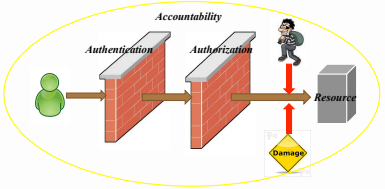
UNIVERSITY OF ONTARIO INSTITUTE OF TECHNOLOGY

Weina Ma
Kamran Sartipi
{Weina.Ma, Kamran.Sartipi}@uoit.ca

Behavior Pattern Based Security Enhancement in Large Distributed Systems

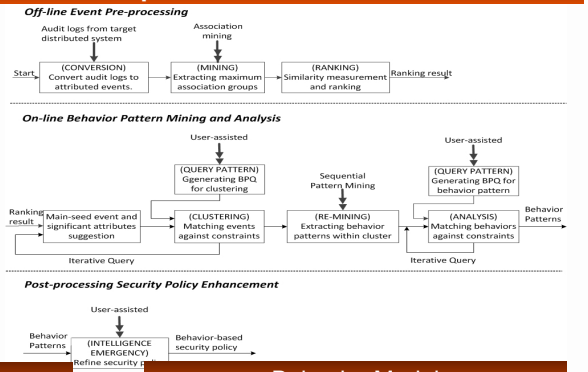
Motivation

- Unusual behaviors of authorized users cannot be detected by normal security mechanisms.
- Behavioral activities of authorized users must be monitored and controlled to protect users private data and identify malicious behaviors.



Architecture & Applied Techniques

- ***Event Pre-processing.*** Audit logs from targeted system are parsed and converted to attributed events.
- ***Event Similarity Measurement.*** An association-based similarity metric is proposed to measure the relationships between events.
- ***Constraint-based Clustering.*** Using an interactive clustering process, the user incrementally selects initial seed-event from suggested event list; and generates a BPQ (Behaviour Pattern Query)
- ***Behaviour Pattern Extraction.*** Sequential pattern mining is applied on constrained clusters to extract frequent behaviours patterns.
- ***Behaviour Pattern Analyzing.*** Provides salient features and characteristics of discovered behaviour patterns to identify common behaviours.
- ***Security Policy Enhancement.*** Provides recommendations for system security provisioning enhancement in terms of evaluating the gap between existing security policies and recovered behaviour patterns.

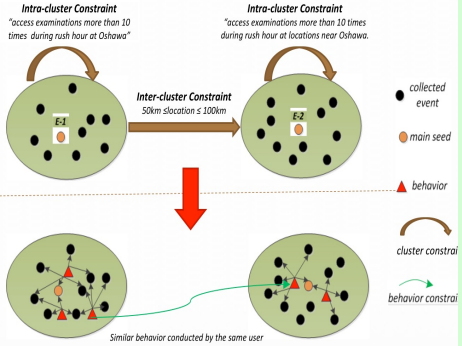


Proposed Approach

This research introduces a new generation of intelligent decision support systems that effectively assist system administrators to get deep insight into the system-user's dynamic behavior patterns, and guides them in their tasks, such as: common behavior and abnormal behavior detection, as well as security policy enhancement. We propose a generic behaviour model that allows the user to map domain specific audit logs onto attributed events and consequently a behaviour model.

An association-based similarity metric is proposed to measure the significance of events based on their attributes. A behaviour pattern query language is designed to allow the administrators to compose queries for identifying complex behaviour patterns. The approach also employs data mining technologies (association mining, sequential pattern mining) and constrained clustering to extract behaviour patterns. We have applied the proposed approach on distributed diagnostic imaging systems, called PACS (Picture Archiving and Communication Systems).

Behaviour Pattern Query

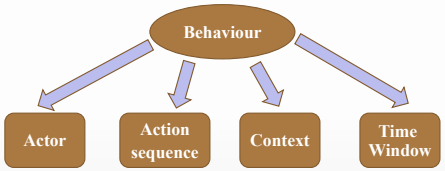



```
BEGIN-BPQ
  CLUSTER: C1
    MAIN-SEED: EVENT E-1
    INTRA-CONSTRAINT:
      Location = 'Oshwa'
      Data_type = 'Examination'
      Date > 2014-01-01; Date < 2014-02-01
      Time > 9:00; Time < 11:00
      GROUP BY User, Date
        HAVING COUNT > 10
  CLUSTER: C2
    MAIN-SEED: EVENT E-2
    INTRAC-CONSTRAINT:
      Data_type = 'Examination'
      Date > 2014-01-01; Date < 2014-02-01
      Time > 9:00; Time < 11:00
      GROUP BY User, Date
        HAVING COUNT > 10
  INTER-CONSTRAINT: CLUSTER C1, CLUSTER C2
    Location > 50km; Location < 100km
END-BPQ
```

Behavior Model

Behavior Pattern is defined as a tuple:

- Actor who issues a behavior.
- Sequence of actions that the actor conducts.
- Context in which a behavior takes place.
- Time Window that the behavior performs.



Experiment

Visualization of Event Association

Totally we collected 695 user access events from a distributed imaging system at Mohawk college within one month.

Event Similarity Metric is defined as:

assoc(ei, ej) = maxgx (\|itemset(gx)\| + w * \|basketset(gx)\|)

- Each node represents an event.
- Each weighted edge represents the association value between two events.

Sequential Pattern Mining Result

Four groups are selected to evaluate the effectiveness of the association-based similarity ranking and constrained clustering to behaviour pattern discovery:

- ***Cluster #1 (whole dataset):*** All events without association mining.
- ***Cluster #2 (raw clustering):*** Select an initial seed event, and collect events that have non-zero similarity with the seed event.
- ***Cluster #3 (constrained clustering):*** Select an initial seed event, and add one intra-cluster constraint "user = U-22 && location = L-6" to this cluster.
- ***Cluster #4 (constrained clustering):*** Select an initial seed event, and add one intra-cluster constraint "time = T-10" to this cluster.