Exercise Sheet #14: Data Stream Infrastructure

February 15, 2013

1 Operator Scheduling

A. Explain and compare thread-based and state-based execution models in a Data Stream Management System.

B. Why is thread-based execution model more costly than state-based execution model of operators?

C. Explain and compare train and superbox scheduling.

2 Load Shedding

You are given a query network of operators. The selectivity and per tuple cost of each operator is given on the figure. Per tuple processing cost is given in terms of CPU cycles and rates of input streams are given in terms of tuples per second.
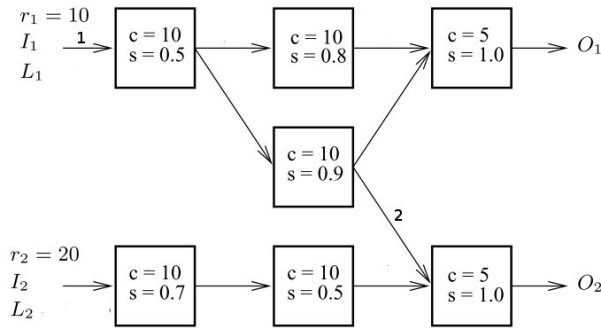

Figure 1: Query network with operator selectivities and costs in CPU cycles

A. What are the trade-offs between placing a drop operator at arc 1 and 2?

B. You are given an excerpt from an input stream of an aggregate query as follows:
   ... 15 5 10 20 30 14 10 15 20 30 10 $\rightarrow$ Average(window-size=3,slide=2) $\rightarrow$ results

   We would like to insert a window-aware drop box before the aggregate operator with a uniform drop probability of 0.5. Give a possible sequence of output results after placing the drop operator.