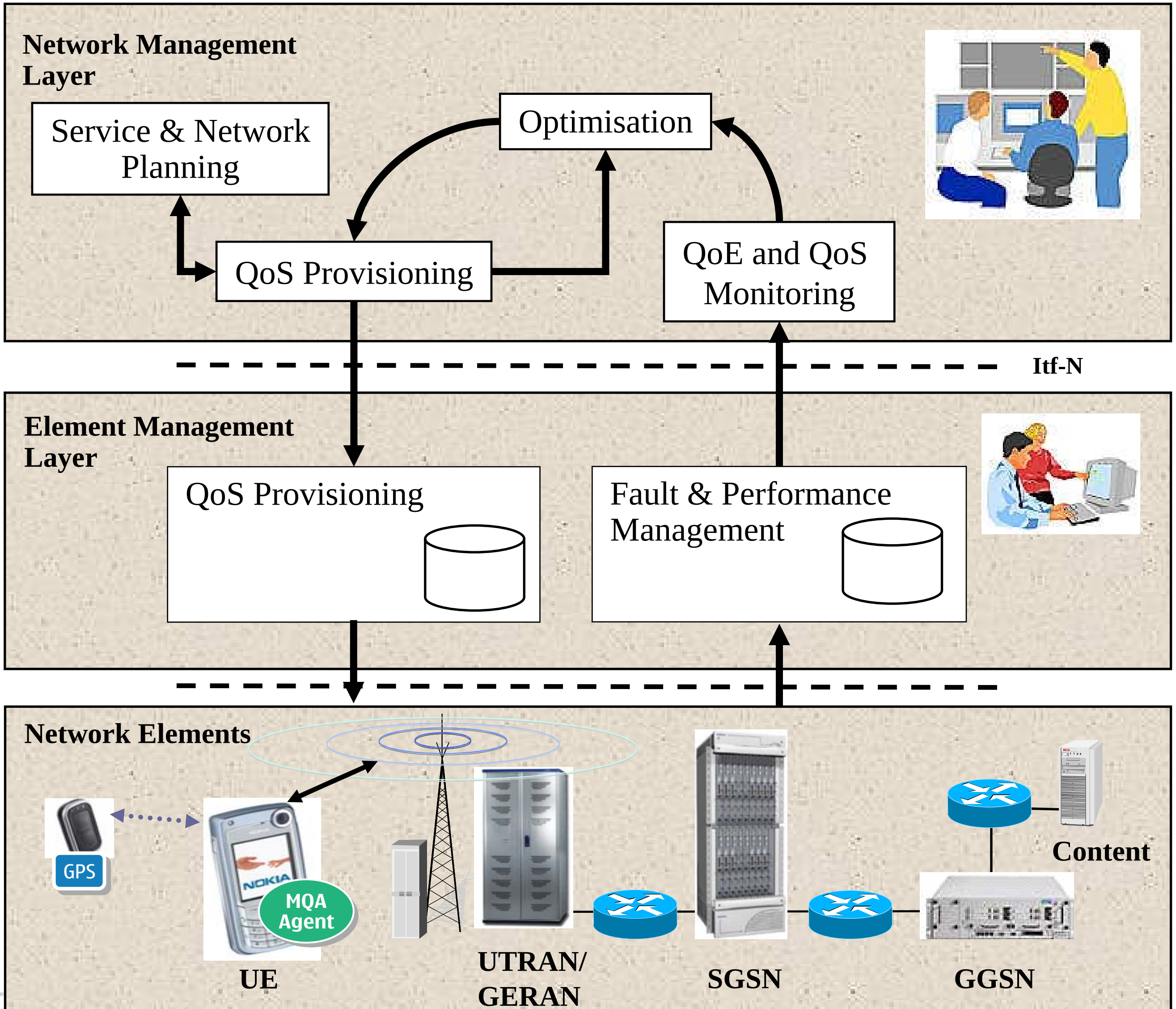
Outline

- Service optimization and architecture
- QoS optimization
  - Radio access networks
  - Core and backbone networks
- Service application performance improvement

QoS management conceptual breakdown

- QoS provisioning
  - Process of configuring and maintaining selected network elements based upon customer service level agreements (SLAs) and observed quality performance
- QoE and QoS monitoring
  - Process of collecting QoE and QoS performance statistics, faults and warnings; these data are then used for generating analysis reports for making changes/upgrades to the network
- QoE and QoS optimization
  - Process responsible for accessing monitored information, processing the data to determine service and network quality metrics, and initiating corrective actions when any of the quality levels is considered unsatisfactory

QoS and QoE optimization concept



Operator processes

Customer

Customer Interface Management Processes

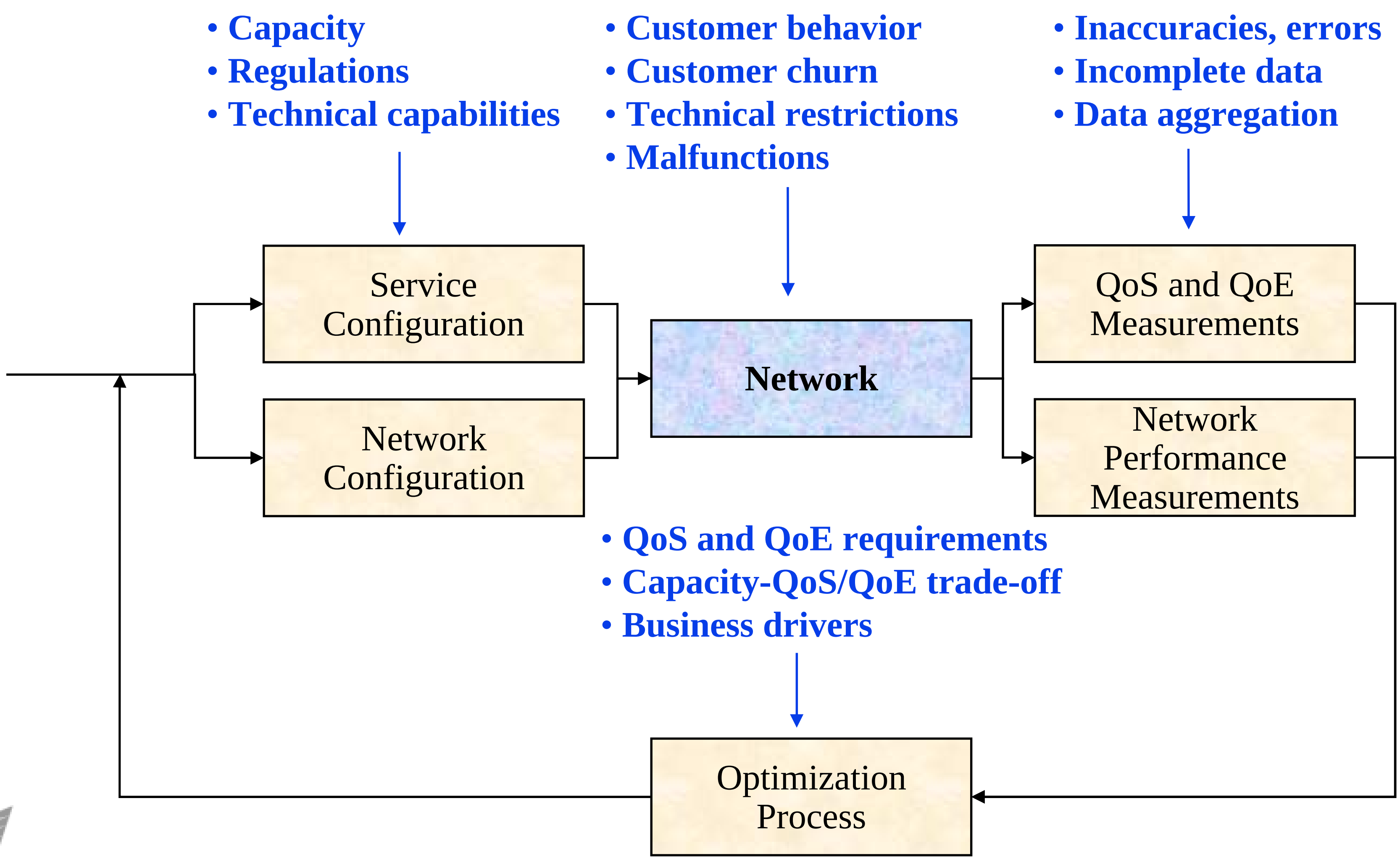
Customer Care Processes
- Sales
- Order handling
- Problem handling
- Customer QoS management
- Invoicing
- …

Service Development and Operations Processes
- Planning and development
- Configuration
- Problem management
- Quality management
- Rating and discounting
- …

Network and Systems Management Processes
- Planning and development
- Provisioning
- Inventory management
- Maintenance
- Data management
- …

Network Element Management Processes

Optimization loop and external factors



Performance metrics

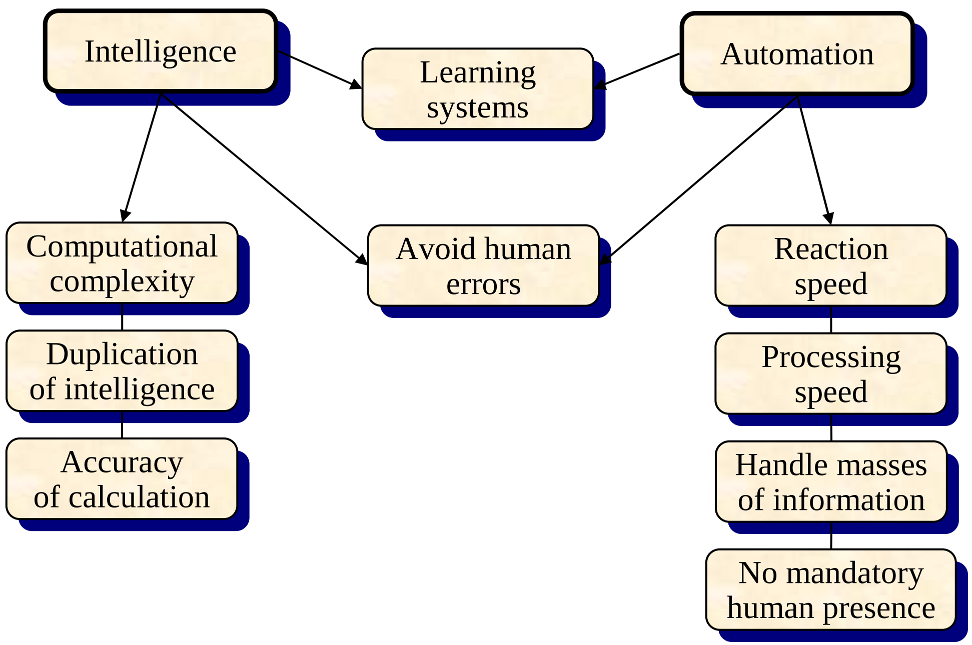
- NMS key performance indicators
- Detailed logging of protocols at different interfaces
- Detailed logging of element performance data
- Drive/walk tests
- Mobile and distributed QoS agents
- …

Optimization process ‘triggers’

- New technology or elements into use
- Changes due to external edge conditions
- Detection of performance deterioration
- Daily service and network operation process
- …

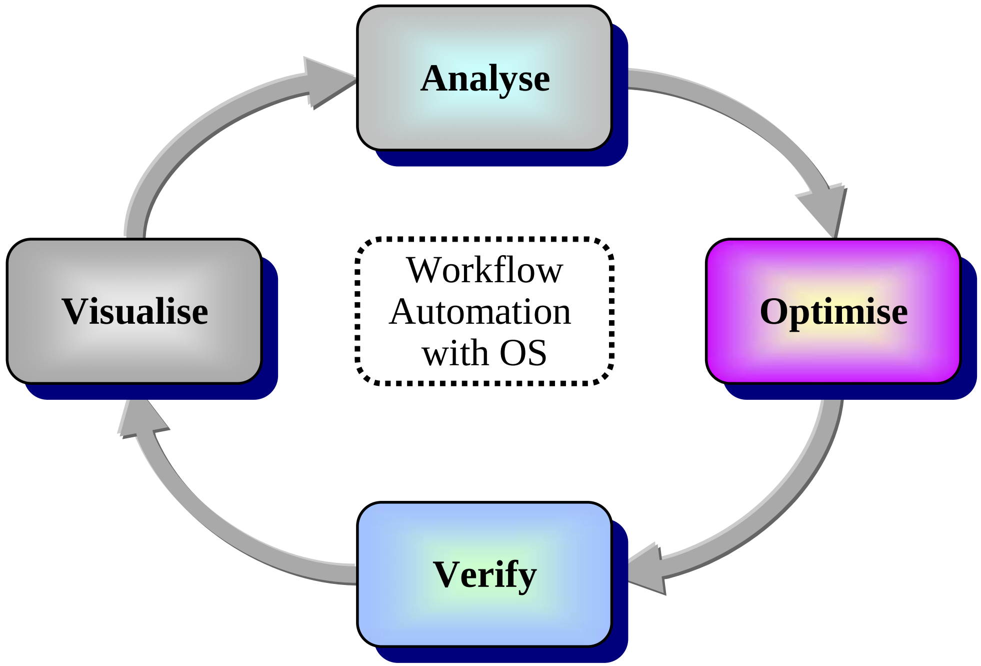
Automation and autonomic (intelligence)

- Automation
  - Reduce human-made repetitive and routine tasks
- Autonomic
  - Model expert decision-making and reasoning (knowledge)



Network wide optimization workflow

- Fast service creation, introduction and provisioning, and improved quality of service at a lower cost
- Proper workflow support and automation of customer care and operational support processes



An approach to QoS optimization

- Users (services) satisfaction criteria: **definition**
- Fitness (objective) function: **definition**
- Input traffic mixes and traffic volume: **selection**
- Crucial QoS parameters: **sensitivity analysis**
- Solution space search algorithm (genetic): **selection**

⇨ Trends and optimal parameter settings

⇨ Design of feasible method to QoS optimization

User satisfaction: definition

- **Speech calls and video calls (GB)**
  - The user does not get neither blocked nor dropped
- **SWIS (GB)**
  - The user does not get neither blocked nor dropped
  - No re-buffering occur during the session
- **PoC (NGB)**
  - The user does not get neither blocked nor dropped
  - No re-buffering occur during the session
- **Streaming (GB and NGB)**
  - The user does not get neither blocked nor dropped
  - No re-buffering occur during the session
- **Dialup (http, emails, ftp) (NGB)**
  - The user does not get neither blocked nor dropped
  - Active session throughput >= 64 kb/s
- **WAP (NGB)**
  - The user does not get neither blocked nor dropped
  - Active session throughput >= 32 kb/s
- **MMS (NGB)**
  - The user does not get neither blocked nor dropped
  - Active session throughput >= 8 kb/s

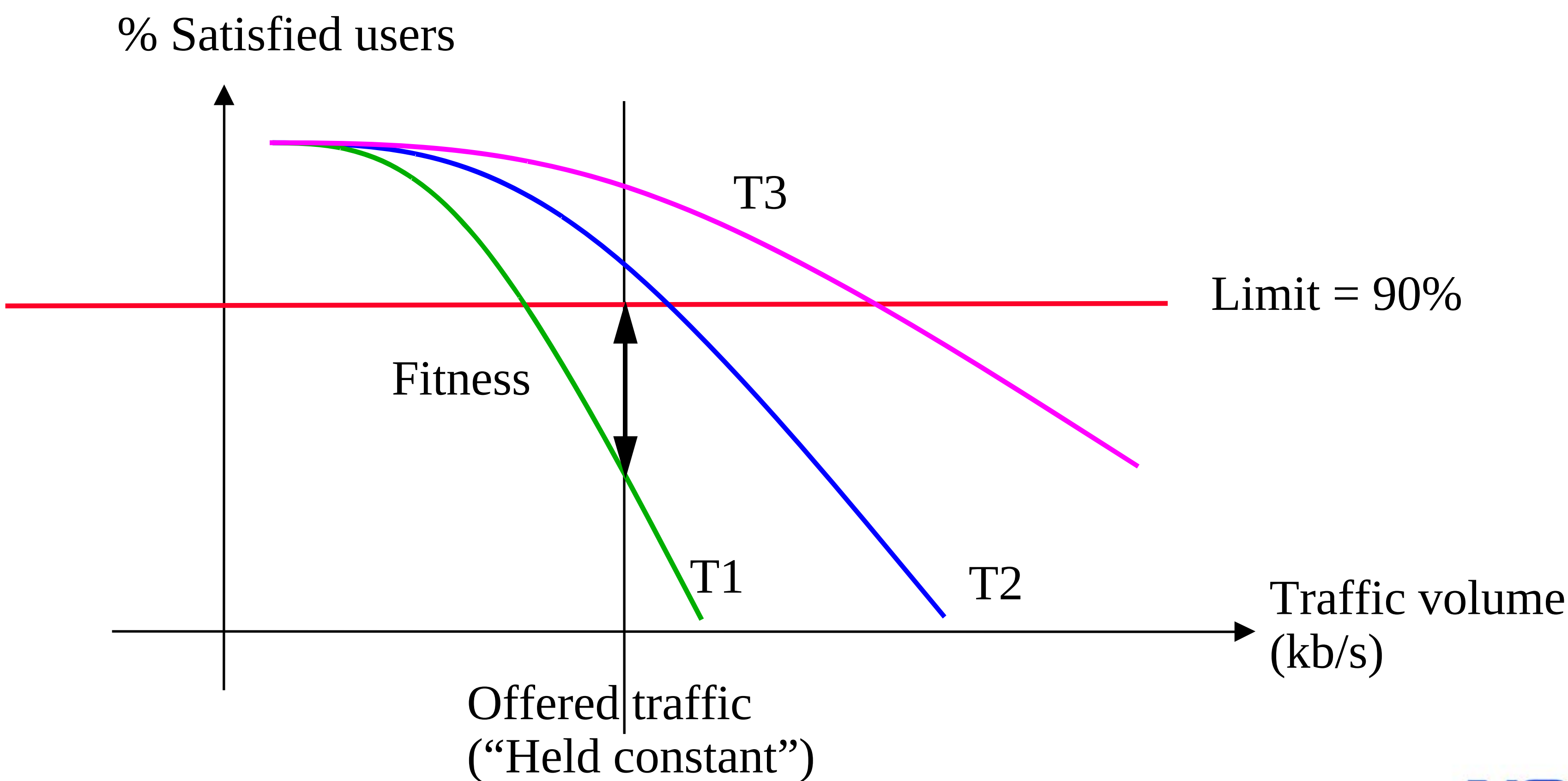
Allowed parameter values

- Population: 10 member vectors of 14 components

Parameter	THP	Range	Service
Minimum allowed bit rate	T1	8, 16 kb/s	PoC
	T2	64, 128 kb/s	Streaming
	T3	32, 64, 128, 144, 256, 384 kb/s	WAP/MMS
	T4	64, 128, 144, 256, 384 kb/s	Dialup
Maximum allowed bit rate	T1	8, 16 kb/s	PoC
	T2	64, 128 kb/s	Streaming
	T3	32, 64, 128, 144, 256, 384 kb/s	WAP/MMS
	T4	64, 128, 144, 256, 384 kb/s	Dialup
Inactivity timer	T1	1, 2, 5, 10, 20, 30 s	PoC
	T2	1, 2, 5, 10, 20, 30 s	Streaming
	T3	1, 2, 5, 10, 20, 30 s	WAP/MMS
	T4	1, 2, 5, 10, 20, 30 s	Dialup
Minimum allocation time	Not differentiated	1, 2, 5, 10, 15, 20 s	All services
Maximum queuing time	Not differentiated	1, 2, 5, 10, 15, 20 s	All services

Fitness function (1/2)

- Cost function of the genetic algorithm: distance actual percentage of satisfied users of the worst performing service and the corresponding target value



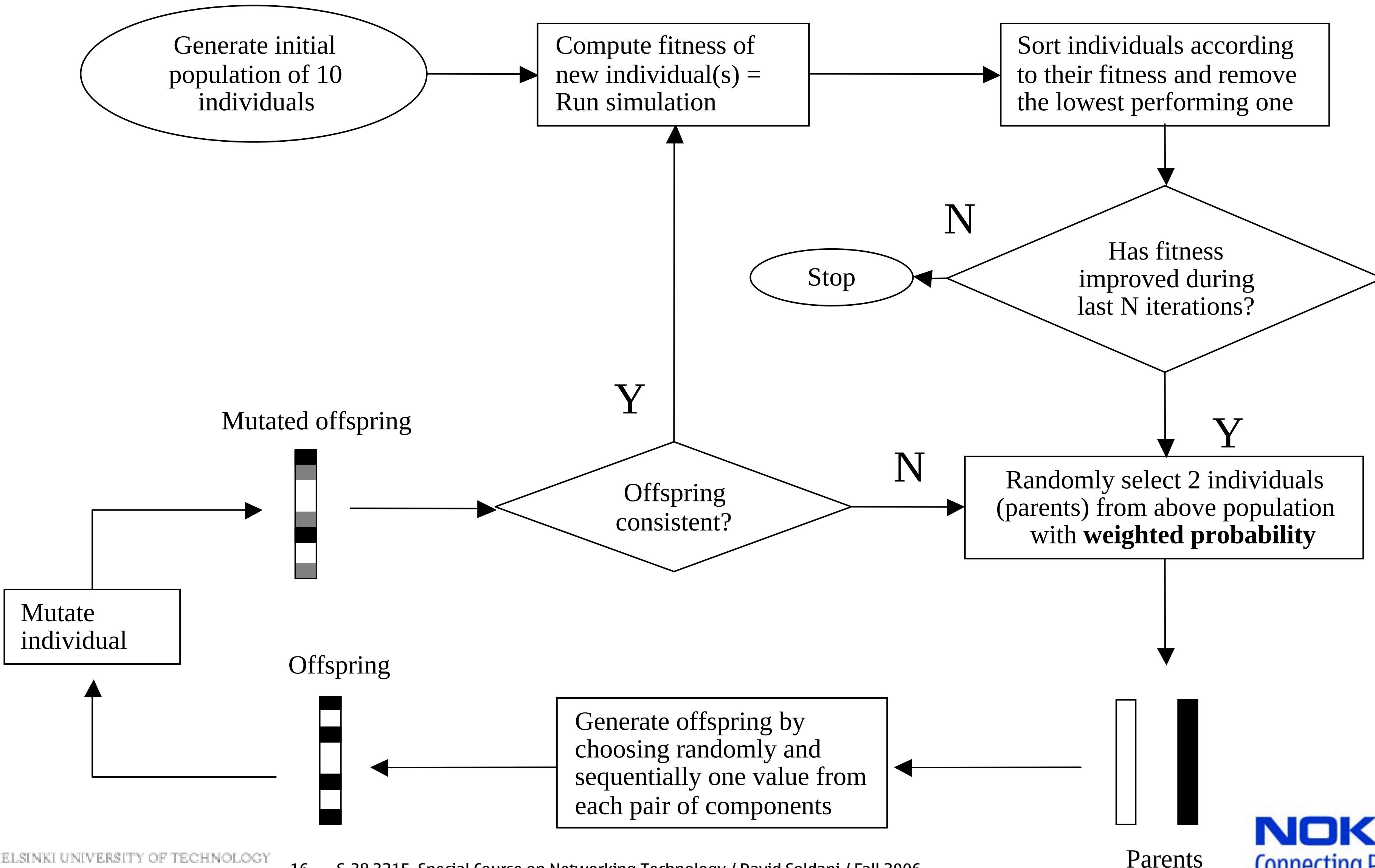
Fitness function (2/2)

- The lower the fitness value the better the network configuration
- **Fitness < 0** means that service performance is **above the target**
- **Fitness = 0** means that service performance is **on target**

$$Satisfaction(i) = \frac{\dfrac{SatisfiedUsers(i)}{max\{ServiceType(i),1\}} - TargetSatisfaction(i)}{\sqrt{\dfrac{TargetSatisfaction(i) \cdot \left(1 - TargetSatisfaction(i)\right)}{max\{ServiceType(i),1\}}}}$$

$$Fitness = -\min_{\{i=1,\ldots,N\}} \{Satisfaction(i)\}$$

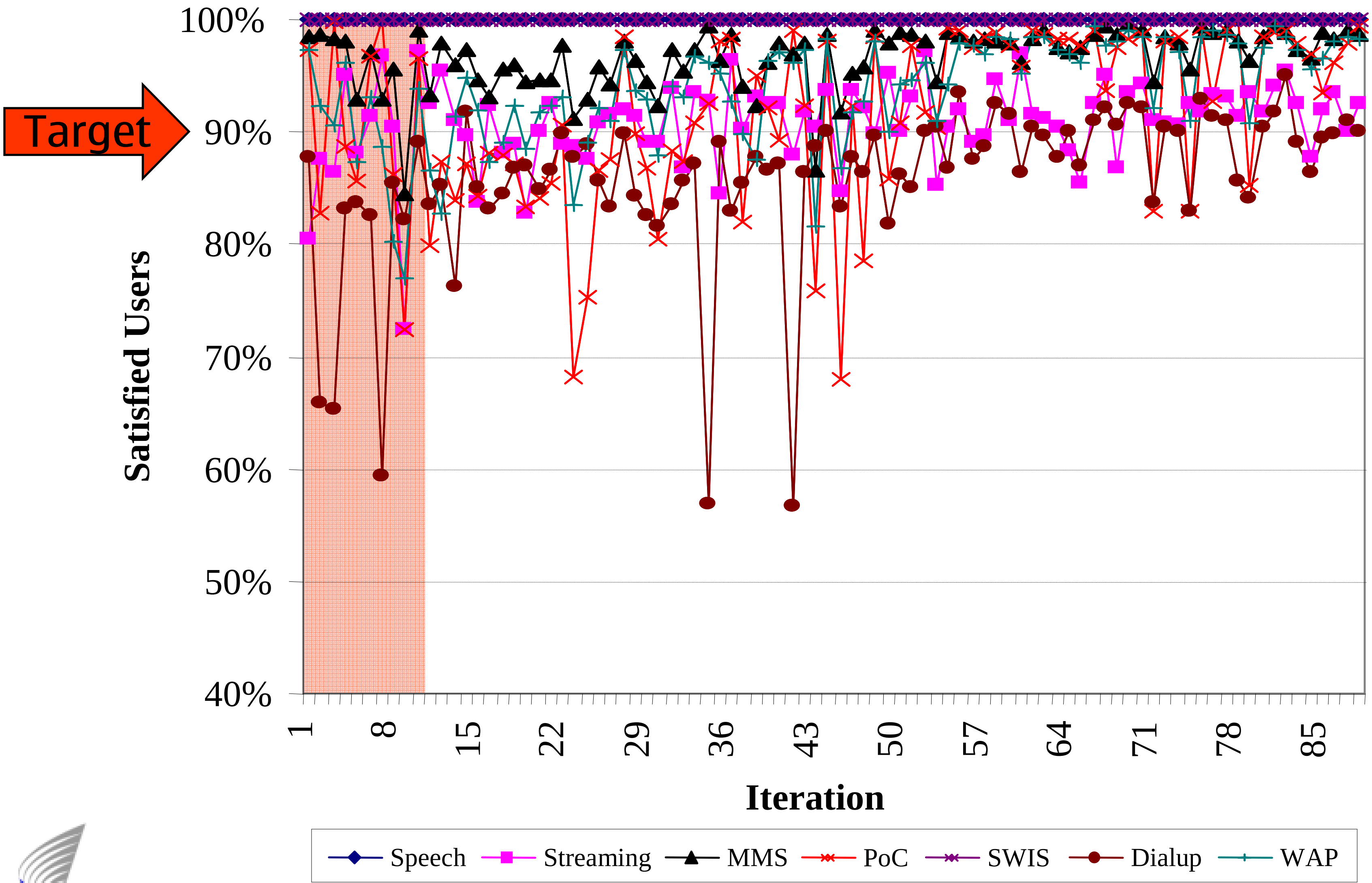
Genetic approach to QoS optimization



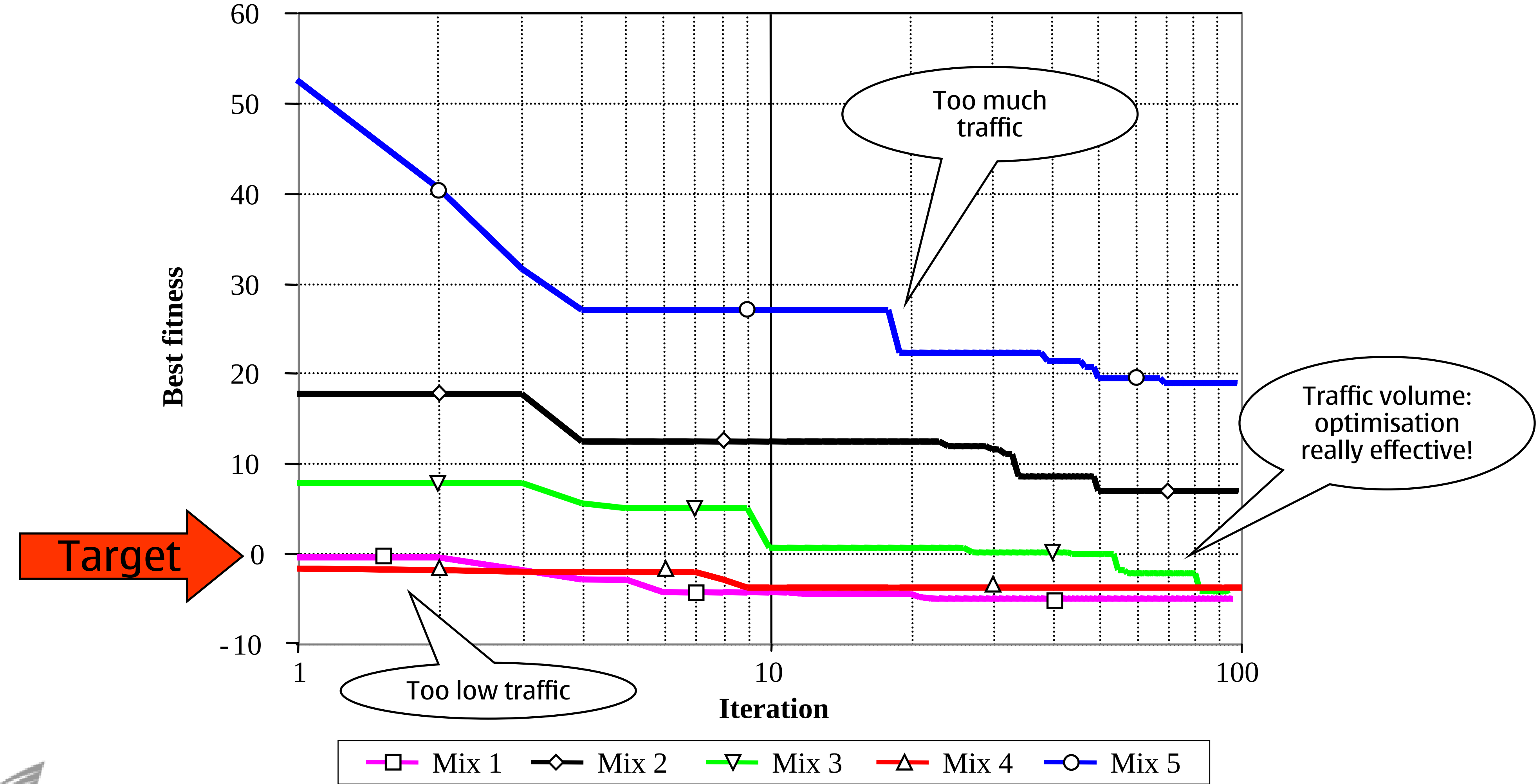
Traffic mixes

Service	Proportion of calls (%)				
	Mix 1	Mix 2	Mix 3	Mix 4	Mix 5
Speech	32	44	40	40	20
Video	0	3	0	2	7
SWIS	3	0	10	5	10
PoC	14	9	10	25	18
Streaming	2	10	10	4	12
WAP	38	17	10	10	13
MMS	8	5	10	7	5
Dialup	3	12	10	7	15

Mix 3: % of SU vs. iteration



Mix 1-5: best fitness vs. iteration number



Optimal parameter settings

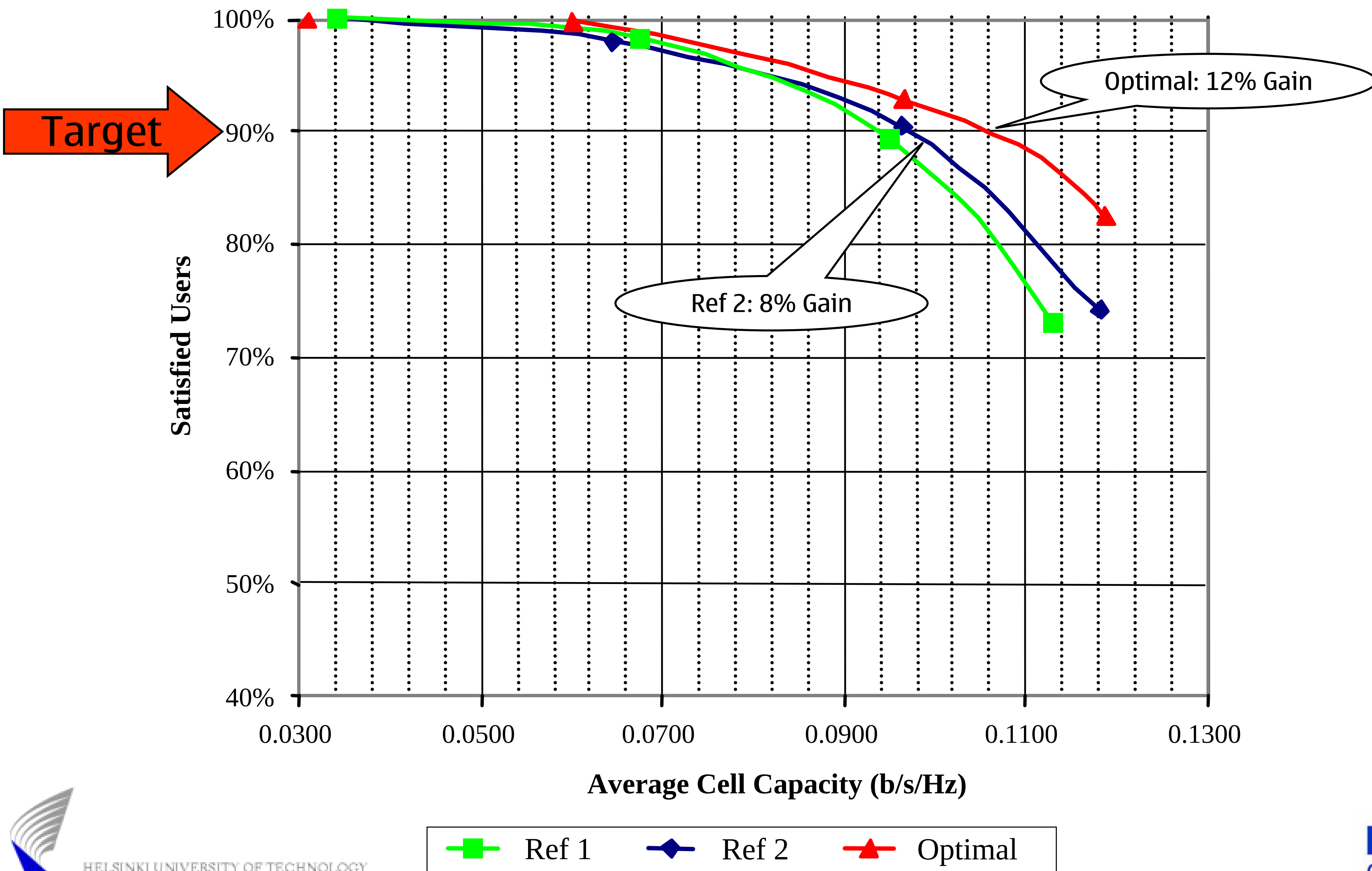
Mix	Min Allowed BR (kb/s)				Max allowed BR (kb/s)				Inactivity timer (s)				Min alloc. time (s)	Max queuing time (s)
	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4		
1	8	128	32	64	16	128	64	144	1	1	2	5	5	10
2	8	64	32	64	16	64	384	128	1	5	30	1	20	5
3	16	64	32	64	16	128	144	144	2	1	5	1	20	20
4	8	64	64	128	16	128	256	144	1	2	5	5	2	10
5	8	64	32	128	16	128	144	256	2	1	1	20	15	5
Ref 2	8	64	32	64	16	128	144	144	1	1	5	5	15	10
Ref 1	64	64	64	64	64	128	144	144	5	5	5	5	15	10

Ref 1:
Undifferentiated

Ref 2:
Differentiated

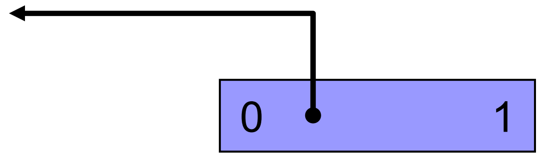
Optimal settings

Mix 4: performance improvements



Fuzzy optimization

- Definition of a delay control parameter (mapping function) for
  - Minimum allowed bit rate (*THPi*)
  - Maximum allowed bit rate (*THPi*)
  - Inactivity timer (*THPi*)
  - Minimum allocation time
  - Maximum queuing time



- The higher the parameters values the shorter the expected packet transfer delay for a specific THP

$$\text{Parameter}(DCP\_THPi) = (max - min)*(1 - DCP\_THPi) + min$$

DCP_THPi = delay control parameter of *THPi*

max = maximum value of the range

min = minimum value of the parameter range

- Small values of *DCP* are expected to produce shorter packet transfer delays than large values

Discrete control rule

FOR *i* = 1 To #*THP* {

Good sample

  IF $ST(i) \geq 5 / (1 - TS(i))$ THEN {

Poor

   IF *Satisfaction*(*i*) < –2 THEN decrease *DCP_THPi* by a step

   ELSE IF *Satisfaction*(*i*) > 2 THEN increase *DCP_THPi* by a step

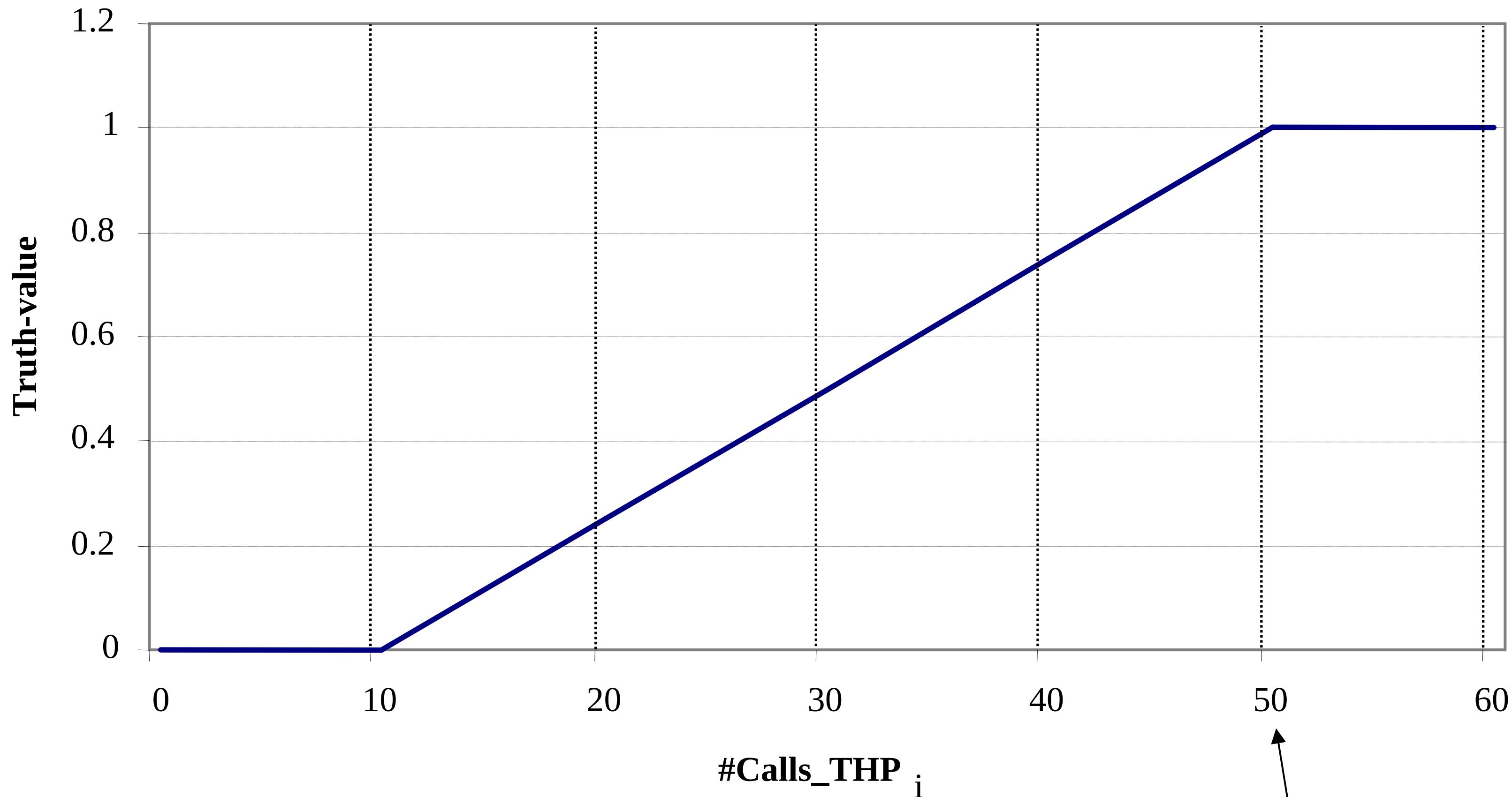
  }}

Excessive

ST(*i*) = Service Type *i* (# users of service type *i*)

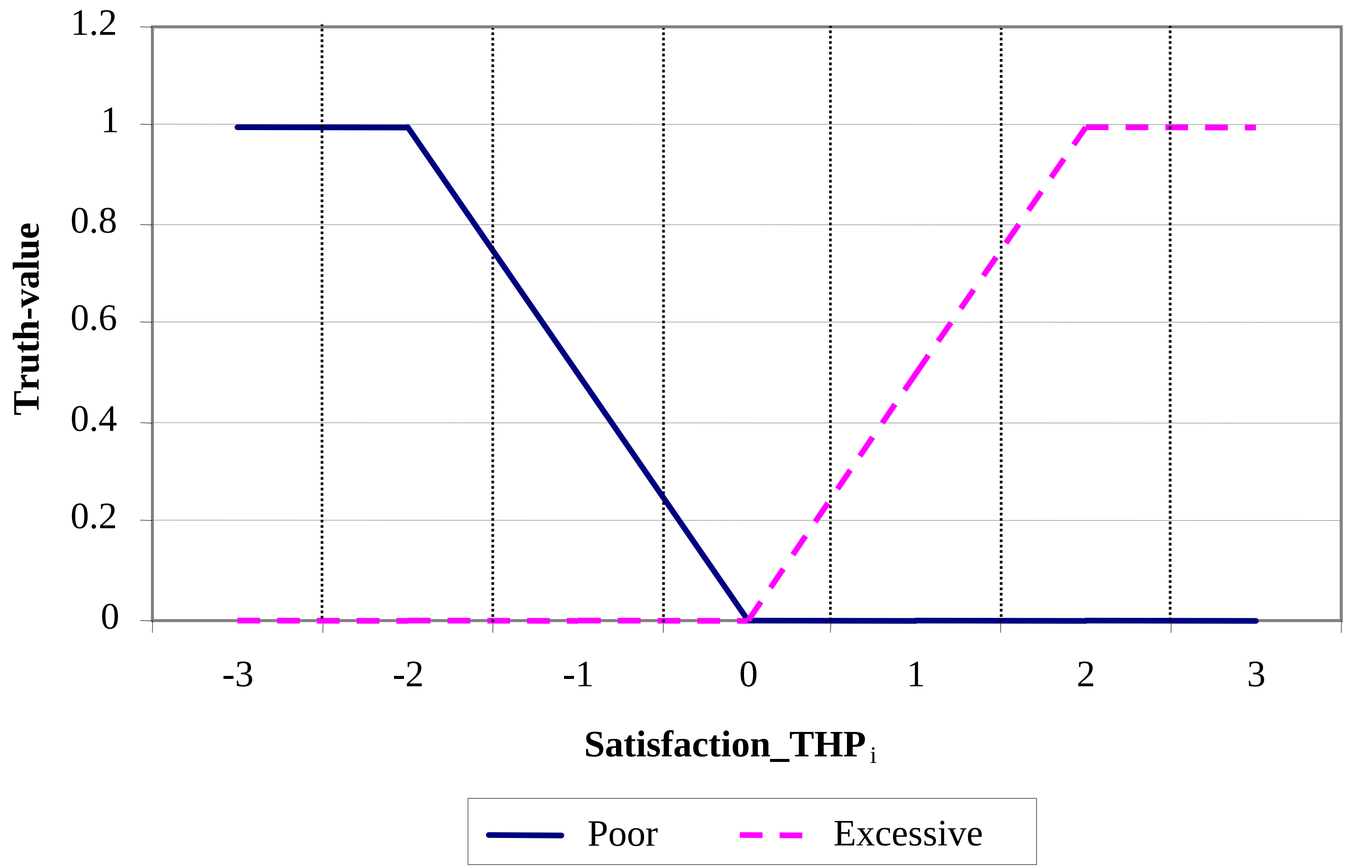
TS(*i*) = Target Satisfaction (90%)

Rule "fuzzyfication": ***Good Sample***



$ST(i) \geq 5 / (1 - TS(i))$

Rule "fuzzyfication": ***Poor, Excessive***



DCP_THPi vs. Fuzzy truth values

Assumption: **AND = Min (*X*,*Y*)** **OR = Max (*X*,*Y*)**

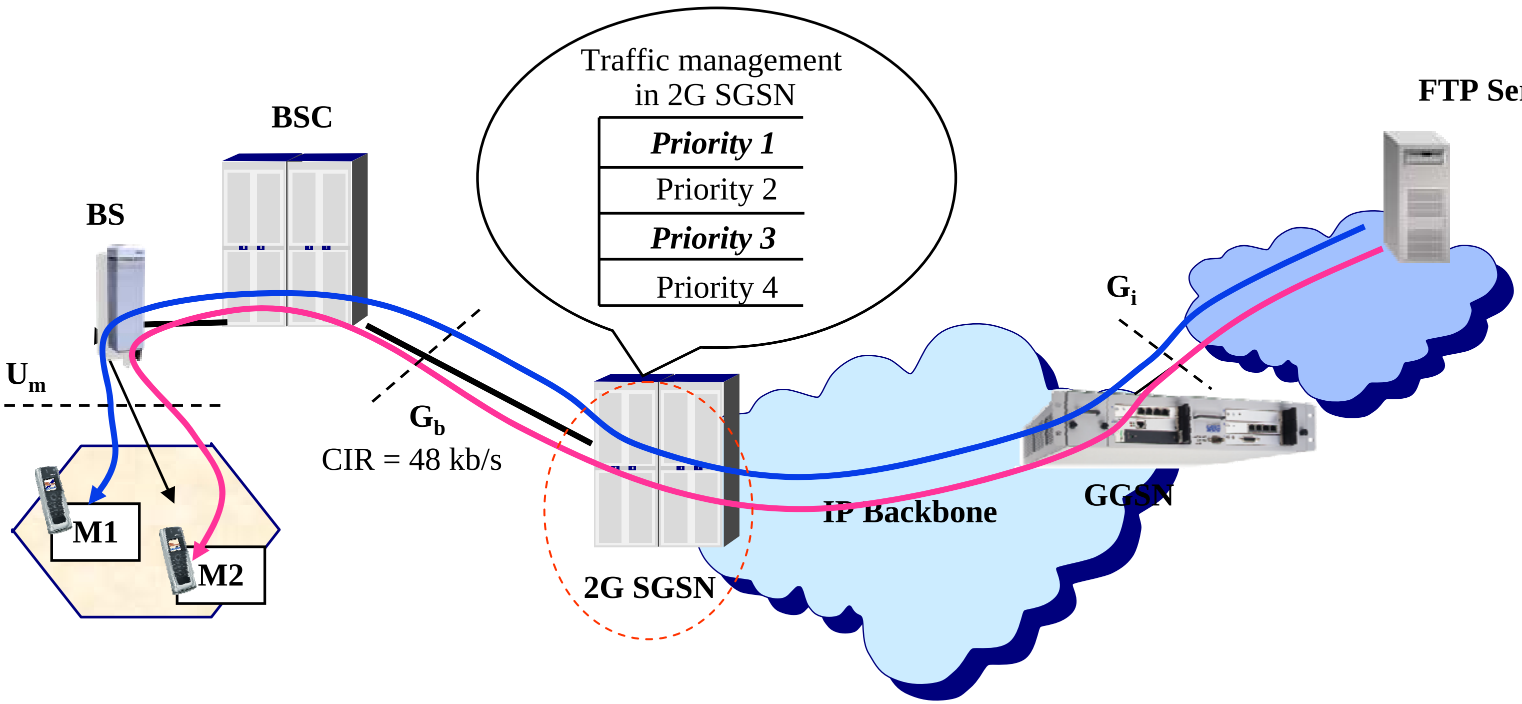
- *DownStep* = – 0.2 * MIN(GoodSample(*N*), Poor (*S*))
- *UpStep* = + 0.2 * MIN (GoodSample(*N*), Excessive(*S*))

⇒ **DCP_THP(*i*) = DCP_THP(*i*) + *UpStep* + *DownStep***

Fuzzy control can respond to small deviations from the target with small control steps, allowing faster solutions to satisfaction problems

Ex: Optimization in core network

- Four queues and congestion at Gb interface
- PVC with Committed Information Rate (CIR): 48 kb/s (42 kb/s)
  - M1 (4/2 TSL): FTP on THP1
  - M2 (3/1 TSL): FTP on THP3

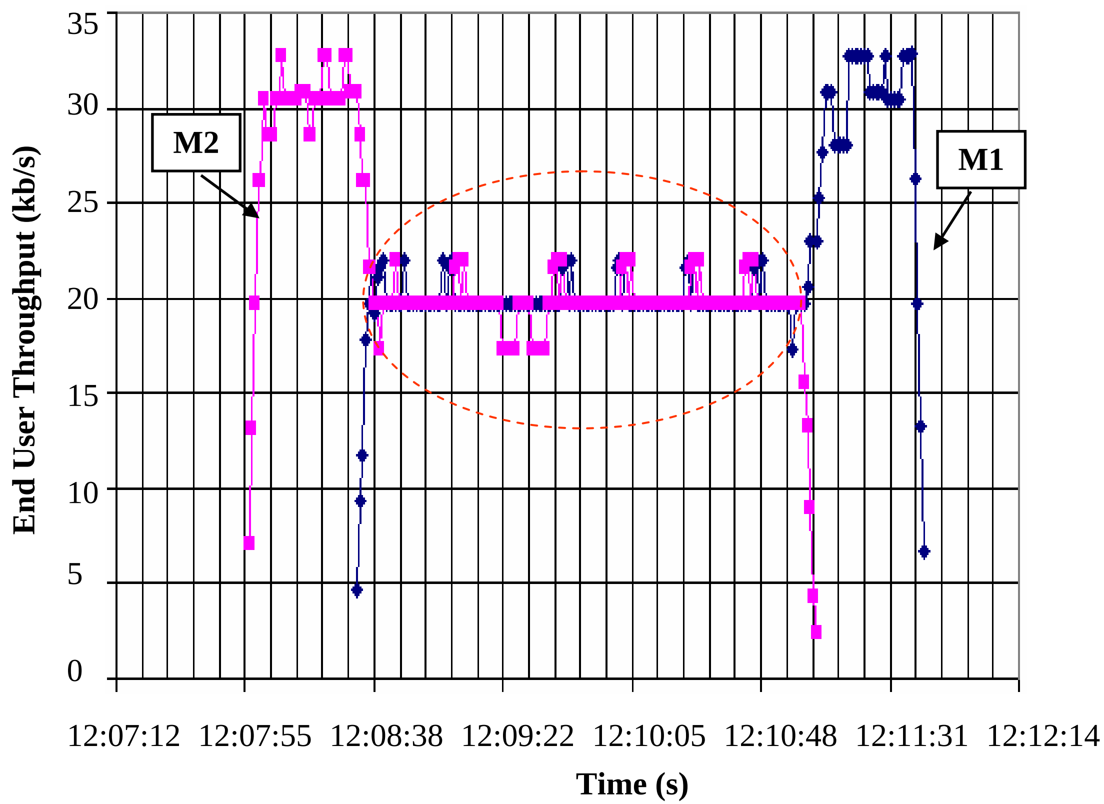


WFQ: weights assigned to priority queues

Queues

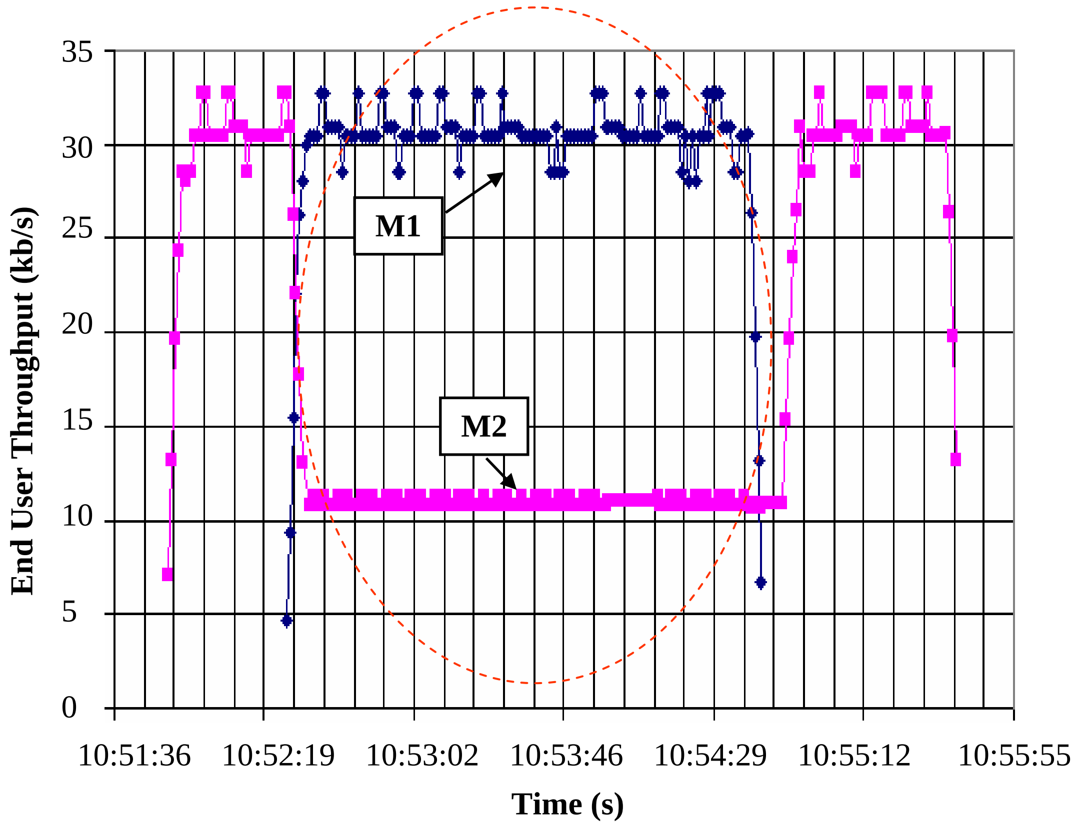
Priority	Weights Set 1 (%)	Weights Set 2 (%)	Weights Set 3 (%)
1	25	50	70
2	25	30	15
3	25	15	10
4	25	5	5

Weight Set 1: User throughput



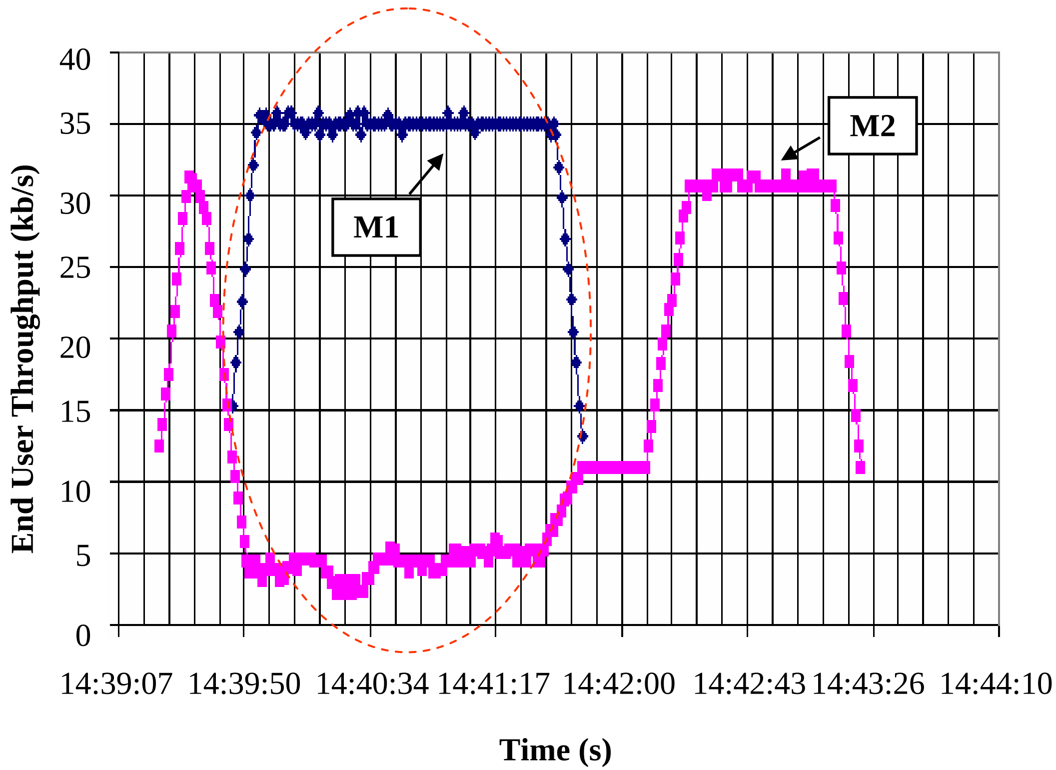
50/50

Weight Set 2: User throughput



77/23

Weight Set 3: User throughput

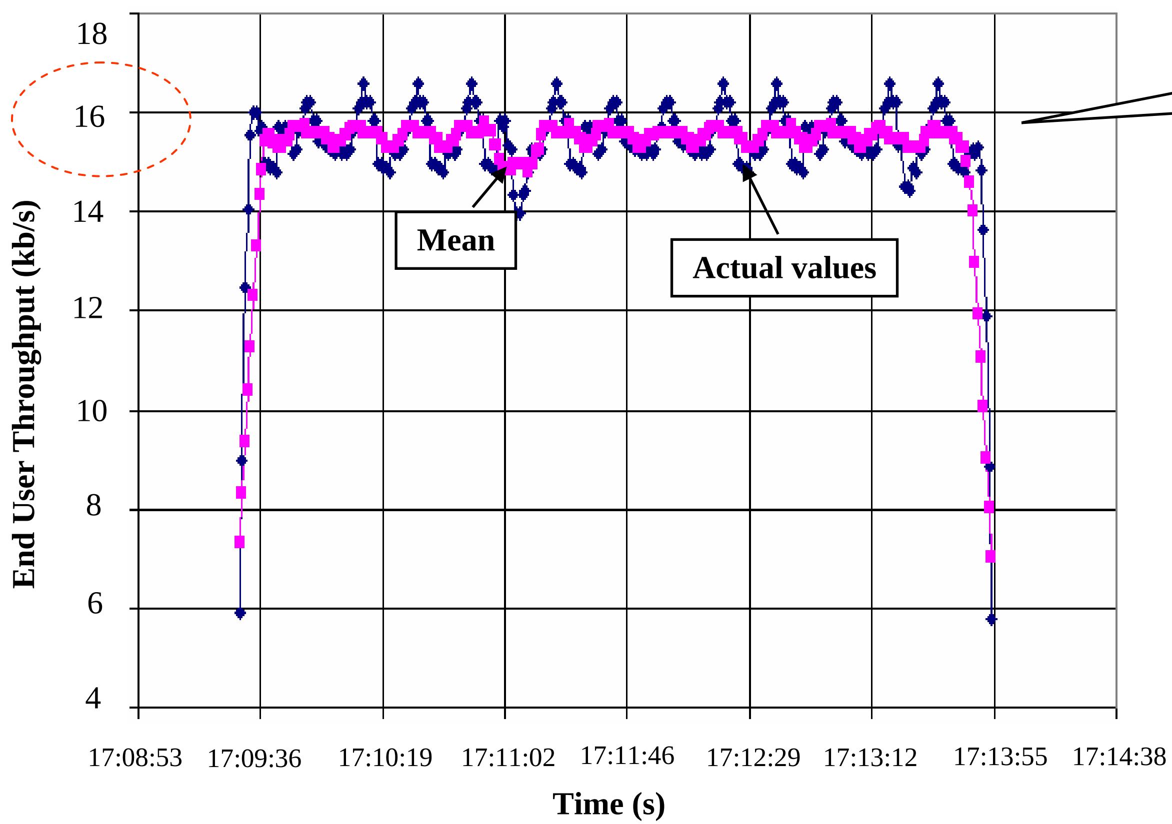


87/13

DSCP marking in SGSN and GGSN

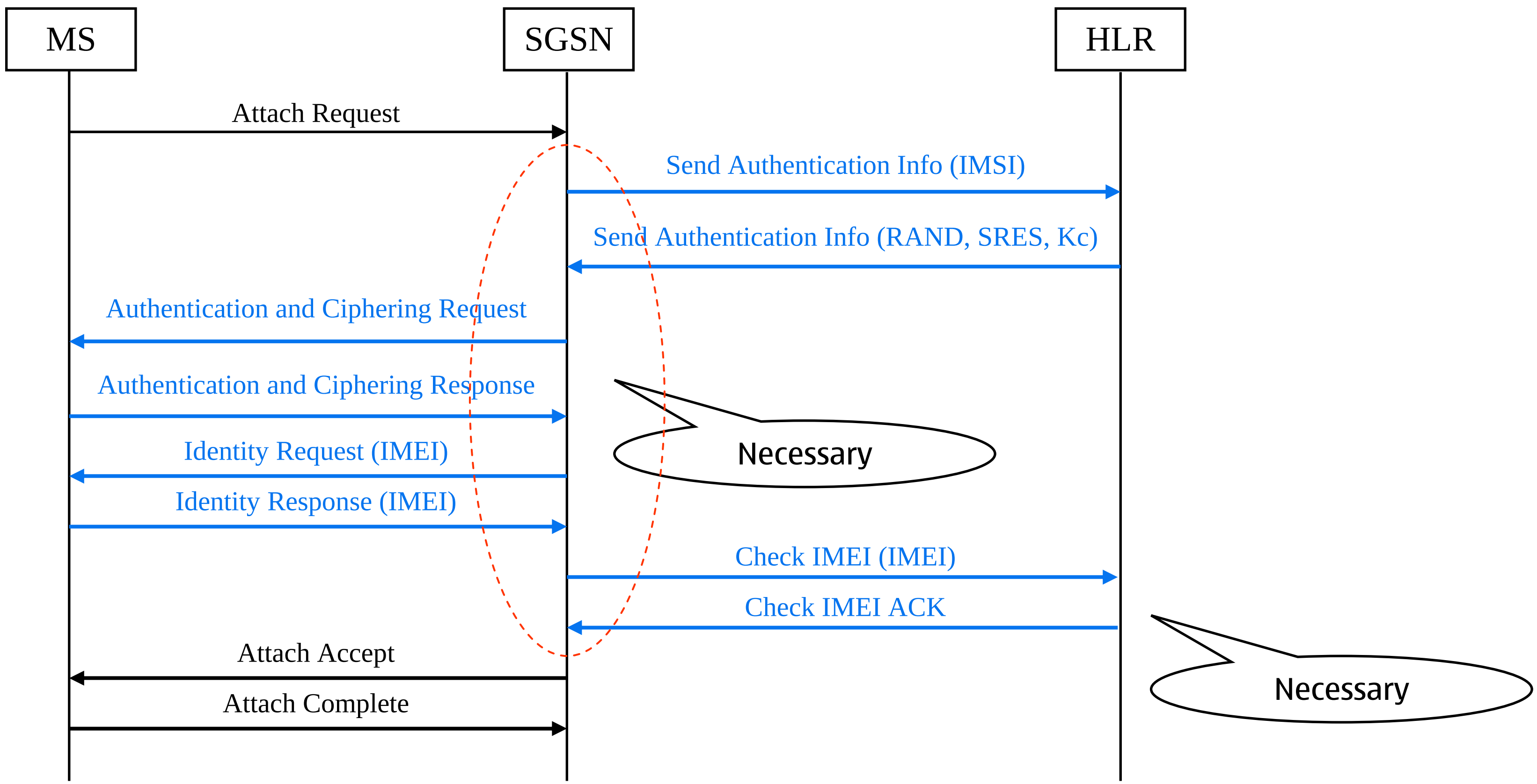
Classifier			Action	
Traffic Class	**Traffic Handling Priority (THP)**	**Allocation Retention Priority (ARP)**	**Per-Hop-Behavior (PHB)**	**DSCP (HEX)**
Conversational	-	ARP1	EF	B8
Conversational	-	ARP2	EF	B8
Conversational	-	ARP3	EF	B8
Streaming	-	ARP1	AF41	88
Streaming	-	ARP2	AF42	90
Streaming	-	ARP3	AF43	98
Interactive	THP1	ARP1	AF31	68
Interactive	THP1	ARP2	AF32	70
Interactive	THP1	ARP3	AF33	78
Interactive	THP2	ARP1	AF21	48
Interactive	THP2	ARP2	AF22	50
Interactive	THP2	ARP3	AF23	58
Interactive	THP3	ARP1	AF11	28
Interactive	THP3	ARP2	AF12	18
Interactive	THP3	ARP3	AF13	38
Background	-	ARP1	BE	00
Background	-	ARP2	BE	00
Background	-	ARP3	BE	00

Ex: Effects of policing and shaping at GGSN

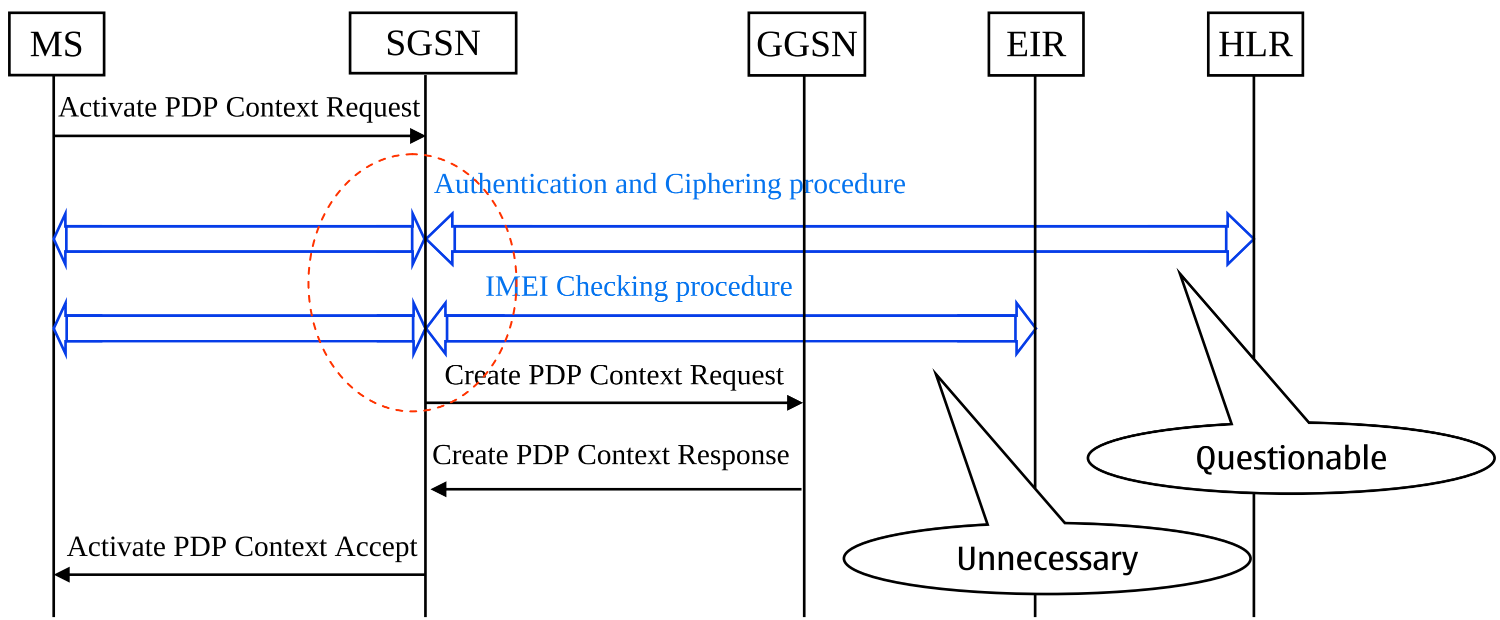


QoS Profile:
Max BR = 16 kb/s

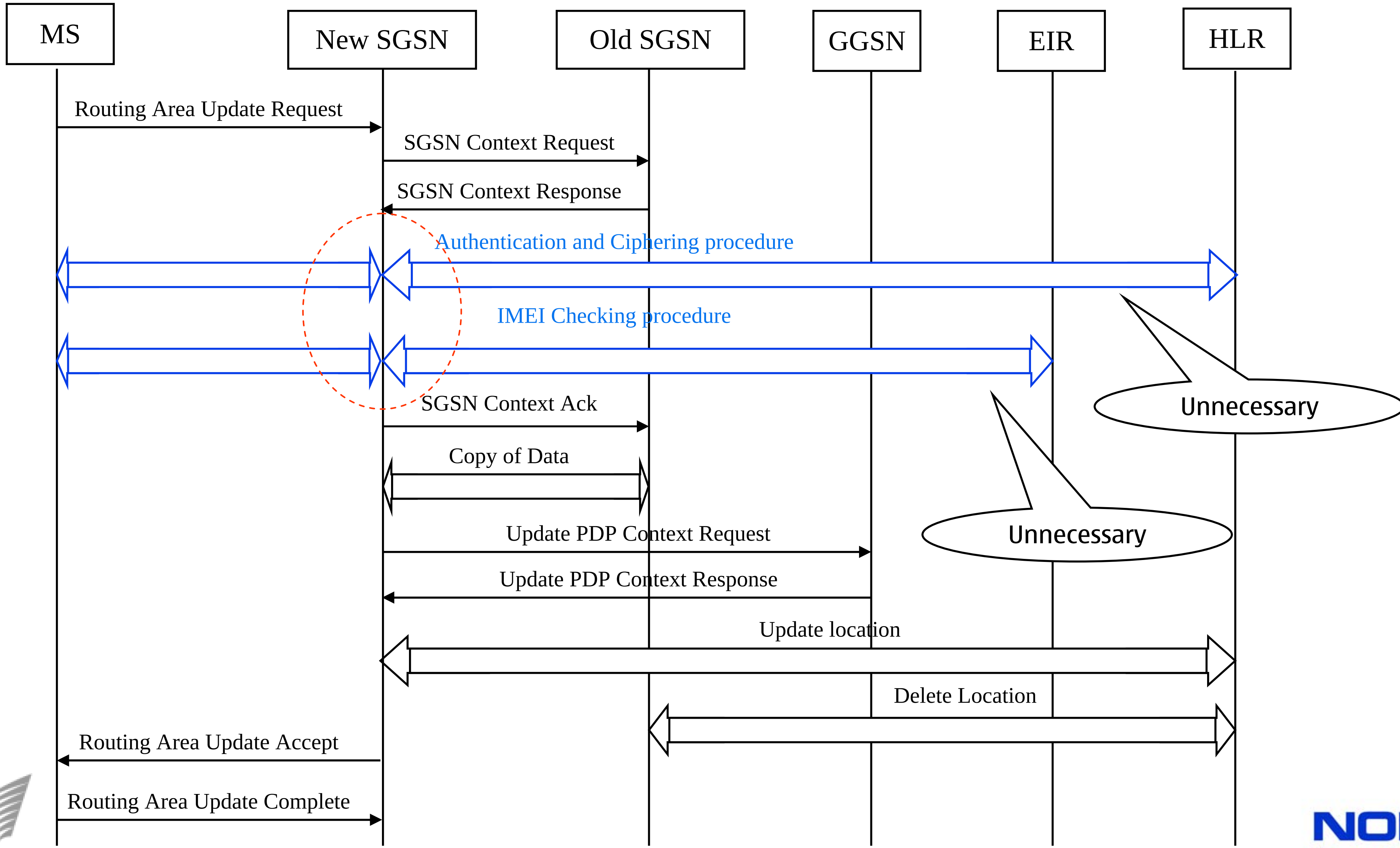
Security functions during “GPRS attach”



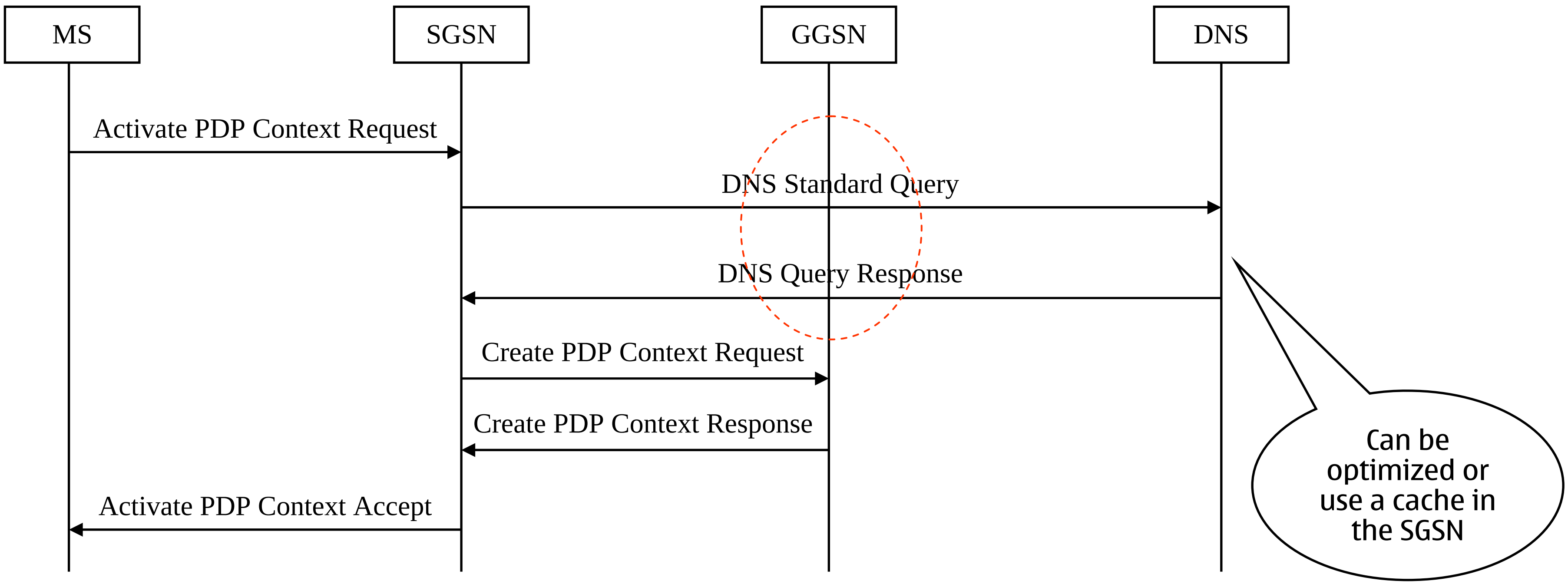
Security functions during “PDP context activation”



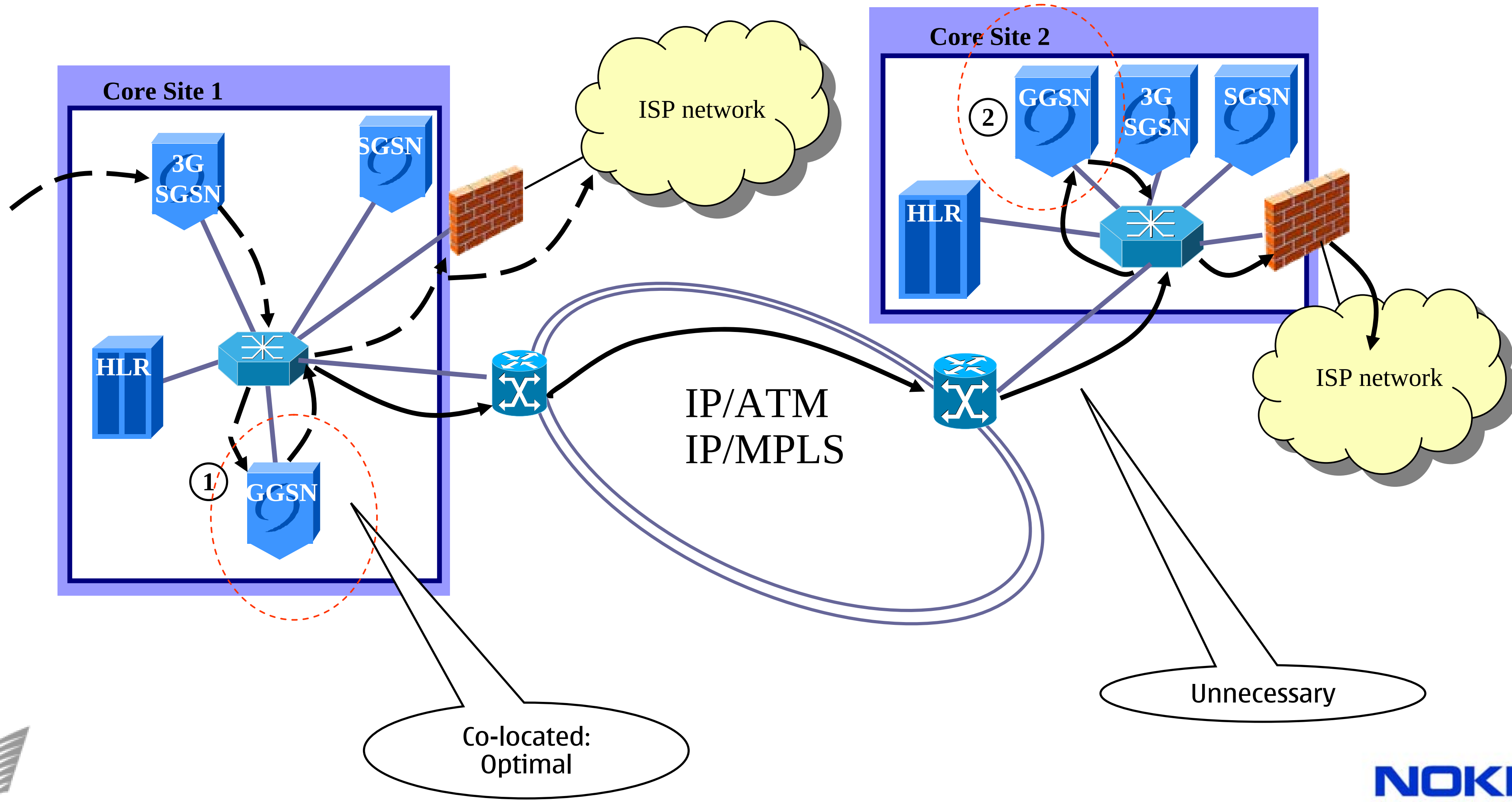
Security functions during “RA update”



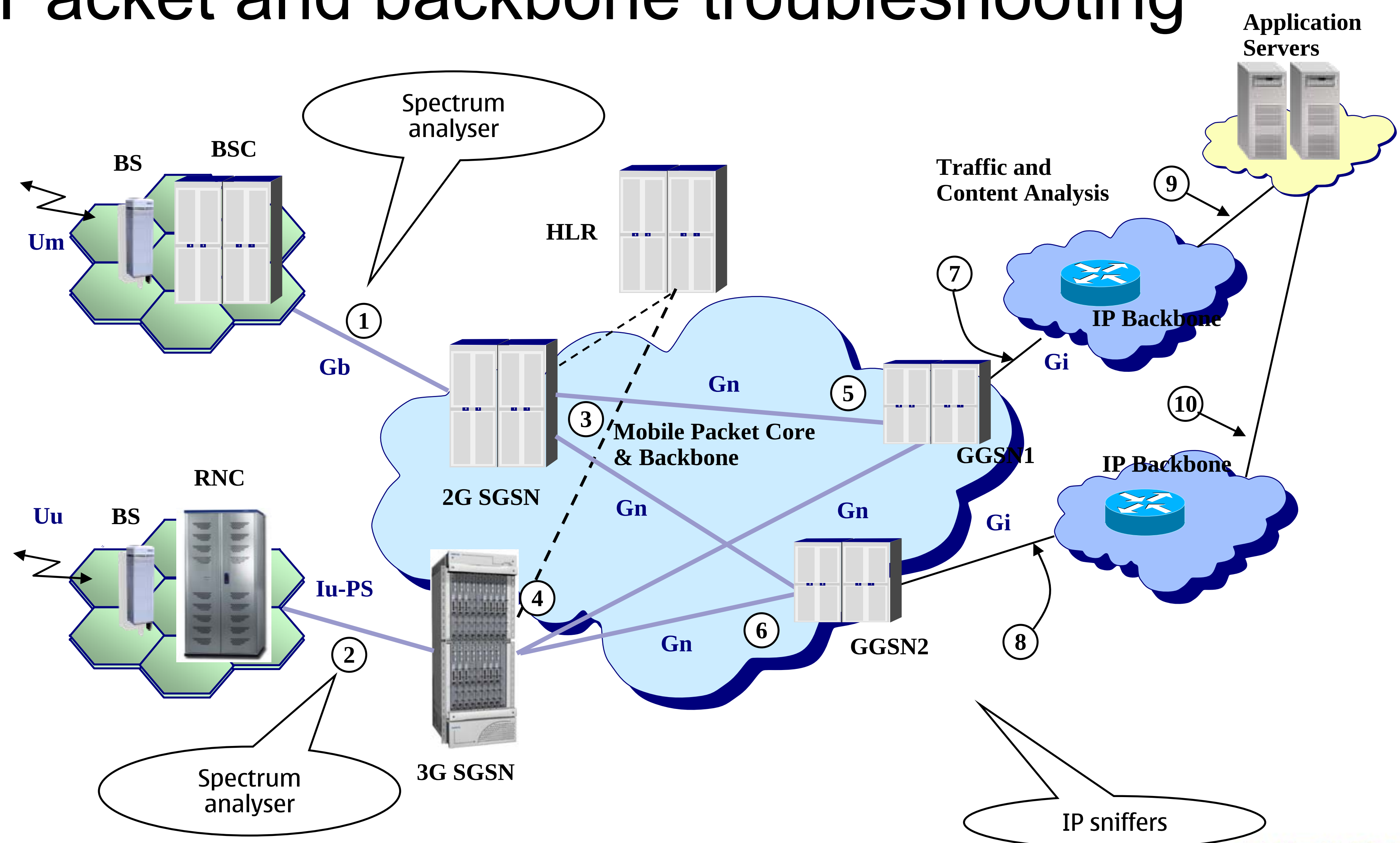
“PDP context activation” procedure



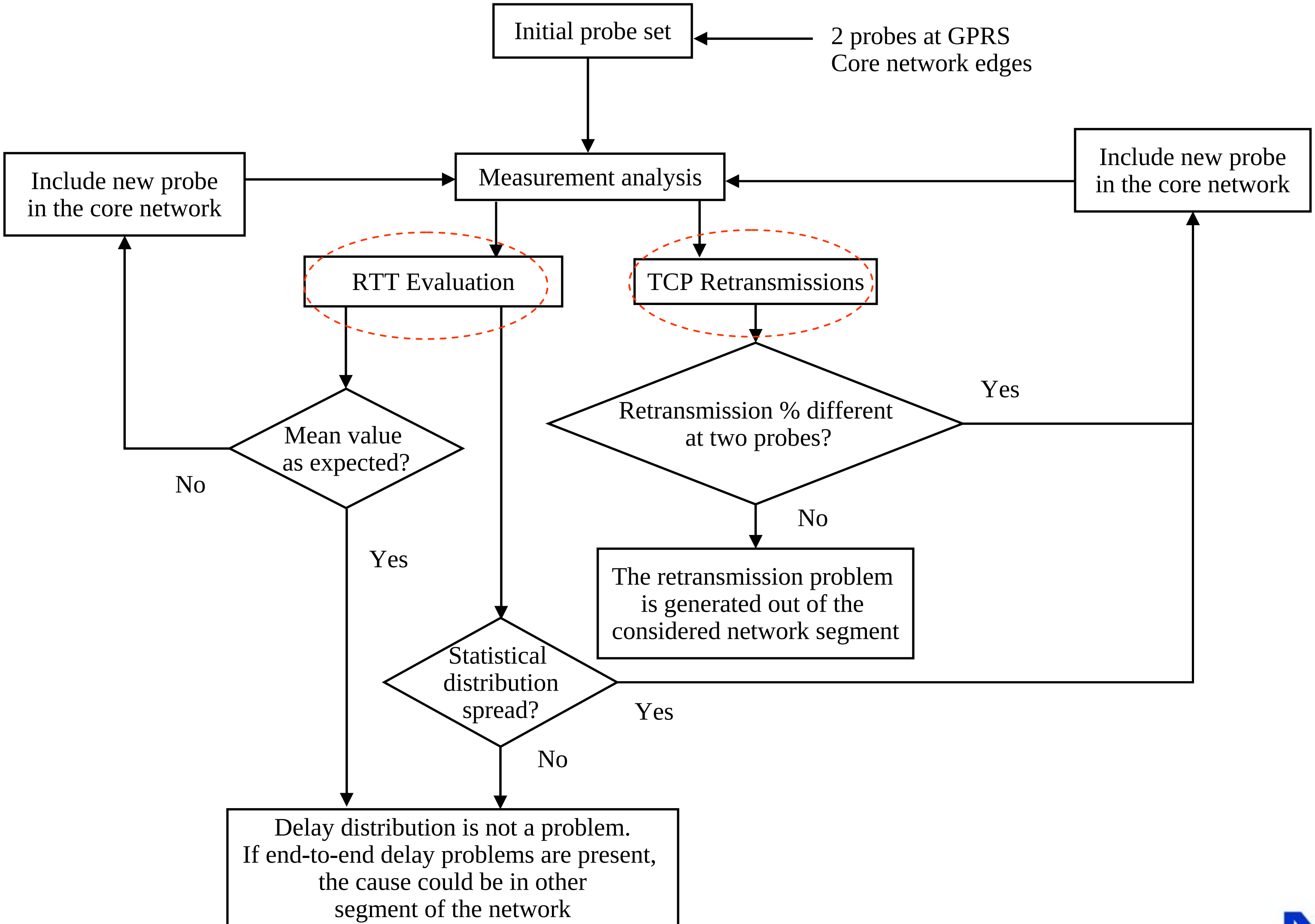
Ex: Geographical links



Packet and backbone troubleshooting



Ex: optimization process



Impact of parameter settings

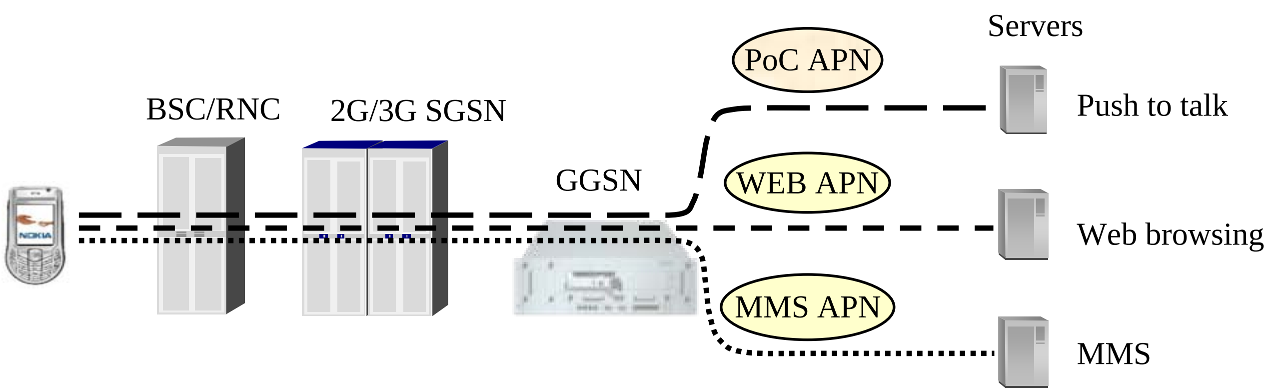
Mapping of Applications to Particular Subsets of QoS Attributes

- MMS (NGB): TC = background, ARP = 3

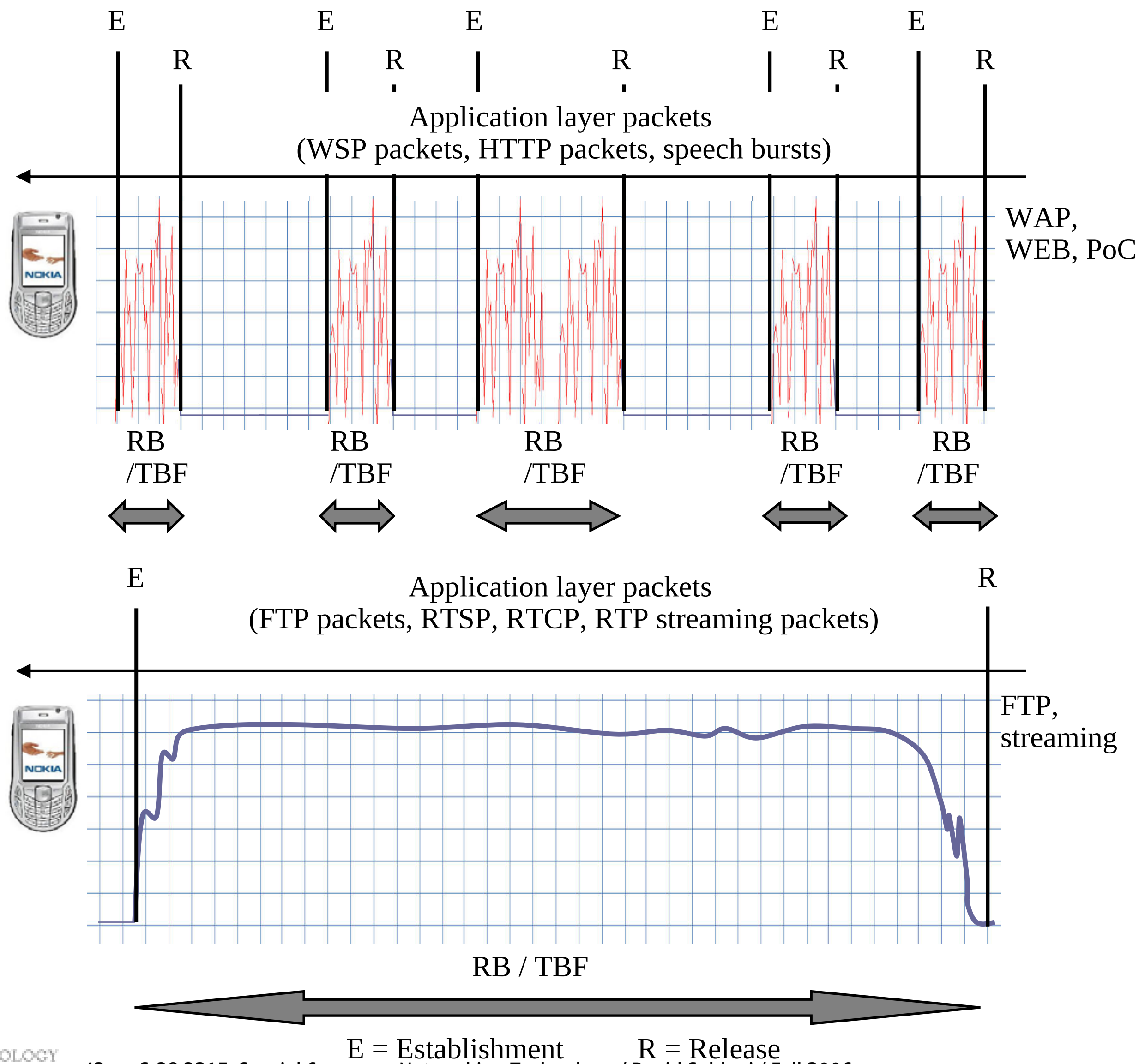
- Web browsing (NGB): TC = Interactive, THP = 3, ARP = 3)

- PoC (GB): TC = Streaming, ARP = 1

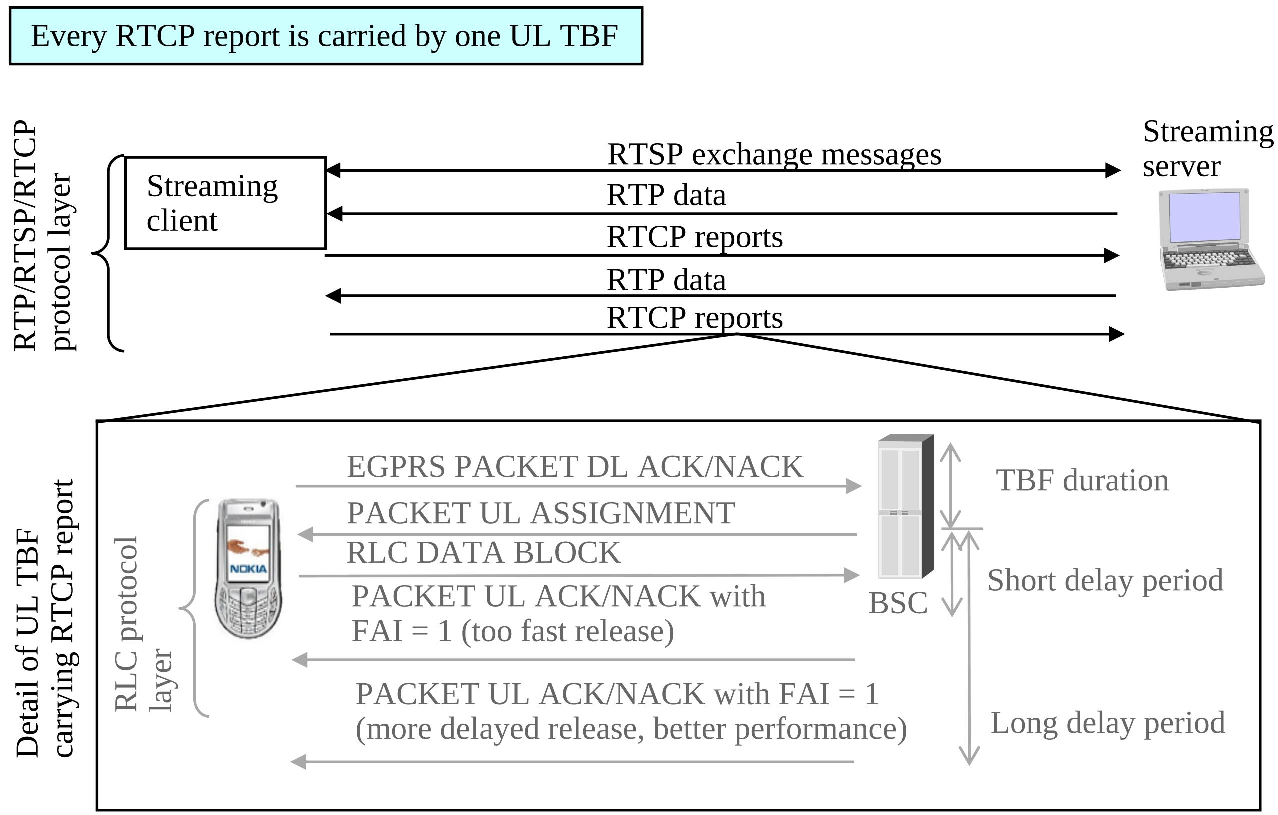
Consistent QoS Provisioning Across All Network Elements



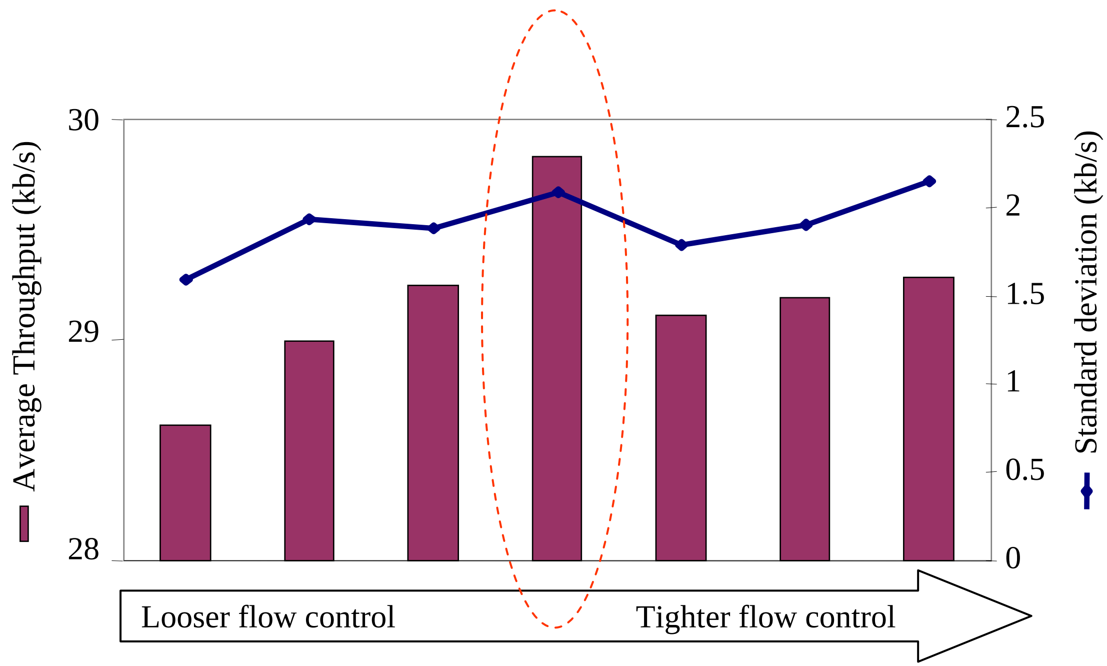
Impact of traffic characteristics



Impact of QoS functions: flow control (1/2)

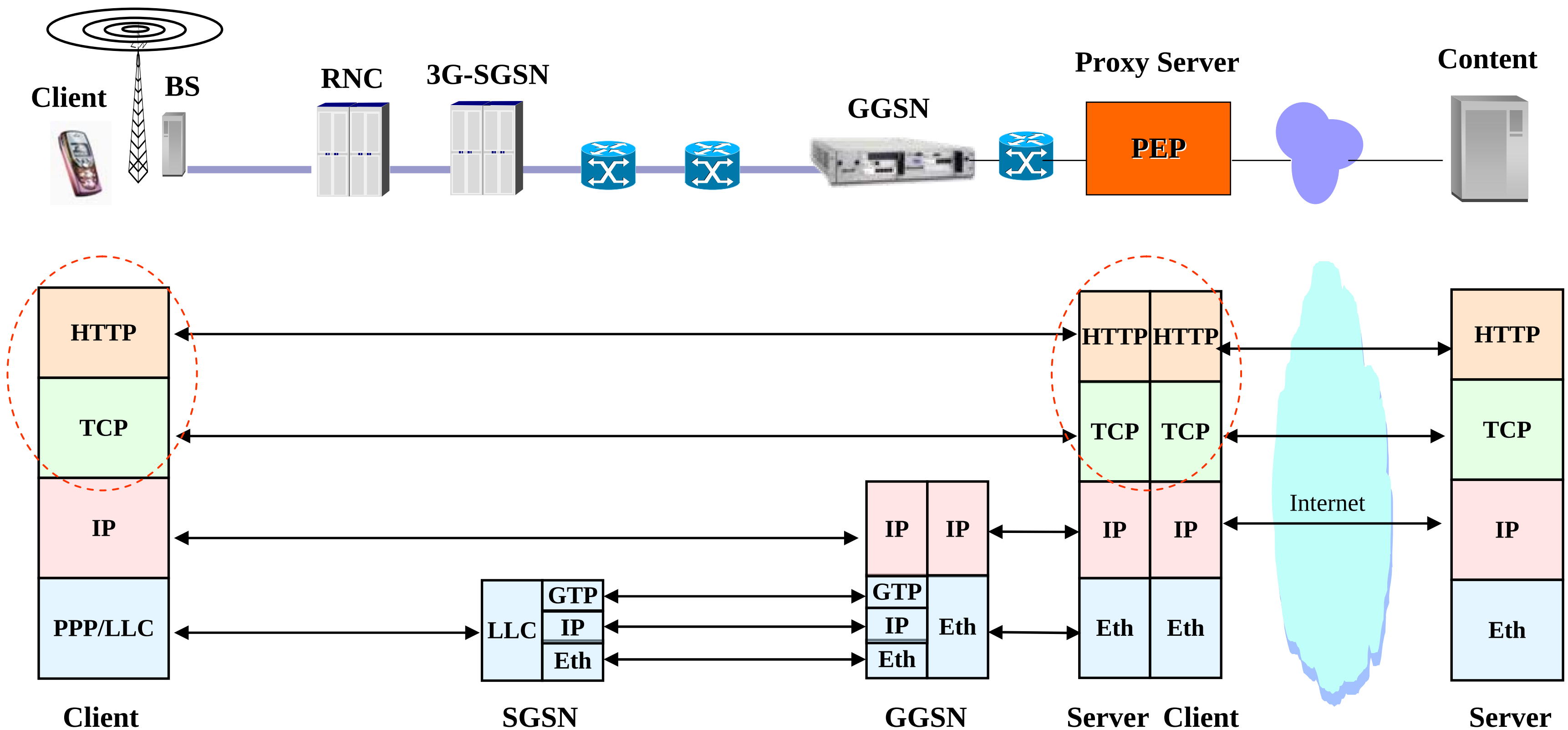


Impact of QoS functions: flow control (2/2)



Impact of performance enhancing proxies

- PEP acts on Layer 4 and above of the OSI stack, between the HTTP and TCP protocol stack by e.g. compressing the contents, http pipelining, content manipulation and optionally protocol conversion



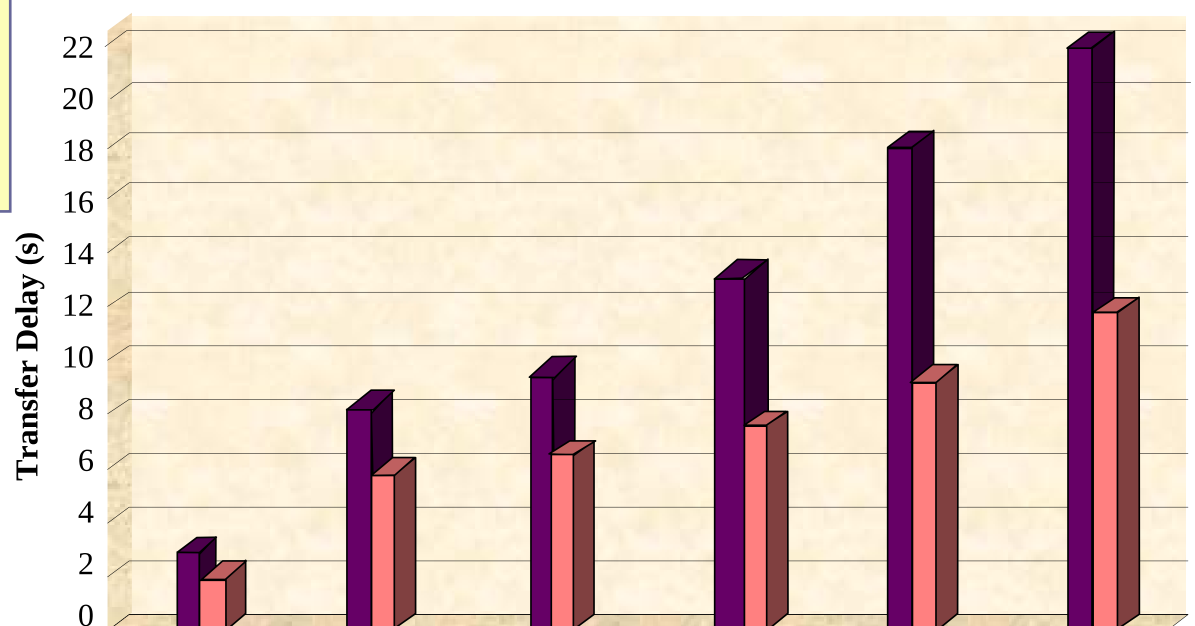
Ex: impact of PEP on Web browsing

3G: DL bearer – 128 kb/s

1. Htm Page 1: **12 kB**
2. Htm Page 2: **55 kB**
3. Htm Page 3: **85 kB**
4. Htm Page 4: **120 kB**
5. Htm Page 5: **190 kB**
6. Htm Page 6: **275 kB**

WEB Browsing - Comparison: Time of downloading

Influence of Proxy Server



Page number	1	2	3	4	5	6
3G	2.14	7.93	9.04	13.91	17.16	21.57
3G with Proxy Server	1.66	5.53	5.98	7.15	8.66	11.48

Average delay (s)

References

- D. Soldani, M. Li and R. Cuny (eds.), **QoS and QoE Management in UMTS Cellular Systems**, John Wiley and Sons, June, 2006, 460 pp.
  - http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470016396.html
  - http://www.connecting.nokia.com/NOKIA/nns.nsf/a/78786C61AB5A7C5AC225718F0026BAA3

    (Contact Mr. Geoff Farrell @ Wiley gfarrell@wiley.co.uk)

See also:

  - http://lib.tkk.fi/Diss/2005/isbn9512278340/