Quality of Service for Next Generation Networks

Amit Mukhopadhyay, Dong Sun

Bell Laboratories, Lucent technologies

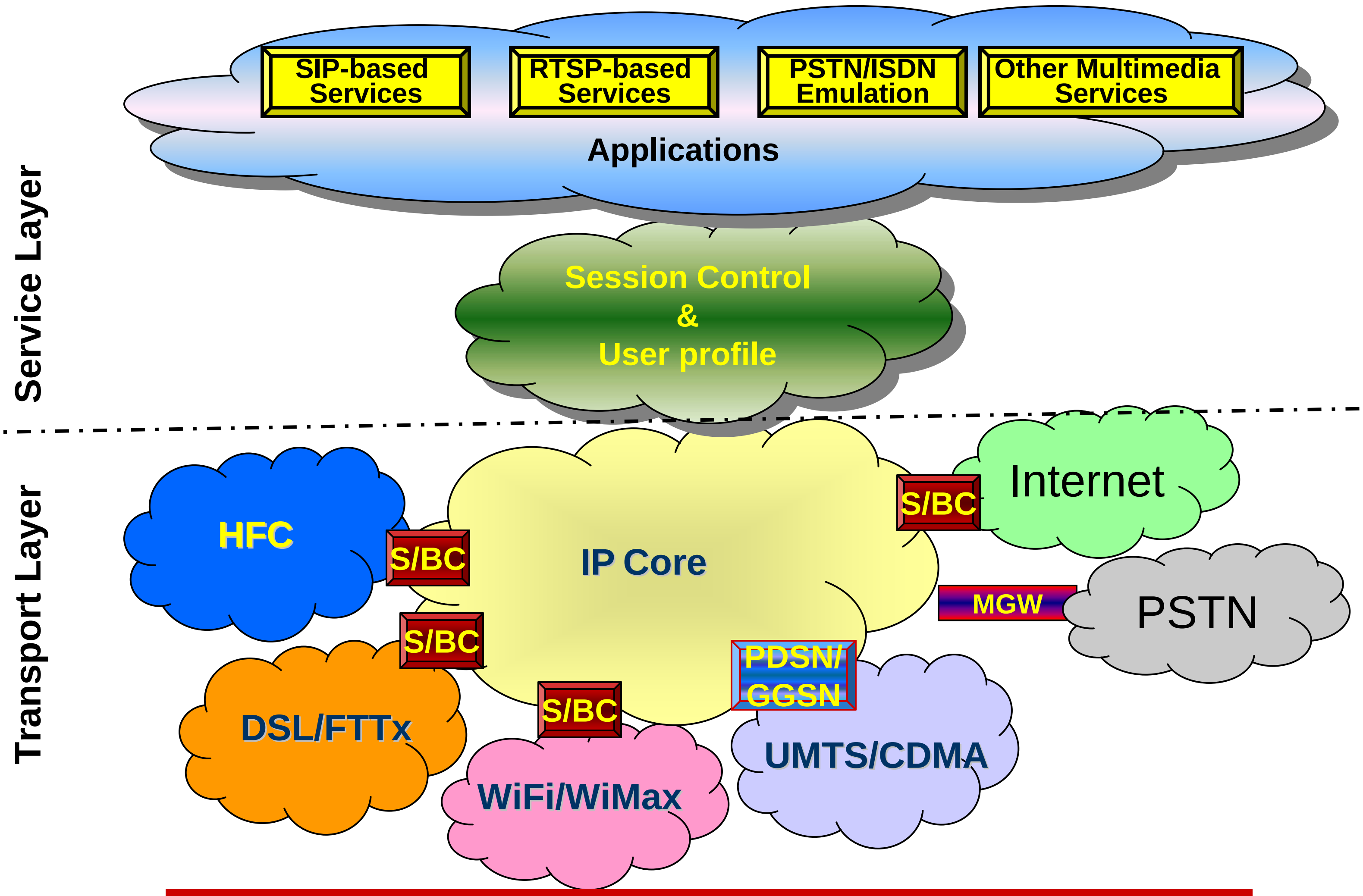
amitm@lucent.com; **dongsun@lucent.com**

Lucent Technologies
Bell Labs Innovations

Agenda

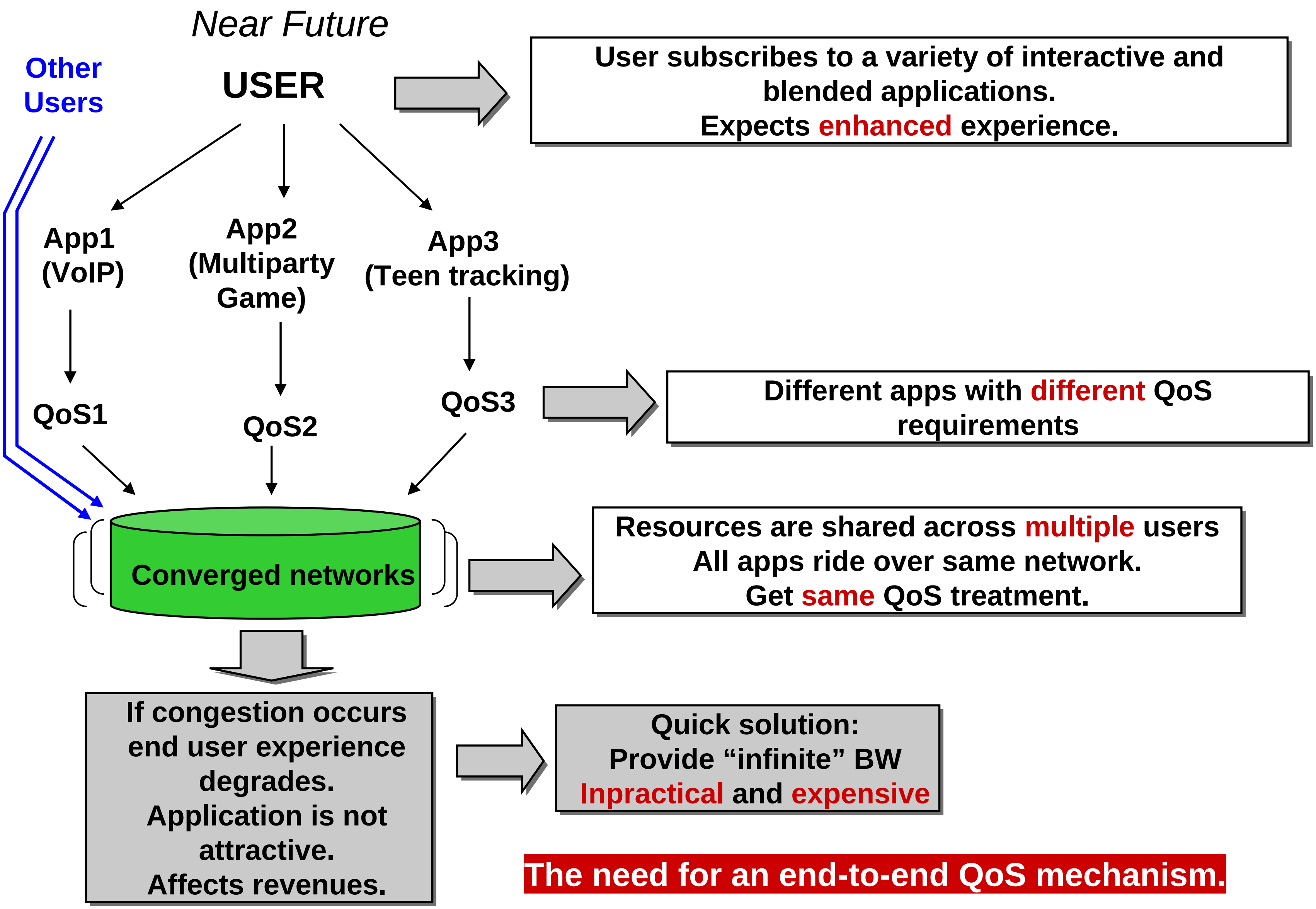
- Introduction to NGN and QoS
- Key Components of QoS
- Existing Mechanism/Architecture
- Challenges of End-to-end QoS
- Current Standards Developments
- Summary
- Appendix

NGN Architecture



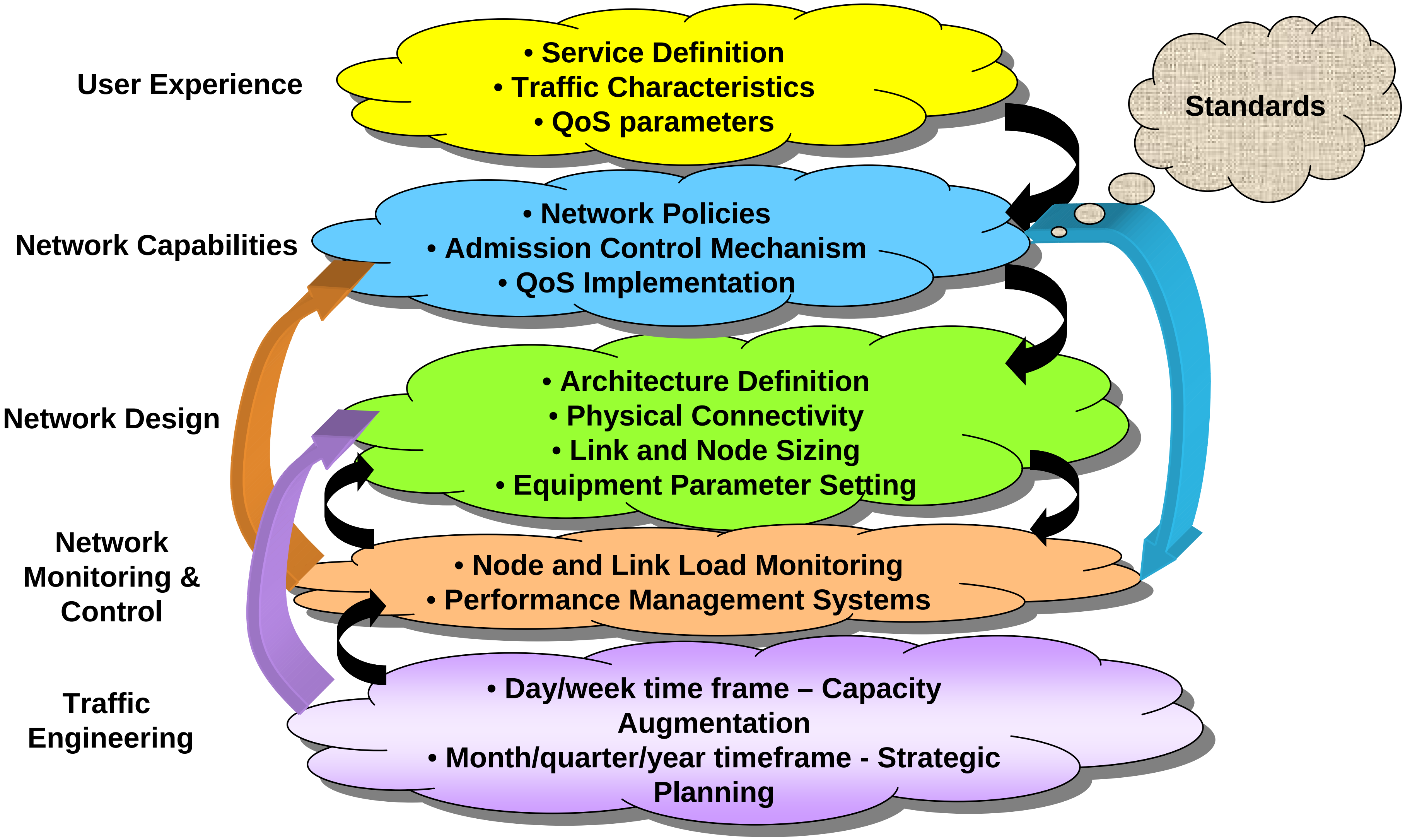
Multiple applications supported over a common infrastructure

The Challenge



The need for an end-to-end QoS mechanism.

Key Components of End-to-End QoS

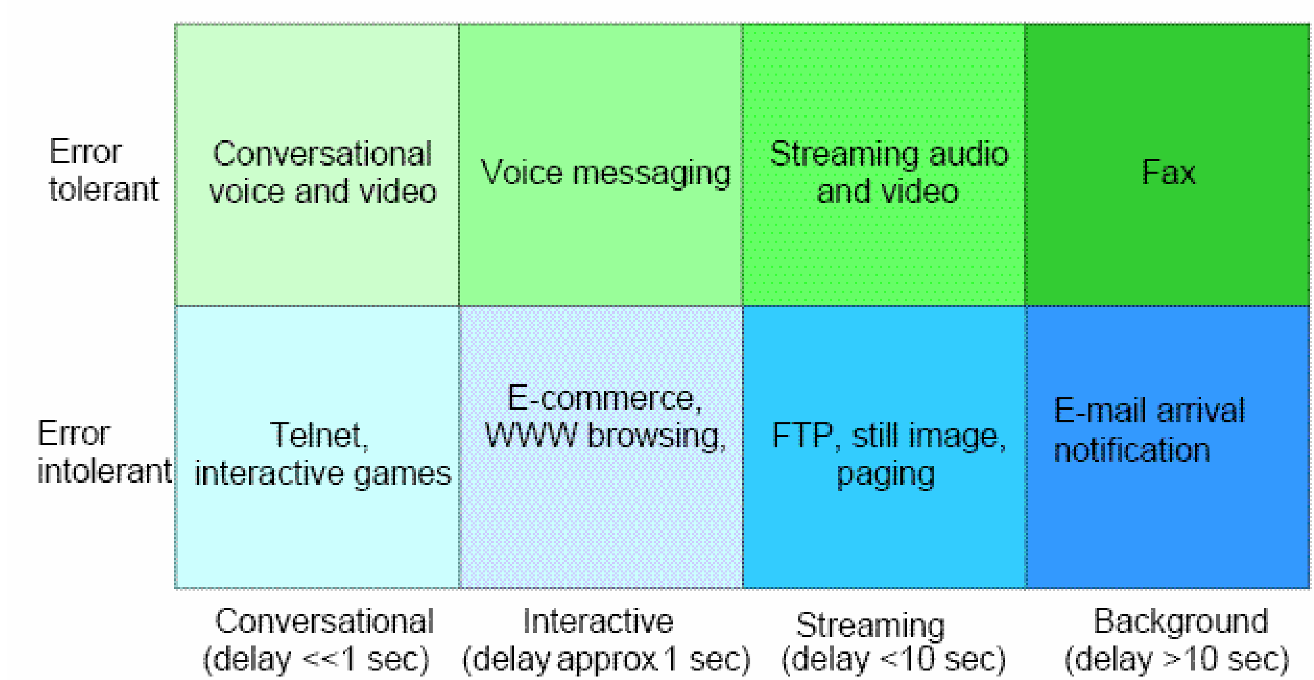


6 Key Components

User Experience

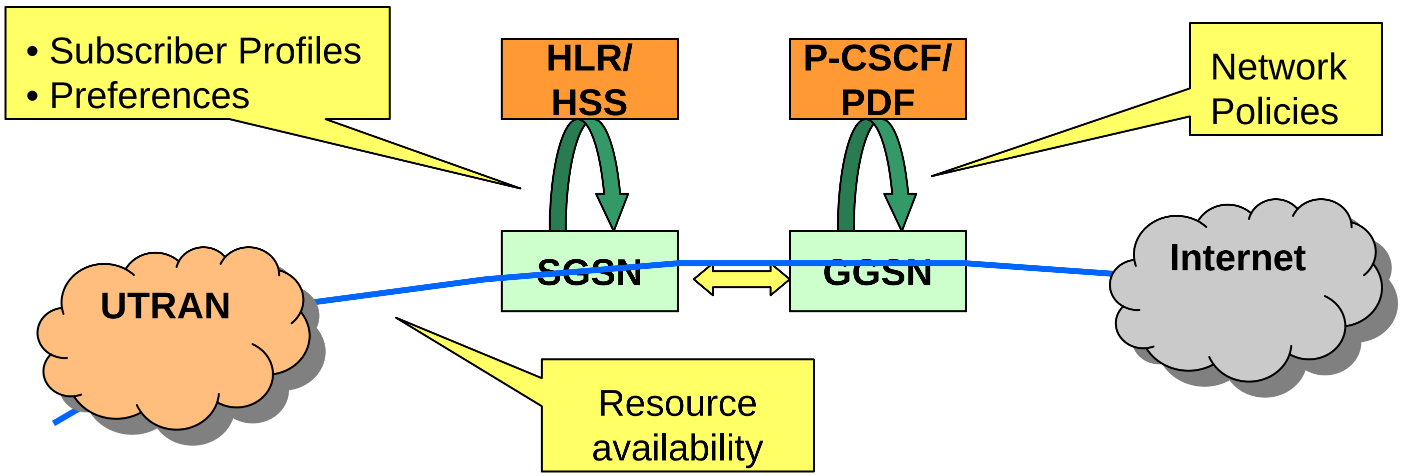
Example Services and their Characteristics

Traffic Class	Conversational	Streaming	Interactive	Background
Typical Characteristics	▪Symmetric bidirectional ▪Delay and jitter very important	▪Mostly unidirectional ▪Jitter very important	▪Asymmetric traffic ▪Sensitive to delay, errors	▪Not time sensitive ▪Sensitive to errors
Sample Applications	VoIP, video conferencing, peer-to-peer gaming	Audio/video streaming, infotainment (weather, traffic etc.)	Web browsing. server-based gaming, messaging	File transfer, e-mail



Different applications have different delay, jitter and bit error rate tolerance

Network Capabilities
Policies

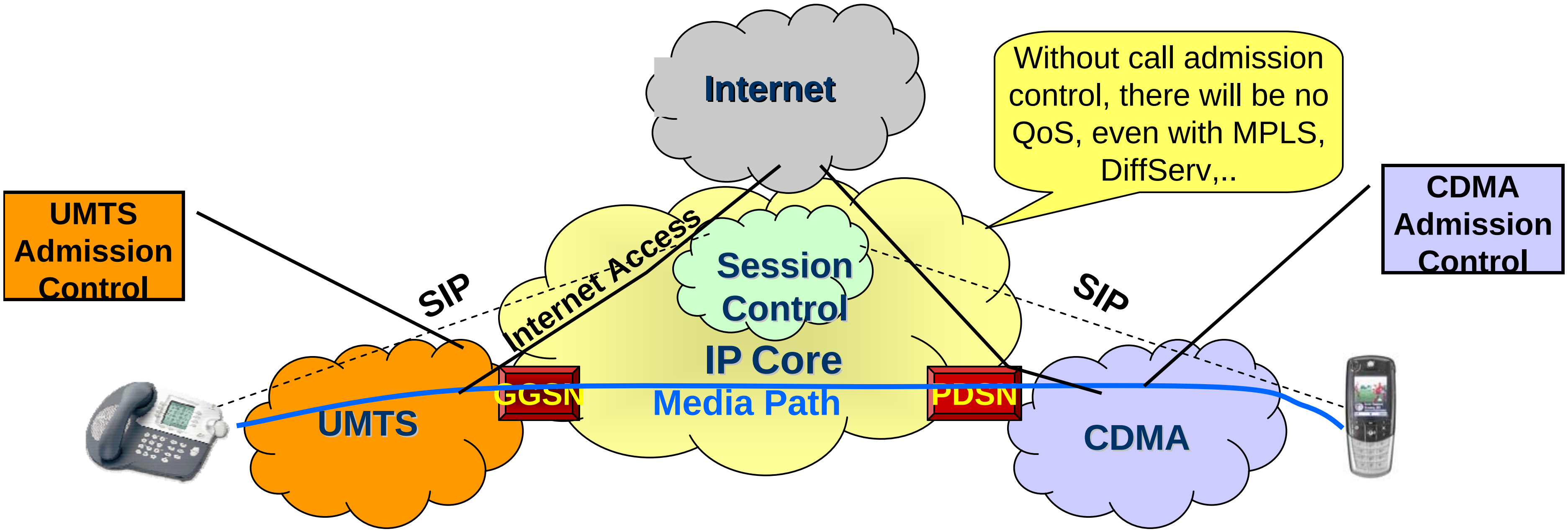


- **Policies key component of Next Gen Networks**
- **Policy Types:**
  - User profiles/subscriptions (stored in HLR/HSS)
  - Network Policies (e.g., codec selection based on network load)
  - Security
  - Services
- **QoS will be impacted by specific policies allowed/selected dynamically**

Challenge is to identify impact of policies on architecture, design, monitoring/management

Network Capabilities

Admission Control



- **What: Network decision to allow setting up of a new session based on multiple criteria**
- **How:**
  - Over provisioning – no “control” >> how much to overprovision what is the cost?
  - IntServ model – reserving bandwidth for every session >> not scalable in core networks
  - End-to-end probing – only an instantaneous view of network >> no control once flow is admitted
  - Pre-provisioned trunks – bandwidth partitioning >> not efficient
  - Pre-provisioned trunks with flexible bandwidth >> efficient but complex
  - Admission control with link level measurements >> normalized view of link load

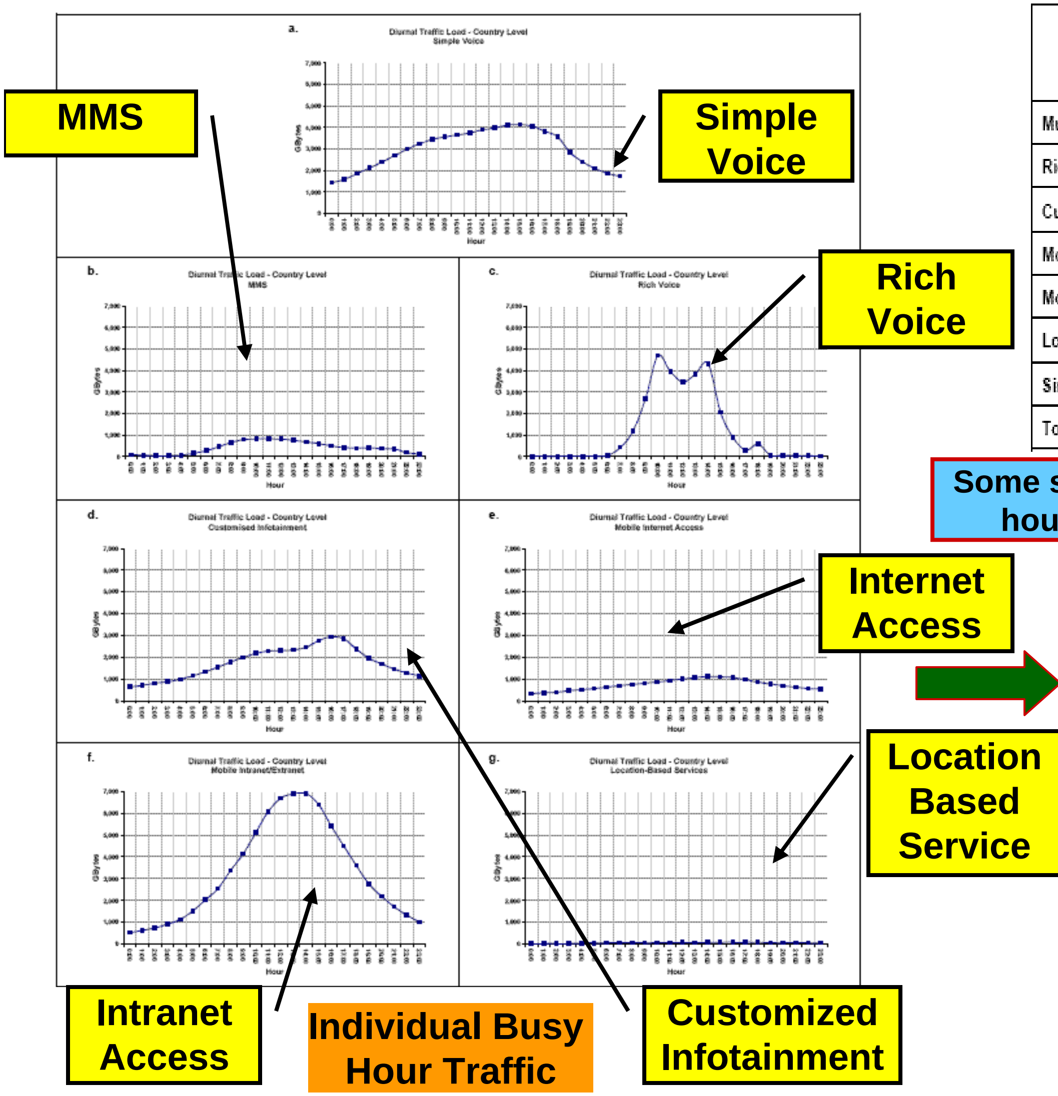
End-to-end Admission Control: Possible today only in some access networks

Network Capabilities
QoS Implementation Mechanisms

- **Classification** (e.g., identify VoIP)
  - 802.1d – NC, VO, VI,…
  - ATM – CBR, RtVBR,…
  - DiffServ – AF, EF, …
- **Queuing** (e.g., which queue to put VoIP packets, what should be the size)
  - Priorities associated with queues **(VoIP is high priority), Q0, Q1,…**
  - Buffer space **(short buffer for VoIP to avoid queuing delay)**
- **Scheduling** (e.g., which queue to process next and how many packets)
  - Priority Queuing, Weighted Fair Queueing,…
  - Order of extracting packets from different queues
  - Proportion of the egress interface bandwidth used by a queue
- **Buffer Overflow Control**
  - Random Early Detection (RED) and WRED (Weighted RED) and packet discard
  - Randomizing packet loss, as opposed to "tail drop"
  - For some traffic types, dropped packets are preferable to delays

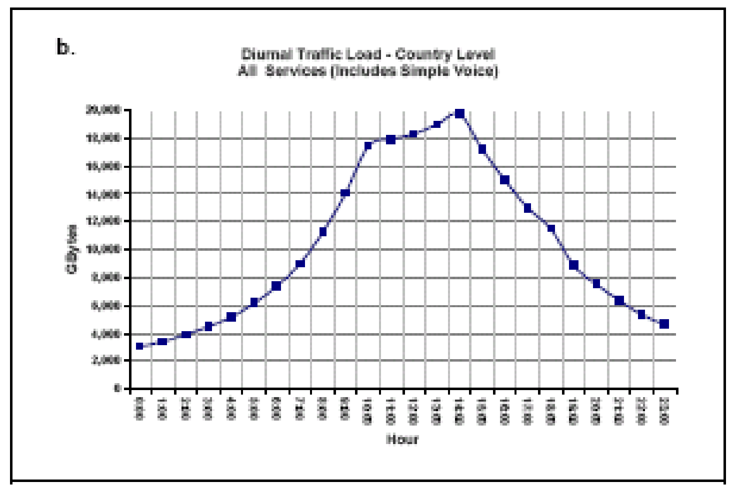
Multiple mechanisms and enablers with many details …

Network Design
Traffic Characterization



Service Category	Total Daily Traffic for the Representative Country (TBytes)	Total Traffic in Service Category Busy Hour (TBytes)
Multimedia Messaging Service	9.92	1.02
Rich Voice	28.66	4.73
Customised Infotainment	42.04	2.95
Mobile Internet Access	17.87	1.13
Mobile Intranet/Extranet Access	78.14	6.70
Location-Based Services	1.30	0.09
Simple Voice	71.28	4.26
Total	249.21	20.88

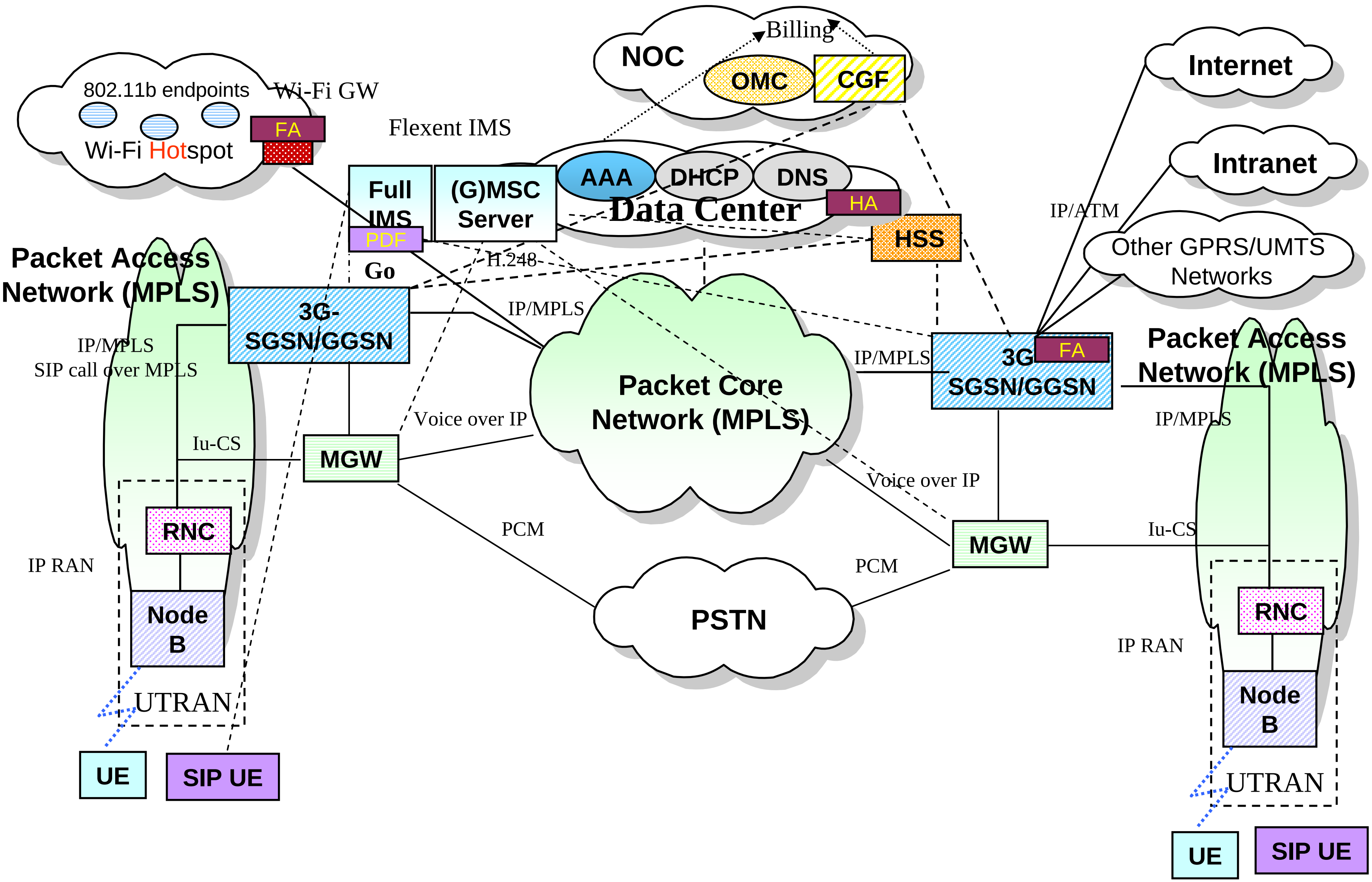
Some services with high traffic volume in their individual busy hour may not generate high traffic at the aggregate level



Aggregate Busy Hour Traffic

Common resources for aggregate BH, Individual service resources for individual BH

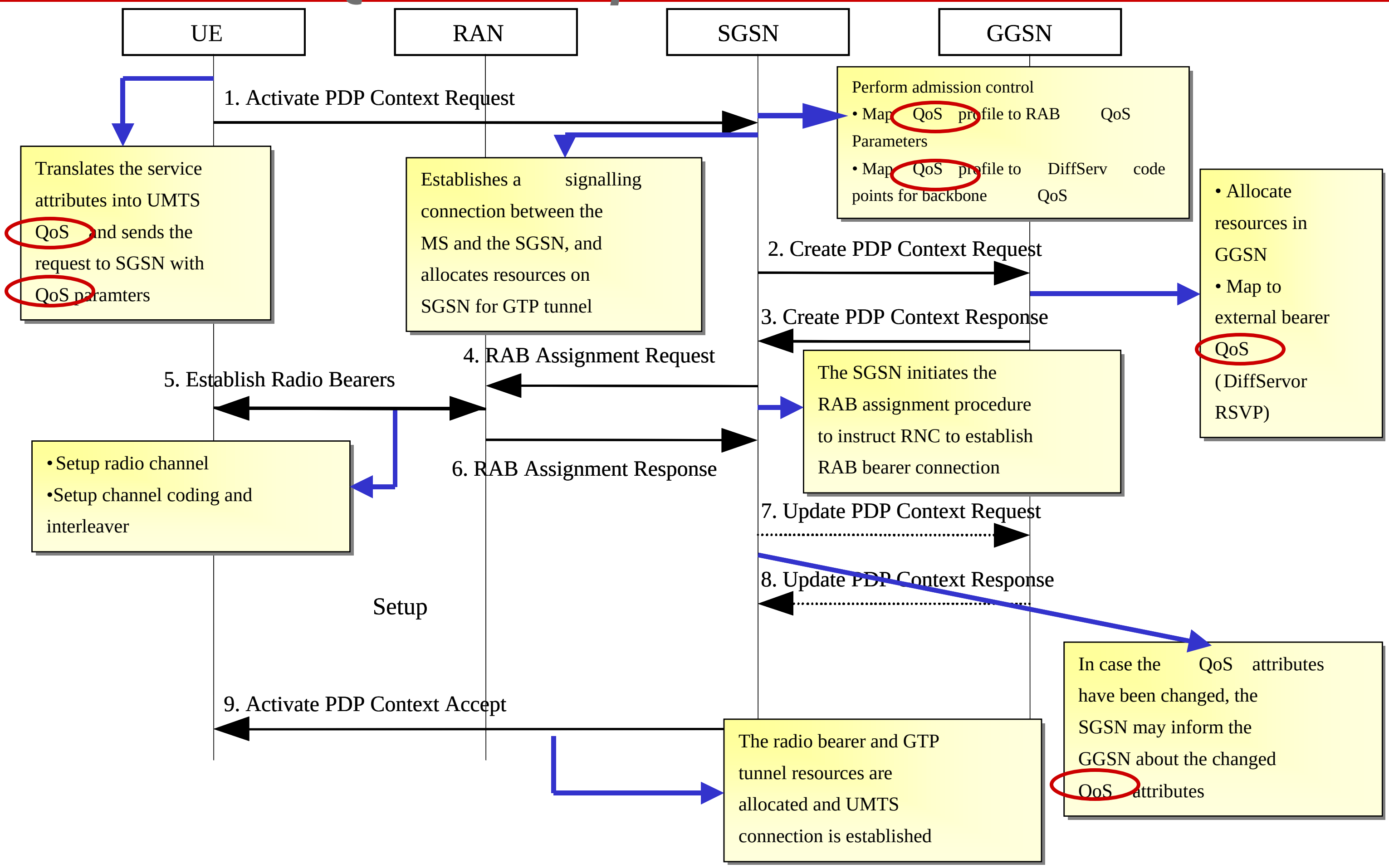
UMTS Example
Release 4/5 Architecture



Each Service Provider may come up with own implementation architecture

UMTS Example

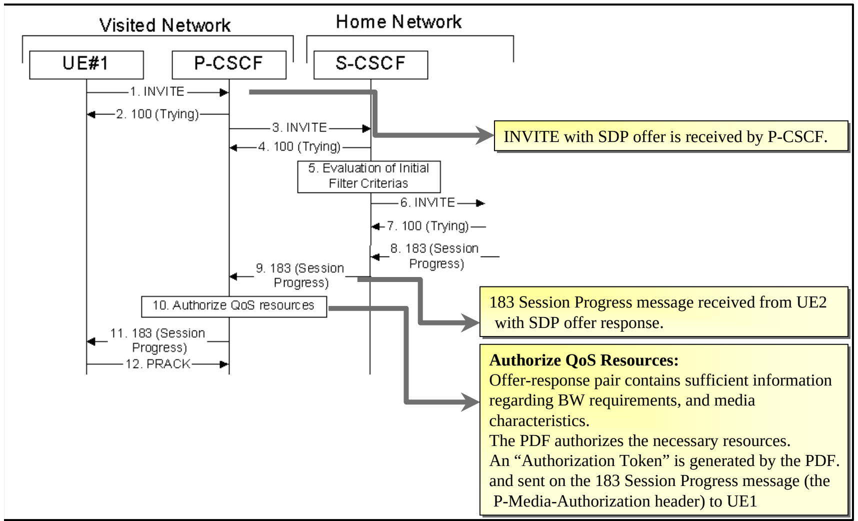
Bearer Service QoS Set Up Process – Generic Packet Service



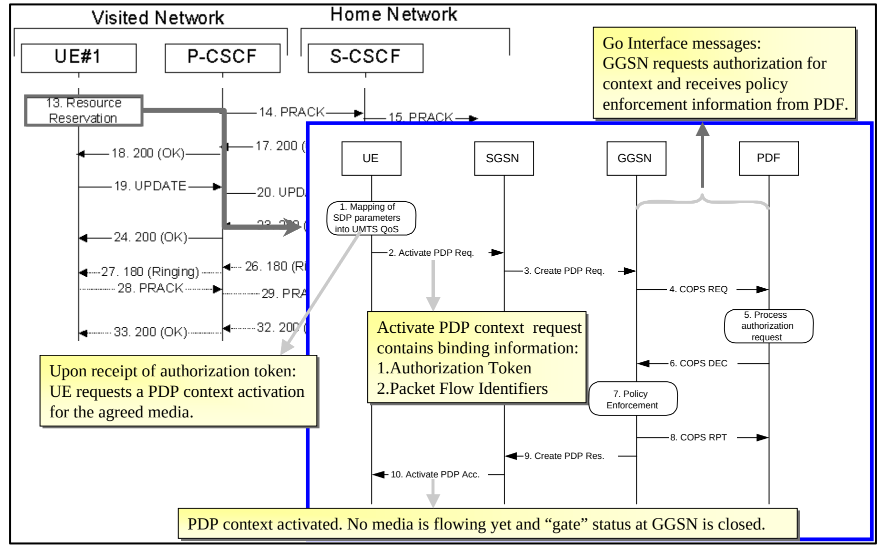
QoS procedures well established in UMTS standards

UMTS Example

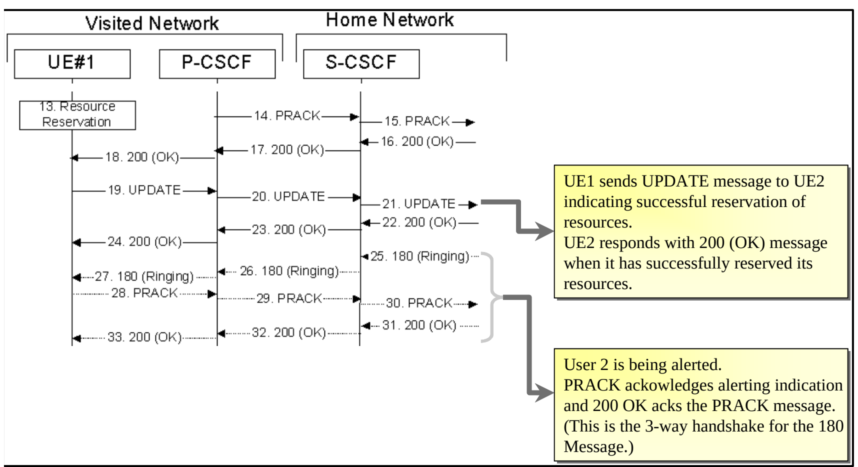
Bearer Service QoS Set Up Process – with Release 5 IMS



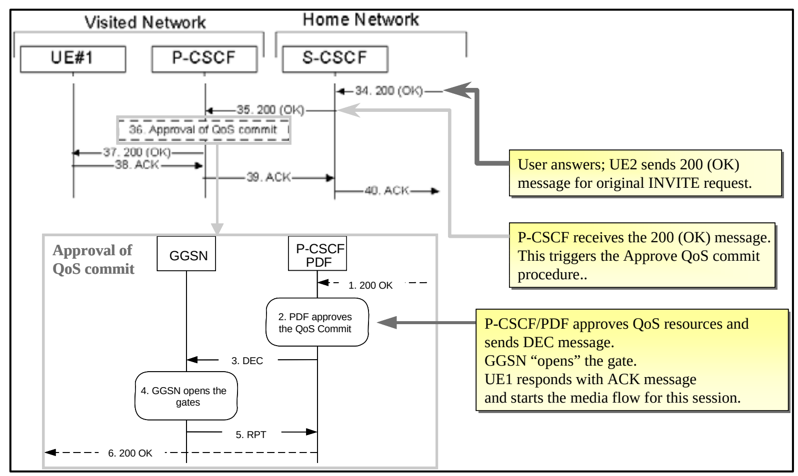
Step 1 - Resource Authorization



Step 2a -Resource Reservation (Start)



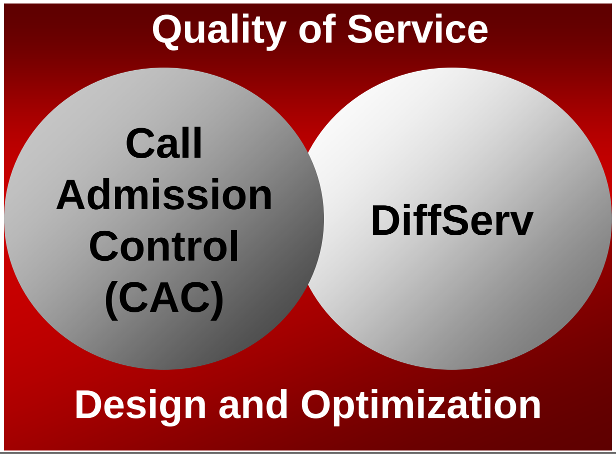
Step 2b - Resource Reservation (Complete)



Step 3 - Approval of QoS Resources

Well defined at UMTS layer, but still needs mapping to transport layer

Call-Admission-Control, DiffServ and QoS



Call Admission Control (CAC)

CAC admits a new call/service into the network based on its requirements and network resource availability so that the QoS for existing calls/services is not impacted

- Congestion affects ***All Calls*** - not just new ones, so CAC is needed to guarantee QoS
- CAC helps balance QoS while increasing network utilization

DiffServ

DiffServ provides a way for the network to differentiate among different service types to control QoS.

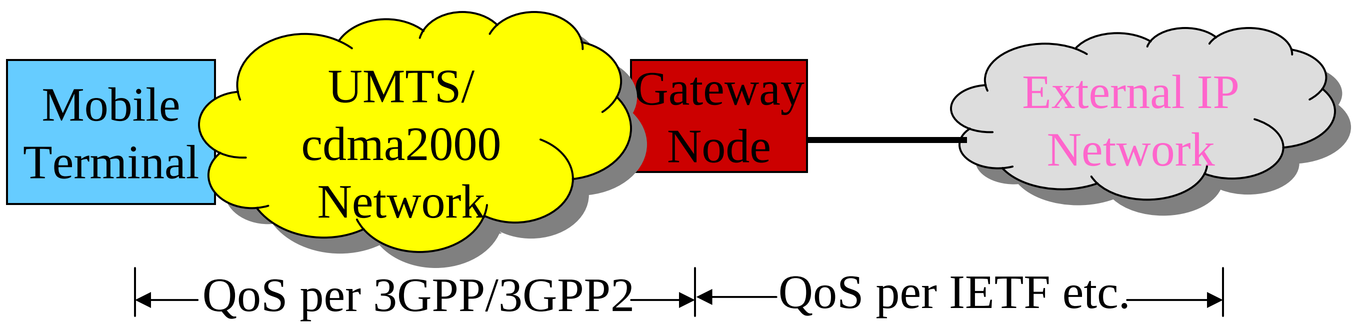
- QoS characteristics such as bandwidth, latency, jitter and packet loss can be supported at different service levels for different service types
- Service packets are marked with Type of Service (TOS) priority
- TOS priority is used to differentiate service types

- **CAC gates admission of calls**
- **DiffServ maintains QoS for those admitted**

Standards

Importance for Multi-Service Networks

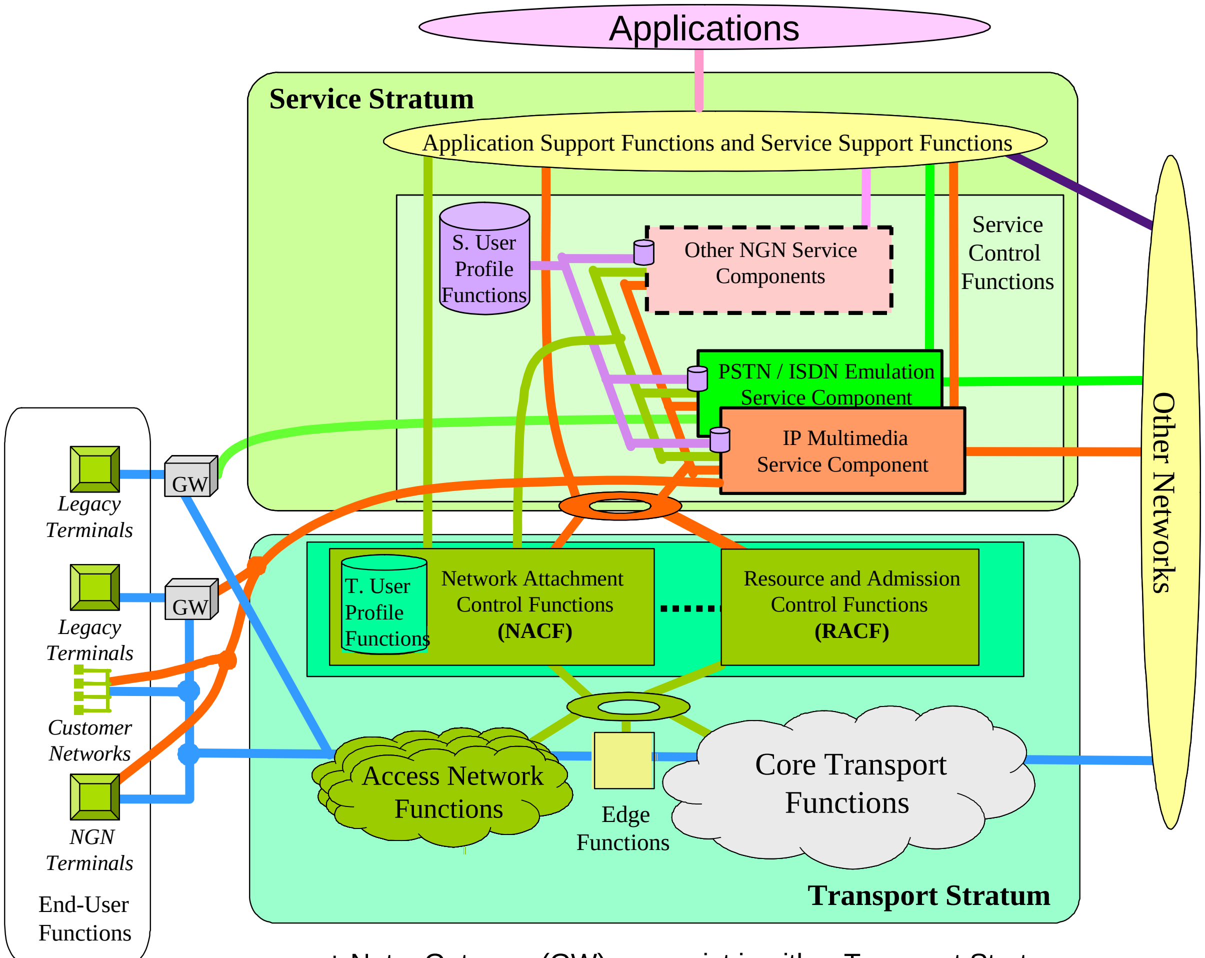
- Identify Standards used to implement end-to-end QoS and identify areas where additional standards would improve design process or implementation



- **Admission control in UMTS/CDMA2000 is provided by the underlying bearer network.**
- **No such mechanism in generic IP networks to go across network boundaries**

Hence the need for Resource and Admission Control Function (RACF) across networks

RACF in Next Generation Networks



* Note: Gateway (GW) may exist in either Transport Stratum or End-User Functions.

Resource and Admission Control Functions (RACF) serves as an intermediary between NGN services and transport networks and makes them work independently

Challenges

- The Internet “pipe” provided by SPs does not seem to be providing enough revenue growth and is increasingly perceived to be a commodity business
- The very same Internet pipes also enable competitors to support services as “best-effort” IP flows
  - Examples: “parasitic” VoIP, video telephony, gaming, video streaming, web portals, etc.
- Demanding applications such as IPTV and VoD need finer granularity (per user/app) of resource control
- Uniform resource management across diversity of applications (IPTV, VoIP, VoD) and transport types (xDSL, PON, Ethernet, MPLS, IP ..)

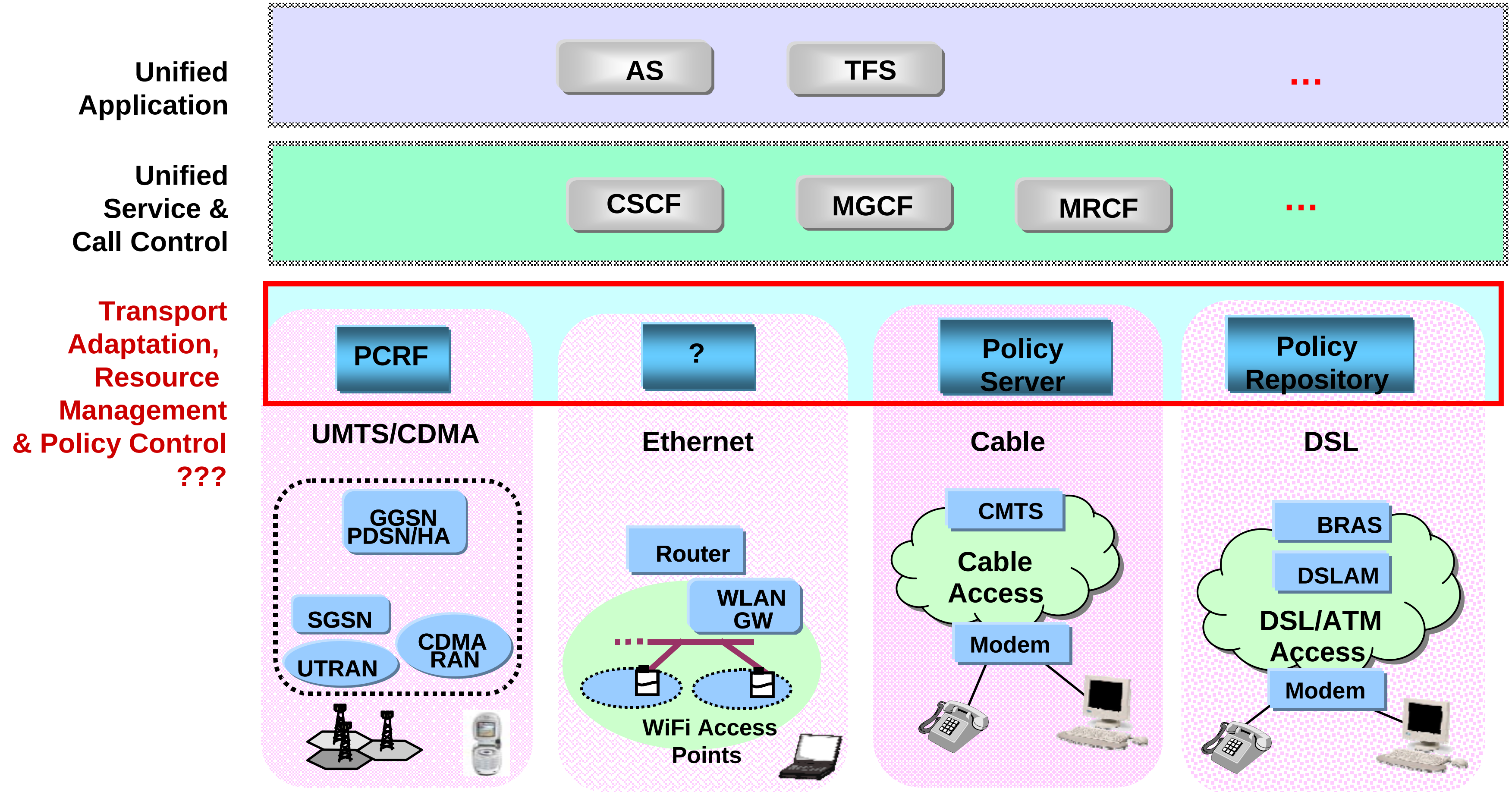
Current “pipe” model and QoS mechanisms are inadequate to support emerging multimedia applications and business model

Service and Revenue Opportunity

- Provide personalized service packages based on
  - Date rate (bps),
  - time of day,
  - QoS,
  - Service type (SDTV or HDTV)
  - Capacity consumption (bytes per month)
- Support occasional in time services, e.g.
  - Temporary service upgrade such as higher bandwidth for limited time such as “turbo button” for web browsing/downloading
- Improve revenue growth
  - QoS Differentiation of SP provided services and parasitic provider’s services
  - Usage/capacity Monitoring
  - Usage limiting for abusive applications

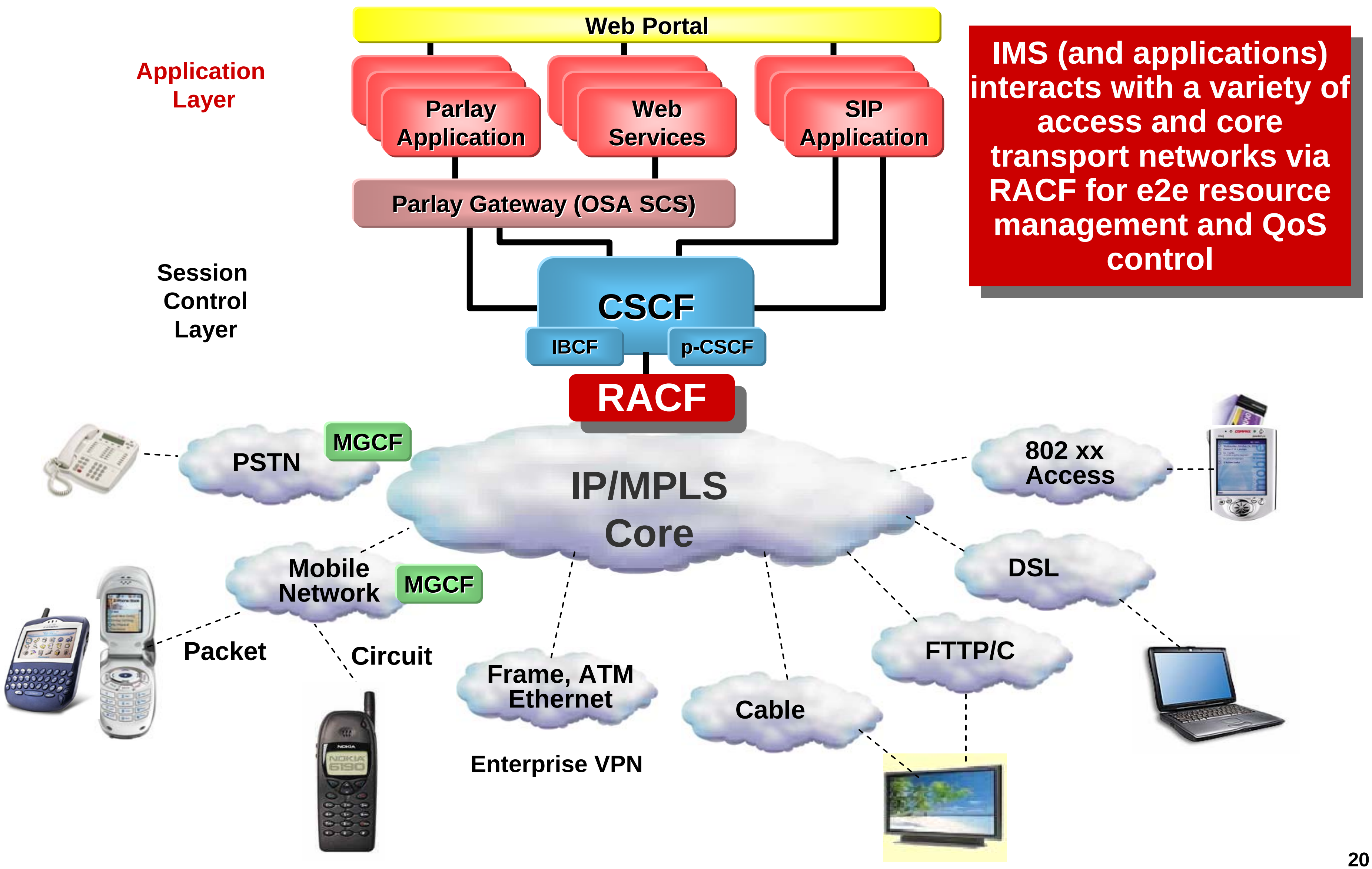
Service aware resource management capability is crucial for SPs to succeed in revenue growth and service differentiation

Unifying Resource Management

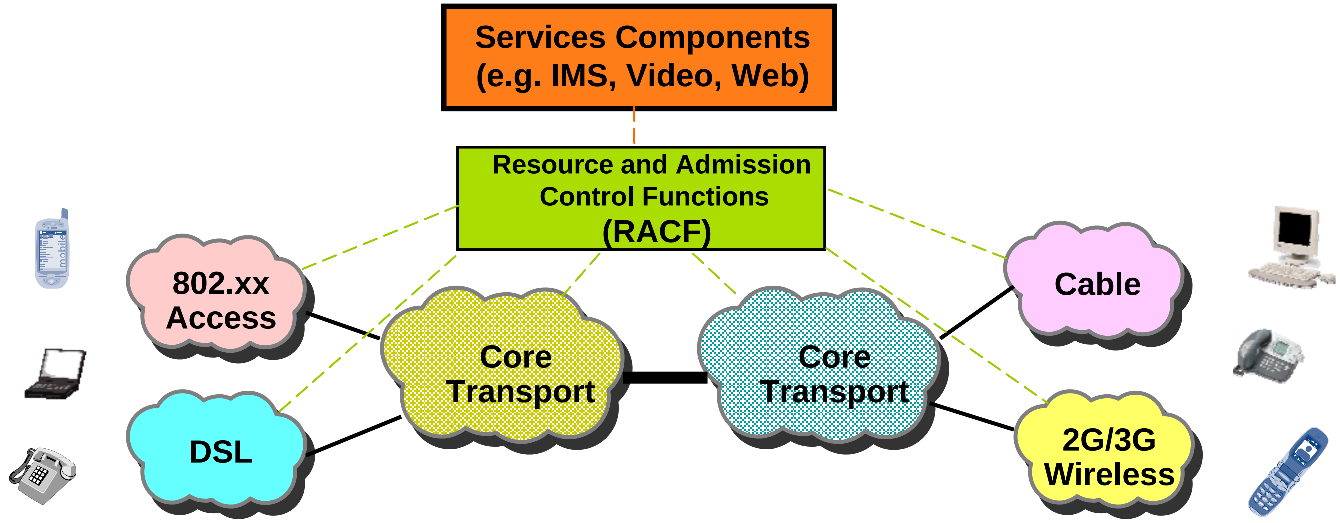


Unified resource management in support of Fixed Mobile Convergence is desirable

Target IMS Architecture



RACF's Job



- Dynamic management of a variety of resources (e.g., bandwidth or IP addresses) across varied transport networks—different technologies, administrative domains, ownerships—to achieve end-to-end QoS and provide border control
- Service independent mechanism for transport resource management common to various applications (e.g., IMS, IPTV and web based service etc.)
  - Admission control for managing network congestion
  - Policy-based arbitration of *many-to-many* relationship

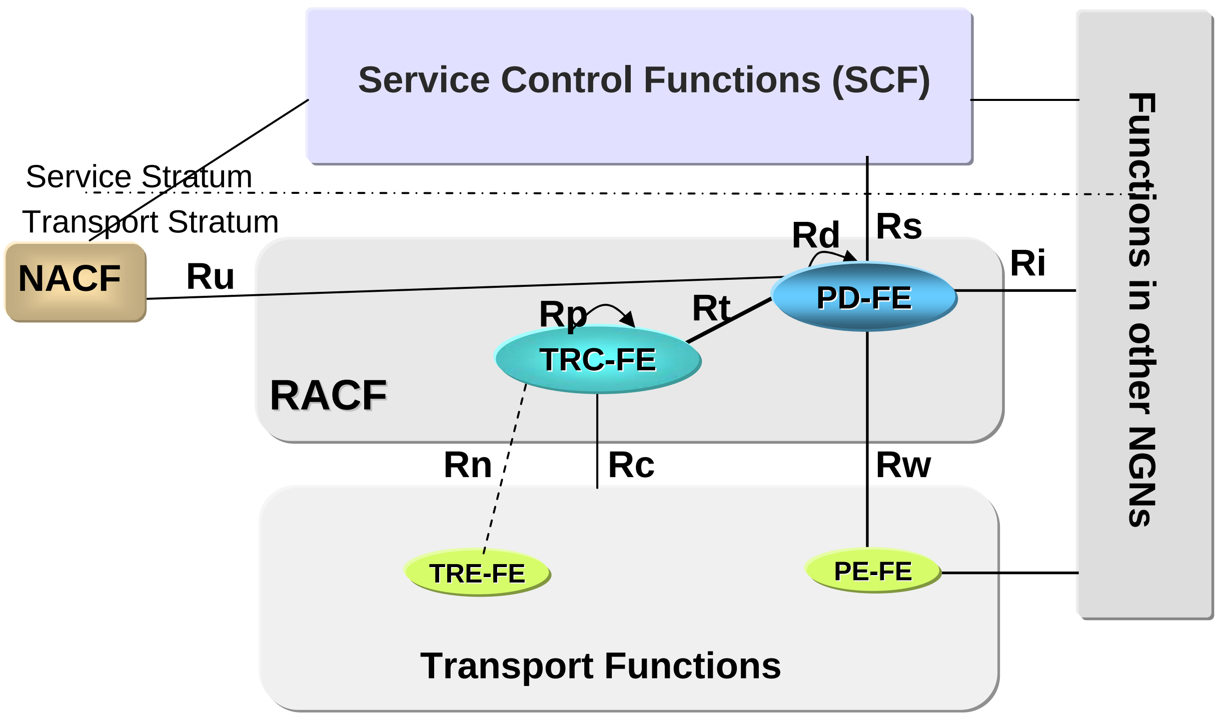
RACF provides an end-to-end solution in support of policy based resource management

Status of Resource Management Standards

- Work is being done in several SDOs based on similar rationale with different scope
  - ETSI/TISPAN RACS Release 1– limited to broadband access networks
  - ITU-T RACF– including the access as well as core network, intended for a variety of network technologies e.g fixed and mobile, core transport (e.g. MPLS)
  - 3GPP PCC and 3GPP2 SBBC are doing similar work, but mainly focus on certain wireless access networks with specific technologies
- ITU-T RACF Release 1 (Recommendation Y.2111) was consented in July ’06

Multiple SDOs are embarking on the subject with different scope and focus

ITU-T RACF Architecture



- Within the NGN architecture, the RACF acts as the arbitrator for resource authorization and allocation between Service Control Functions and Transport Functions.
- PD-FE/PE-FE covers the decision and allocation of transport resources based on user/network policies such as time of day, priority, etc.,
- TRC-FE covers admission/traffic control, within access and core transport, based on resource availability and QoS requirements

PD-FE — Policy Decision Functional Entity ← service-based, transport technology independent

TRC-FE — Transport Resource Control Functional Entity ← service-unaware, transport-technology dependent, network-segment specific

PE-FE — Policy Enforcement Functional Entity
TRE-FE — Transport Resource Enforcement Functional Entity
NACF — Network Attachment Control Functions

RACF isolates specific attributes of services and transport networks

RACF Highlights

- RACF provides Transport Resource Management Capabilities
  - Application-driven (network-independent) “real-time” control
  - Management of transport resources within networks (access or core) and at network boundaries
  - Policy-based authorization and allocation of the resources supporting
    - End-user equipment of varying QoS control capabilities
    - *Push* and *pull* models for policy control
    - Multiple transaction models for resource authorization, reservation and commitment
    - A combination of resource management methods based on accounting, measurement and reservation
- RACF interfaces to Service Control Function (e.g. SIP Proxy Server or IMS) to allow an Application to request resources including:
  - QoS (BW Guarantees, per flow traffic shaping/policing, priority, …)
  - NAPT control and NAT Transversal capabilities
  - Gate control and other border control functions
- RACF can interface across network boundaries to support a variety of business models
  - Addresses Session Border Control Issues
  - Will integrate flow based charging capabilities (future)

RACF covers end-to-end QoS control as well as border control functions

RACF Key Elements

- PD-FE – Policy Decision Functional Entity
  - –Apply network policies to resource management requests from Service Control Functions
  - –Given an IP address pair and required BW, determine if the given flow can be supported in the network
  - –Manage resources along the flow path including NAPT Transversal and Gate Control
- TRC-FE – Transport Resource Control Functional Entity
  - – "Connection Admission Control"
  - –Monitor network resource utilization and network topology to manage path bandwidth availability (reservation and/or monitor)
- PE-FE – Policy Enforcement Functional Entity
  - –Provides media path functions such as gate control / Firewall
  - –NAPT translation and Transversal
  - –Per flow policing and QoS-marking
  - –Can report usage status to Service Control

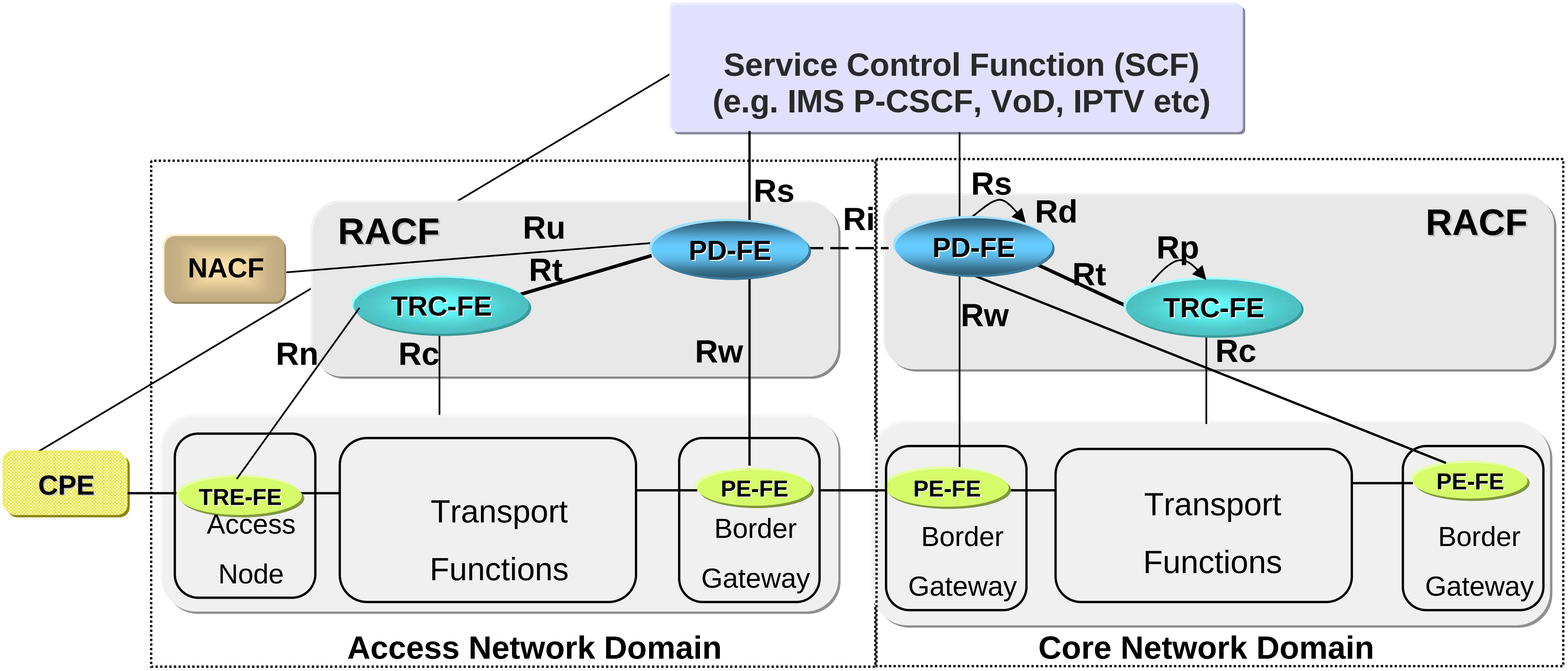
RACF consists of 2 elements: PD-FE and TRC-FE for policy control and resource control

Key Reference Point Requirements

- Rs: PD-FE－SCFs
  - For SCFs to request transport resource authorization and control
  - Information exchanged: *session ID, media descriptor*, *application QoS requirements*, *priority*, *gate or NAPT control policy*, *authorization token*, etc.
- Rw: PD-FE－PE-FE
  - For PD-FE to apply controls to PE-FE concerning NAPT, hosted NAT traversal, gating, bandwidth, packet marking, etc.
  - Information exchanged: *media descriptor*, *DSCP value*, *bandwidth committed*, *bandwidth authorized, authorization token, gate control command*, *NAPT control command, usage information,* etc.
- Rt: PD-FE－TRC-FE
  - For PD-FE to request resource availability check by TRC-FE
  - Information exchanged: *media descriptor*, *bandwidth, other network QoS requirements*, *network path,* etc.

3 main interfaces between service control, PD-FE/TRC-FE and enforcement element

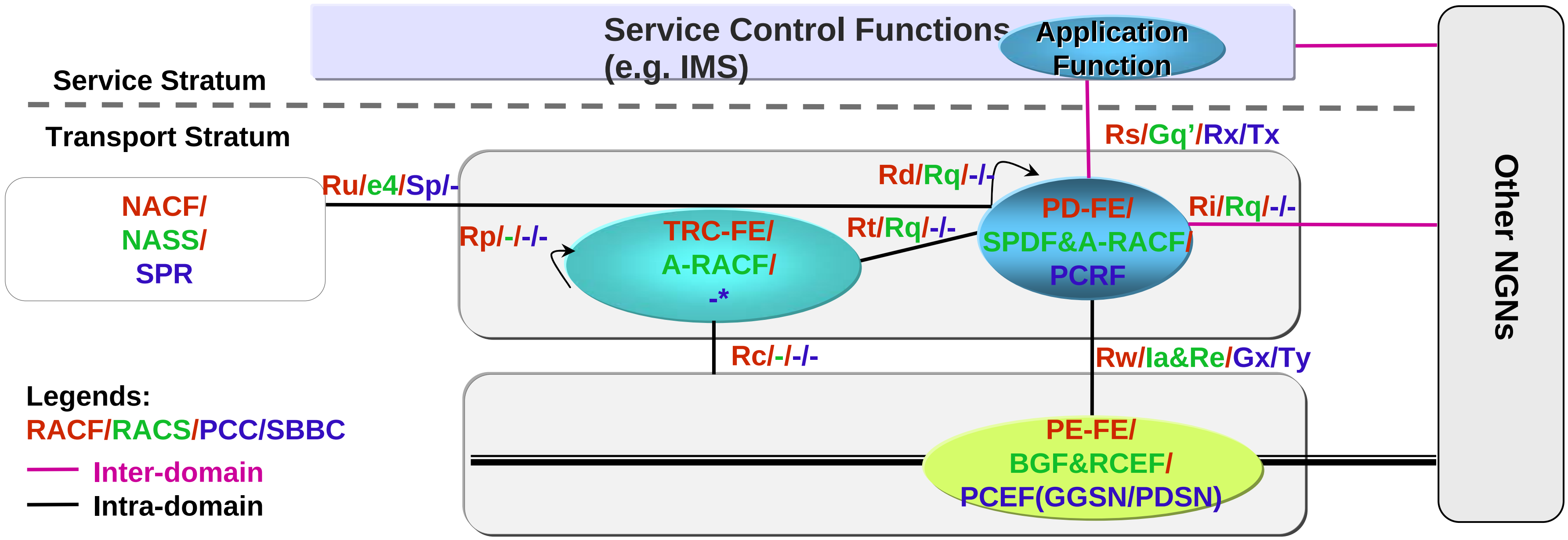
RACF Example Implementation Architecture



- Each operator administrative domain should have its own PD-FE for policy control
- Multiple PD-FE and TRC-FE instances are allowed in the same domain
- PD-FE and TRC-FE can be centralized or distributed, can be a standalone device or integrated with other network devices
- TRC-FE may control the TRE-FE for aggregation transport QoS

The deployment of RACF depends on SP's network configuration, service requirements and business model

QoS Reference Model for all Standards Bodies



	ITU-T – Rel 1	TISPAN – Rel 1	3GPP – Rel 7	3GPP2 – Rev B
Policy Decision Function	PD-FE - Policy Decision Functional Entity	SPDF - Service-Based Policy Decision Function A-RACF - Access Resources & Admission Control Function (Partial)	PCRF - Policy & Charging Rules Function	PCRF - Policy & Charging Rules Function (PDF & CRF)
Transport Resource Control Function	TRC-FE - Transport Resource Control FE	A-RACF – (Partial)	GGSN/SGSN/RNC/Node-B (Embedded, GPRS only)	PDSN/PCF/BSC (Embedded, CDMA only)
Policy Enforcement Function	PE-FE - Policy Enforcement FE residing in network devices (e.g. DSLAM/BRAS, GGSN/PDSN, border gateway)	BGF - Border Gateway Function (e.g. core Border node) RCEF - Resource Control Enforcement Function (e.g. IP Edge)	PCEF - Policy & Charging Enforcement Function (e.g. GGSN, TrGW)	AGW (PEP –Policy Enforcement Point) (e.g. PDSN)

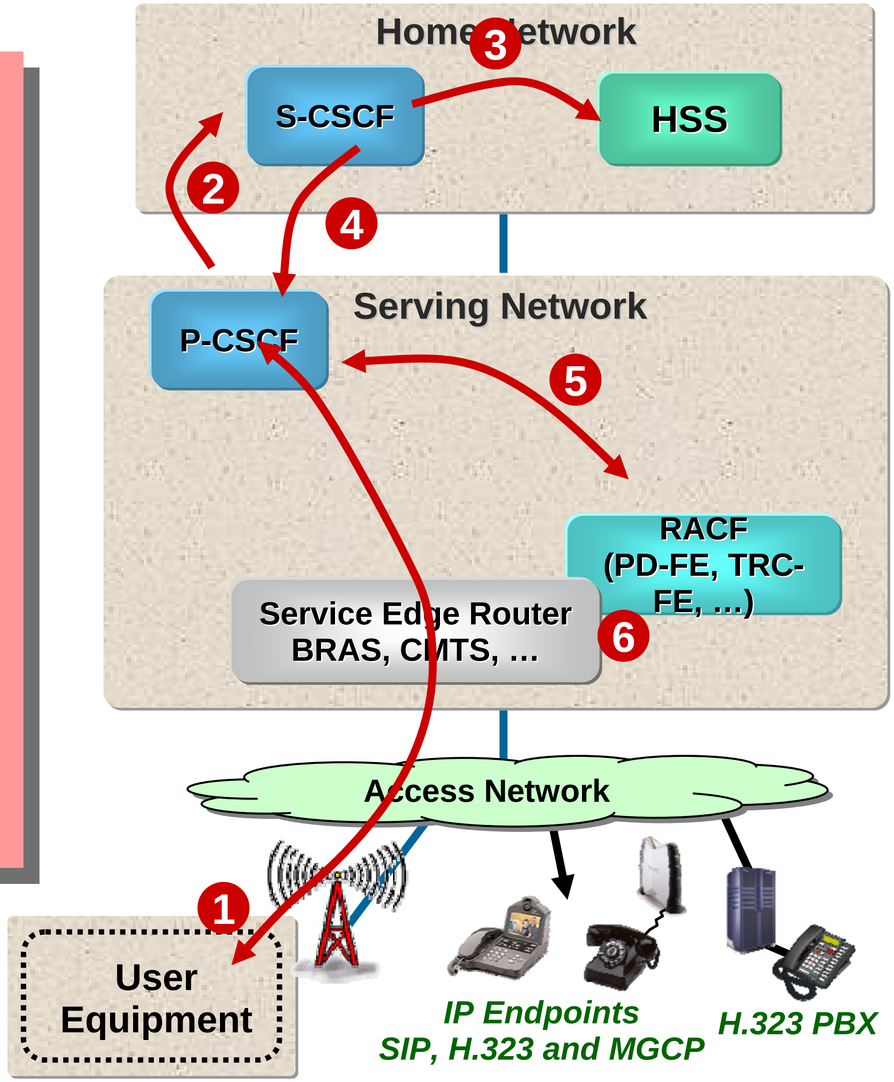
intra- domain *inter-domain*

Interface: ITU-T/TISPAN/3GPP/3GPP2

Combined view of ITU-T, TISPAN, 3GPP, 3GPP2 QoS reference models

QoS Call Flows for Wireline Networks

1. UE initiates service with media parameters in SIP INVITE
2. P-CSCF verifies that UE is registered and forwards INVITE to Home S-CSCF
3. Home S-CSCF verifies that the requested service has a valid subscription – session is routed to called network elements
4. Assuming all is well in called network, P-CSCF receives SIP 183/200 message
5. P-CSCF checks to see if resources are available with PDP (in RACF) - PD-FE pushes QoS policy to TRC-FE to request QoS
6. RACF can verify request based on local resource availability and push policy down to Service Edge Router.

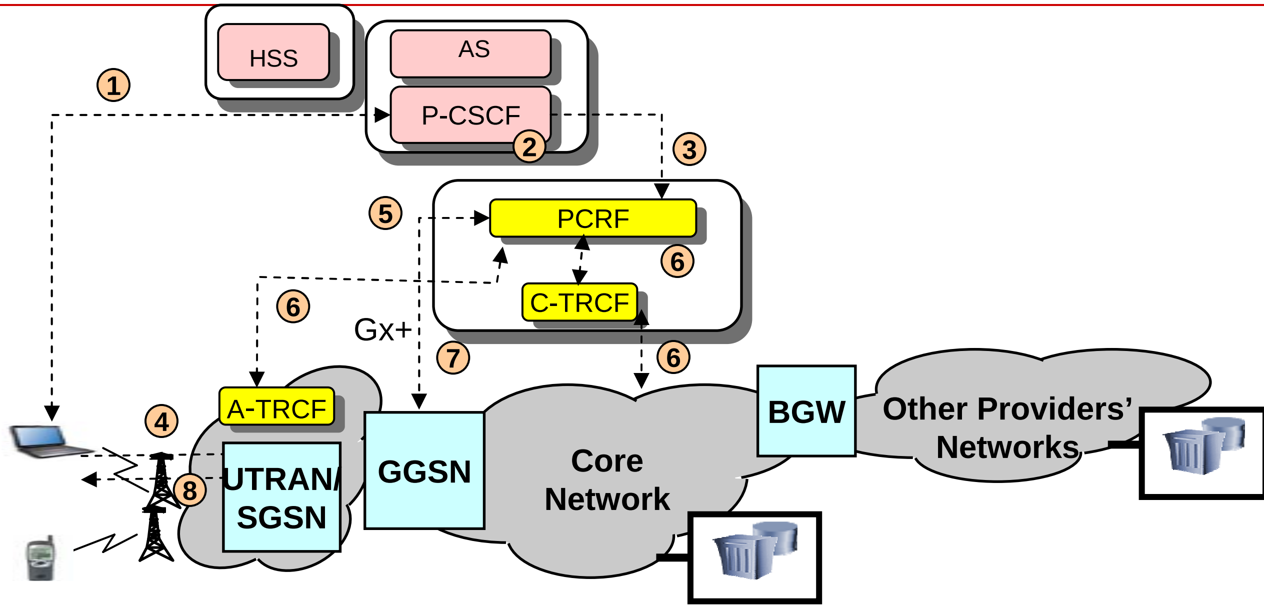


QoS parameters come from either:

- SDP parameters from the UE
- Subscription info in the HSS

Interactions among P-CSCF, RACF and network element for E2E QoS

Example for UMTS



1. User initiates SIP session. Results in SIP INVITE to P-CSCF
2. Based on negotiation between the user equipment and AS, P-CSCF determines how much bandwidth is required for the SIP session
3. P-CSCF issues a request to PCRF for Service level QoS authorization
4. User terminal initiates a Bearer Resource Establishment request to GGSN
5. GGSN requests PCRF to get authorization
6. PCRF checks with A/C-TRCF(s) for resource availability. A/C-TRCF(s) will check resources with the UTRAN, SGSN and Core Network.
7. PCRF sends authorized QoS to GGSN for policy enforcement
8. GGSN responds to User terminal with Bearer Resource confirmation.

Interactions among PCRF, A-TRCF and C-TRCF will lay the foundation for E2E QoS

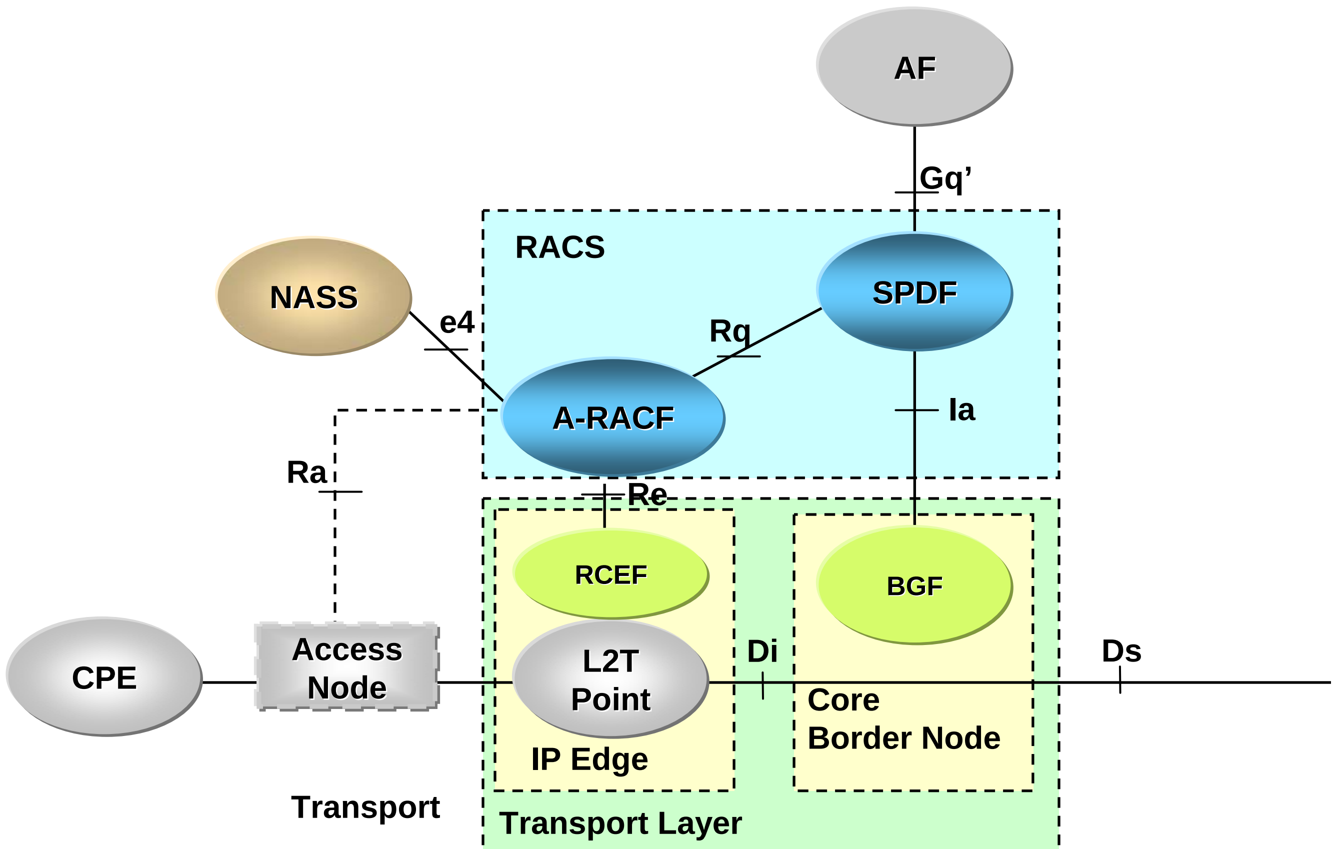
Summary

- Service based, dynamic policy based control of network resources is viewed as crucial to the profitability of telecom operators
- Fine granularity of QoS enables NGN/IMS to support a large range of QoS sensitive applications
- End-to-end service quality assurance is essential for the fixed mobile convergence in support of anytime, anywhere ubiquitous services

Appendix

ETSI TISPAN Resource and Admission Control Subsystem (RACS)

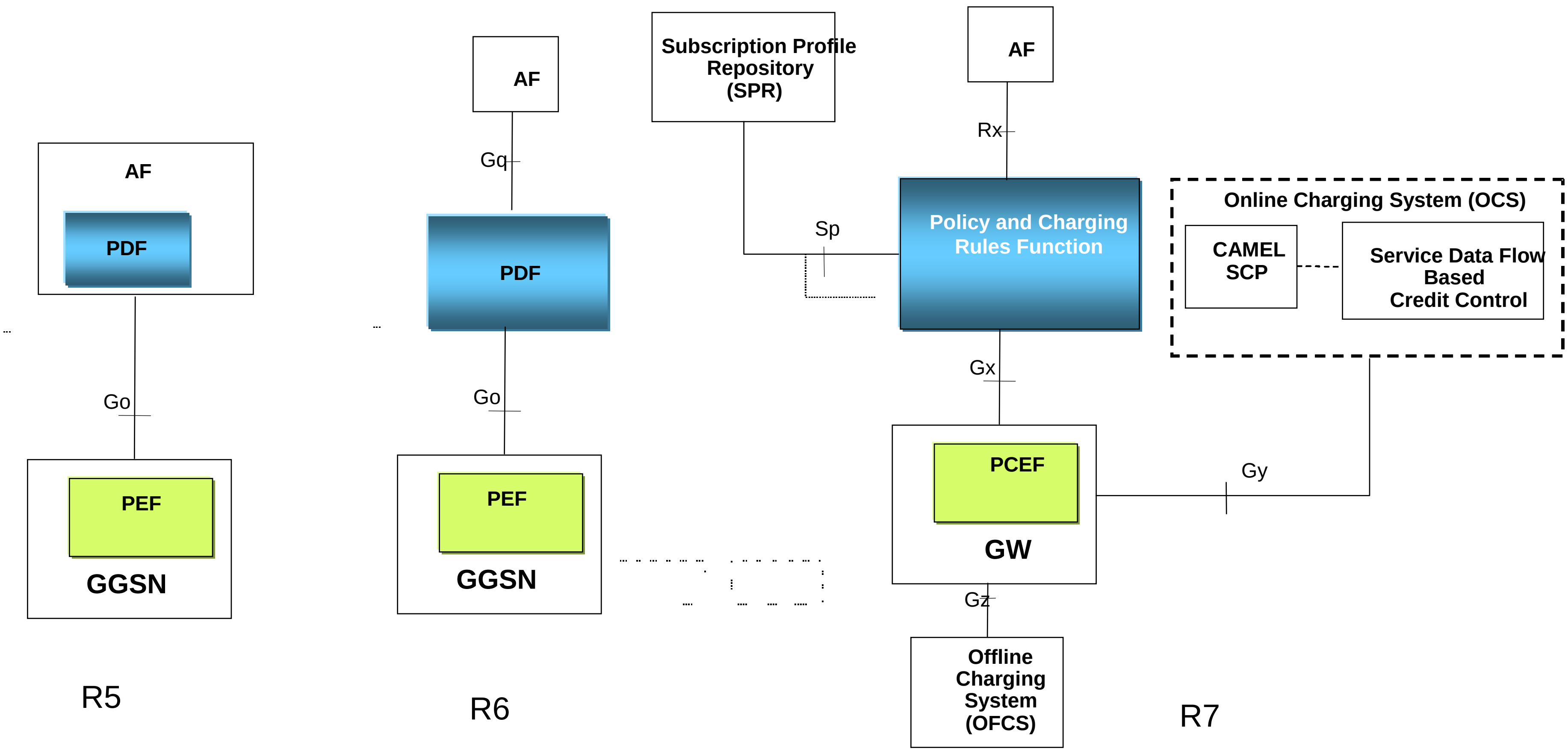
(V1.6.8, 12/05)



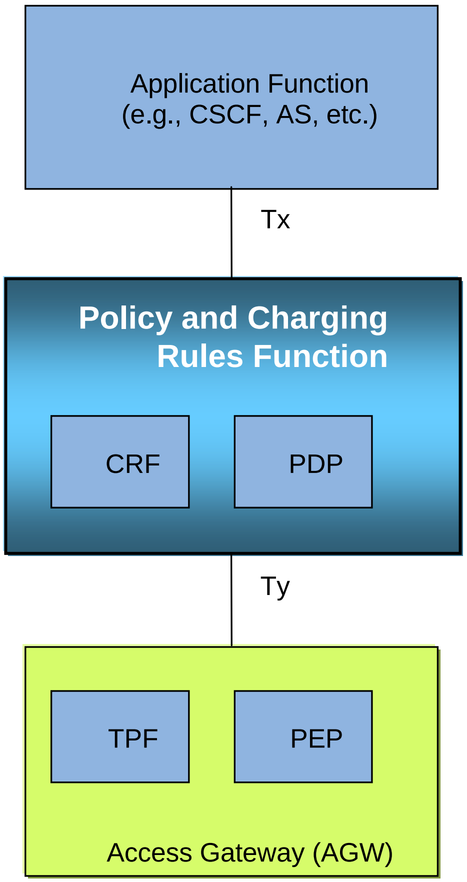
AF - Application Function
NASS - Network Access Attachment Functions
L2T - Layer 2 Termination
CPE - Customer Premise Equipment

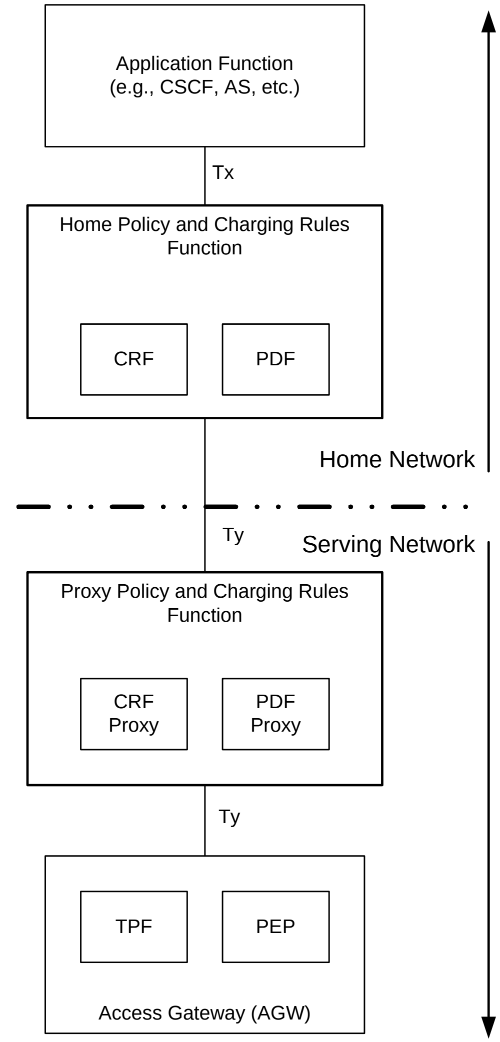
SPDF - Service Policy Decision Function
A-RACF – Access Resource and Admission Control Function
RCEF – Resource Control Enforcement Function
BGF - Border Gateway Function

3GPP PCC Architecture



3GPP2 SBBC Architecture (Rev. B)





Broadband Access