

# End-to-End QoS in Mobile Networks

**an holistic 'engineering' view**

IEEE ComSoc / IST

Paulino Corrêa

António Carlos Fernandes

8<sup>th</sup> January 2013

|||

|||

WIKIPEDIA



## Quality of service

---

From Wikipedia, the free encyclopedia

### Definition

*“(...) In the field of computer networking and other packet-switched telecommunication networks, the traffic engineering term refers to **resource reservation control mechanisms** (...)*

*Quality of service is the **ability to provide different priority to different applications, users, or data flows, or to guarantee a certain level of performance to a data flow.** (...)*

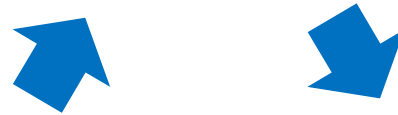
***Quality of service** guarantees are **important if the network capacity is insufficient**, especially for real-time streaming multimedia applications such as voice over IP, online games and IP-TV, since these often require fixed bit rate and are delay sensitive, **and in networks where the capacity is a limited resource**, for example in cellular data communication.”*

# Why QoS ?

Three different angles ...



Quality of Experience  
QoE



Network  
Scalability

Network  
Efficiency



# Why QoS ?

Three different angles ...



User perception of the service quality, in his location, with his terminal

Quality of Experience  
QoE

Effective network development (=CAPEX) to match usage growth

Network Scalability



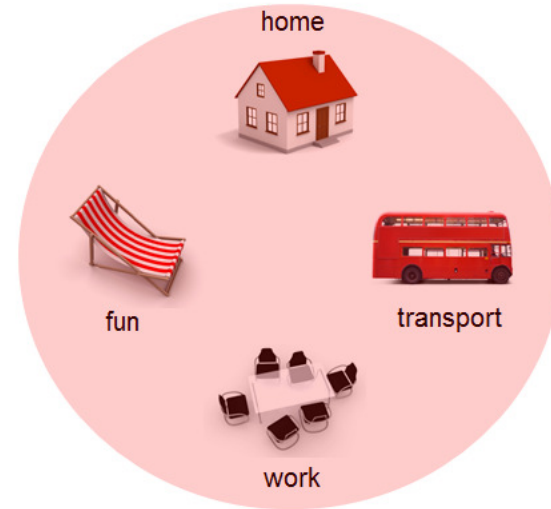
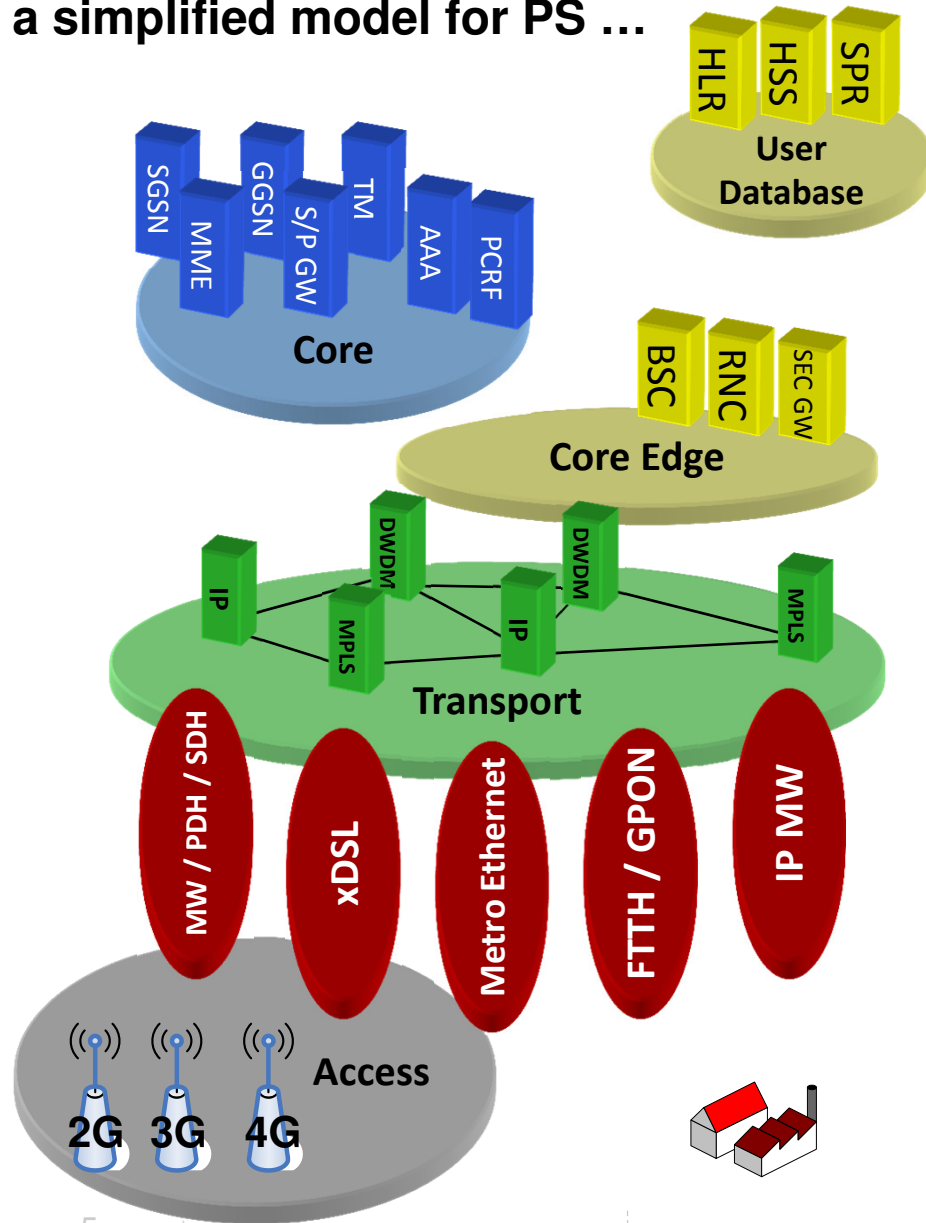
Efficient use of the existing network resources, to the fullest, in stable conditions

Network Efficiency



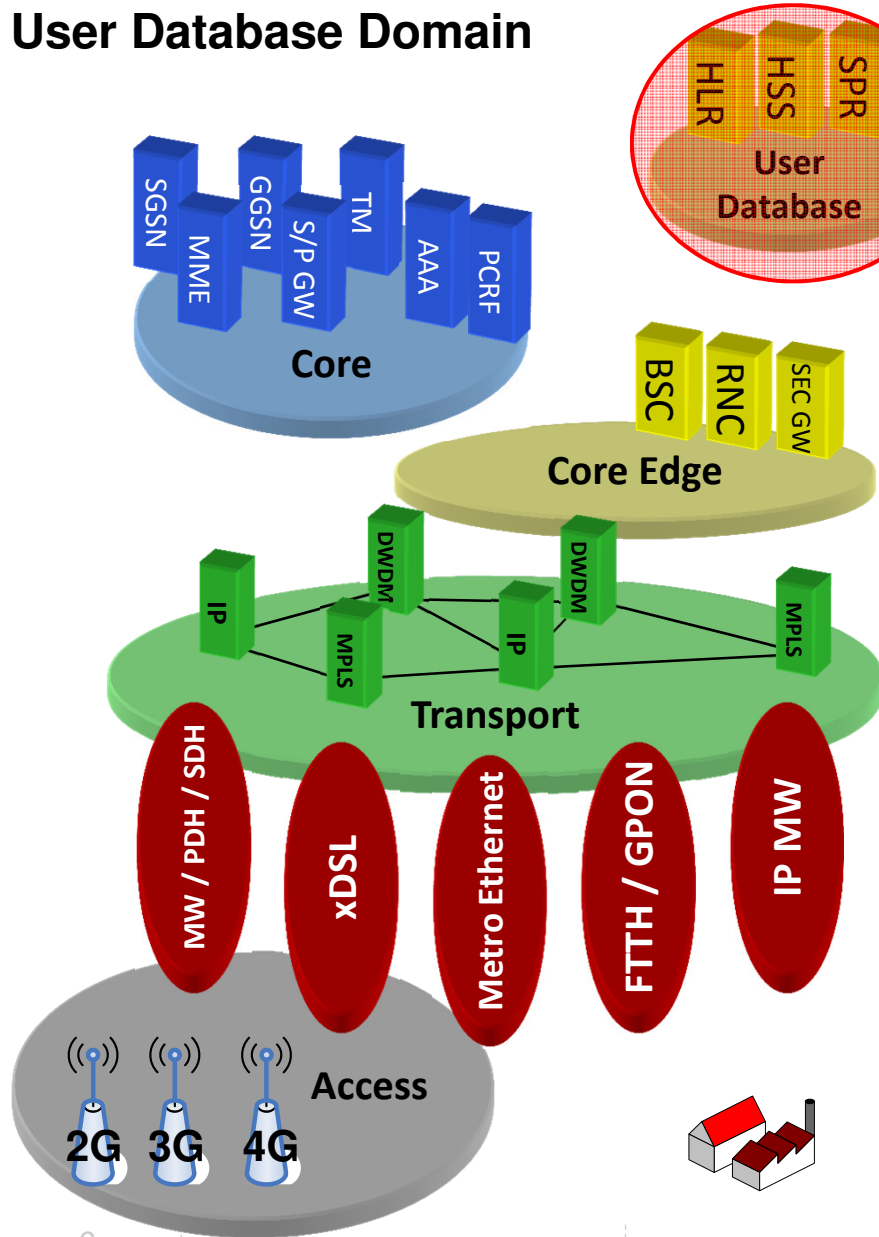
# Mobile Network Domains

a simplified model for PS ...



# Mobile Network Domains

## User Database Domain



Specific provisioning allows **static** differentiation of users

User parameters are defined here for specific services on a per user basis

*example:*

21 Mbps in APN x GOLD  
14 Mbps in APN y SILVER

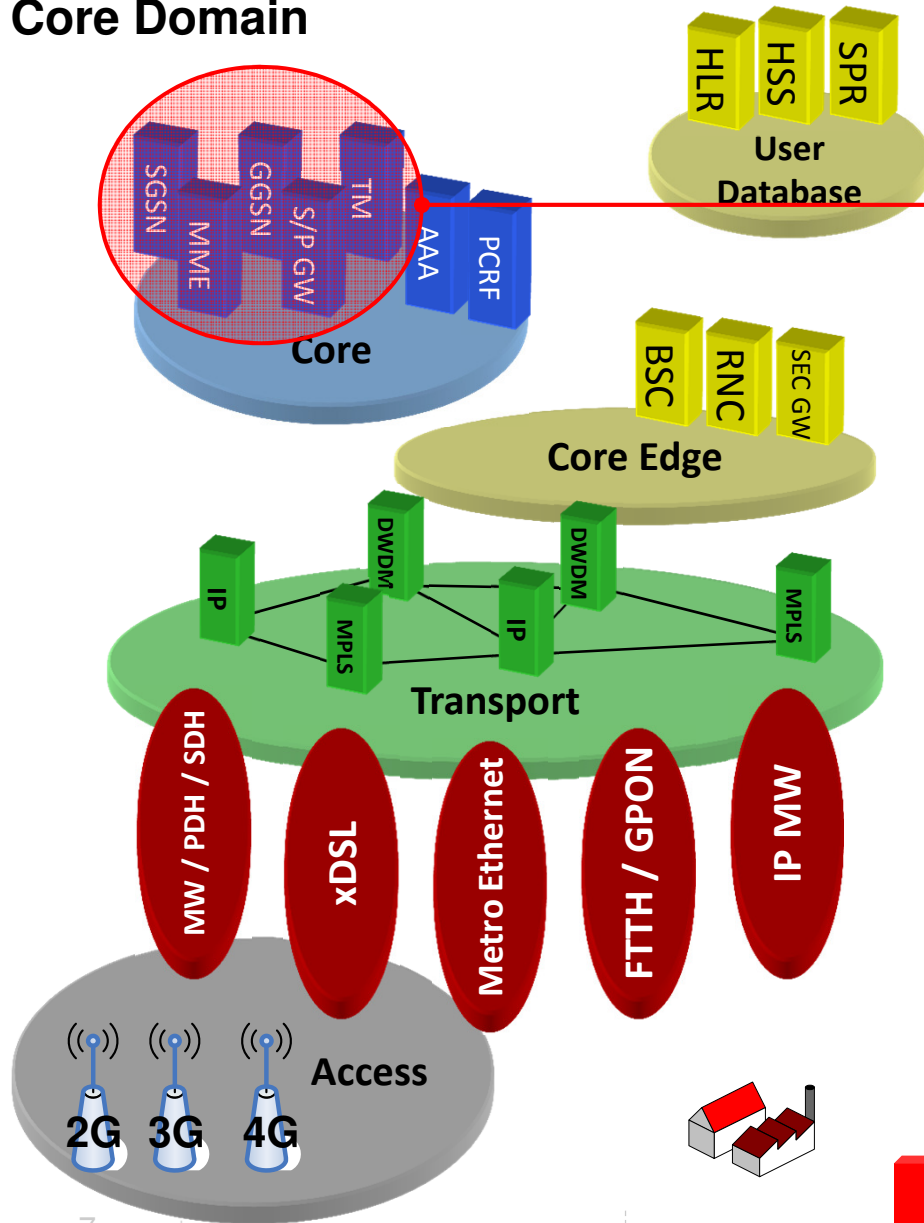
*These parameters are passed on to the other domains when the user first attaches or establishes a new bearer (e.g. 3G PDP)*

**The granularity here is the user (IMSI/MSISDN) and the service package (APN)**



# Mobile Network Domains

## Core Domain



There is no provisioning in this domain, but user awareness and context is stored. **User plane traffic is processed** and application/protocol awareness is possible.

It is also here that mobility management is done and specific triggers to the access domain are launched (3GPP QoS model).

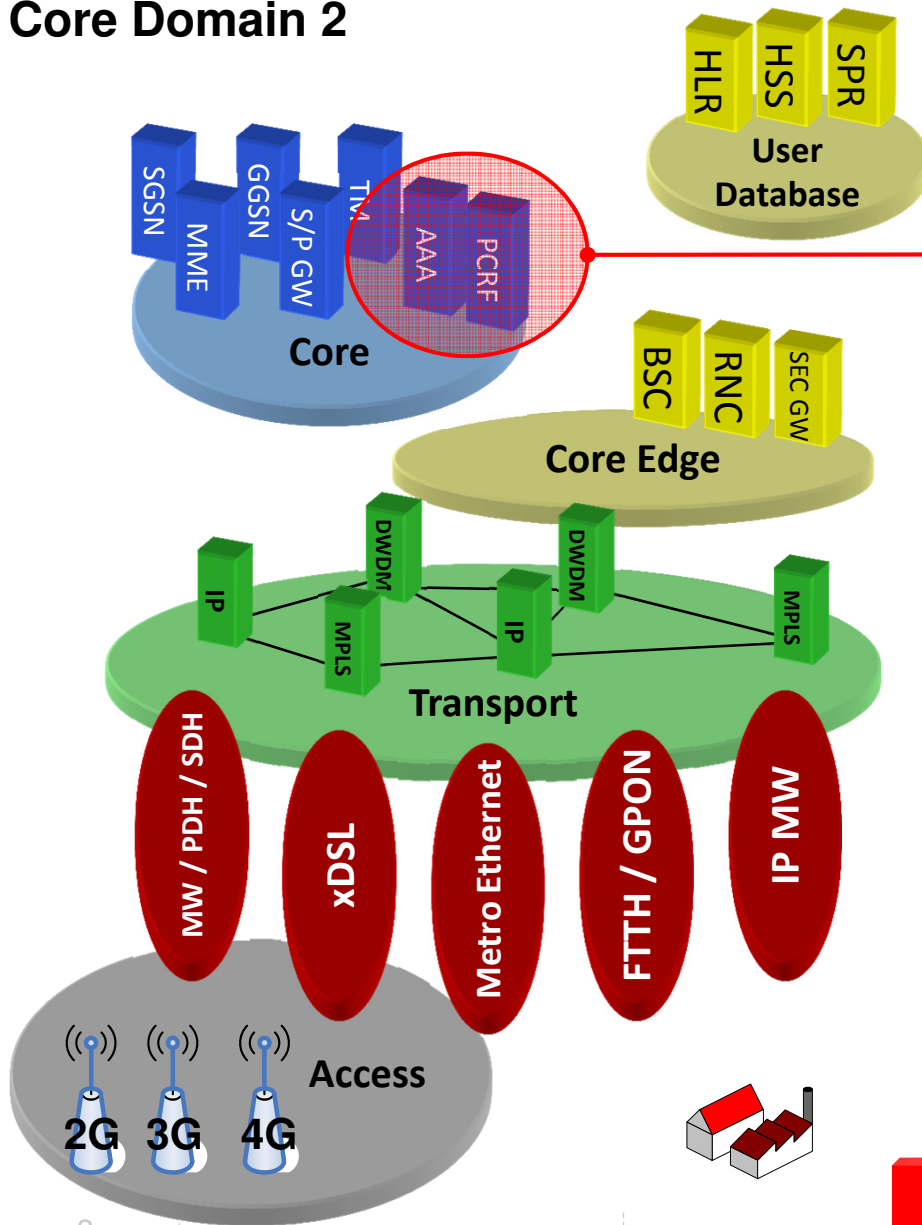
User plane processing allows user IP packet marking (e.g. DSCP) if the right triggers are configured

*The granularity here is the user (IMSI/MSISDN), the service package (APN) but also the application, protocol or IP flow*

subscriber awareness	application awareness	access network awareness	user terminal awareness
----------------------	-----------------------	--------------------------	-------------------------

# Mobile Network Domains

## Core Domain 2



Within the core domain there are also highly customisable functions such as the **AAA**, **Radius** and **PCRF**. These can be used to provide many triggers to implement specific service logic or network specific (e.g. protection) policies depending on:

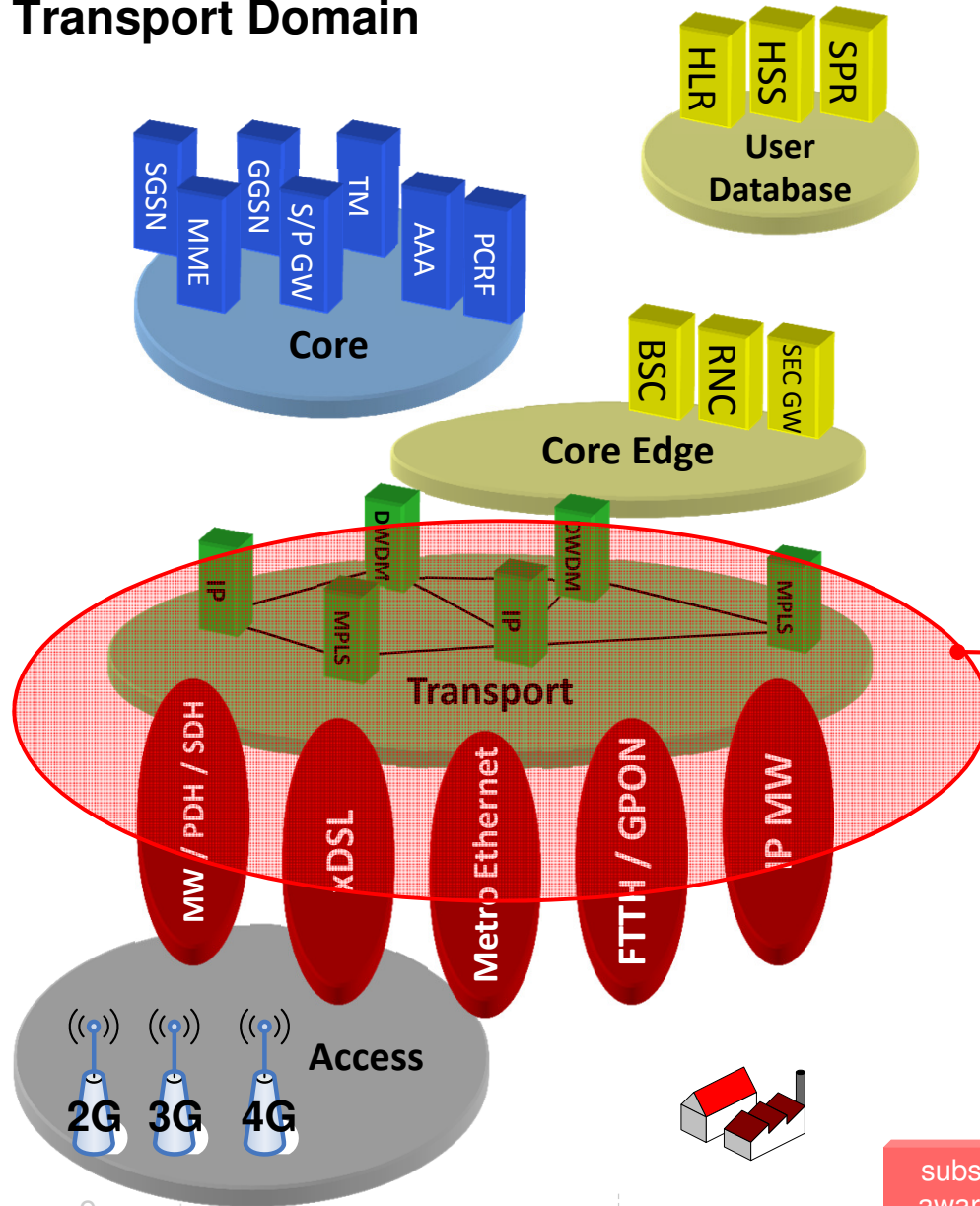
- User*
- Terminal*
- Location*
- Radio access technology*
- Roaming status*
- Past usage*
- Network conditions*
- (...)

subscriber awareness    application awareness    access network awareness    user terminal awareness



# Mobile Network Domains

## Transport Domain



In the transport domain there is no direct user awareness but **global policies are possible** (e.g. prioritise voice over data in IP backhaul). If traffic is correctly marked by the core more effective traffic management is a possibility.

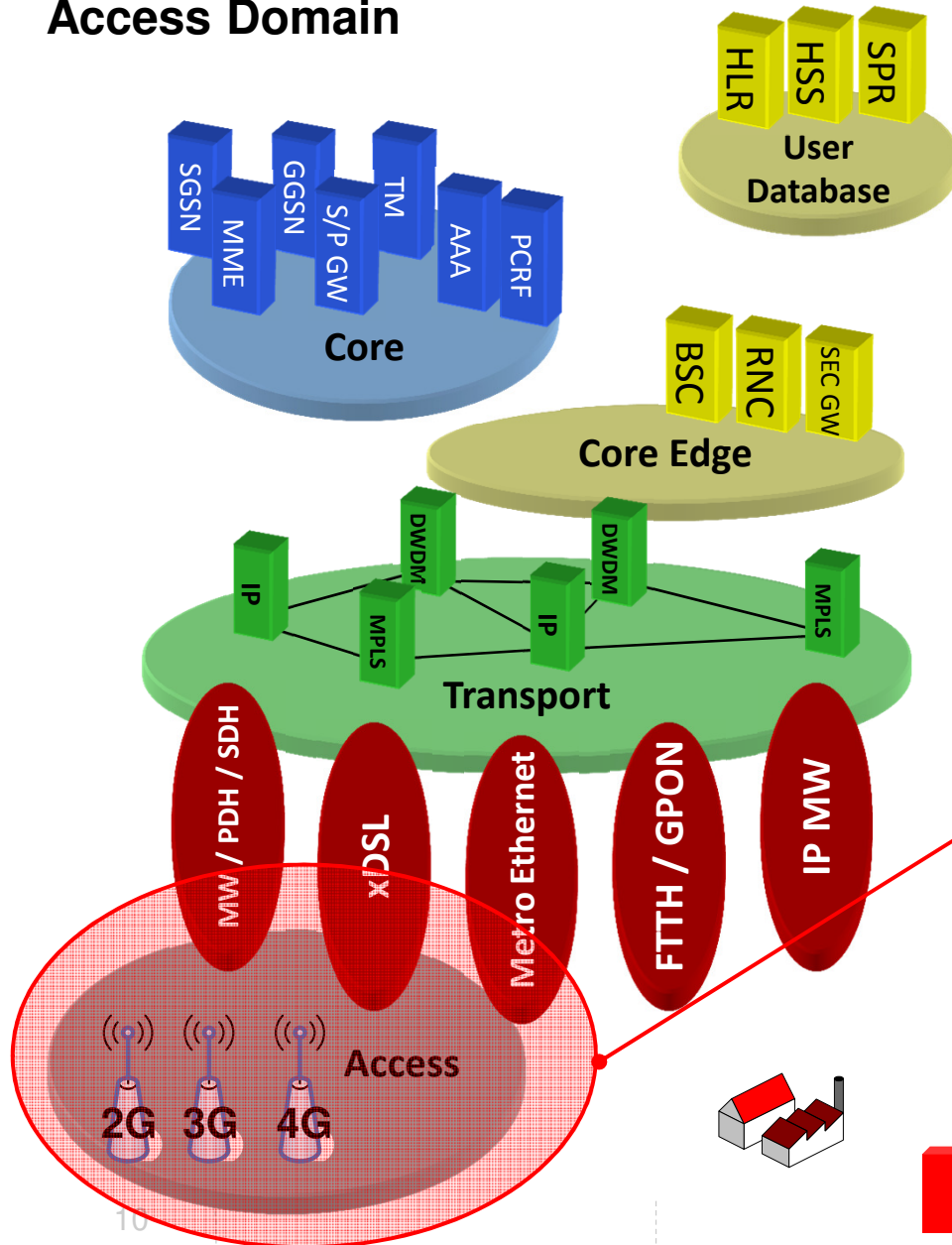
*example:*

**IP backhaul routers and switches can react to DSCP and ethernet P-bit markings**

- subscriber awareness
- application awareness
- access network awareness
- user terminal awareness

# Mobile Network Domains

## Access Domain



In a mobile network typically this is the most expensive resource (spectrum). Congestion management is thus critical and widely used (e.g. RNC load control).

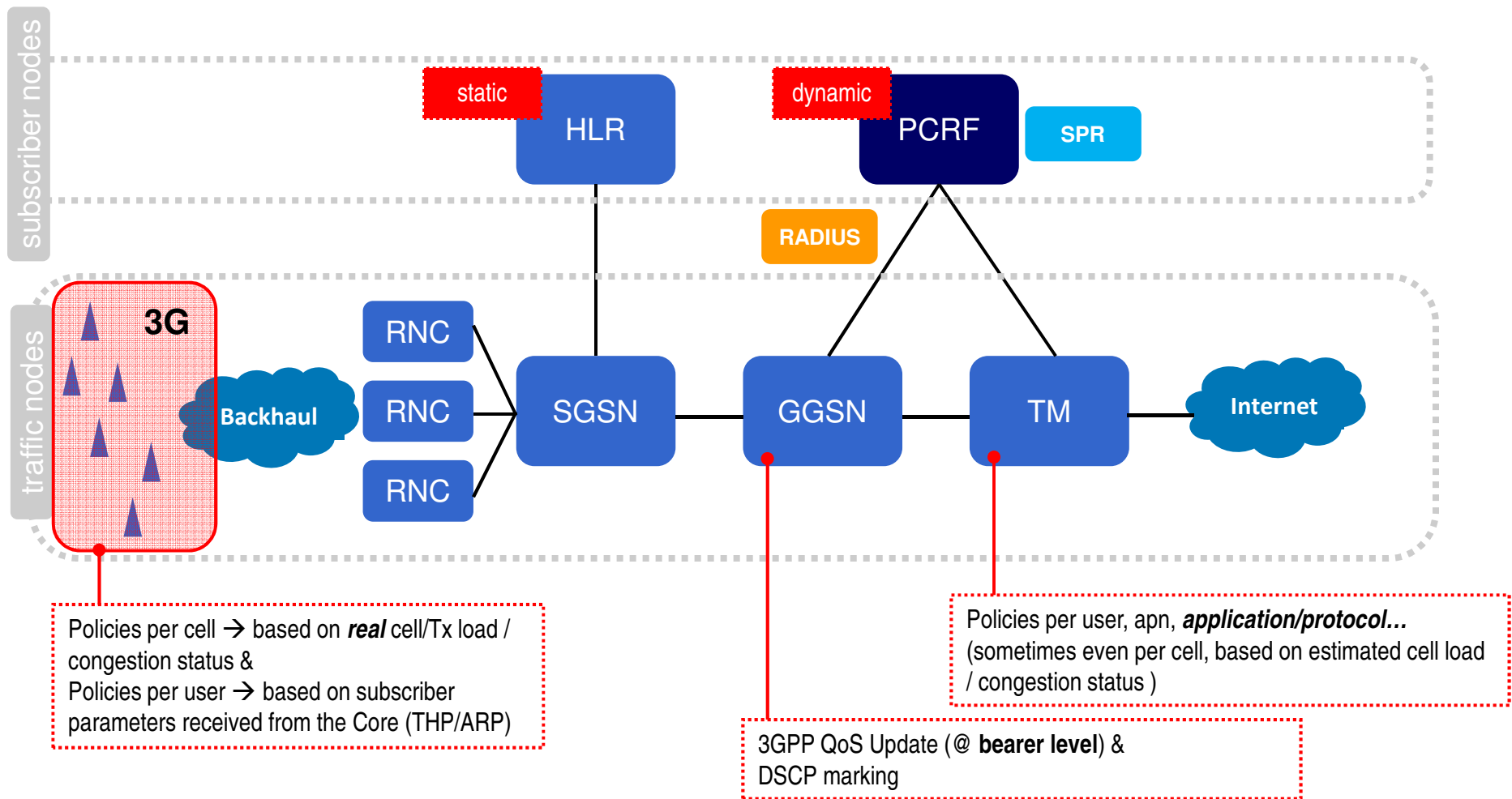
It can be enhanced if specific triggers received from the core domain are used to influence nodeB scheduler

Example:

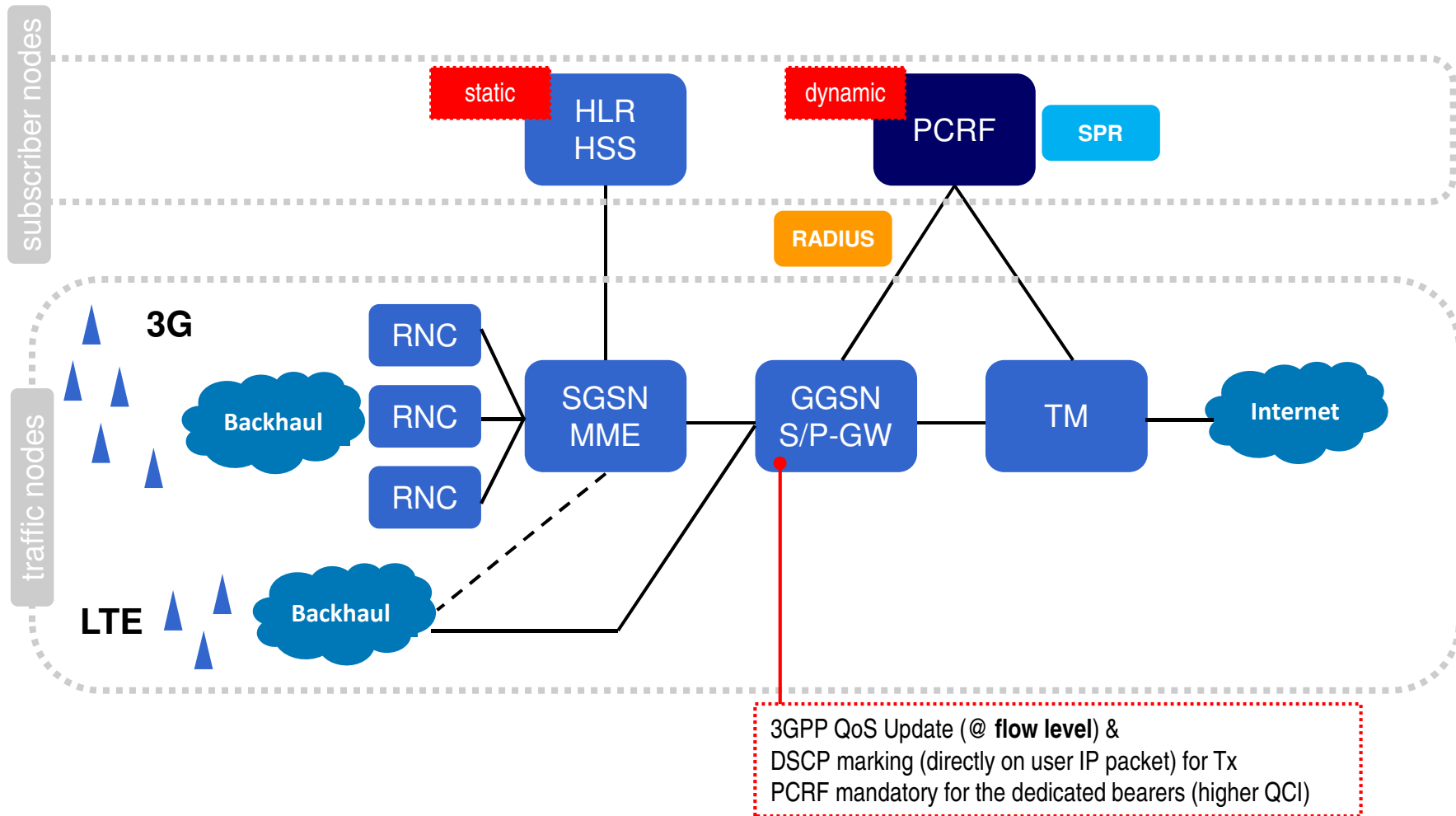
***In case of sector congestion prioritise all traffic from user A over user B***

- subscriber awareness
- application awareness
- access network awareness
- user terminal awareness

# Setting up the e2e architecture



# Setting up the e2e architecture



# User profiles

## 3 very different, very real mobile data users

[A] Alice

- Active video user plus mobile apps (social networks, IM, other)
- “Always on” smartphone with data intensive tariffs
- High mobility, sporadic volume, ubiquitous usage



[B] Bob

- Bandwidth “hogger” – very high volumes, constantly downloading
- Nomadic usage in 2-3 specific hotspots
- Laptop terminal, possibly tablet



[C] Carl

- Professional profile
- Medium to high usage but not constant
- Highly critical – customer with high ARPU

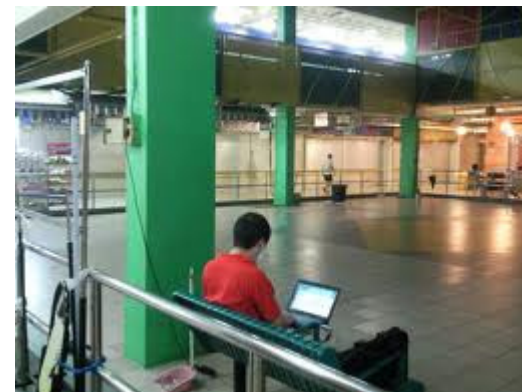


# #1 Use Case – Nomadic Description

- Shopping mall food court (or airport, university campus, etc)
- 1/3 Alices, 1/3 Carls, 1/3 Bob
- 1/3 of Bobs may “kill” the network if they are exposed to better radio conditions and using demanding file transfer applications.

Problem:

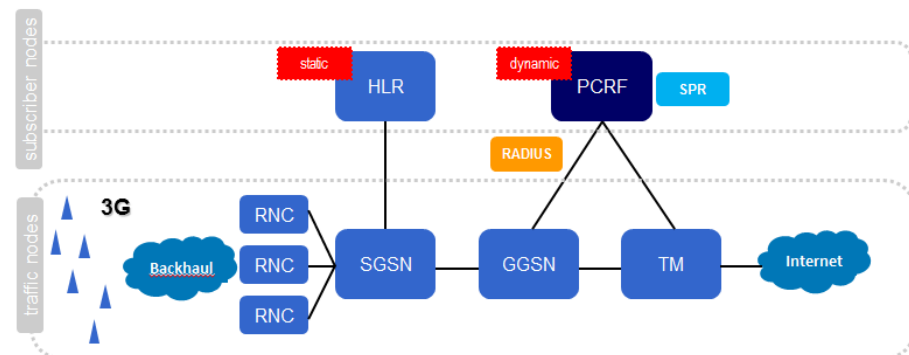
- Manage Bobs as best as possible, to prevent them to exhaust the NW!



# #1 Use Case – Nomadic Actions

- What do we know?
  - User profile
  - Device type & capabilities
  - Location (\*)
  - Type of applications (real-time, non real-time, bursty, etc)
  - Cell load
- Bob is using download managers/aggressive P2P to get the latest available “technical documentation” with excellent radio coverage

- What can we do?
  - Apply G/S/B, meaning:
    - If there are resources, let Bob use them!
    - Bob can even be in background mode, not looking into the laptop
  - What else?
    - Bob can actually shift from Silver to Bronze along his monthly billing cycle depending on his overall data usage



## #2 Use Case – Mobility Description

- Urban environment, users accessing different sorts of content (Video, IM, Email, Web browsing, Social NWs)
- Moving around, in the streets, public transportation, in the car...
- 45% Alices, 45% Carls, 10% Bobs



Problem:

- Youtube and Web is slow and my email is not flowing (attachments and replying)!!

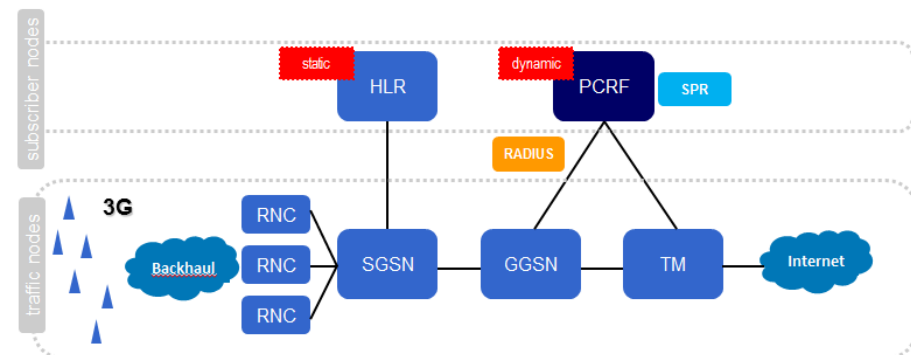




## #2 Use Case – Mobility Actions

- What do we know?
  - User profile
  - Device type & capabilities
  - Location (\*)
  - Type of applications (real-time, non real-time, bursty, etc)
  - Cell load
- Alice #1 and #2 have same HLR profile (both silver, middle THP/ARP)
- #1 is using predominantly Video, the other is IM
- Carl is doing email, opening attachments, enterprise/gold subscriber (with already the highest THP/ARP)

- What can we do?
  - Decrease dynamically Alice's #2 THP/ARP (bronze)
  - Apply G/S/B (Carl will prevail with higher QoS)
  - Adjust, in real-time, the codec/rate of Alice's #1 video to ensure:
    - Less BW consumption
    - Smooth streaming, no stalling



## #3 Use Case – Mass events

### Description

- Large concentration of users in one specific place
- Constraints on capacity increase
- Everyone is taking photos, uploading to social NW, doing IM
- Sports reporters (Carls) are covering the event, generating a lot of traffic
- 90% Alices, 10% Carls, 0% Bobs

Problem:

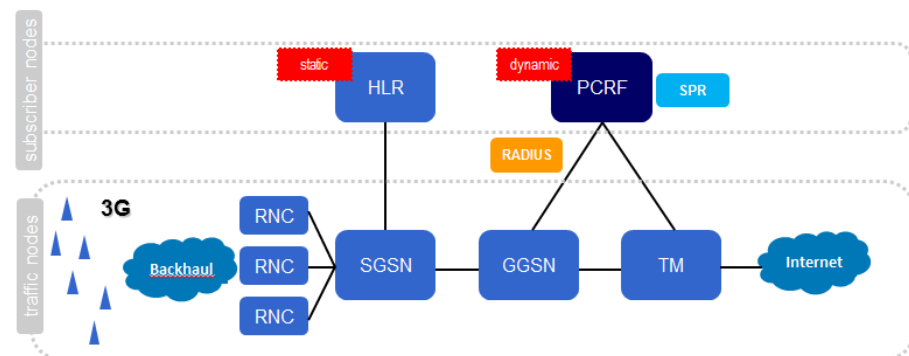
- Do your best!



# #3 Use Case – Mass events Actions

- What do we know?
  - User profile
  - Device type & capabilities
  - Location (\*)
  - Type of applications (real-time, non real-time, bursty, etc)
  - Cell load
  - Time of the event

- What can we do?
  - Block background applications for the duration of the event (in the dedicated cell sites)
  - Apply DSCP IP marking (differentiating real-time applications positively)
  - Apply more aggressive compression policies for video
- What else?
  - Move users to different access types if available (eg: wi-fi) using ANDSF\*

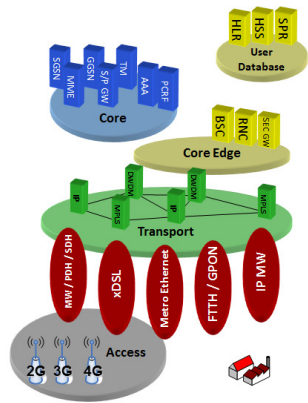


# Evolution Outlook

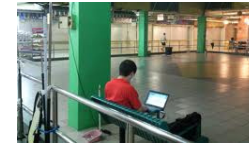
Other concepts in the radar .. Or that should be there !

- **Closed Loop Analytics**  
*Extract, analyse, influence*
- **Enhanced QoE policies**  
*Adapting policies to factors such as screen size, first time an application is used, etc.*
- **Intelligent resource management** and activation  
*Follow massive groups of users e.g. from stadium to metro station*
- User **terminal and network interaction** (signalling) protection  
*Avoid signalling overload from uncontrolled causes such as OTTs*
- **Content delivery**  
*Place the content where it is most effective*
- **Location accuracy** and update (towards the core)  
*Going for real-time location tracking*
- **Application awareness in the access**  
*Differentiate applications directly in the 'last mile'*
- **Heterogenous networks and ANDSF**  
*Use the right access type at the right occasion, .seamlessly*



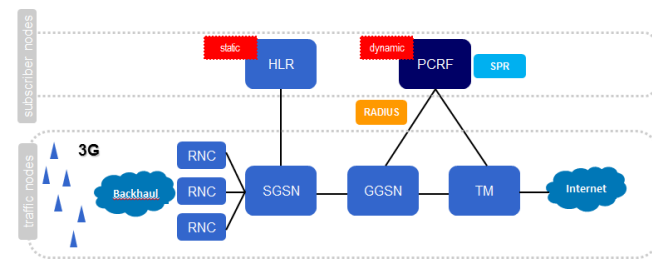


Quality of Experience  
QoE



Network  
Scalability

Network  
Efficiency



THANK YOU