CprE 419 Lab 9:

Standard Operators in Streams Processing Language (SPL)

Shuo Wang

Figure 1 shows the streams dataflow graph for my code.
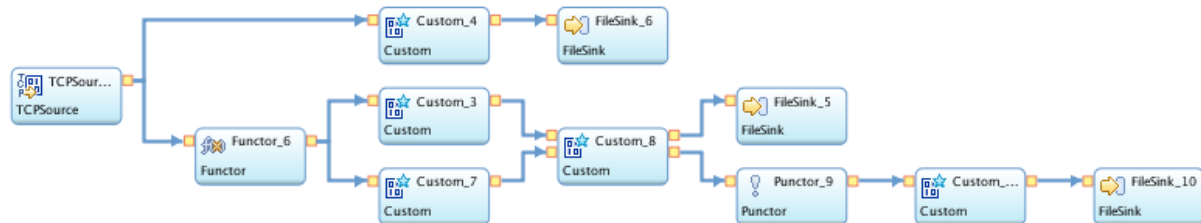

FIGURE 1 Streams Dataflow Graph

TCPSource generates raw data

Custom_4 and FileSink_6 write the raw data in "SourceData.txt". The schema is:
{date, time, CallerID, CalleeID, Duration, CallStartTime in seconds, CallStartTime in minutes, CallEndTime in seconds}
Where   CallEndTime in seconds = hour*3600+minute*60+second
          (e.g. 11:13:09 => 11*3600+13*60+9=40389)
          CallStartTime in seconds = CallEndTime in seconds − Duration
          CallStartTime in minutes = rounddown(CallStartTime in seconds / 60)
Recording "SourceData.txt" is for debugging and checking purpose.

Functor_6 calculated CallStartTime in seconds, CallStartTime in minutes and CallEndTime in seconds. (similar to what Custom_4 does)

Custom_3 checks the suspect type A and B. Custom_7 checks the suspect type C.

Custom_8 combines the stream from Custom_3 and Custom_7.

FileSink_5 archives the historical suspect alerts in "SuspectRecords.txt", for debugging and checking purpose.

Punctor_9 adds punctuation to the stream after a certain time. (the time point needs to be specified manually in the code)

Custom_... aggregates all the suspect alerts by each CallerID and makes output at punctuation.

FileSink_10 writes the final suspect list in "suspects.txt"

I run the code for 47 minutes from 21:43 to 22:30 April 7, 2015, and the final suspect list is:

```
Number of Suspects 4
239769310 C
239788154 BCA
239793121 B
239718753 BA
```

Appendix: files in submission .zip

1. lab9.spl is the IBM Infosphere code for lab 9
2. "SouceData.txt" is the archived stream data, as described above.
3. "SuspectRecords.txt" is the archived historical suspect alerts, as described above.
4. "suspects.txt" is the final list for the police.