**Introduction**

We aim to develop methods to understand and identify insider threats to databases and sensitive data posed by malicious insiders within large organizations.

**Approach**

- Using query intent to develop profiles for members of the organization
- Computing query similarity for modeling intent
- New tools for hierarchical clustering and statistical inference for complex structures such as queries
- Construct user profiles and compute profile proximity as shown in Figure A
- Identify normal temporal behavior drift in user activity to reduce false positives
- Detecting deviations from user profiles as potential insider threat

**Query Log Exploration**

- Surveyed various query similarity detection techniques
- Tested against realistic workloads obtained from a financial organization
- Created a benchmark for query similarity assessment
- Developed multiple methods for extracting relevant high-order features from queries as shown in Figure B
- Create interactive summaries of query groups as shown in Figure D

**Database Activity Monitoring**

- We construct user profiles by accumulating the extracted features for each user for a given period of time
- People who work in the same role in the organization can have different work habits, styles, and priorities
- Hence, the expectation of drift in behavior changes for different roles, and for different people as shown in Figure E
- The behavior patterns of tasks can change the temporal drift of a profile as can be seen in Figures E and F

**Threat Modeling**

- Tested multiple threat models on the detection mechanism
- Modeled an insider threat ontology (See Figure C) for transforming anomaly detection into misuse detection

**Publications**

- G. Kul, D. Luong, T. Xie, P. Coonan, V. Chandola, O. Kennedy, S. Upadhyaya, Summarizing large query logs in Ettu, Arxiv 2016

**Products**

- EttuBench – A SQL Query Similarity Metric Benchmark, [https://github.com/UBOdin/EttuBench](https://github.com/UBOdin/EttuBench)