Identifying user behavior patterns from audit logs is valuable for system security of monitoring authorized users. Due to restricted access to production event-logs, security and privacy issues, and high costs of real datasets, synthetic event-log datasets are crucial in designing and evaluating data analytics approaches. A controlled event-log simulation environment provides the data analysts various synthetic dataset containing embedded interesting patterns and features. The produced testing datasets reduce the algorithm evaluation time.

We developed an interactive data exploration environment to such a design-generate-visualize-analyze-optimize process. Design: statistical characteristics (distribution), association pattern, sequence pattern. Generate: produce a dataset that contains predefined attributes and patterns. Visualize: extract simplified workable information from generated dataset. Analyze: verify the differences between generated dataset and expected dataset. Optimize: refine input parameters.

We proposed a synthetic event-log generator that effectively assists data analysts in designing scenario-driven event-logs with embedded user behavior patterns, and visually analyzing the quality of the generated datasets. The toolkit includes three layers:

- **Behavior pattern representation layer**: allows data analysts to design interesting features and patterns that will be injected into the dataset.
- **Dataset generation layer**: creates datasets that are controlled by data size, data distribution, and the designed behavior patterns.
- **Dataset visualization layer and analysis layer**: provides an interactive exploration environment for visual analysis of the quality of generated datasets.

We developed a toolkit that can produce the following visual graphs for analyzing the dataset:

- **Sequence overview**: sequence is an ordered list of events performed by one person per day.
- **Frequent sequential patterns**: are subsequences that appear frequently among all user sequences.
- **Clustering based on sequence similarity**: divides the frequent sequential patterns into a number of clusters.
- **Clustering representatives**: explores the representative patterns of each cluster.

Design an event dataset to simulate user-system interactions in distributed medical imaging systems. Each event has 6 attributes, where Event=<User, Location, Action, Patient, Date, Time>. Table II defined attribute distributions. Table III defined 9 typical user behavior patterns that constitute ordering, timing, and sequencing. Produced 30,000 events with randomly selected attribute values but following predefined distribution; predefined behavior patterns are inserted into the events.