Stream and Complex Event Processing
Final Evaluation

G. Cugola  E. Della Valle  A. Margara

Politecnico di Milano  Vrije Universiteit Amsterdam

cugola@elet.polimi.it  a.margara@vu.nl
dellavalle@elet.polimi.it
The general approach

• The course ends with a project developed by students, i.e., you :-) ...
• ...which you present in the final “lesson”...
• ...where you also comment the works of the others
• This “project” requires you to apply some of the technologies and methodologies learnt during the course to a practical case study
The rules

• The project must be developed in groups of 2 students
  • No more than 2, no less than 2 😊

• Each group must use two of the systems described during the course to implement a well precise case study (more on this later)

• Which systems to choose:
  1. Esper
  2. Another system at your choice
    • T-Rex or C-SPARQL
    • Or one of the other systems described but not shown in practice (i.e., SPARQL-Stream, CQELS, ETALIS, StreamBase)
      • The latter choice will be appreciated

• Course organizers must approve the systems you choose (to avoid every group choose the same systems)
The rules

• During the last lesson you are expected to:
  • Present your solution
  • Compare the two systems (when applied to the specific case study) in terms of:
    • Expressiveness, performance, ease of use, ...
  • In general, report about your experience/lessons learnt
  • Be prepared to demo your solution (you could be asked to)
  • Answer to questions coming from the course organizer or other students
The case study

• Analyze data in real time
• In particular, the data acquired during a soccer game

DEBS 2013 Grand Challenge
Case study: The Data

• Data originates from sensors located near the players' shoes (1 sensor per leg)
  • The goal keeper is equipped with two additional sensors, one at each hand
• The ball has its own sensor
• The sensors in the players' shoes and hands produce data at 200Hz
• The sensor in the ball produces data at 2000Hz
Case study: The Data

- The event schema is following:
  
  \( \text{sid, ts, x, y, z, } |\text{v}|, |\text{a}|, \text{vx, vy, vz, ax, ay, az} \)

- where \( \text{sid} \) is the sensor id
- \( \text{ts} \) is a timestamp in picoseconds
- \( \text{x, y, z} \) describe the position of the sensor in mm (more on this later)
- \( |\text{v}| \) (in \( \mu \text{m/s} \)), \( \text{vx, vy, vz} \) describe direction by a vector with size of 10,000. Hence, the speed of the object in x-direction in SI-units (m/s) is calculated by:
  
  \[ \text{v'}_x = |\text{v}| \times \text{vx} \times 10^{-4} \times 10^{-6} \]

- Similarly, \( |\text{a}| \) (in \( \mu \text{m/s}^2 \)), \( \text{ax, ay, az} \) describe the absolute acceleration and its constituents in three dimensions (the acceleration in m/s\(^2\) is calculated similar to that of the velocity)
  
  - The acceleration does not include gravity, i.e. \( |\text{a}| \) is zero when the ball is at a fixed position and not 9.81 m/s\(^2\)
Case study: The Data

• The playfield (7 players per team) and the coordinate system

  ![Coordinate System Image]

  • May approximate to: (0,33965), (0,-33960),(52483,33965), (52483,-33960)

• Timing:
  • 1st half: Start 10753295594424116, End 12557295594424116 (no ball 12398*)
  • 2nd half: Start 13086639146403495, End 14879639146403495
Case study: The Data

• In addition to the sensor data a separate data stream for referee events is available
  • It includes the time when a game was paused/resumed and the time and player_ids for substitutions
• The mapping between player ids and team ids as well as between sensor id and player id is provided in a metadata file
• Finally, a video of the entire game is provided
Case study: Ball possession rule

- Calculate the ball possession for each of the players
- A player obtains the ball whenever the ball is in his proximity (less than one meter away) and he hits (the ball acceleration peaks) it
  - A ball is hit if its (transmitter) distance from a player's foot (transmitter) is less than 1 meter and its acceleration reaches a value of minimal 55 m/s²
- The ball will stay in player’s possession until another player hits it, the ball leaves the field, or the game is stopped
  - The ball may leave the player proximity and will still stay in his possession
- The ball possession is calculated as time between the first ball contact (hit) and the last ball contact (hit)
Case study: Ball possession rule

- The per player ball possession stream should contain following information:
  
  ts, player_id, time, hits

where:

- ts is the latest time stamp of the event which lead to the update of the ball possession,
- player_id is the player identifier,
- time is the total time of the ball possession for a given player, and
- hits is the total number of ball contacts of the given player
Case study: Shot on goal

- Detect when a player shoots the ball in an attempt to score a goal
- A “shot on goal” is defined as any shot that would hit (or closely miss) the goal of the opposing team
  - This includes unsuccessful attempts that are, e.g., blocked by a player or saved by the goal keeper
  - Shot on goal is a state. Exit conditions are (a) the ball leaves the field, or (c) the direction changes such that the goal area would no longer be hit
- The goal areas are defined as rectangles with the following coordinates:
  - Goal area team 1: \( x \in [22578.5, 29898.5], y = 33941.0, z < 2440.0 \)
  - Goal area team 2: \( x \in [22560.0, 29880.0], y = -33968.0, z < 2440.0 \)
- Note that the hit distorts the speed values of the ball. Data is preprocessed by a Kalman-filter and stabilizes over time. The computation of the projection may take this into account
Case study: Shot on goal

• For the duration of the shot (i.e. as long as the state "shot on goal" is active) the result stream should be updated with motion values of the ball and the ID of the shooting player:
  
  ts, player_id, x, y, z, |v|, vx, vy, vz, |a|, ax, ay, az

• The result stream should be updated with the frequency of the sensor data until an exit condition occurs
Case study: Oracles

• We will provide data files containing ball possession statistics, and shots on goal statistics

• These statistics have been manually created and can serve as an aid in validating the respective rule results