Proposal Defense

Doctor of Philosophy in Computer Science

“Optimizing Operators for Temporal and Spatiotemporal Data” by Rakan Alseghayer

Date: May 11, 2022
Time: 9:30AM – 12:00PM
Place: https://pitt.co1.qualtrics.com/jfe/form/SV_ba47SHsjsbX0ZyC

Committee:
- Panos K. Chrysanthis, Professor, Department of Computer Science, School of Computing and Information
- Alexandros Labrinidis, Professor, Department of Computer Science, Schools of Computing and Information
- Kirk Pruhs, Professor, Department of Computer Science, School of Computing and Information
- Mohamed Sharaf, Associate Professor, Computer Science & Software Engineering, United Arab Emirates University

Abstract:
Sensors, mobile devices, and localization technologies have become affordable and highly available in the last decade, which led to the development of smart solutions and applications. A common characteristic of these solutions and applications is that they operate on temporal and/or spatiotemporal data that is often in the form of endless data streams. To harvest the concealed value in this data, online analytic methods have been developed. Two fundamental operators supporting analytical queries are correlation and join. Motivated by two widespread applications, in this dissertation we are proposing a framework to optimize data streams/timeseries correlations for monitoring temporal events, and an execution query strategy to optimize spatiotemporal (trajectories) joins for contact tracing. There is an increasing demand for real-time monitoring of large volumes of data streams that are produced at high velocity. Typically, pairs of most recent data streams need to be correlated within a specified delay target in order for their analysis to lead to actionable results. We propose to address this need by: (i) segmenting data streams into micro-batches; and (ii) leveraging incremental sliding window computation, priority scheduling, and caching techniques, to avoid unnecessary recomputations and I/O. Furthermore, we propose to devise and evaluate exploration strategies that effectively steer the processing of data stream correlations based on the monitoring objective. Trajectories are inherently personal information and generated on users’ mobile devices. We claim that contact tracing applications that require joining of trajectories belonging to different users need to process the join on the users’ mobile devices to protect data privacy and reduce costly data movement. We propose to optimize spatiotemporal joins on mobile devices by exploiting in-memory access structures that combine spatial indexing with interval trees to encode the locations and the durations at each location of the trajectories. We also propose to develop data movement protocols among mobile devices that preserve privacy, achieve scalability, and reduce energy consumption. Finally, we propose to formalize contact tracing as an ST-Aggregate join query that considers the duration of contacts cumulatively. We will evaluate the effectiveness of our approaches both theoretically and empirically using real-life data.