dvjClause09McFairness03.fm

March 10, 2003 8:52 am

Multichoke fairness

This discusses multichoke fairness.

9.1 Problematic fairness frame

The format of the proposed fairness frame is shown in Figure 9.1, and problems are illustrated below:.

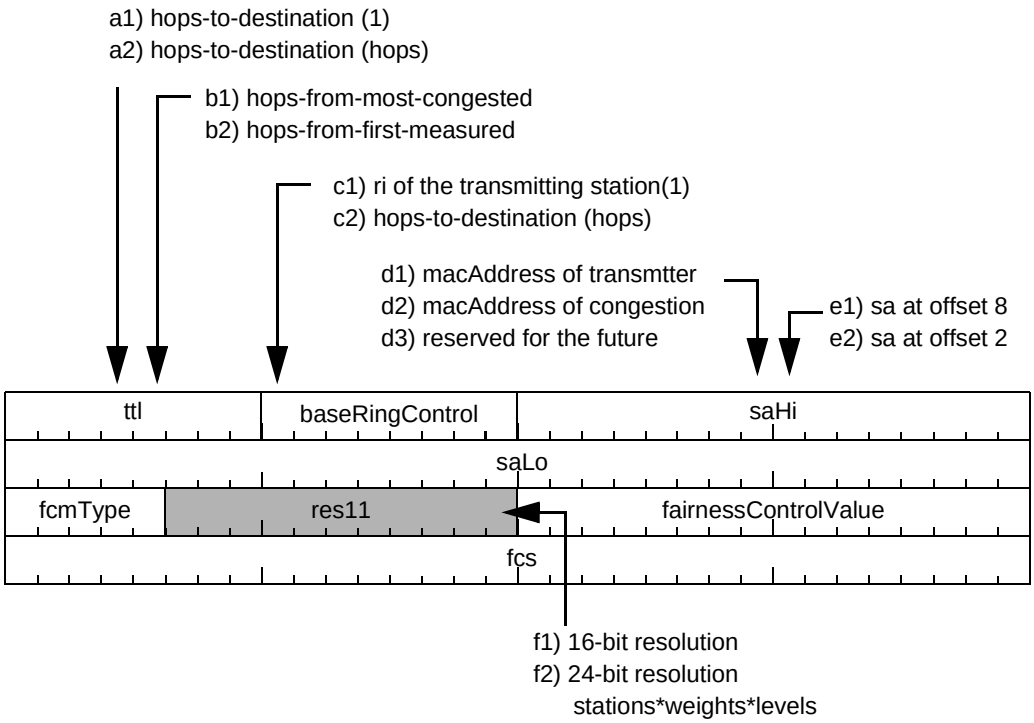


Figure 9.1—Problematic fairness frame

a) How should the ***ttl*** field be defined?
   1) Current&consistent: distance from the source transmitter.
   2) Fairness useful: distance from the congestion point.

b) The ***ttl*** field provides two distinct indications, with no distinctions:
   1) Hops from the worst of all congestion points (previous have been sampled).
   2) Hops from the worst of sampled congestion points (starts at relieved congestion point).

c) How should the ***ri*** bit be defined?
   1) Current&consistent: ringlet0/ringlet1 of the source transmitter.
   2) Fairness useful: ringlet0/ringlet1 of the congestion point.

d) What does the ***sa*** (source address) field represent?
   1) Current&consistent: sa of the source transmitter.
   2) Fairness useful: not applicable (hops from congestion is more useful).

e) What is the ***sa*** (source address) field offset?
   1) Current&consistent: 8, as is true for control and data frames.
   2) Fairness useful: not applicable (hops from congestion is more useful).

f) What is the ***fairnessControlValue*** resolution?
   1) Current: 16, since others complained when it was larger.
   2) Desired: 24, as needed due to the following contributions:
      i) 8 bits: to encode the 255 stations may be contributing
      ii) 8 bits: the fairness weight resolutioin
      iii) 8 bits: sufficient transmission-level resolution.

g) How do we support multichoke congestion?
   1) With distinct multichoke messages, sent periodically.
   2) With single-choke messages, with tweeked start-points and additional parameters.

9.1.1 Proposed solution

The format of the proposed fairness frame is shown in Figure 9.2.

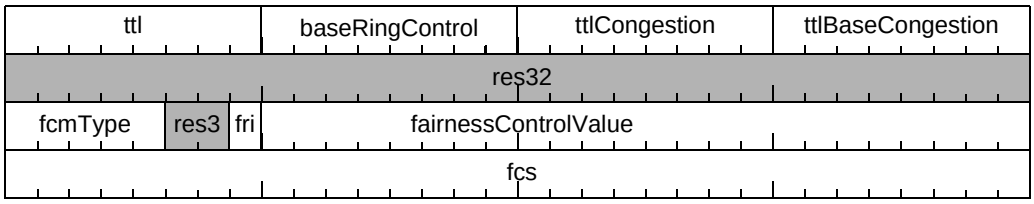


Figure 9.2—typeLengthValue format

The 8-bit ***ttl***, 8-bit ***basrRingControl***, 5-bit ***fcmType***, and 32-bit ***fcs*** field are specified in 8.3.1, 8.3.2, 9.5.1, and 8.3.9 respectively. The 32-bit ***res32*** and 2-bit ***res3*** fields shall be reserved.

The 8-it ***ttlCongestion*** (time to live, congestion) field is set to MAX_STATIONS at the first-measured congestion point and decremented when passing through same-station fairness control units.

The 8-it ***ttlBaseCongestion*** (time to live base, congestion) field is set to *ttlCongestion* at the worst-case congestion point.

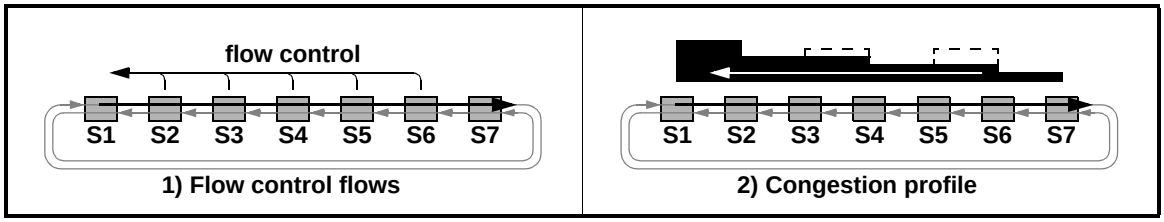
The ***fri*** (fairness ringlet identifier) bit specifies the source-location of congestion information, where 0 and 1 correspond to ringlet0 and ringlet1 respectively.

9.2 Fairness operation

9.2.1 Flow-rate indications

The multichoke fairness protocol is intended to efficiently support nonblocking virtual queues, by distributing congestion information with a per-hop granularity. Flow control information for station S1-to-S7 transfers is returned from stations S2-through-S6, as illustrated in the left side of Figure 9.3.

Figure 9.3—Flow controlled stations



This flow control information is intended to provide a flow-rate profile of the downstream stations, as illustrated in the right side of Figure 9.3. That profile is monotonic decreasing plot of congested flow-rates versus hop-count distance. In this illustration, the flow-rate for the S1-to-S3 link is reduced by the limited flow rate over the S2-to-S3 link. The flow-rate on the S3-to-S4 link (in this example) is larger and does not affect the flow-rate profile; however, the S3-to-S4 link flow rate is masked and not visible to the S1 station.

Station S1 uses the flow-rate profile to limits its flow to no more than the worst-case (i.e. the smallest) flow-rate between itself and its selected destination station, on a per frame basis. The profile allows transmissions to be evaluated on a hop-count distanced basis. Thus, moderate-rate transmissions can be allowed over the S1-to-S4 or S1-to-S6 paths, despite flow-rate blocking of lower-rate S1-to-S7 transmissions.

Such flow-rate profiles could be generated by having each station broadcast its congestion information in the upstream station. Such distribution strategies would unnecessarily suffer from excessive overheads, since the overhead would increase in proportion to the number of stations *N*. Sending fairness messages less frequently, could compensate for the cumulative effect of broadcast messages, but would have the undesirable effect of increasing the message-delivery latencies and there the system response times.

A different approach is therefore taken: flow-rate messages are point-to-point periodic messages. These periodic messages pass into a station, the flow-rate information from that station is merged, and that merged information is sent to the next downstream station. The merging process is not lossless (the larger S3-to-S4 flow-rate remains hidden), but maintain the important monotonic-decreasing flow-rate profile information.

The key features for supporting this feature include the following:

a) Dithered start. As currently defined, fairness messages “start” at a relieved congestion point. Dithering the starting point, to start at a pseudo-random point, this would light-up dark spots.
b) Supplemented. Sufficient storage space (two 16-bit values) would enable soft startup, by providing additional parameters, perhaps such as ***newFlow***&***topFlow,*** as illustrated in the following subclause.

9.2.2 Rate comparisons

Rate parameters, as illustrated in Figure 9.4.

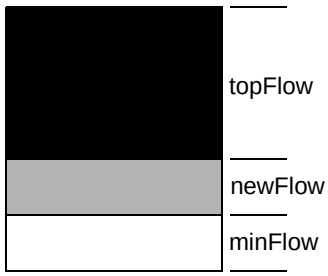


Figure 9.4—Fairness parameter values