광대역 이동통신망에서의 트래픽 관리 기술

장 영 민

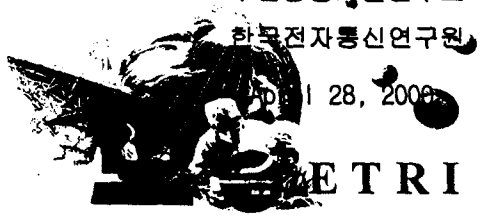
(042)860-5454

yjang@etri.re.kr

무선방송기술연구소

한국전자통신연구원

April 28, 2006



Contents

Wireless Networking Trends

▣ Wireless Networking Trends:

- ◈ Networks and services:
  - wired networks ===> wireless networks (IMT-2000, PCS, OBP satellite , wireless LAN), mobile computing
- ◈ Multimedia broadband applications :
  - dense Wavelength Division Multiplexing (DWDM) allows 64 wavelengths in a single fiber (IP over WDM using OOI)
  - IEEE 802.11 WLAN 2Mbps (1997), 11Mbps (1999), 54Mbps (200X)
  - IMT-2000 2Mbps (2001)
- ◈ Circuit Switching (CS) ===> Packet Switching (PS)
  - CS: telephone circuit, a fixed route, and guaranteed QoS
  - PS: Internet, a dynamic route, and statistical sharing of resources
  - if users are bursty (e.g., On-Off), then PS is advantageous
  - several VCs (virtual circuits) use a common VPs (virtual paths)
  - QoS requirements (CLR, delay, delay variation)
  - data over voice ===> VoIP (voice over data): IP needs QoS and signaling
- ◈ STM/Packet (IP, ATM) hybrid ===> All IP
  - Internet Engineering Task Force (IETF) is the center of action
  - attendance at ATM Forum and ITU is down

Wireless Networking Trends(Cont.)

- ◈ Seamless communications between all radio access networks (RANs)
  - public RAN, 802.11 WLAN, WPAN, satellite networks
- ◈ Asymmetric bandwidth / Multicast and multi-channel
- ◈ Hard QoS ("Hard-wire" link)===> Soft QoS ("Soft-wire" link)
  - scarcity and highly variable availability of bandwidth resources
  - link and channel characteristics : time varying: channel error
- ◈ Limited Mobility ===> Full mobility
  - variable capacity, reconfigurable end-to-end links
- ◈ Limited power consumption
  - power efficient communication protocols
  - expensive channel and limited computational capability
- ◈ Higher frequency band (e.g. 2GHz ~ 60 GHz, mm-wave)
- ◈ SDR (software defined radio) approach
  - flexible and open approach
- ◈ QoS Differentiation Not Integration
  - DiffServ , RSVP
- ◈ Long propagation delay and small buffer (OBP Satellite)

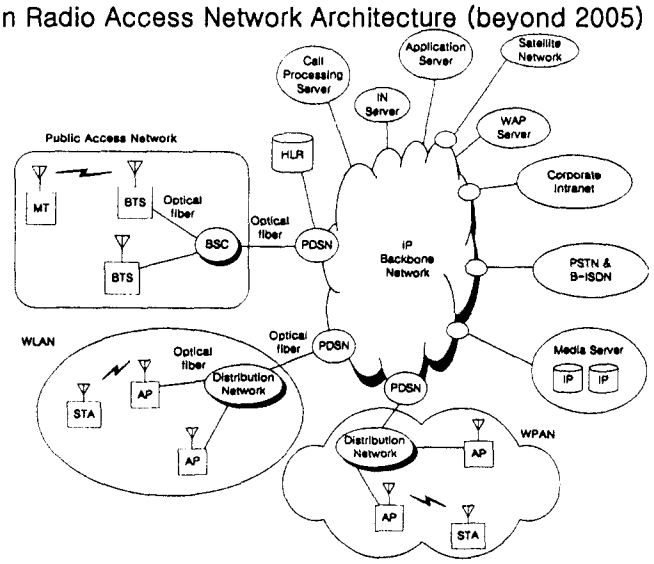
Wireless Networking Trends(Cont.)

▣ Public Radio Access Network에서 3세대 (IMT-2000)와 4세대간의 기술 비교 *(See 대한전자공학회 텔레콤지 1999년 12월호)*

주 서비스	음성전화 및 저속 멀티미디어 서비스	데이터 및 고속 멀티미디어서비스
중심주파수 / 주파수 이용효율	2GHz / 70%	밀리미터파 (20~60GHz) / 100% 이상
셀 구성 및 무선 제어	Macro / Micro 계층 셀 기지국- 제어국간 ATM 망	Micro / Pico 계층 셀 기지(무선)국- 제어국간의 IP over WDM 망
호 제어	ATM 지원 호 제어 Hard QoS 기반 호 수락제어	IP 지원 호 제어 Soft QoS 기반 호 수락제어
무선접속 프로토콜	고정 / 예약 할당방식 ARQ	지능형 동적 예약 할당방식 Hybrid ARQ
정보전송속도 / 오류율	144k~2Mbps / 10^{-3}~10^{-6} 급	2M~150Mbps / 10^{-6}~10^{-10} 급
트래픽 모델	Poisson 모델	Bursty 및 self-similar 모델
무선접속 물리계층	CDMA 전송 FDD / TDD 단일 / 다중반송파 OCQPSK / QPSK 길쌈 / 연접 / 터보코드 부분적 가변 변복조 직/ 병렬간섭신호 제거 적응형 안테나	CDMA / TDMA TDD /FDD 혼용 다중반송파 OFDM / OCQPSK / QAM 터보코드 적응형 실시간 가변 변복조 병렬간섭신호 제거 적응형 안테나
Roaming	이종간에 제한적인 Roaming	글로벌 Roaming
보안 기술	비밀키/ 공개키 혼용	공개키 기반 인프라 방식

Wireless Networking Trends(Cont.)

▣ 4th Generation Radio Access Network Architecture (beyond 2005)



Traffic Management Tech.

- ▣ QoS Architecture
  - ◈ Network congestion
  - ◈ Cause: fluctuations in the availability of allocated network resource to a connection
  - ◈ Solution: combined application-network effort
    - diverse data rate and error rate
    - diverse delay, jitter tolerance, and mobile device capabilities
  - ◈ Two possible approaches for QoS architecture
    - fixed-level of QoS in the network (does not support many application)
    - very high degree in pre-provisioning (economically impossible)
  - ◈ Two main purposes of general QoS architecture
    - avoid congestion
    - elimination of congestion over some reasonable time
  - ◈ Components of general QoS architecture
    - traffic characterization
    - CAC
    - resource reservation
    - cell scheduling
    - QoS/VC routing

Traffic Management Tech. (Cont.)

- ◈ Multicast applications
- ◈ Multilayered services
- ◈ Substream filtering
- ◈ Substream scheduing and error control
- ◈ All substreams have different QoS requirement
- ◈ Adaptive and power/spectrum efficient manner to meet individual substream QoS

- ▣ Digital Video
  - ◈ user level: max (quality) and min (cost)
  - ◈ encoding: max (quality) and min (bit rate)
  - ◈ transfer : max (utilization) and min (queuing delay)
  - ◈ Bit rate control : max (uniformity of quality) and min (bit rate variability)
  - ◈ Error control: max (error recovery) and min (overhead)
  - ◈ Video coding:
    - energy compaction: interpolative (B), predictive (P), and intraframe (I)
    - quantization: scalar or vector quantization
    - entropy coding: huffman, Lempel-Ziv coding, arithmetic coding and etc.

Traffic Management Tech. (Cont.)

- Three General Service Classes
  - Deterministic quality guarantee (CBR)
    - delay or loss never exceed a given threshold
    - deterministically bound (all traffic flow)
    - enough capacity is reserved
    - blocking only session level, not cell level
    - problem: low utilization of resource
  - Statistical quality guarantee (VBR)
    - P(delay > X) or CLR < Y
    - both session and cell level blocking
    - traffic descriptors / measurement of actual network load
    - connection request characterizes the traffic that will be sent and states its desired QoS
    - allow the new connection if QoS requirement will be met
    - the network subsequently monitors the traffic of the new connection to verify that it indeed behaves specified
  - No quality guarantee (UBR)
    - call admitted, cell level loss
    - Mbone (internet Multicast backBone) / some video applications
    - not pose stringent delay limits, allow ample buffering

Traffic Management Tech. (Cont.)

- ATM Traffic Management
  - The Goal of Traffic Management
    - provide differentiated QoS for network applications
    - promote the efficient use of network resources
    - required even with under load / measurement-based CAC
    - protect the network and the end-system from congestion
    - congestion problem will not be solved when memory/links/processors become cheap
      - congestion is a dynamic problem / static solutions are not sufficient
      - bandwidth explosion ==> more unbalanced networks
  - QoS Parameters (negotiated):
    - CLR / maxCTD (max. cell transfer delay) / CDV (cell delay variation)
  - QoS Parameters (not negotiated):
    - CER (cell error ratio) / SECBR (severely errored cell block ratio) / CMR (Cell misinsertion rate)
  - Service Categories
    - real-time traffic (CBR, rt-VBR), non-real-time (nrt-VBR,UBR,ABR, and GFR)
    - ATM FORUM TM Spec. Ver. 4.1 (March, 1999): ABR (TCP/IP, LAN traffic) flow control only . (See *http://www.atmforum.com/atmforum/specs/approved.html)*

Traffic Management Tech. (Cont.)

▣ Traffic Contract

- ◆ Traffic parameters
  - characteristics of traffic source
  - PCR, SCR, MCR, MBS, MFS
- ◆ Source traffic descriptor
  - a set of traffic parameters of the ATM source
- ◆ Connection traffic descriptor
  - specifies the traffic characteristics of the ATM connection
  - the source traffic descriptor +the CDVT+the conformance definition

▣ A Good Traffic Management Technique

- ◆ Better traffic characterization
- ◆ Measurement/estimation-based
- ◆ Dynamic and adaptive control
- ◆ Multiple class traffic
- ◆ More accurate estimation of QoS measures
- ◆ Simple and implementation

Traffic Management Tech. (Cont.)

▣ Functions for Traffic Management

- ◆ Functions of Traffic/Congestion Control (Reactive Control)
  - CAC (connection admission control) -- reject connection (busy tone)
    - BW allocation (deterministic multiplexing and statistical multiplexing)
    - BW computational algorithms (EB method, Gaussian approximate method)
  - UPC (Usage parameter control) -- cell tagging and discard -- Optional
  - CD (Selective Cell Discarding) -- discard for CLP=1 -- Optional
  - TS (Traffic Shaping) -- cell spacing/scheduling schemes
    - FIFO, priority, queuing (weighted round robin / weighted fair queuing and its variant)
  - EFCI (Explicit Forward Congestion Indication)
  - RM using Virtual Paths -- more than one VPC between O-D pair
  - Frame (AAL PDU) discard -- GFR service, rather than cell discard and dropping randomly
    - Early packet discard (EPD)/ Partial packet discard (PPD) /random early discard (RED)
  - Generic Flow Control -- ITU-T I. 150 and I. 361
  - ABR Flow Control -- ATM Forum Traffic Management Spec., feedback control (including VS/VD, Point-to-Multipoint, and support of VPs)

Traffic Management Tech. (Cont.)

▣ Congestion Control and Flow Control

- ◈ React to congestion
  - many unack'ed cells may mean long end-end delays, congested network
  - network itself may provide send with congestion indicator
- ◈ Avoid congestion: dynamic control
- ◈ Flow control: speed and resource match of sending and receiving entities
  - a sliding window (explicit flow control)
  - delayed ACK's slow down sender (implicit flow control)
- ◈ Congestion control: action taken in response to network layer (and below) congestion
  - throttling sender is but one solution to congestion
  - buffering cells
  - buffers full: cell lost (discarded) / lots of buffering: excessive delays
  - in face of loss or long delay, end-end timeout and retransmit mechanism make things worse
  - QoS guarantee (probability distribution and hard guarantee)
  - overload control (transient approach vs. steady state)
  - two types of congestion control: reactive (closed-loop) and preventive (closed-loop + open-loop) scheme

Traffic Management Tech. (Cont.)

▣ Congestion Control

- ◈ Window-based congestion control
  - send a cell with sequence number in window / control strictly end-to-end
  - on timeout, decrease congestion window size or increase timeout value
- ◈ Network-indicated congestion control
  - network marks (flag) cells passing through congested node
    - receiver see congestion indication (CI) and tells sender to slow down
    - longer feedback time
    - send already may have sent lots of cells before CI return from receiver
  - upon detecting congestion, congested node sends choke cells to source
- ◈ Rate-based congestion control
  - congestion control difficult in high speed networks
    - thousands of cells "in the air" propagating cross satellite
    - when congestion occurs, too late to react
    - avoid congestion by regulating flow of cells into networks: Leaky Buckets
- ◈ Congestion control by buffer pre-allocation
  - lack of buffering within network is fundamental cause of congestion
  - avoid congestion by allocating buffers to end-end connection
  - generally larger delay problem

Traffic Management Tech. (Cont.)

▣ Credit-Based Schemes

- ◆ Per-link, per-VC window flow control / origin from the sliding-window protocol
- ◆ Separate VC queue per link
- ◆ Queue length-based credit
- ◆ Credit >= link cell rate * link RTT
- ◆ Problems:
  - credit loss : credit resynchronization (periodically exchange count)
  - reserve the entire RTT worth of buffers : adaptive buffer allocation
  - save buffer but require ramp-up time

▣ Rate-Based Schemes

- ◆ Two approaches
  - binary bit scheme - feedback via congestion bit in cells / evolved from DECbit congestion control (many standards since 1986)
  - rate scheme - feedback via RM cells / changes explicit rate
- ◆ Combined schemes (binary feedback scheme, EPRCA, OSU, ERICA)
  - switch mark congestion in either data or RM cell
  - every 31th data cell, destination returns congestion information in RM (resource management) cell
  - rate-based prefer to credit-based because does not require per-VC queuing

Traffic Management Tech.(Cont.)

▣ Bit/Decbit scheme

- ◆ Binary feedback scheme
- ◆ Source: send data cells with EFCI bit
- ◆ Switch: measures queue length, congestion status bit = f(queue length)
- ◆ Destination: monitors EFCI bit, sends RM cell to source periodically with congestion indication bit (CI)
- ◆ Additive increase/multiplicative decrease
- ◆ Problems:
  - loss of RM cell with CI bit set results in continuous increase
  - queue length is no good indication
  - in high speed network, longer convergence time
  - oscillatory behavior in steady state
  - solutions: add extra NI(No Increase) bit to signal increase: PRCA scheme

▣ The Explicit Rate Scheme

- ◆ Explicit rate scheme: explicit rate to source
- ◆ Explicit vs. binary
  - straight forward policing
  - fast convergence time/robust
  - application can adapt to rate change: video
  - more information is better

Traffic Management Tech.(Cont.)

▣ MIT Scheme

- ◆ Explicit rate feedback mechanism --- Baseline for other schemes
- ◆ Explicit rate only. / max-min optimality
- ◆ Source:
  - send a RM cell (current cell rate, desired rate and reduced bit) every *n* th data cell
  - if reduced bit is clear, demand a higher desired rate
  - if reduced bit is set, uses the current rate as the desired rate
- ◆ Switch:
  - maintain a list of all VCs and their last desired rates
  - monitor all VC rates and compute a fair share
- ◆ Destination:
  - if desired rate < fair share, grant the desired rate
  - if desired rate > fair share, reduce to the fair share and reduce bit is se
- ◆ Problems:
  - O(n)
  - not compatible with current ATM FORUM standard
  - does not use measured source rate
  - no policy during transient period/errors in feedback

Traffic Management Tech.(Cont.)

▣ EPRCA

- ◆ Queue length only
- ◆ EFCI for data cell and a RM cell after every 31 data cells
- ◆ Source: EFCI for data cell and a RM cell (ER+ CCR+ CI) after every 31 data cells
- ◆ Switch: exponential weighted average rate and fair share
- ◆ Destination:
  - queue threshold based control
  - if Th1 < queue length < Th2, then intelligent marking (selective marking)
- ◆ Advantage: O(1)
- ◆ Problems: fairness / queue growth rate / still need queue length to get whole picture

▣ OSU

- ◆ Use input rate instead of queue length
- ◆ Use measured loads rather than declared loads.
- ◆ Count the number of active sources
- ◆ Backward congestion notifications can not be used to increase
- ◆ O(1)

▣ ERICA (Explicit Rate Indication for Congestion Avoidance):

- ◆ Input rate and queue length (load factor)
- ◆ Fair share first to avoid transient overloads

Traffic Management Tech. (Cont.)

- Selection Criteria for Congestion Control Schemes:
  - Scalability -- LAN, MAN and WAN
  - Optimality -- max-min allocation
  - Fairness -- fairness index using normalized allocation
  - Implementation (cost/performance tradeoff) --suitable for every switch Arch.
  - Simulation/testing configuration -- "Parking Lot" configuration for fairness
  - Traffic patterns -- persistent/ staggered/ bursty sources
  - Implementation, space and time complexity : O(1)
- 5 Ways to Improve ABR over Satellite
  - Increase the limit on the number of outstanding cells before decreasing
    - large Transient Buffer Exposure (TBE) parameter
    - the size was increased from 8 bit to 24 bit to accommodate satellite paths
  - Use larger Rate Increase Factor (RIF)
    - RIF=1 ===> Fast transient response
  - Implement Backward Explicit Congestion Notification (BECN)
  - Use larger ACR Decrease Time Factor (ADTF)
  - Implement Virtual Source/Virtual Destination (VS/VD)

Traffic Management Tech.(Cont.)

- Traffic Management in TCP/IP
  - Loss based:
    - If a segment times out, TCP reduce its window to 1
    - TCP starts with a window of 1 and increase if nothing is lost
    - not turned for large windows in satellite networks
  - Selective drop policies, e.g., RED affect fairness not throughput
  - Selective ack. helps in satellite networks
  - Explicit congestion notification (binary feedback) is being introduced but needs new algorithms
  - MPLS (MultiProtocol Label Switching) may help for TCP/UDP
- Four TCP Congestion Control
  - Slow start
  - Congestion avoidance
  - Fast retransmit
  - Fast recovery

Traffic Management in WCDMA Systems

▣ Soft QoS Control in Wireless Networks

- ◆ Soft QoS control procedure
  - check the availability of resources on the terminal's coverage area at connection setup (soft CAC), the new connection is accepted if sufficient resources are estimated to available for the connection to operate within the service contract without affecting the services of other ongoing connections (soft CAC controller)
  - while the connection is in progress, dynamic BW allocation is performed to match the requirements of interactive VBR traffic. When congestion occurs, the soft QoS control mechanism reallocates BW among connections (satisfaction estimator)
- ◆ Needs negotiated and soft QoS requirement (fading and mobility) to modify and establish radio access bearer
  - sudden variations in BW availability due to terminal moving and VBR interactive multimedia connection
  - source-initiated and network-initiated (due to handover) BW negotiation based on the concept of a connection's softness profile
  - achieve high utilization and maintain satisfactory application performance even during network congestion
  - Internet: RSVP, ATM: Q. 2963 signaling
  - the softness profile: allow an efficient match of application requirements to network resource profile , QoS-fair allocation (achieve uniform satisfaction) of resources

Traffic Management in WCDMA Systems (Cont.)

▣ CAC in IMT-2000 WCDMA Systems

- ◆ 요구된 call의 QoS를 만족하기 위한 필요한 자원 결정
- ◆ Multi-cell 환경하에서 home-cell and neighboring cells 의 Interference 영향 결정
  - interference from the reference cell and adjacent cells
  - traffic activity parameters
  - required E_b/I_o for each traffic (e.g., 7 dB for voice, 9 dB for data)
- ◆ BER 요구사항(=0.001)을 만족시키기 위한 required E_b/I_o을 만족하는 QoS measure로 Outage probability에 따라 CAC 결정
- ◆ Needs negotiated and soft QoS requirement (fading and mobility) to modify and establish radio access bearer
- ◆ Session Management level QoS parameters
  - SM level QoS parameters are negotiated between Terminal, SGSN, and GGSN in PDP Context Activation Procedure
  - delay class, reliability class, priority class, mean and peak throughput, traffic class, delivery order, delivery of erroneous SDU, max SDU size, max and guaranteed bit rate for uplink and downlink, transfer delay, residual BER, and SDU error ratio

Traffic Management in WCDMA Systems(Cont.)

▣ Radio Resource Management and Strategies *(See http://www.3GPP.org)*

- ◆ (Soft) Handover control
  - based on the measurement results reported by the UE/RNC and various parameters set for each cell
  - among cell-types in multicell structure
  - inter-mode and inter-system
  - causes: up/down link quality, signal strength, distance, change of service, better cell, traffic distribution (load control), O&M intervention, and preemption
- ◆ Admission control
  - CAC is a function of "required type of service" and "current system load"
  - Service_type
    - derived from QoS
    - premium (high-priority, PSTN domain, voice)
    - best-effort (low-priority, IP-domain, Web)
- ◆ Radio access bearer (RAB) control
  - control the UE and system resource
  - control the RAB based on traffic volume measurement
  - RAB rate (service rate)가 변경되면 TrCH(transport CH) and PhyCH (Physical CH) 형상을 재구성
  - increase uplink date rate ==> TrCH type switching (Common==> Dedicated)

Traffic Management in WCDMA Systems(Cont.)

▣ Radio Resource Management and Strategies (Cont.)

- ◆ Dynamic resource allocation
  - code allocation strategies for FDD
    - different codes are allocated to different connections.
    - OVSF (orthogonal variable spreading factor) code tree
  - DCA for TDD
    - a physical channel is characterized by a combination of its carrier frequency, time slot, and spreading code
    - resource allocation to cells (slow DCA)
    - resource allocation to bearer services (fast DCA)
- ◆ Power management
  - variable rate packet transmission
    - UL/DL power를 전파 환경에 따라 가변하여 interference를 줄임.
    - 측정결과==> 채널환경 악화 ==> power control ==> 송출 전력 증대 ==> rate 감소 (TFC in MAC) ==> 버퍼에 일시 저장 후 송출
  - site selection diversity power control (SSDT)
    - soft handover case, reduce the interference: best cell (=primary cell_id)만 접속

Traffic Management in WCDMA Systems(Cont.)

▣ Radio Resource Management and Strategies (Cont.)

◈ Radio link surveillance
- mode control strategies for TX diversity
- two categories: open loop mode and closed loop mode
- criteria for mode control: radio channel conditions

◈ Codec mode control: AMR mode control
- AMR speech codec
  - multi-rate speech code with eight source rates: 4.75Kbps ~12.2 Kbps
- AMR adaptation
  - 무선환경 (acceptable load, FER) 변화에 따라 speech rate 변경 (each 20 ms speech frame): RNC (Radio Resource Management) 주관
- location of the transcoder: Core network in WCMA

Analytic Approaches

▣ Traffic Models

◈ Poisson model: telephone call

◈ Short-range dependence model (Markov chains): easy to analyze, exponentially decaying ACF
- On-Off model : single voice and video source
- MMPP (Markov Modulated Poisson Process): Aggregated voice source
- IPP (Interrupted Poisson Process)
- MMDP (Markov Modulated Deterministic Process)

◈ AR processes: difficult to approximate the ACF
- AR model : video sources,

◈ Long-range dependence model: only statistical analysis
- self-similar or fractal models: internet traffic

▣ Multi-layer States of Terminals

◈ Subscription (months)

◈ Call or connection (3 min.)

◈ Burst (msec)

◈ Cell or packet (usec)

Analytic Approaches (Cont.)

- CTMC Queuing Models for Phase-dependent Arrivals & Service Rates
  - PGF (probability generation function) approach - Exact analysis, difficult to solve
  - Matrix geometric technique - Exact analysis, easy to solve, but complex calculation
  - Fluid-flow model - simple to calculate and approximation
    - estimating the traffic tail distribution of the aggregate traffic
      - central limit approximation: the law of large numbers
      - large deviation theory using the Chernoff bound : upper bound
      - the refined large deviation approximation: sharpened by a factor

- Buffer Dimensioning
  - Deterministic burstiness curve
    - rate envelope multiplexing, lossless multiplexing
    - not sufficient to provide any useful type of QoS guarantees
  - Probabilistic burstiness curve
    - fluid model
    - MMPP/MMDP/1/K model

Conclusions

- Wireless Networking Trends:
  - IP-based core networks and radio access networks
  - OBP satellite, WLANs, WPAN, and public cellular networks (including the IMT-2000)
  - Super-exponential increase in data traffic and VoIP ==> Traffic Engineering and QoS over IP
  - Effective and efficient end-to-end QoS provisioning in wireless networks
- Wireless Traffic Management and Resource Management Tech.:
  - Static BW allocation models lack flexibility, and thus cope poorly with multimedia interactivity and session mobility
  - Convey application-level QoS information to the network
  - Explicit rate based traffic management in ATM is required
  - Correlated or long-range dependent traffic models
  - Traffic management is key for the success of ATM ABR services
  - Soft resource management, admission policy , and bandwidth re-negotiation
  - Transient fluid-flow model, Chernoff bound, and Large deviation approximation
  - Mobility-enhanced Soft (adaptive) QoS management / Probabilistic buffer dimensioning
  - Measurement-based CAC based on Individual QoS measures
  - Physical layer and application-aware approach for video communications