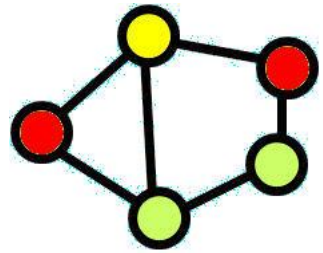
Slide supporting material



Lesson 8: ATM Network, 2nd part

Giovanni Giambene

Queuing Theory and Telecommunications: Networks and Applications
2nd edition, Springer

QoS Support in ATM

ATM: Traffic Management for QoS Support

- In ATM networks **flow control and error control are not operated at intermediate nodes, but only end-to-end at AAL level**. Suitable techniques must be used to prevent congestion.
- In circuit-switched networks, congestion control is simply operated during the set-up phase of the end-to-end link (it is sufficient to check the availability of all the links along the source-to-destination path).
- In ATM networks, **congestion control** is more complicated because:
  - Sources may produce **variable bit-rates** that make unpredictable their loads;
  - **Links are shared** by different paths having a variable congestion level.
  - There can be different traffic sources with **different characteristics and QoS requirements**.
  - Each traffic flow must have **guaranteed a given bandwidth** (equivalent bandwidth = mean bit rate plus some margins) in the different links of its virtual path in order to fulfill its QoS levels in terms of delay, packet loss, etc.

ATM: Traffic Management for QoS Support (cont'd)

- The traffic can be with guarantee of QoS or without guarantee of QoS.
- Referring to QoS-guaranteed traffic, two different types of techniques can be adopted:
  - **Preventive control**: Preventive control is used both to decide whether a new connection can be admitted in the network (**Connection Admission Control, CAC**), to smooth its traffic (**traffic shaper**), to monitor the input traffic produced on a connection to avoid that unacceptable traffic peaks or overloads are produced (**traffic policer** of Usage Parameter Control, UPC), and to **schedule different traffic flows** at multiplexers. Finally, also **buffer management** functions are part of preventive control.
  - **Reactive control**: Reactive control is an action taken when a congestion event has occurred. The problem of such approach is that it implies a delay (typically, on the order of the round-trip time) before taking a repair action (e.g., slow down of the traffic produced by a source). This is the same approach adopted by the congestion control scheme at transport layer (TCP protocol).

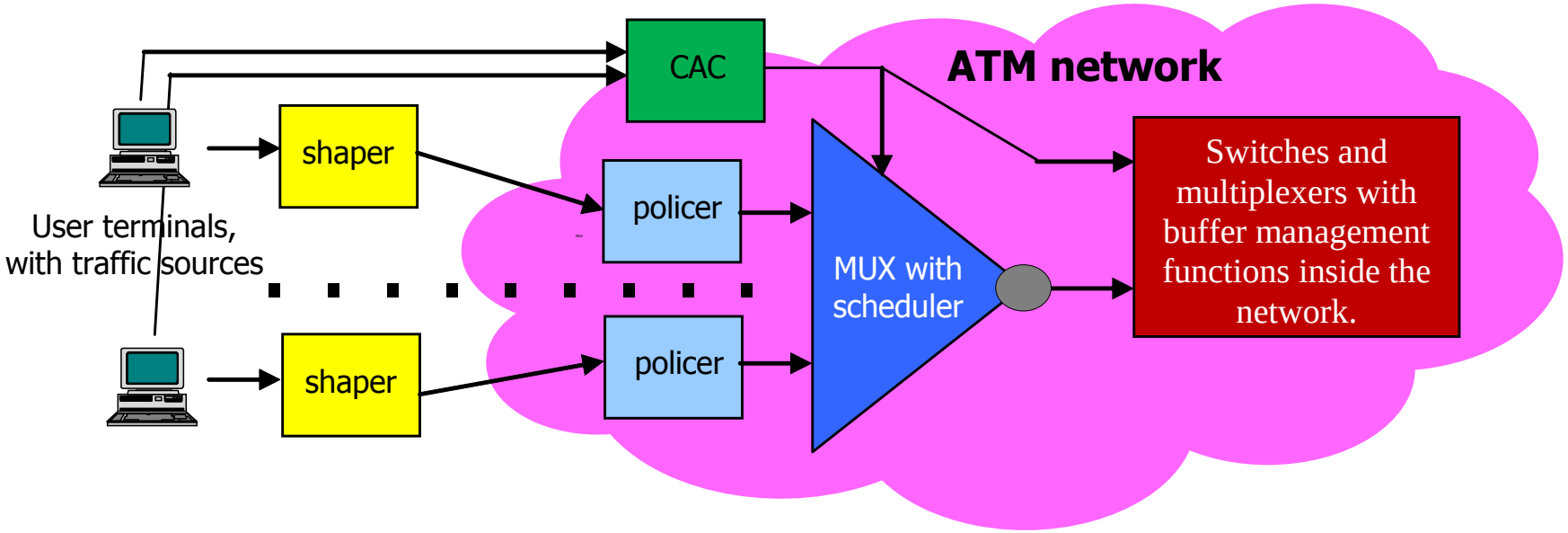
ATM: Traffic Management for QoS Support (cont'd)

- The traffic can be with guarantee of QoS or without guarantee of QoS.
- Referring to ifferent types of techniques
  - **Preventi** both to decide whether a new connectio **nection Admission Control, CAC**), to smooth its traffic (**traf per**), to monitor the input traffic produced on a connection to av hat unacceptable traffic peaks or overloads are produced (**traffic policer** o Usage Parameter Control, UPC), and to **schedule different traffic flows** at multiplexers. Finally, also **buffer management** functions are part of preventive control.
  - **Reactive control**: Reactive control is an action taken when a congestion event has occurred. The problem of such approach is that it implies a delay (typically, on the order of the round-trip time) before taking a repair action (e.g., slow down of the traffic produced by a source). This is the same approach adopted by the congestion control scheme at transport layer (TCP protocol).

More details on policers, shapers, CAC, etc. will be provided later in this lesson.

ATM: Traffic Management for QoS Support (cont'd)

- In a typical ATM access network, we have different traffic sources, each regulated by a traffic shaper, a policer, a CAC block, policers to monitor the traffic loads injected into the network by the different sources, and a multiplex with a scheduler to regulate the resource sharing on the access link.



ATM Service Classes and Characteristics

- The following services have been defined **with respect of the bit-rate** that the network is able to provide to the related traffic flows:
  - Constant Bit-Rate (CBR)
  - Variable Bit-Rate (VBR)
    - real-time VBR
    - non-real-time VBR
  - Available Bit-Rate (ABR)
  - Unspecified Bit-Rate (UBR)
  - Guaranteed Frame Rate (GFR).

	CBR	rt-VBR	nrt-VBR	ABR	UBR
Bandwidth guarantee	Yes	equivalent	equivalent	minimum	No
Real-time traffic	Yes	Yes	No	No	No
Data bursty traffic	No	Yes	Yes	Yes	Yes
Congestion notification	No	No	No	Yes	No

- A fixed bandwidth is **reserved** in the network for CBR sources; instead, an **equivalent bandwidth must be available** on all the links of the path to accept an rt-VBR or an nrt-VBR traffic source. ABR sources have guaranteed a minimum end-to-end bandwidth. Finally, **no capacity is guaranteed** for UBR sources.

ATM Service Classes and Characteristics

- The following services have been defined **with respect of the bit-rate** that the network is able to provide to the related traffic flows:
  - Constant Bit-Rate (CBR)
  - Variable Bit-Rate (VBR)
    - real-time VBR
    - non-real-time VBR
  - Available Bit-Rate (ABR)
  - Unspecified Bit-Rate (UBR)
  - Guaranteed Frame Rat

	CBR	rt-VBR	nrt-VBR	ABR	UBR
Bandwidth guarantee	Yes	equivalent	equivalent	minimum	No
Real-time traffic	Yes	Yes	No	No	No
Data bursty traffic	No	Yes	Yes	Yes	Yes
Congestion notification	No	No	No	Yes	No

- A fixed bandwidth is **r**
  instead, an **equivaler**
  links of the path to ac
  ABR sources have gua
  Finally, **no capacity i**

The ABR traffic class is the sole to support a reactive congestion control by means of the EFCI bit (e2e) in the cell header (forward traffic) and the use of RM cells sent back from the destination to instruct the source to reduce the traffic injection (forward + backward mechanisms).

ATM Service Classes and Characteristics

Increasing priority level at the scheduler

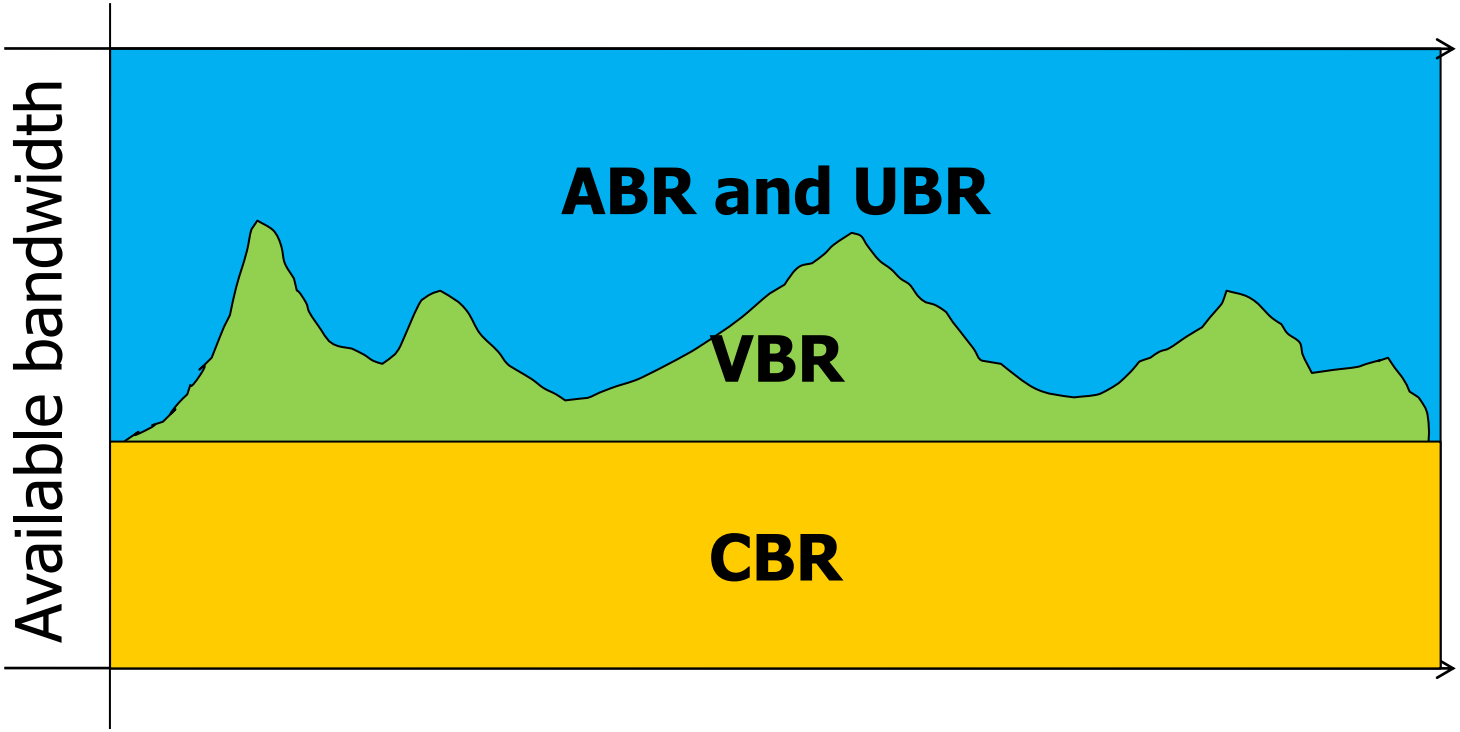
- The following services have been defined **with respect of the bit-rate** that the network is able to provide to the related traffic flows:
  - Constant Bit-Rate (CBR)
  - Variable Bit-Rate (VBR)
    - real-time VBR
    - non-real-time VBR
  - Available Bit-Rate (ABR)
  - Unspecified Bit-Rate (UBR)
  - Guaranteed Frame Rate (GFR).

Note that there is a mapping between ALL protocols (ALL1, ALL2, …, ALL5) and ATM service classes (CBR, rt-VBR, …, UBR). For instance, ALL1 corresponds to CBR; ALL2 corresponds to rt-VBR; ALL5 can support any service except CBR traffic, etc.

- A fixed bandwidth is **reserved** in the network for CBR sources; instead, an **equivalent bandwidth must be available** on all the links of the path to accept an rt-VBR or an nrt-VBR traffic source. ABR sources have guaranteed a minimum end-to-end bandwidth. Finally, **no capacity is guaranteed** for UBR sources.

ATM Service Classes and Characteristics (cont'd)

- The picture below describes how the different traffic classes share a given link capacity. There is a **priority** in the service of the traffic classes: first CBR is allocated, then rt-VBR is allocated. The remaining bandwidth is shared by ABR and UBR classes.



ATM Traffic Descriptors

- Traffic descriptors are used to characterize the traffic produced by a given source (i.e., the contractual parameters of the traffic source).
  - PCR denotes the maximum bit-rate allowed to the source;
  - SCR corresponds to the mean bit-rate;
  - The traffic **source burstiness is β = PCR/SCR**.

	Acronym	Definition
Peak Cell Rate	PCR	Maximum rate according to which cells will be sent in the network
Sustainable Cell Rate	SCR	Mean rate (long term) according to which cells will be sent in the network
Minimum Cell Rate	MCR	Minimum acceptable rate of cells in the network
Maximum Burst Size	MBS	Maximum number of cells that may be sent in a burst at the line rate
Cell Delay Variation Tolerance	CDVT	Maximum acceptable difference in the delay of output cells at a node (related to queuing delays)

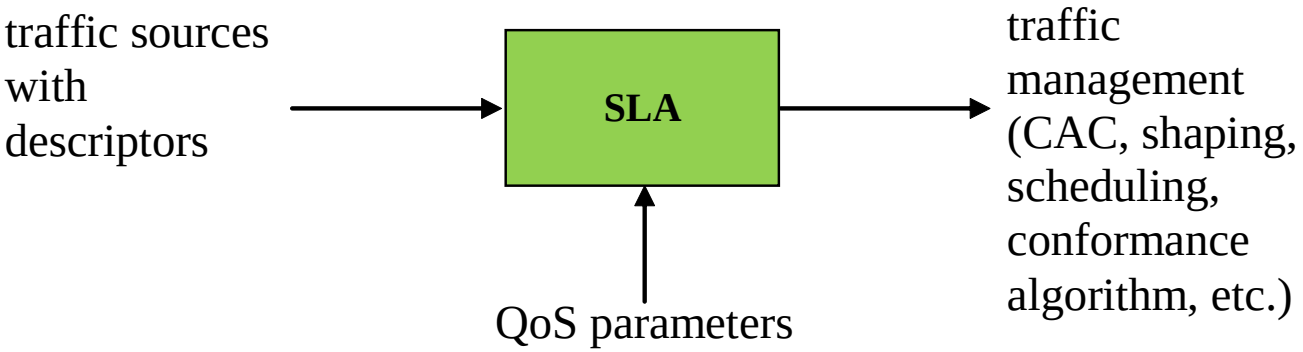
ATM QoS Parameters

	Acronym	Definition
Cell Loss Ratio	CLR	Percentage of lost (or late) cells
Cell Transfer Delay	CTD	End-to-end delay for the transmission of a cell (maximum and mean value)
Cell Delay Variation	CDV	Variance of the end-to-end transmission delay of a cell
Cell Error Ratio	CER	Percentage of erroneous cells
Cell Misinsertion Rate	CMR	Percentage of erroneously delivered cells (routing error) among all the sent cells on a flow

- For traffic with guaranteed QoS levels, the user and the network stipulate a **traffic contract** specifying the expected network QoS under some traffic characteristics defined by the descriptors (e.g., PCR, SCR, MBS, MCR, and CDVT) and a traffic conformance algorithm. **The guaranteed QoS level can be in terms of mean throughput, maxCTD, CDV or CLR**.

- ABR and UBR traffic do not require a description of traffic, have no QoS guarantee (ABR can have guaranteed the MCR). These traffic classes should have no impact on the QoS provided to the other traffic classes with QoS guarantee.

QoS Approach in ATM



SLA = Service Level Agreement

Real Time Services (Inelastic Traffic): CBR and rt-VBR

- If we want to avoid a delay variation (jitter), CBR or rt-VBR are the appropriate choices.
- **CBR** with a fixed data rate continuously available
  - Commonly-used for uncompressed audio and video
    - Video conferencing
    - Interactive audio
- **rt-VBR** best suited to time-sensitive applications that transmit at a rate that varies in time
  - Tightly constrained delay and delay variation
    - The compressed video produces varying-sized image frames and then the data-rate varies.

Non-Real Time Services (Elastic Traffic): nrt-VBR, ABR, UBR, GFC

- Intended for **applications with bursty traffic and limited constraints on delay and delay variation.**
- Greater flexibility, greater use of statistical multiplexing
  - **nrt-VBR**: Application specifies peak cell rate (PCR) and sustainable cell rate (SCR)
  - **ABR**: Application specifies peak cell rate (PCR) it will use and minimum cell rate (MCR) it requires
  - **UBR**: Unused capacity by CBR and VBR traffic is made available to UBR traffic
  - **GFC**: Designed to support IP-based traffic.

Details on Traffic Management Techniques for QoS Support: CAC, Policers, Shapers, Schedulers, Buffer Management, EFCI

A Summary on QoS Support 'Pillars' in ATM

Let us recall the main elements for QoS support:

- **Preventive control**
  - CAC
  - Resource reservation
  - Traffic regulators: policers (UPC) and shapers
  - Traffic scheduling
  - Buffer management
- **Reactive control**:
  - Explicit congestion indication (backward or forward type)

Note that these techniques are not only used in ATM, but also in the Internet. This further motivates our interest to deepen this part.

CAC

- **CAC is a control that is operated by the network at the connection set-up phase in order to verify whether the QoS parameters of both the new connection and the connections already in progress can be fulfilled on the access link.**
- CAC uses the connection traffic descriptors to determine the amount of resources to be allocated on each link along the path from source to destination.
- Two broad groups of CAC techniques can be considered: (i) CAC based on **bandwidth** aspects; (ii) CAC based on **CLR considerations**.
- An example of CAC of the first type:
  - Let us refer to VBR traffic sources (bursty traffic) to be admitted on a given (shared) access link to the ATM network.
  - It would be highly inefficient to reserve the bandwidth corresponding to the PCR value of the VBR connection; hence, it is important to allocate to each VBR flow its equivalent bandwidth B_{eq} that guarantees controlled packet losses and delays: SCR < B_{eq} < PCR.
  - Let C denote the capacity of the link and let B_{eqi} be the **equivalent bandwidth** of the i-th connection on the same link. **The new connection with equivalent bandwidth B_{eq} is admitted (on the link) by CAC only if it fulfills the following condition**:

$$\sum_i B_{eqi} + B_{eq} \leq C$$

CAC

- **CAC is a control that is operated by the network at the connection set-up phase in order to verify whether the QoS parameters of both the new connection and the connections already in progress can be fulfilled on the access link.**
- CAC uses the connection traffic descriptors to determine the amount of resources to be allocated on each link along the path
- Two broad groups of CAC techniques ca **bandwidth** aspects; (ii) CAC based on
- An example of CAC of the first type:
  - Let us refer to VBR traffic sources (bursty traffi ATM network.
  - It would be highly inefficient to reserve the bandwidth corresponding to the PCR the VBR connection; hence, it is important to allocate to each VBR flow its equivalent ba h B_{eq} that guarantees controlled packet losses and delays: SCR < B_{eq} < PCR.
  - Let C denote the capacity of the link and let B_{eqi} be the **equivalent bandwidth** of the i-th connection on the same link. **The new connection with equivalent bandwidth B_{eq} is admitted (on the link) by CAC only if it fulfills the following condition**:

$$\sum_i B_{eqi} + B_{eq} \leq C$$

More details on the formal derivation of B_{eq} are beyond the scope of these slides and of this book.

UPC: Monitoring Each Flow

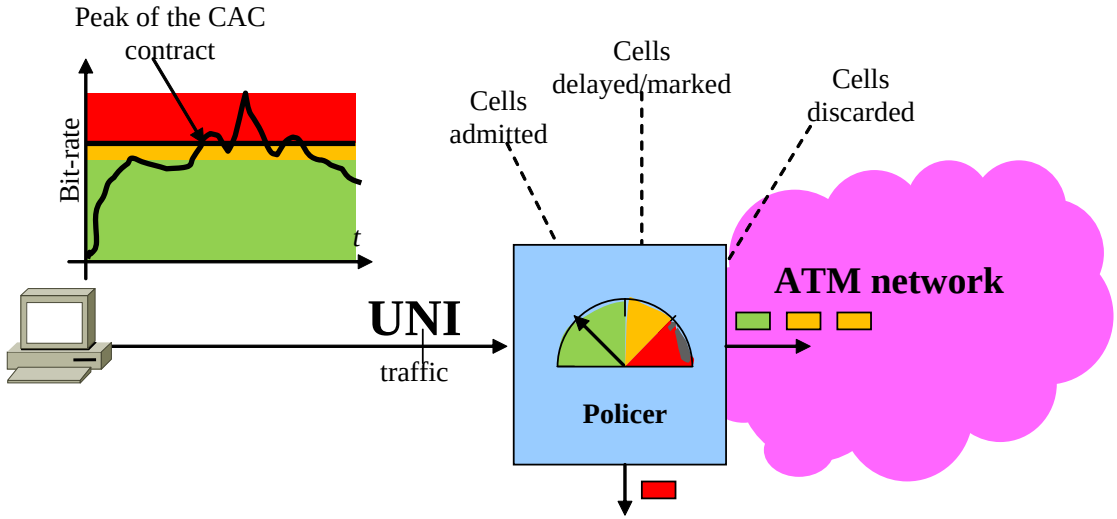
- CAC operates only in the **set-up phase** according to QoS criteria, but it does not guarantee that a traffic source cannot overload the network once admitted (the contract specifies the Service Level Agreement, SLA).
- There could be both temporary bursts of traffic produced by VBR sources or persistent traffic loads that violate the contract agreed with the network that is controlled in the CAC phase.
- **UPC is needed to monitor the traffic injected into the network by an admitted source to protect network resources from malicious users (imagine a Denial of Service attack) as well as from unintentional misbehaviors that could affect the QoS of other connections.**
- UPC must be performed on a connection basis. This could be computationally heavy. In ATM, UPC is defined in the ITU-T Recommendation I.356.

UPC (cont'd)

- **UPC is intended to ensure the conformance of a connection with the negotiated traffic contract (SLA).**
- The connection traffic descriptors contain the necessary information for **conformance testing** at UNI.
- The UPC function is implemented in the **policer** on the **network side of UNI**.
- Based on the UPC check, the network may decide whether or not a connection is compliant and hence whether or not admitting a given cell in the network for this connection (and in case to set the CLP flag).

Traffic Policer

- The policer shall be capable of
  - **Passing a cell**, which is **conformant** to the connection traffic descriptors (grant, SLA).
  - **Discarding a cell** if it is **not conformant** to the connection traffic descriptors; alternatively, if the tagging option is allowed for a connection, the UPC function shall be capable of converting from CLP = 0 to CLP = 1 for a non-conformant cell accepted into the network.



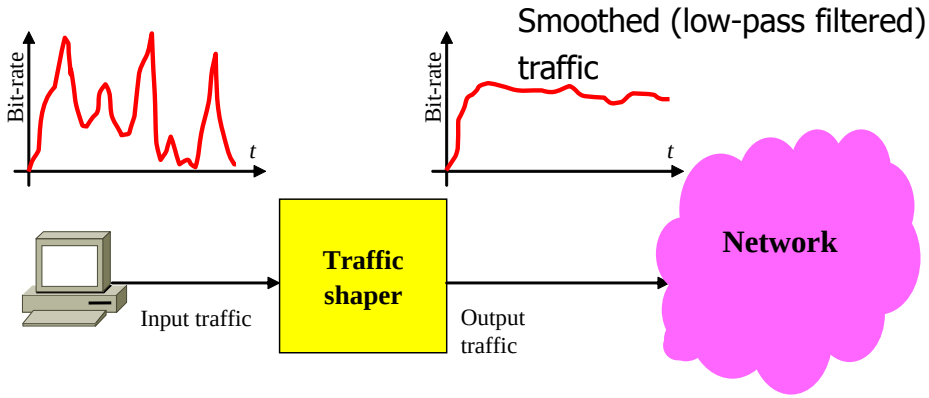
Traffic Policer (cont'd)

- Some examples of policers are listed below:
  - PCR policing (i.e., controlling the maximum bit-rate), suitable for CBR sources;
  - SCR and MBS policing (i.e., controlling the mean bit-rate and the maximum burst size), for VBR sources without limits on PCR;
  - PCR, SCR and MBS/CDVT policing for VBR sources with PCR, SCR and burst (or delay, packet loss) limitations.

Traffic Shaping: Trading Delay for Traffic Burstiness

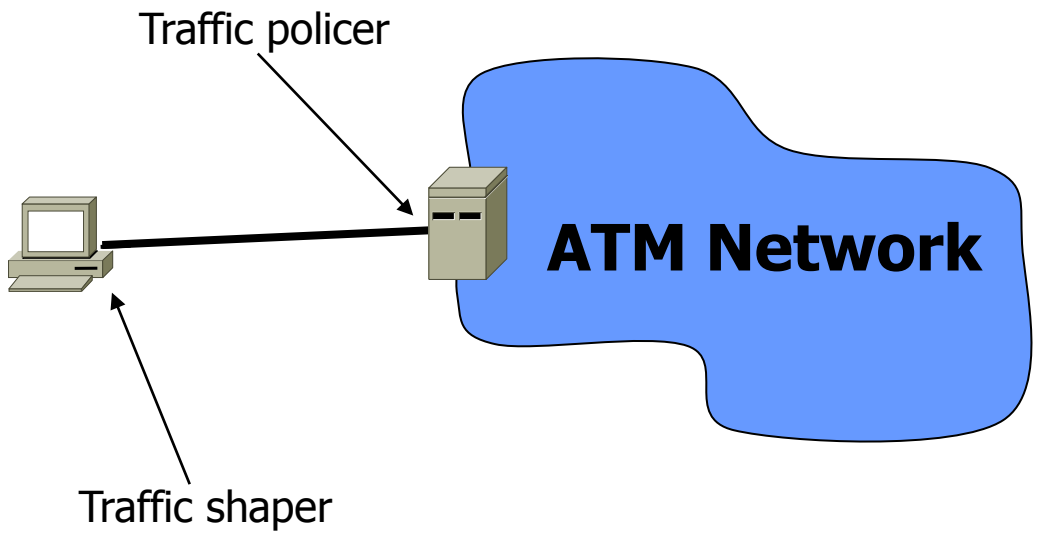
- **Bursty traffic** sent in a network could immediately create congestion (= suddenly fill in the buffers), thus causing **high queuing delays** and **packet losses**. Bursty traffic does not allow an adequate utilization of network resources that are underutilized most of the time (→ **low efficiency**).

- The idea is to regulate the traffic injection of a source so that the traffic entering a network is smoothed (= **almost constant traffic**), queues are not congested and a better utilization of network resources can be achieved. This is obtained at the expenses of extra delays experienced by source packets.



Shapers and Policers used in Tandem in ATM

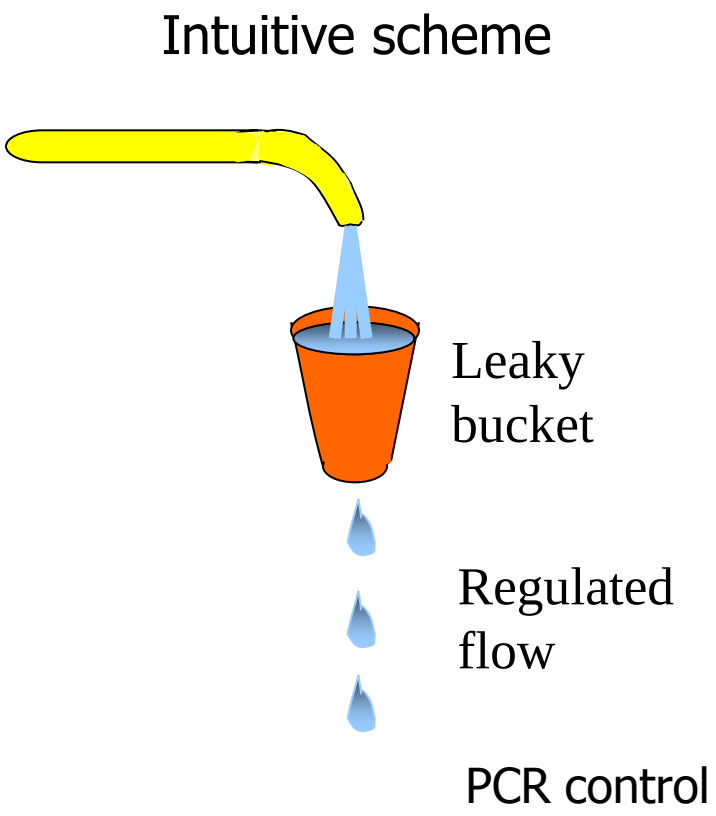
- **Traffic regulators such as shaper and policer typically need to be used in tandem, as shown in the picture.**
- **Shapers and policers are quite similar; the main difference is that shapers need a buffer to queue input traffic.** Instead, policers do not buffer traffic as they are not interested in smoothing it.



The traffic shaper is used on the traffic source side; instead, the traffic policer is used on the network side.

Leaky Bucket Shaper

- The leaky bucket is a mechanism to transmit packets at regular intervals, that is by 'filtering' them with a **G/D/1 FIFO queue**, where the packet transmission is regulated at a given rate.
- The transmission rate is set to the declared effective bandwidth of the traffic source.
- The output traffic has a very reduced **burstiness**. This can be an advantage for the network (regulated injected traffic flow), but it may lead to unacceptable **delays** for the user application if its input traffic is too bursty.
- To support better input traffic burstiness the token bucket traffic regulator could be used.

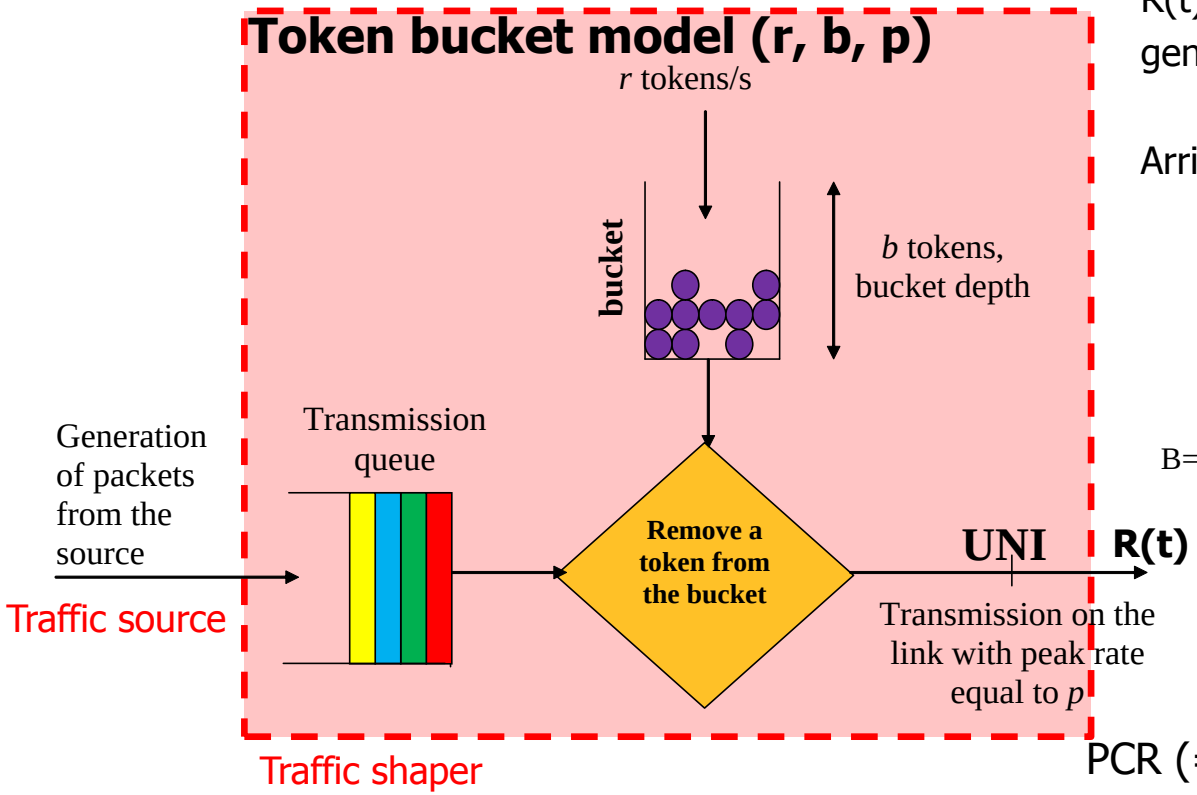


Token Bucket Shaper

- If we are interested to **a shaper that allows passing some degree of burstiness**, instead of the leaky bucket shaper we have to adopt the token bucket shaper.
- This shaper uses both a **token bucket and a data queue**.
- **Tokens** are put into the bucket at a certain rate r. The bucket has a given capacity b of tokens (i.e., bucket depth). If the bucket is full, newly arriving tokens are discarded.
- Each token represents the permission for the source to send a certain number of bits (e.g., one bit or one cell) into the network.
- **To send a packet (cell), the regulator must remove from the bucket a number of tokens that correspond to the packet size.** If not enough tokens are in the bucket to send a packet, the packet either waits until the bucket has enough tokens or is discarded or is marked.
- The largest burst a source can send into the network is roughly proportional to the bucket depth.

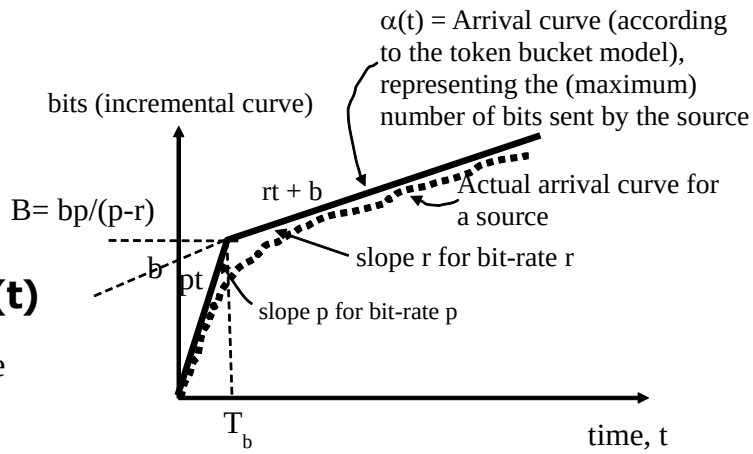
Token Bucket Shaper (cont'd)

- The token bucket permits to guarantee that a mean bit-rate r is maintained. Moreover, the bucket depth b allows some traffic burstiness.
- The peak bit-rate is equal to p.



R(t) is the stochastic process of the bit-rate generated as a function of time t (fluid-flow)

Arrival curve $\alpha(t) = \int_0^t R(t)dt$



PCR (= p), SCR (~ r) and MBS (= b) control

Regulators … not only used in ATM !

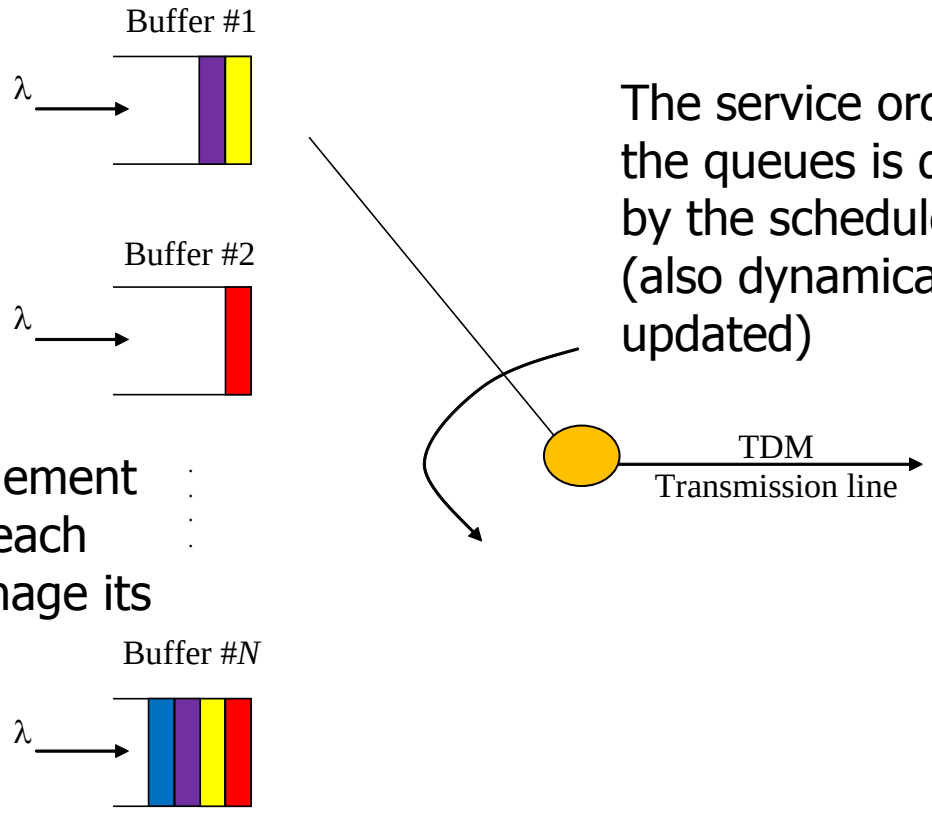
- Shaping and policing functions are widely implemented in different networks to support QoS:
  - Asynchronous Transfer Mode (ATM)
  - Multi-Protocol Label Switching (MPLS)
  - DiffServ (RFC 2475) for IP networks: the DiffServ boundary node uses a meter (**policer**) to measure the traffic conformance and a **shaper/dropper** to delay impulses of traffic.
  - IntServ (RFC 2215) for IP networks: a token bucket approach is used to describe a traffic flow (T-SPEC descriptor); CAC and resource reservation are preformed accordingly.
  - ETSI BSM standard for satellite communications (see Lesson No. 20).

Traffic Scheduling

- Traffic scheduling is a fundamental function that is required in ATM networks **to share the physical transmission resources among competing flows** with QoS requirements.
- Scheduling must guarantee to preserve some form of **priority** among traffic classes. On each transmission link, **different queues are used for the transmission of different traffic classes**; the service order of these queues is decided by the scheduler.
- Different scheduling and priority schemes can be adopted:
  - **Weighted Round Robin (WRR)** where the service of the different queues is according to a cycle; the different queues are serviced for time intervals depending on their weights.
  - **Earliest Deadline First (EDF)** where the priority of each packet depends on the residual lifetime of the packet (a deadline is assigned to each packet).

Hierarchical Scheduler Architecture

FIFO queue management and WRR scheduling of the service of the different buffers.



Buffer management is applied to each queue to manage its congestion.

Buffer Management

- We can consider two buffer management techniques to protect **high-priority cells** (CLP = 0) with respect to **low-priority cells** (CLP = 1):
  - In a **push-out mechanism** all cells are allowed to enter the buffer if there are available rooms. Let us consider a cell arriving at a **full buffer**. If the cell has a low priority, it is discarded. Whereas, if the cell has a high priority, it is discarded only if there is no low-priority cells waiting in the queue (if a low priority cell is found in the queue, it is discarded and the high-priority cell is admitted).
  - The **threshold mechanism** allows all cells to enter the buffer as long as the number of waiting cells is less than a specified **threshold**. When the number of waiting cells exceeds such limit, newly-arriving cells with low priority will be discarded, whereas high-priority cells will be admitted as long as there are available rooms in the buffer.

Buffer Management (cont'd)

- Both **push-out mechanism and threshold mechanism** provide similar performance.
- More refined schemes **control the cells to drop** rather than having them dropped at random.
  - When an AAL5 packet (corresponding to a TCP/IP segment) is divided into cells, the drop of a single cell entails the need to re-send also the other cells of the same AAL5 packet (AAL5 CRC fails).
  - It is convenient **to drop all the cells from the same AAL5 packet** in the presence of congestion in order to avoid that congestion affects multiple AAL5 packets.

Reactive Control: ABR and EFCI

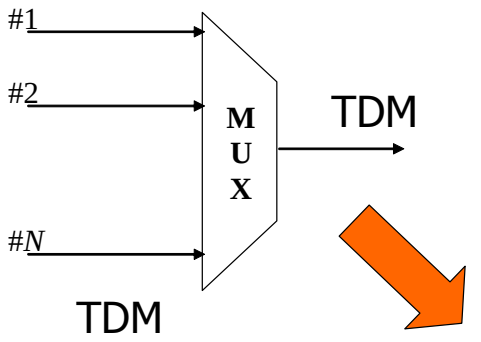
- The second bit of the **PTI field** (in the header of a data cell) is equal to 1 to notify node congestion in the same direction of the cell. This is the **Explicit Forward Congestion Indication (EFCI)**.
  - When a switch becomes congested, it will mark on each VC the EFCI state of all cells being forwarded to the destination. Upon receiving **marked cells**, the destination returns congestion notification cells (**Resource Management, RM, cells**, by means of the PTI field) to the source to inform the source of the congested VC. The source uses this feedback information to decrease (or increase, when congestion ends) the cell transmission rate accordingly.
  - **ABR is the only traffic class that allows the notification of congestion** and that can thus support reactive congestion control.

PTI value	Cell Type	Congestion Notification	AUU
000	Information data	No	0
001	Information data	No	1
010	Information data	Yes	0
011	Information data	Yes	1
100	OAM	-	-
101	OAM	-	-
110	RM	-	-
111	Reserved	-	-

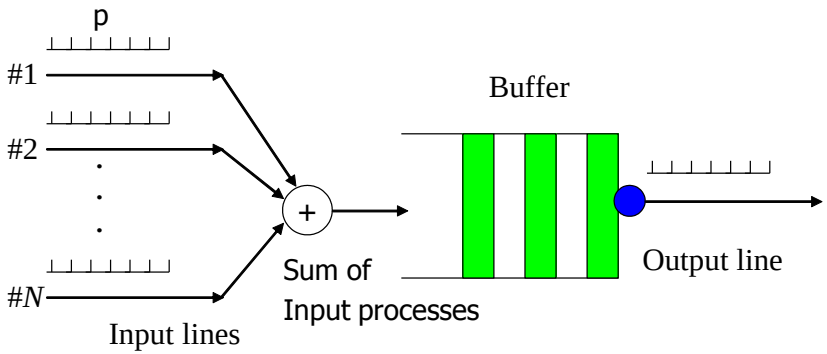
Exercise on TDM and Asynchronous Multiplexing (ATM Case)

Queuing Model for the Asynchronous Multiplexer (ATM)

- We consider an ATM multiplexer that receives N synchronous input time-division flows of traffic. ATM packets (i.e., cells) are stored waiting for transmission in a buffer with infinite rooms. **The arrival process is discrete-time (the system behaves as discrete-time)**.
  - There is only one output flow.
  - One packet needs first to be stored in the buffer before it can be transmitted.
  - Input and output lines are synchronized. They have the same slot duration, T, that permits to convey one packet (i.e., input and output lines have the same speed).
- We need to characterize the mean number of packets in the buffer, N_p, and the mean packet delay, T_p.



Model



ΣBernoulli/D/1 (or Binomial/D/1) Discrete Time Queuing Model

- Arrival process: each slot of an input line conveys a packet with probability p. This behavior is memoryless from slot to slot (Bernoulli arrival process of packets from each input line).
  - The total number of packets that arrive at the ATM multiplexer on a slot basis is according to a **binomial process (= sum of elementary Bernoulli processes)**:

$$\text{Prob}\{n \text{ packets arrive in a slot}\} = \binom{N}{n} p^n (1-p)^{N-n}$$

- **Imbedding instants at the end of the slot of the output transmission line, ξ_i.**
  - n_i denotes the number of ATM cells at the end of the i-th slot of the output line (instant ξ_i^+);
  - a_i denotes the number of ATM cells arrived at the buffer during the i-th slot (due to the assumed synchronization, these arrivals complete at instants ξ_i^-)
- The system is characterized by the same difference equation of the classical M/G/1 queue, even if the arrival process is not Poisson:

$$n_{i+1} = n_i - 1 + a_{i+1} \text{ for } n_i > 0 \text{ and } n_{i+1} = a_{i+1} \text{ for } n_i = 0.$$

ΣBernoulli/D/1 (or Binomial/D/1) Discrete Time Queuing Model

- Arrival process: each slot of an input line conveys a packet with probability p. This behavior is memoryless from slot to slot (Bernoulli arrival process of packets from each input line).
  - The total number of packets t
    according to a **binomial proc**
    processes):

$$\text{Prob}\{n \text{ packets arri}$$

$$\binom{}{n}$$

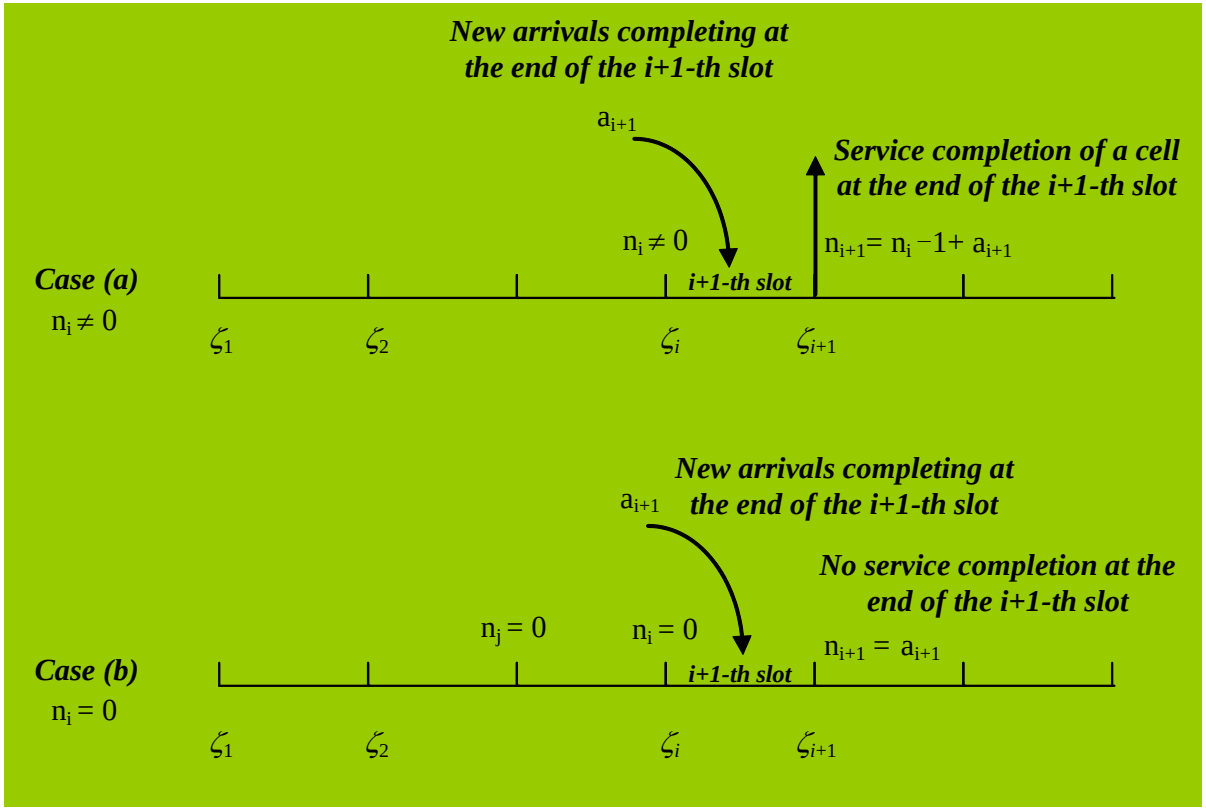
We cannot apply the Pollaczek-Khinchin formula because the arrival process is not Poisson.

- **Imbedding instants at the end of the slot of the output transmission line, ξ_i.**
  - n_i denotes the number of ATM cells at the end of the i-th slot of the output line (instant ξ_i^+);
  - a_i denotes the number of ATM cells arrived at the buffer during the i-th slot (due to the assumed synchronization, these arrivals complete at instants ξ_i^-)
- The system is characterized by the same difference equation of the classical M/G/1 queue, even if the arrival process is not Poisson:

$$n_{i+1} = n_i - 1 + a_{i+1} \text{ for } n_i > 0 \text{ and } n_{i+1} = a_{i+1} \text{ for } n_i = 0.$$

ΣBernoulli/D/1 (or Binomial/D/1) Discrete Time Queuing Model

- Visual representation of relations between quantities n_i, n_{i+1}, a_{i+1}:



Solution

- In this case we have not an M/G/1 queue, but the system is described by the same difference equation (the same hypotheses hold). Hence, we can write the same expression for the mean number of 'objects' in the queue according to which we need to determine A(z):

$$A(z)=\sum_{n=0}^{N}\binom{N}{n}z^n p^n (1-p)^{N-n}=(1-p+zp)^N$$

Binomial PGF

$$A'(1)=Np$$
$$A''(1)=N(N-1)p^2$$

- The buffer stability is guaranteed if Np < 1 cells/slot.

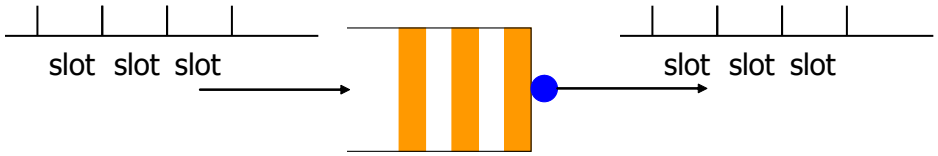
$$N_p=A'(1)+\frac{A''(1)}{2[1-A'(1)]} \qquad \begin{bmatrix}\text{cells in}\\ \text{the buffer}\end{bmatrix}$$

$$T_p=\frac{N_p}{A'(z=1)}=1+\frac{(N-1)p}{2(1-Np)} \qquad [\text{slots}]$$

Exercise on Traffic Shaping for ATM

Leaky Bucket Shaper Analysis in the ATM Case

- We refer to a leaky bucket shaper that regulates the traffic so that it can send a cell every time T. We have a Time Division Multiplexing (TDM) line with slots of duration T both at the input and at the output of the regulator; these lines are synchronous.
  - The cell arrival process (i.e., the packets coming from the input line) is characterized as follows:
    - A slot carries a message with probability q; otherwise it is empty.
    - Each message is formed of a random number of cells with PGF L(z); note that a message is composed of at most L_{max} cells.
  - It is requested to evaluate the following quantities:
    - The mean delay experienced by a cell from input to output of the regulator;
    - The burstiness degree of the output traffic to be compared with that of the input traffic.



Solution

- The leaky bucket shaper can at most allow the transmission of one cell every time T. Our input process is not Poisson (but **compound Bernoulli**), but still **memoryless**: we can still apply the classical M/G/1 solution method for the analysis of this regulator. This queue is actually 'M'/D/1.
- We **imbed our study at the instants of slot ends of the output TDM line from the shaper** (this type of imbedding instants permits to avoid to adopt the differentiation that will be discussed in Lesson No. 9).
  - Let n_i denote the number of cells in the regulator at the end of the output i-th slot;
  - Let a_i denote the number of cells arrived at the regulator during the i-th (output) slot.
- We can thus write the classical M/G/1 difference equation where we have **to determine A(z)**, the PGF of the number of cells arrived at the regulator in a slot.
  - The cell arrival process is compound Bernoulli. Therefore, a_i corresponds to the PGF L(z) with probability q and corresponds to the PGF z^0 with probability 1–q:

$$A(z) = qL(z) + 1 - q$$

Solution (cont'd)

- Mean number of cells in the buffer and mean cell delay:

$$N_c = qL'(1) + \frac{qL''(1)}{2[1-qL'(1)]} \quad [\text{cells}] \qquad T_c = \frac{N_c}{A'(1)} = 1 + \frac{L''(1)/L'(1)}{2[1-qL'(1)]} \quad [\text{slots}]$$

- We compare input and output traffic burstiness, β_{in} and β_{out}, to study the effects of the leaky bucket regulator.
  - $\beta_{in} = L_{max}/[qL'(1)]$: the peak cell rate is equal to L_{max}/T; the mean cell rate is given by $qL'(1)/T$.
  - β_{out}: the peak cell rate is equal to $1/T$; the mean cell rate is obtained considering that the output cell rate is $1/T$ when the regulator is non-empty [with probability $1-P_0 = A'(1)$] and it is 0, otherwise. Consequently:

$$\beta_{out} = \frac{1/T}{(1-P_0)/T + P_0 \times 0} = \frac{1}{qL'(1)}$$

  - $\beta_{in} = L_{max}\,\beta_{out}$: **the leaky bucket regulator reduces the burstiness of the input traffic at the expenses of extra delays.**

Thank you!

giovanni.giambene@gmail.com