Annex J

(informative)

Implementation guidelines

J.1 Weighted fairness

J.1.1 Overview

Weighted fairness with spatial reuse is a goal of the RPR protocol. The MAC fairness model regulates the flow aggregates from a given ingress node, while allowing the client to regulate independent end-to-end flows within that aggregate. Flow aggregates may be subpartitioned by the client, based on their destination station, to ensure maximal spatial reuse opportunities.

To better understand the implications of these fairness strategies, multiple traffic-load scenarios are illustrated. For each scenario, the desired and possible problematic behaviors are defined.

J.1.2 Scenarios

The following scenarios demonstrate the aggregate flow rates, expressed as a share of the total capacity. In each case, a tandem segment of a ring is depicted. The input traffic is assumed to be constant rate; if not otherwise specified, the offered input rate equals the link capacity for each depicted flow.

Within the scenario illustrations, the percentage number associated with a flow represents the percent of the nominal link bandwidth that is consumed by that flow. The numbers above the flow line correspond to the desired bandwidth (from a theoretical perspective); the numbers below the flow line correspond to the bandwidths predicted for a suboptimal implementation.

J.1.2.1 Scenario 1: Parking lot

The first parking lot scenario illustrates the desire to avoid overly throttling one source over another, based on the distance of the source from the most congested (choke-point) link, as illustrated in figure J.1 and figure J.2. The objective is to evenly partition the available choke-point bandwidth between each of the contending upstream stations.

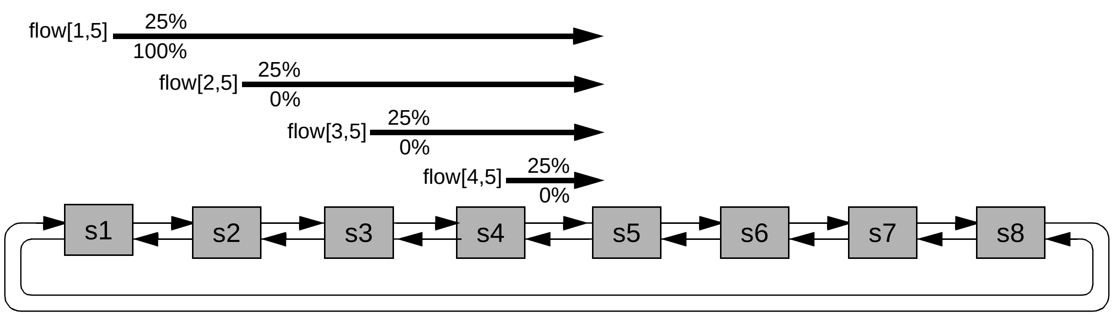

Figure J.1—Scenario 1a: Parking lot

Concern: Station 1 consumes the entire link bandwidth, due to transit-queue transmission precedence.
Applicable: This problem is addressed by single-choke and multi-choke fairness algorithms.
Solution: Throttle upstream stations with congestion information sent from downstream stations.

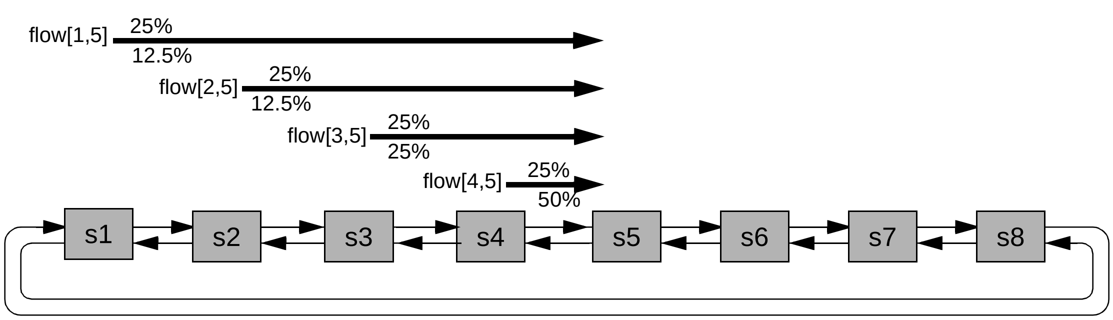

Figure J.2—Scenario 1b: Parking lot

Concern: Station 1 traffic is overly restricted, due to others' alternate selections of transit-queue and stage-queue transmissions. This is an example where local fairness does not ensure global fairness.
Applicable: This problem is addressed by single-choke and multi-choke fairness algorithms.
Solution: Communicate global knowledge by passing congestion information beyon the upstream station.

J.1.2.2 Scenario 2: Parallel parking lot

The second parking lot scenario illustrates the desire to support spatial reuse of subaggregate flows, as illustrated in figure J.3. The objective is to evenly partition the available choke-point bandwidth between each of the contending upstream stations, without throttling nonconflicting flows.

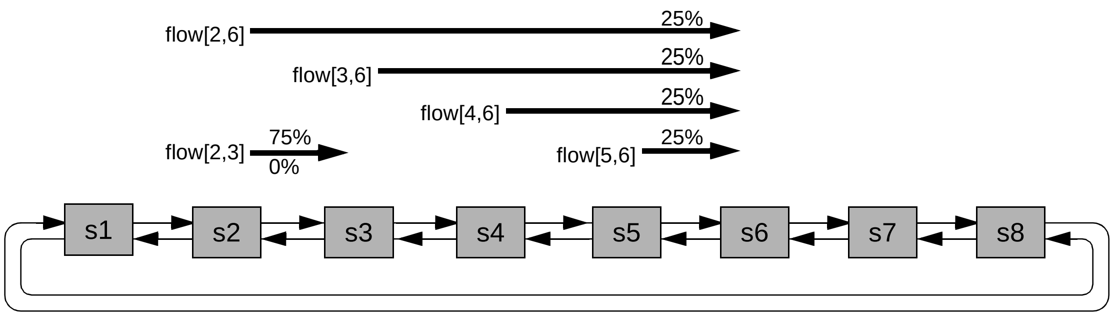

Figure J.3—Scenario 2: Parallel parking lot

Concern: The station2-to-station3 transmissions are throttled by station5-to-station6 congestion indications.
Applicable: Occurs when the hop-count and congestion level are not both reported to the client.
This problem is addressed by single-choke and multi-choke fairness algorithms.
Solution: Two capabilities are required to avoid near-side starvation:
a) The MAC provides hop-count as well as congestion-level information to the client.
b) The client provides distinct virtual destination queues (VDQs) for several hop-count distances.

J.1.2.3 Scenario 3: Upstream parallel parking lot

The next parking lot scenario illustrates the desire to support spatial reuse of nonconflicting flows, as illustrated in figure J.4. The objective is to evenly partition the available choke-point bandwidth between each of the contending upstream stations, without throttling nonconflicting pass-through flows.

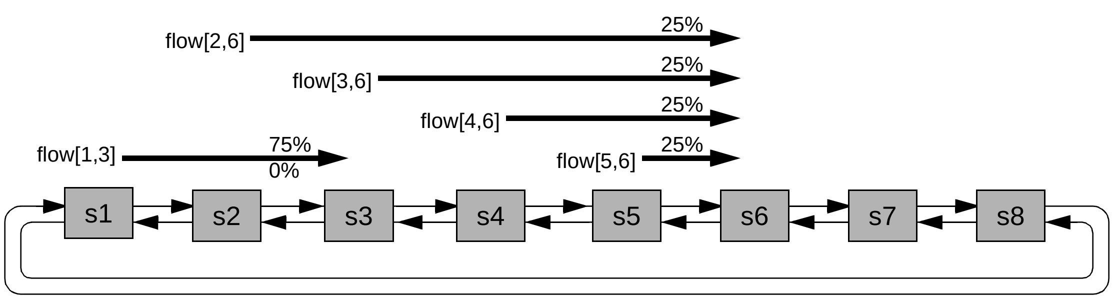

Figure J.4—Scenario 3: Upstream parallel parking lot

Concern: The station1-to-station3 transmissions could be unnecessarily throttled.
This topology also exhibits the oscillatory behaviors of scenario 7 (see J.1.2.7).
Applicable: Occurs when the hop-count and congestion level are not both reported to the client.
This problem is addressed by single-choke and multi-choke fairness algorithms.
Solution: Two capabilities are required to avoid near-side starvation:
a) The MAC provides hop-count as well as congestion-level information to the client.
b) The client provides distinct virtual destination queues (VDQs) for several hop-count distances.

J.1.2.4 Scenario 4: Multi-flow parking lot

This parking lot scenario illustrates the effect of supporting weighted aggregate flows, as illustrated in figure J.5. The objective is to evenly partition the available choke-point bandwidth between each of the contending upstream stations' aggregate flows, rather than equally weighting individual flows within the aggregate.

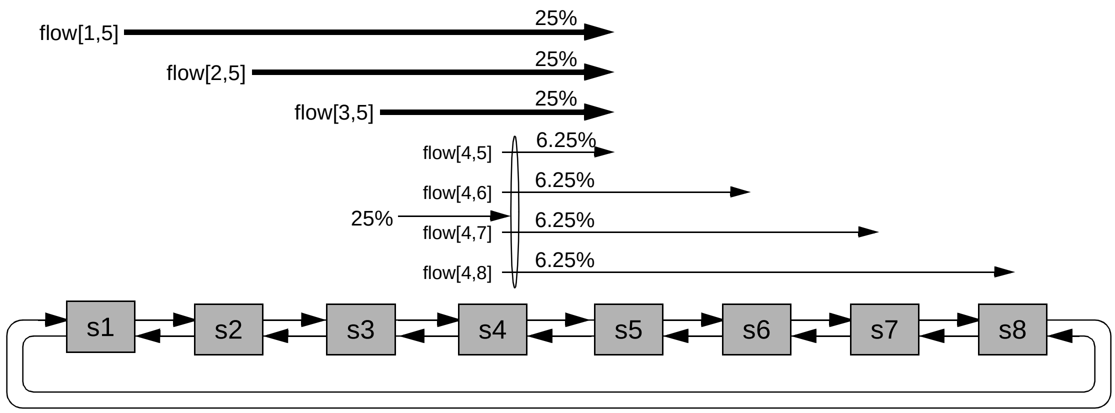

Figure J.5—Scenario 4: Multi-flow parking lot

Concern: Insufficient information is available for the client to allocate per-destination flow bandwidths. This could be the effect of using an on-off MAC-to-client flow-control indication.
Applicable: This appears to not occur with either single-choke or multi-choke fairness.
Solution: Provide a hop-count based congestion indication to the client.

J.1.2.5 Scenario 5: Dual-exit parking lot (multiple choke points)

This dual parking lot scenario illustrates an effect of having multiple choke points, as illustrated in figure J.6. The objective is to evenly partition the available choke-point bandwidth based on primary as well as secondary choke-point locations, so that station2 and station3 evenly distribute the 40% residual bandwidth available on the station4-to-station5 link.

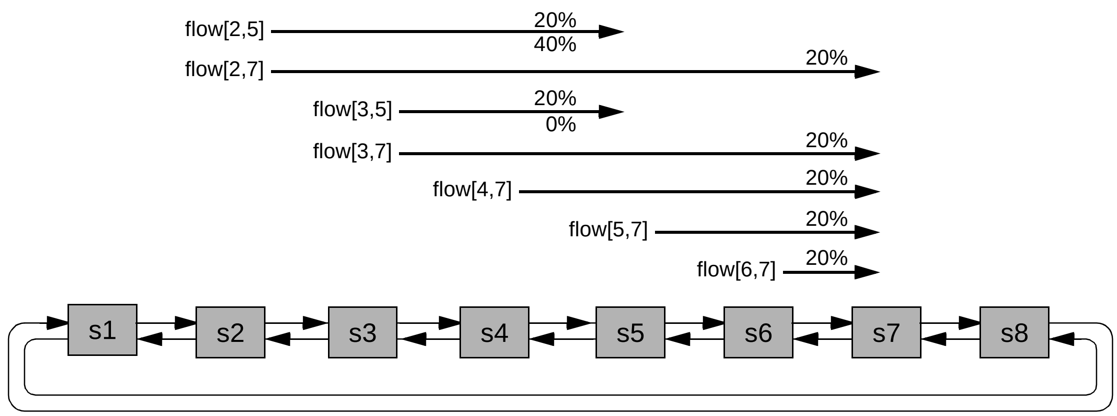

Figure J.6—Scenario 5: Dual-exit parking lot (multiple choke points)

Concern: Flows into the less congested station4-to-station5 hop observe only the most-congested indication from station7.
Applicable: This appears to happen with single-choke but not multi-choke fairness.
Solution: Each station must be aware of all choke-point conditions, not just the worst congestion point.

J.1.2.6 Scenario 6: Migrating choke point

This migrating choke-point scenario illustrates another effect of having multiple choke points, as illustrated in figure J.7. In this example, station5 is the original choke point that forces bandwidth to be divided equally between flow[1,8], flow[2,8], flow[4,7], and flow[5,6].

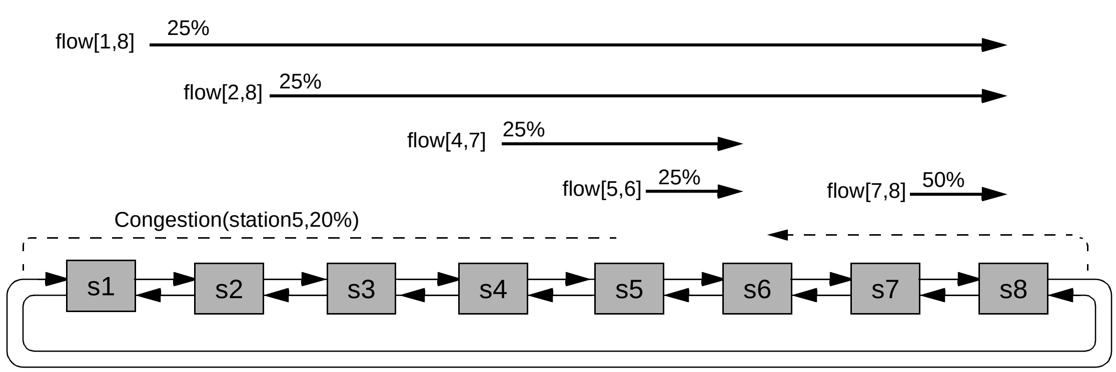

Figure J.7—Scenario 6a: Stable station5 choke point

Reductions in flow[5,6] and flow[4,7] cause station5 to deassert its congestion indication, as illustrated in figure J.8. A stale version of that assertion remains observable to station7, inhibiting station7's assertion of its less-congested indication. Thus, nearly a full ringlet circulation time can pass between 1) the sensing of a not-congested indication at station1&station2 and 2) The reassertion of a station7-asserted congestion indication. During that time, flow[1,8] will be momentarily unconstrained before flow[2,8] congestion reduces its flow (not illustrated); station7 remains starved.

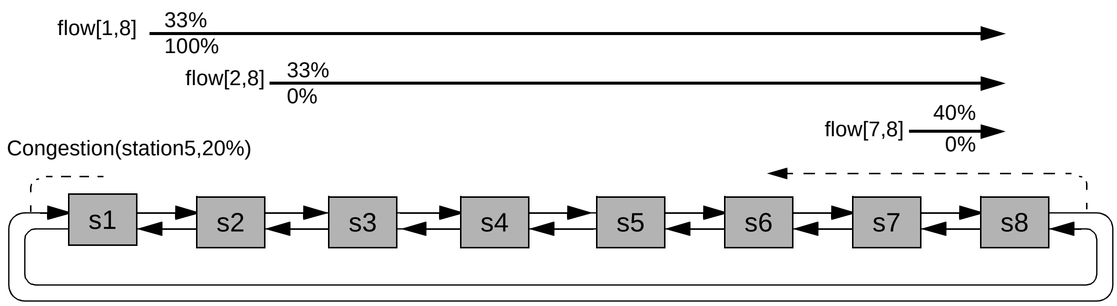

Figure J.8—Scenario 6b: Migrating choke point

The transient condition will eventually stabilize and normal fairness algorithms will apply, as illustrated in figure J.9.

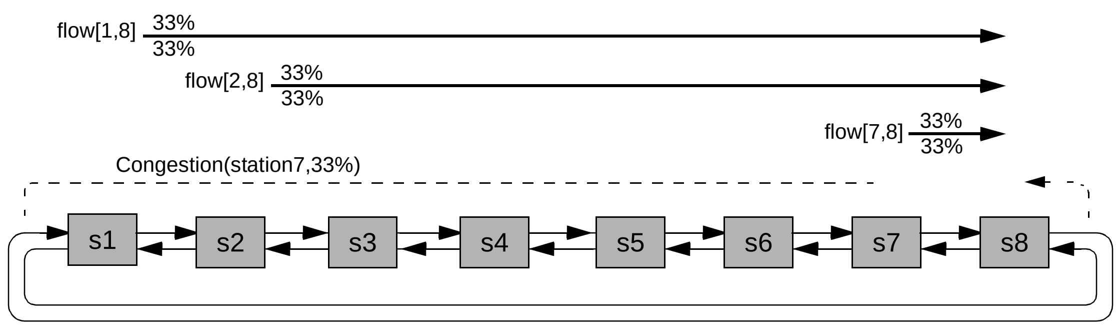

Figure J.9—Scenario 6c: Stable station7 choke point

Concern: The congestion indication temporarily vanishes when station5 congestion is relieved, as illustrated in figure J.8. This occurs during the distributed computation of the new most-choked indication.
Applicable: Addressed by both single-choke and multi-choke fairness.
Solution: The solution differs for single-choke and multi-choke fairness:
a) Single-choke. A slow ramp time avoids excessive flow[1,8] and flow[2,8] transmissions.
This is more effective when the ramp time more than exceeds the ringlet-circulation time.
b) Multi-choke. Each station has knowledge of all congestion conditions, allowing theworst-case congestion location(s) to be determined without this worst-case distributed-congestion computation glitch.

J.1.2.7 Scenario 7: Choked high/low bandwidth pairs

Editors' Notes (DVJ): *To be removed prior to final publication.*

The following discussion on TTL aging better belongs in the overview (with details in clause 6), but is being retained in this annex until new placement location is confirmed.

This scenario illustrates an effect of having time-delayed transmission of flow-control indications, as illustrated in figure J.10. In this scenario, station1 is a high bandwidth source trying to send steady traffic at the highest possible rate; station5 is lower-bandwidth source providing an offered load of approximately 5% of the link-bandwidth.

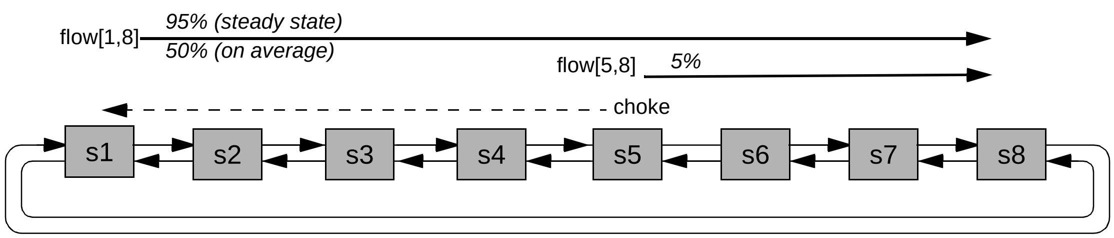

Figure J.10—Scenario 7: Choked high/low bandwidth pairs

This scenario can produce oscillatory flow conditions, as illustrated in Figure J.11. At time t_A, flow[5,8] is active in the absence of upstream-flow conflicts; flow[1,8] is stopped in the absence of a station7-sourced choke indication. At time t_B, flow[5,8] is stopped due to upstream-flow conflicts; flow[1,8] is active in the absence of a station7-sourced choke indication. At time t_C, flow[5,8] is stopped due to upstream-flow conflicts; flow[1,8] is active in the absence of a station7-sourced choke indication. At time t_D, flow[5,8] is active in the absence of upstream-flow conflicts; flow[1,8] is active in the absence of a station7-sourced choke indication. This cycle repeats indefinitely.

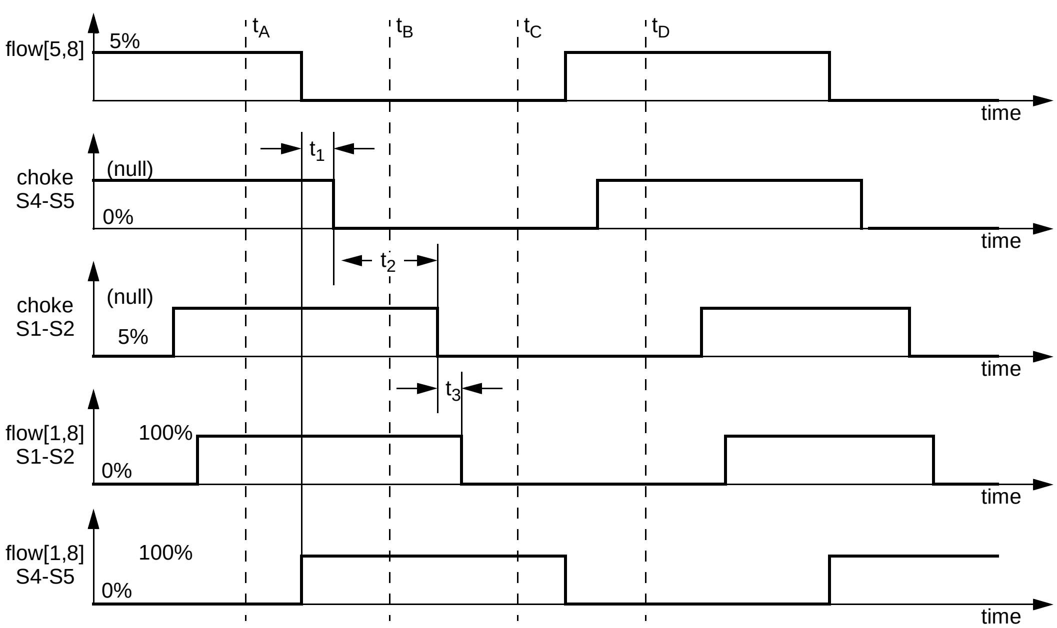

Figure J.11—Scenario 7a: Oscillatory flows&indications, fast ramp

Within Figure J.11: time ***t1*** is the station5's delay between sensing and reporting of congestion; time ***t2*** is the propagation delay between station5 and station1; time ***t3*** is station1's delay between receiving station5's congestion condition and inhibiting the excessive station1 data transmissions.

Concern: The flows from station1 are unnecessarily and severely limited in a cyclical fashion.
Applicable: This could happen with either single-choke or multi-choke fairness protocols.
The problem exists with mono-queue and dual-queue designs, with distinct t1 time parameters:
a) For a mono-queue design, t1 is based on the rate-averaging interval.
b) For a dual-queue design, t1 is based on the congestion-depth threshold of the STB.
Solution: A multi-choke fairness solution would be to provide linear flow-rate and congestion levels, to support a more linear (as opposed to on-off) feedback control mechanism.

A longer single-choke ramp-up can avoid problems associated with round-trip oscillations. However, the ramp-up protocols exhibit their own oscillatory behaviors, as illustrated in Figure J.11. In this case, station1 is periodically shut off by station5 and has to ramp up slowly.

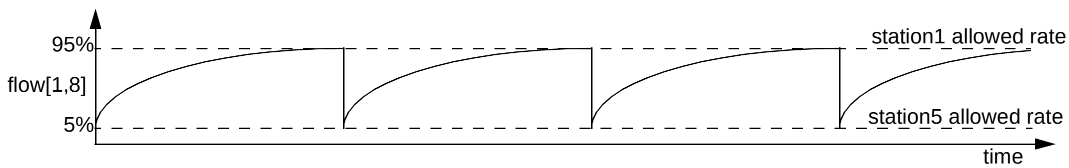

Figure J.12—Scenario 7b: Oscillatory flows&indications, slow ramp

J.1.2.8 Scenario 8: Rotating choked pairs

The rotating choked pairs scenario illustrates an effect of measuring transient rates, as opposed to overall byte transfers, as illustrated in figure J.13. In this scenario, two pairs of overlapping high/low transmission segments (as described in J.1.2.7) are present. The high/low station2-to-station4 and station5-to-station7 segments produce cyclical choke-point indications, *choke3* and *choke6* respectively.

The objective is provide station1 with a flow[1,8] allowance equal to the average time-averaged flows from station2 and station5..

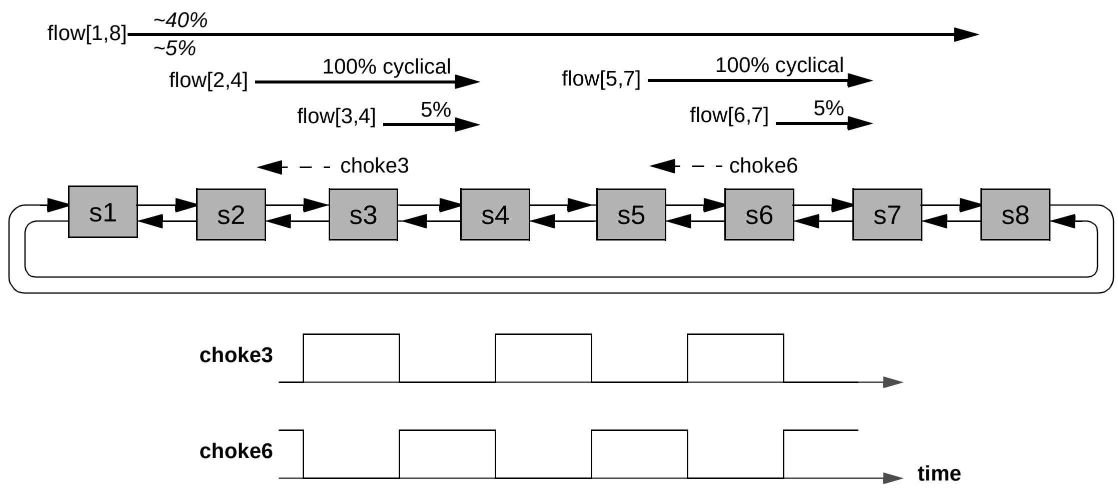

Figure J.13—Scenario 8: Rotating choked pairs

Concern: Station1 is unnecessarily throttled due to constant observation of a 5% worst-case choke indication, based on alternate observations of choke3 or choke6 conditions.
Applicable: Could occur with single-choke fairness protocols.
Solution: Base flow progress on total-bytes-transferred, not an interval-dependent bytes-per-second rate. The total-bytes-transferred is more easily supported by multi-choke fairness.