Scheduling on Links with Changing Parameters

Subir Varma

Aperto Networks
1637 South Main Street
Milpitas, CA 95035

1.0 Description of Proposed Invention

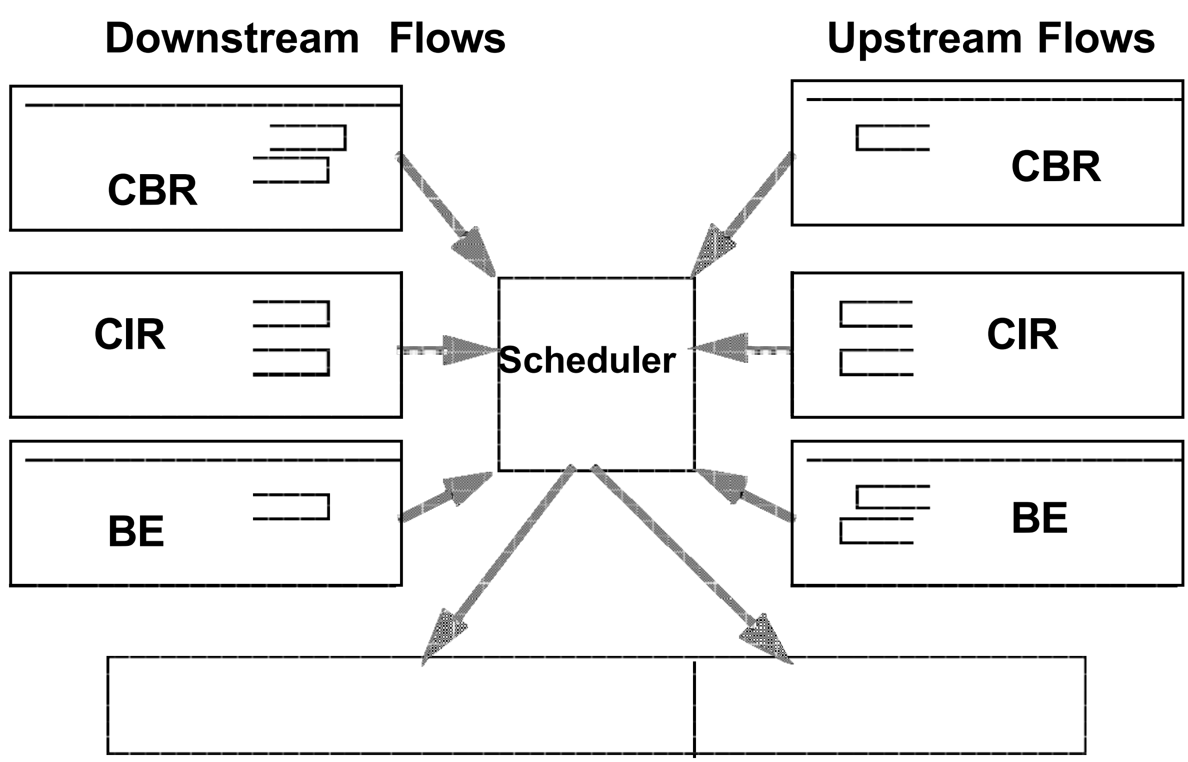
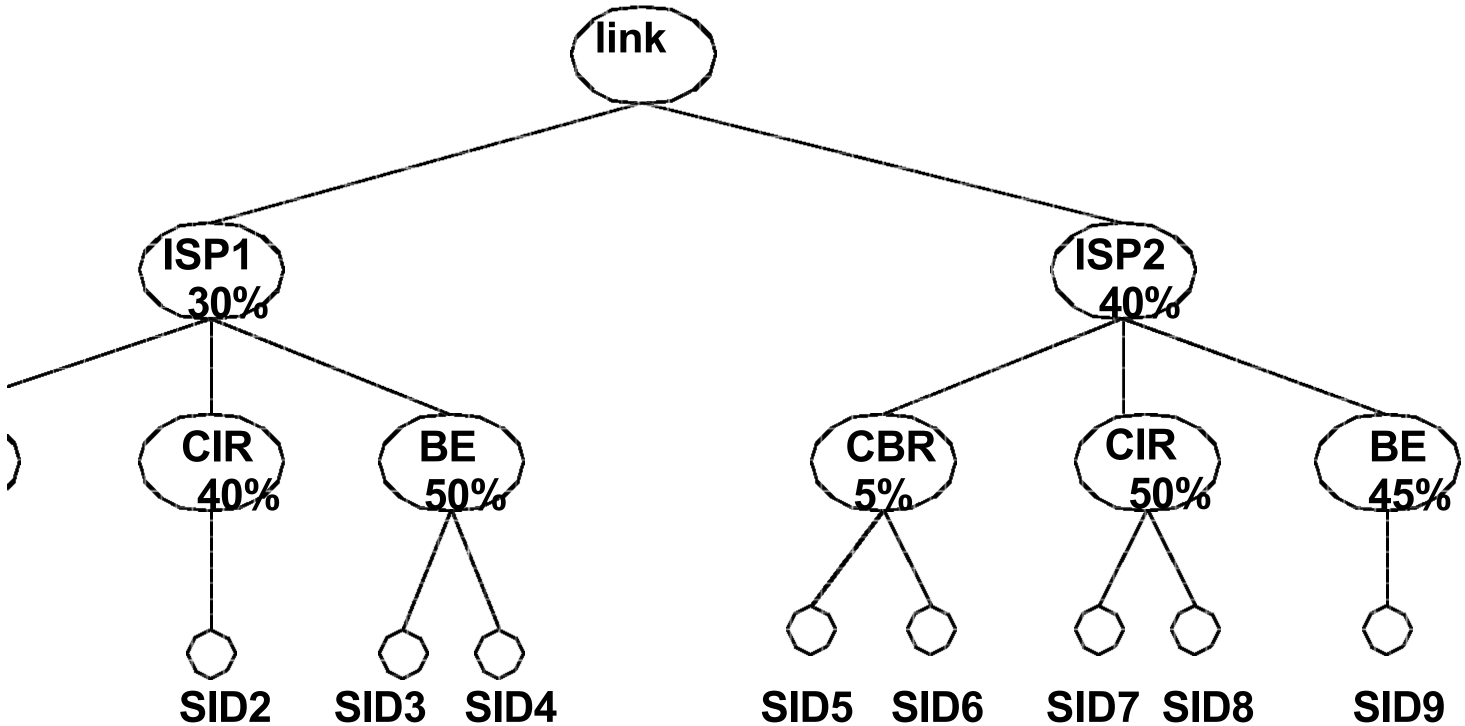


FIGURE 1.

FIGURE 2.


link
ISP1
30%
ISP2
40%
CIR
40%
BE
50%
CBR
5%
CIR
50%
BE
45%
SID2
SID3
SID4
SID5
SID6
SID7
SID8
SID9

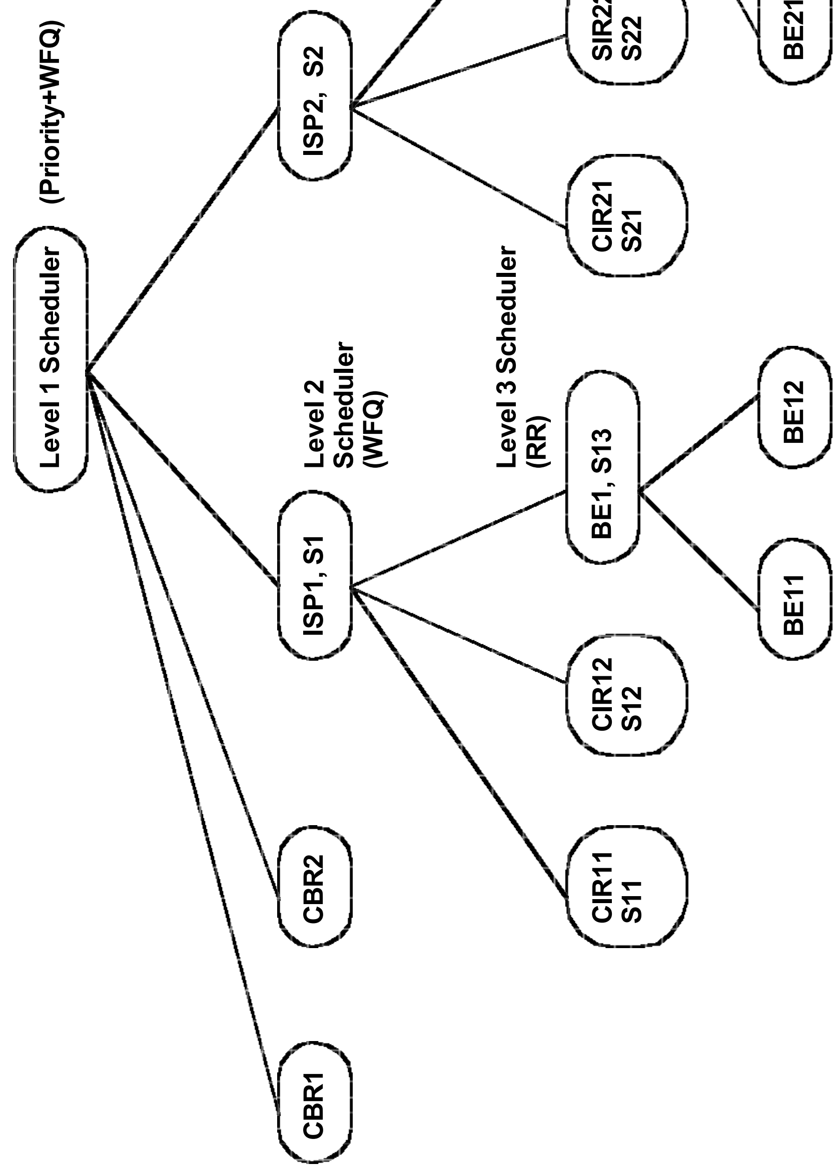

Level 1 Scheduler
(Priority+WFQ)
CBR1
CBR2
ISP1, S1
Level 2
Scheduler
(WFQ)
ISP2, S2
Level 3 Scheduler
(RR)
CIR11
S11
CIR12
S12
BE1, S13
CIR21
S21
SIR22
S22
BE2, S23
BE11
BE12
BE21
BE22
BE23

FIGURE 3.

1.1 Scheduler Features

- For both Upstream and Downstream directions, supports traffic flows belonging to different service classes, each with its own set of CoS/QoS parameters
- Makes efficient use of the available BW. If a flow is not making full use of its guaranteed BW, the unused portion should be made available for other flows
- Selects the position of the data slots within a TDD frame so as to minimize co-channel interference within a cell
- Takes into account the variable symbol rates and constellation types (per CPE)
- Takes into account WPDU error events and ARQ protocol operation
- Takes into account wireless related channel impairments, such as fading
- Integrate the various MAC management messages within the overall TDD frame structure

The basic unit of scheduling is a flow, which is a layer 2 entity and maps one-to-one with a SID. Flows exist in both directions, and may be set up statically at CPE initialization or dynamically at any time later. Each flow belongs to a pre-defined Service Class and is associated with its own set of CoS/QoS parameters. Each CPE has at least 1 BE upstream flow and 1 BE downstream flow associated with it. The scheduler periodically examines all backlogged flows (up and down), and allocated time slots on the channel to selected flows (Fig. 1). It selects the flows on the basis of CoS/QoS parameters for the flows.

1.2 Admission Control

CBR and CIR connections are subject to admission control:

- CBR connections: The sum of the peak rates of the admitted connections should not exceed the fraction allocated to CBR traffic. If the rate is exceeded, then the packet is dropped (or the call is blocked)
- CIR connections: The sum of the minimum gauranteed rates of the admitted connections should not exceed the fraction allocated to CIR traffic. If the rate is exceeded, then only BE service is provided to the SID

SIDs of the CBR and CIR type may be established at the of CPE initialization, however the scheduler should not reserve BW for them until they become active. When the classifier discovers a packet from one of these SIDs, then it performs admission control and reserves BW for the SID. If the SID becomes inactive, as detected by the fact that it has not transmitted packets for some interval T, then the BW reserved for that SID is de-allocated.

Note: In case of dynamic SID setup and teardown, the BW allocation and de-allocation is done at the time of connection setup and teardown respectively.

Assume that R_{CIR} bps is reserved for the CIR traffic class. Then if there are n CIR connections with minimum guaranted rates rates R1,...,Rn bps active, then the following condition should be satisfied:

$$r1 + r2 + ... + rn < R_{CIR}$$

If the admission of a new connection causes this condition to be violated, then that connection is given only BE service. The relation beteen Ri and ri is as follows: Let Rmax bps be the maximum link rate (=19.2 mbps using 16QAM and 4.8 msymbols/s symbol rate), and let R(i), i = 1,...K be the other link rates. Then

$$ri = (Rmax/R(i)).Ri$$

where R(i) is the link rate for flow i

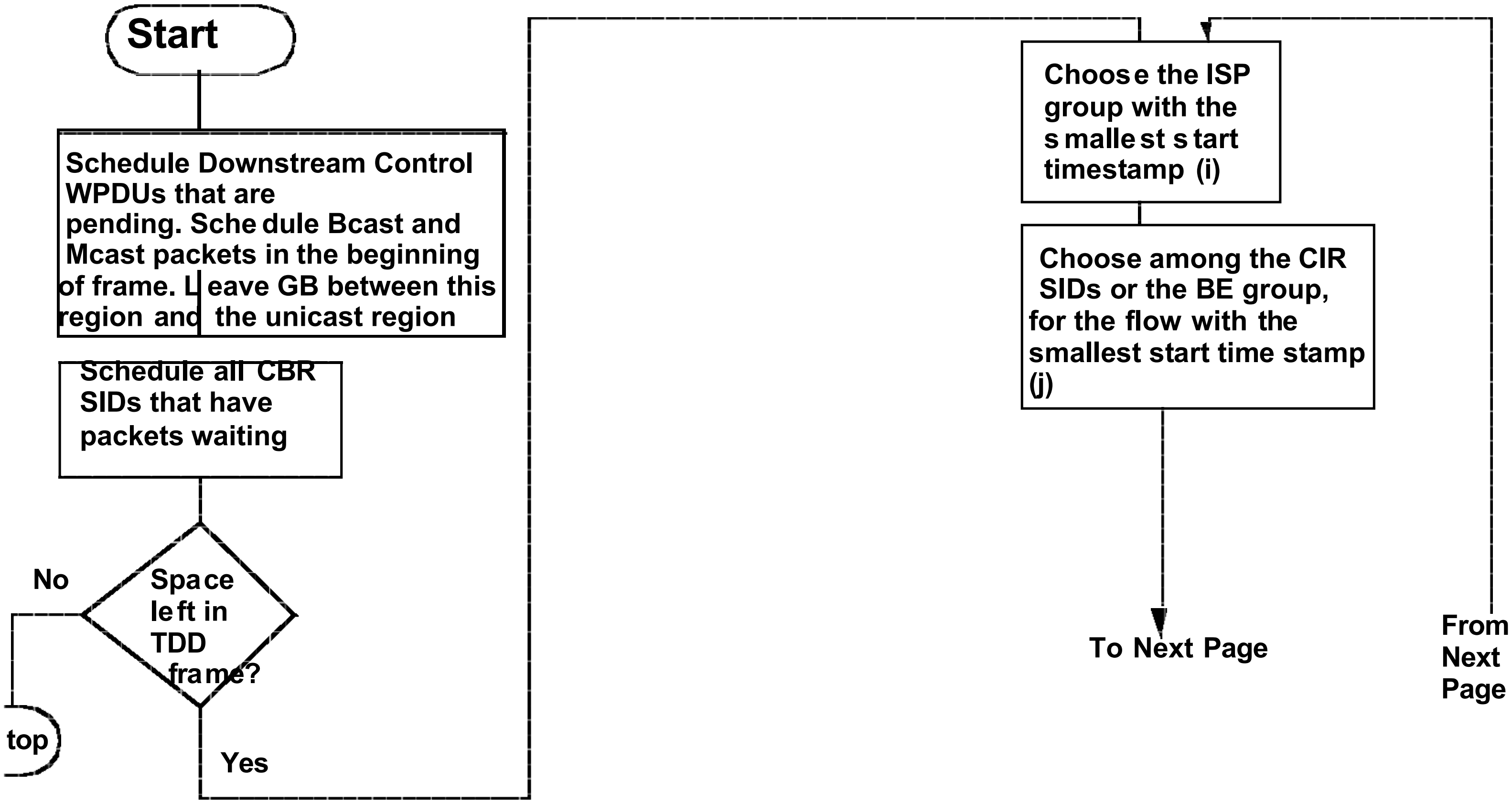
Justification: The bit rate allocated to a CPE is scaled up to the bit rate it would have if the CPE were to operate at maximum symbol rate using 16QAM.

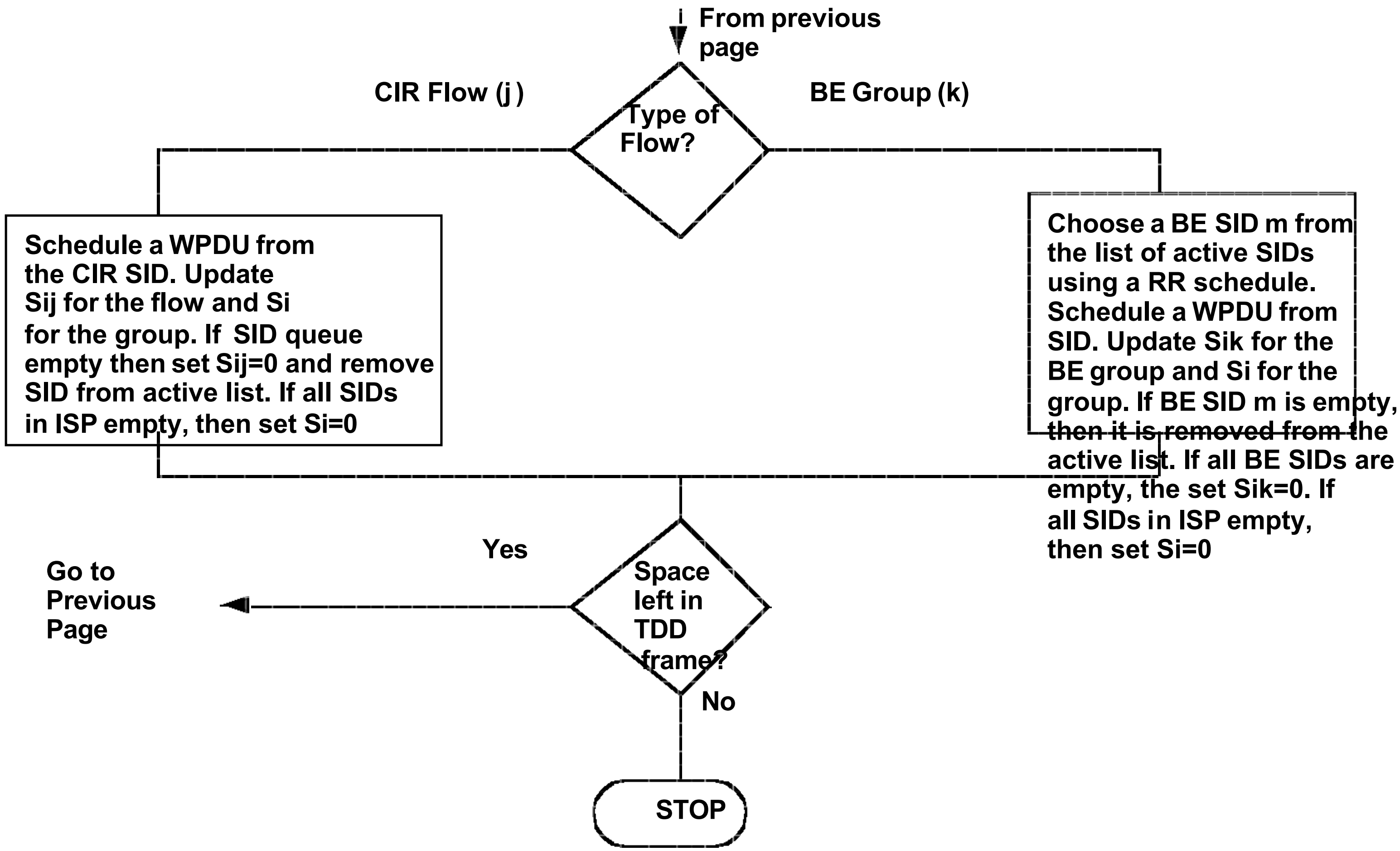
1.3 Scheduler Structure

There are three levels of scheduling (Fig. 3) -

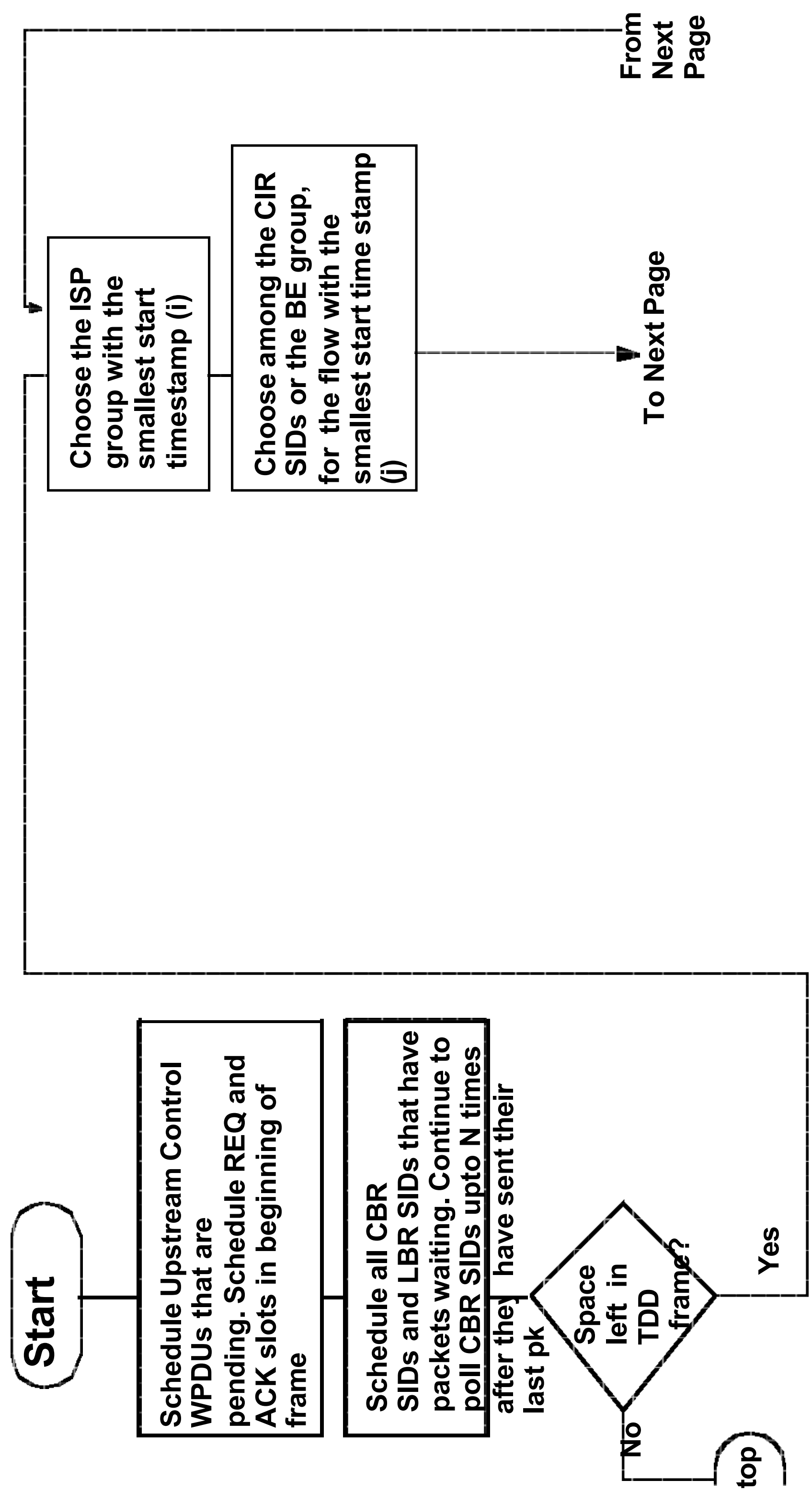
- The level 1 scheduler: When invoked, this scheduler chooses among the MAC Control WPDUS, CBR SIDs and the ISPs, as the source for the next WPDU. The CBR SIDs may span multiple ISPs. The MAC control WPDUs have the highest priority, followed by CBR SIDs which have higher priority then the ISPs. the choice among the ISPs is made on the basis of a Fair Queueing scheme
- The level 2 scheduler: There is one instance of this scheduler for each ISP. When invoked, this scheduler chooses among the CIR SIDs and the BE class belonging to the ISP, as the source for the next WPDU. The choice is made on the basis of a Fair Queueing scheme
- The level 3 scheduler: There may be one or instances of this scheduler per ISP. When invoked, this scheduler chooses among the BE SIDs that belong to the ISP. The choice is made on the basis of a Round Robin scheme, that tries to allocate approximately equal time slices to each active BE SID

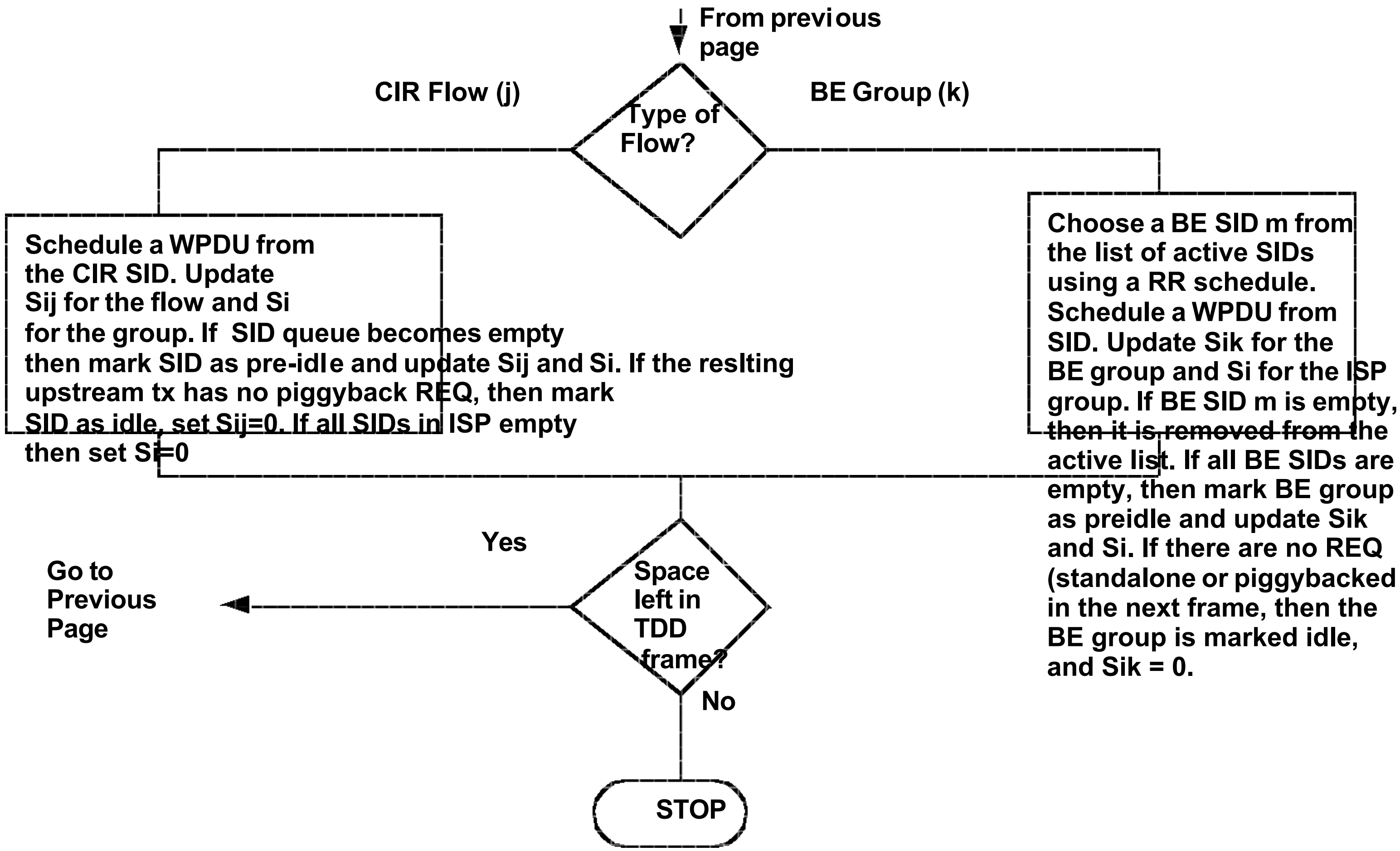
1.4 Scheduler Operation - Downstream MAP Construction



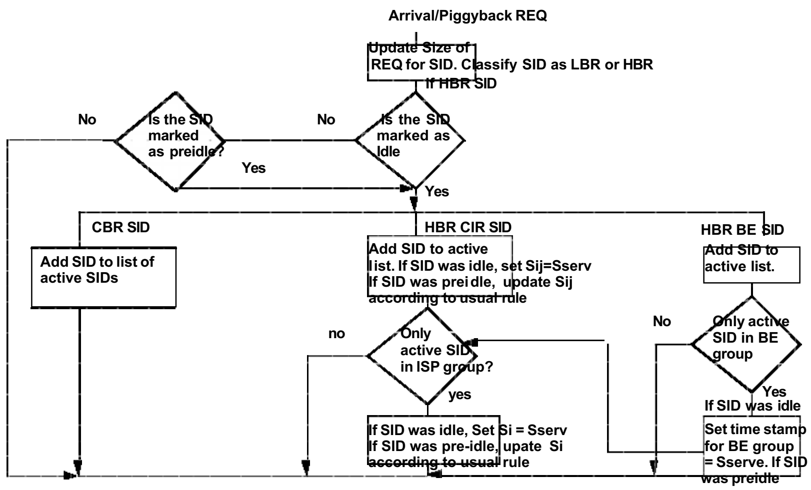

From previous page
Type of Flow?
CIR Flow (j)
BE Group (k)
Schedule a WPDU from the CIR SID. Update Sij for the flow and Si for the group. If SID queue empty then set Sij=0 and remove SID from active list. If all SIDs in ISP empty, then set Si=0
Choose a BE SID m from the list of active SIDs using a RR schedule. Schedule a WPDU from SID. Update Sik for the BE group and Si for the group. If BE SID m is empty, then it is removed from the active list. If all BE SIDs are empty, the set Sik=0. If all SIDs in ISP empty, then set Si=0
Space left in TDD frame?
Yes
Go to Previous Page
No
STOP

1.5 Scheduler Operation - Upstream MAP Construction

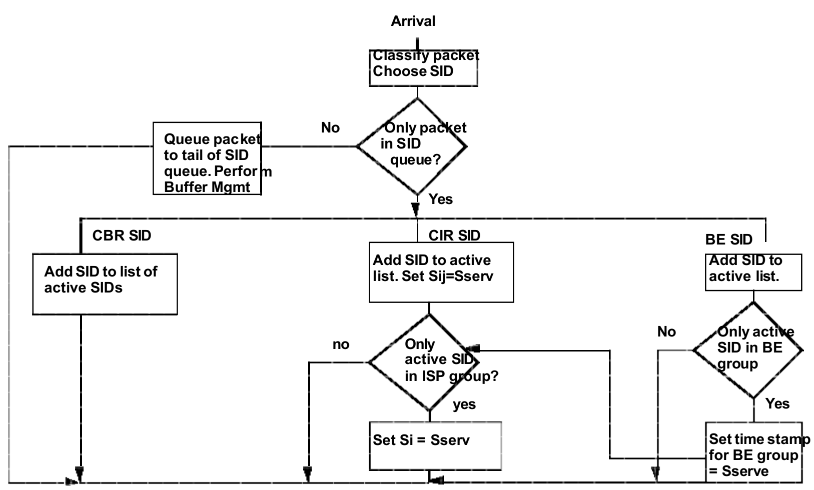



From previous page
CIR Flow (j)
Type of Flow?
BE Group (k)
Schedule a WPDU from the CIR SID. Update Sij for the flow and Si for the group. If SID queue becomes empty then mark SID as pre-idle and update Sij and Si. If the reslting upstream tx has no piggyback REQ, then mark SID as idle, set Sij=0. If all SIDs in ISP empty then set Si=0
Choose a BE SID m from the list of active SIDs using a RR schedule. Schedule a WPDU from SID. Update Sik for the BE group and Si for the ISP group. If BE SID m is empty, then it is removed from the active list. If all BE SIDs are empty, then mark BE group as preidle and update Sik and Si. If there are no REQ (standalone or piggybacked in the next frame, then the BE group is marked idle, and Sik = 0.
Yes
Go to Previous Page
Space left in TDD frame?
No
STOP

1.6 Upstream REQ Arrival



1.7 Downstream Packet Arrival



1.8 Scheduling CBR Flows

CBR flows are always subject to peak rate regulation, on a per SID basis. CBR traffic is given the highest priority during the allocation in each TDD frame, and as much of the CBR traffic is allocated in the frame, as there is space available. If the CBR flows are not dynamically setup and tore down, then the admission control and BW allocation is done on the basis of presence and absence (within some timeout period) of packets from the CBR SIDs.

- Dynamic Connection Setup and Teardown Case: The connection may specify the precise interval between packets and the size of the packets. In this case the scheduler must allocate WPDUs at the precise intervals (using the UGS service class for the upstream)
- Non dynamic connection setup and teardown case: The downstream case is straightforward. In the upstream case, the NRTP class should be used

Example: Given a VoIP flow, in which 20 bytes of voice data are generated once every 20 ms (rate = 8 kbps). At the time of flow set up, compute size of the WPDU payload (= voice data (8) + RTP header (12) + UDP header (8) + IP header + Ethernet header), and compute the resulting WPDU payload rate based on this number. Reserve the resulting CBR rate on the upstream and downstream links

1.9 Scheduling Best Effort Flows

Serving BE flows with a RR type discipline that tries to serve an equal numberof bytes from each BE SID leads to the following problem: SIDs to CPEs with low CRI take a disproportionately large amount of ticks, thus degrading the performance of SIDs on CPEs with higher CRIs.

Strategy: Let N be the number of ticks required to transmit the max. size WPDU using the highest CRI. Then the number of ticks allocated to the lower CRIs is approximately in the region around N, under the constraint that an integral number of FEC codewords can be accomodated within it.

1.10 Scheduling CIR Flows

The set of SIDs that are given minimum upstream rate guarantees are placed in a separate group, and each SID in this group has a time stamp associated with it. These timestamps are computed based on the minimum BW guarantees to these SIDs, as well as their current rate of BW consumption. For each upstream frame, and there is a certain number of active CIR SIDs waiting for grants. For each of these SIDs the scheduler knows the total number of outstanding requests and the timestamp associated with that SID. The scheduler allocates grants to the SIDs in the current frame, in increasing order of timestamp priority, until the frame is full. After each grant, the scheduler updates the timestamp of the SID to which the grant was made and recomputes the minimum over the modified set. The grants are in units of MaxWPDU size, or smaller. Each grant is given a Sequence Number for ARQ purposes. When a CIR flow starts, the scheduler checks whether there is enough BW left over in the CIR allocated region for the new flow. If no, then the flow is given BE service (and the network management agent is informed about this).

The scheduler executes the following algorithm during each scheduling cycle:

1. Choose the CIR flow with smallest start timestamp, and allocate L_{MAC} bytes from that flow, which requires $N_T(f)$ ticks. Reduce the tick allocation for the CIR traffic class in the TDD frame by $N_T(f)$ ticks.

- If the flow has less than L_{MAC} bytes, then recompute the number of ticks required
- If the number of ticks left is less than required, then compute the number of bytes that can be accomodated in those many ticks, and allocate them (the number of ticks should be large enough to accomodate at least one FEC codewode worth of data)

2. Compute the service interval SI for the allocation and Update theTimestamp for the flow as follows:S_f is the start timestamp for SID f, while F_f is the finish timestamp. SI is set equal to the number of bytes allocated times the weight associated with that SID
3. If the region reserved in the TDD frame for the CIR flow is used up and/or all flows are empty, then stop. Otherwise go back to step 1.

Integration with the ARQ protocol: The scheduler assigns Sequence Numbers to all upstream WPDUs. It does not discard the REQ until it correctly receives the data. In case there is an error, then it directs the CPE to re-transmit (upto n times before it finally discards the REQ). Since the WFQ timestamp is per SID, if a REQ has to be re-transmitted, it does not affect the rules for assigning timestamps to the flow. For example a REQ may be associated with TS1 in frame m, and may also be associated with timestamp TS2 in frame m+1, in case it has to be re-transmitted.

Let S_{serv} be the start time stamp of the WPDU that was served last. If a packet arrives to a SID queue which is currently empty, then its Start time stamp is set to $S(f) = \max(S_{serv}, F(f))$.

The complexity of choosing the next CIR SID to be served = O(log N), where N is the number of active SIDs. For example, for 300 CIR SIDs per channel with 50% activity ratio, the complexity is O(8). With 8 CIR WPDUs per TDD frame, the number of accesses to memory required to schedule the CIR portion of the TDD frame = 64. The size of the timestamp should be sufficient to accomodate the largest Service Interval. In our the system the largest service interval is 73,728 (corresponding to the smallest BW allocation of 64 kbps) so that 17 bits is the minimum required. If we choose the Size of timestampto be 32 bits then this is more than sufficient to support the bandwidth granularity in the system.

2.0 How the Proposed Invention Extends the Current Art

- Existing schedulers assume that the capacity of the underlying link is fixed. In our case, the capacity of the link is a function of the CPE to which the burst is being sent to, or being received from, hence existing scheduling theory cannot be used. The proposed scheduler has several new features that enable operation over such a link possible.