SFU SIMON FRASER UNIVERSITY

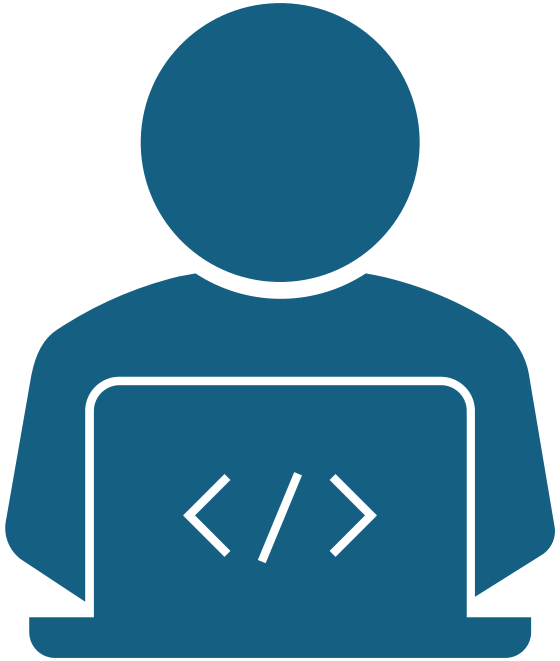
GenAttackTracker: Real-Time SCADA-based Cyber Threat Detection Through Scoring and Bayesian Model Integration

Fatemeh Movafagh and Uwe Glässer

Computing Science, Simon Fraser University

British Columbia, Canada

Presenter: Fatemeh Movafagh (fma44@sfu.ca)



SFU SIMON FRASER UNIVERSITY

Presenter Bio

Fatemeh Movafagh is a *PhD student* and *Research Assistant* at the *Software Technology Lab*, School of Computing Science, Simon Fraser University, British Columbia, Canada. She works under the supervision of *Prof. Uwe Glässer*. Her research focuses on *cyber intelligence*, *threat analysis*, and *critical infrastructure security*, with expertise in *anomaly detection*, *time series analysis*, and *machine learning* for *securing operational technologies* and *supervisory control systems*.

Introduction

Operational Technology (OT) & SCADA Vulnerabilitie

Evolving Cyber Threats in Critical Infrastructure (CI)

Introduction: Research Aim

- **Research Question:**
  - How can secondary threat intelligence sources enhance real-time detection of security breaches in SCADA systems?

- **Methodology:**
  - Utilizing Bayesian inference and dynamic anomaly scoring to continuously update and improve situational awareness.

- **Contribution:**
  - GenAttackTracker framework

Online Anomaly Detection

- **Supervisory Control Data**
  - **Time-series data**
  - **Anomalies =** deviation from expected normal behavior

- **Challenges in Anomaly Detection**
  - Diverse Causes of Anomalies
  - Identifying True Threats
  - Real-time Detection

Suspicious Activity Markers

Contextual data points that provide additional insights into potential cyber threats.

Examples:

Unusual data transfer activity.

Login attempts from suspicious locations.

Communication through non-standard ports.

Abnormal spikes in traffic (e.g., SMTP, DNS).

...

Bayesian Analysis

Continuously updates the probability of an attack as new data becomes available.

Why Bayesian?

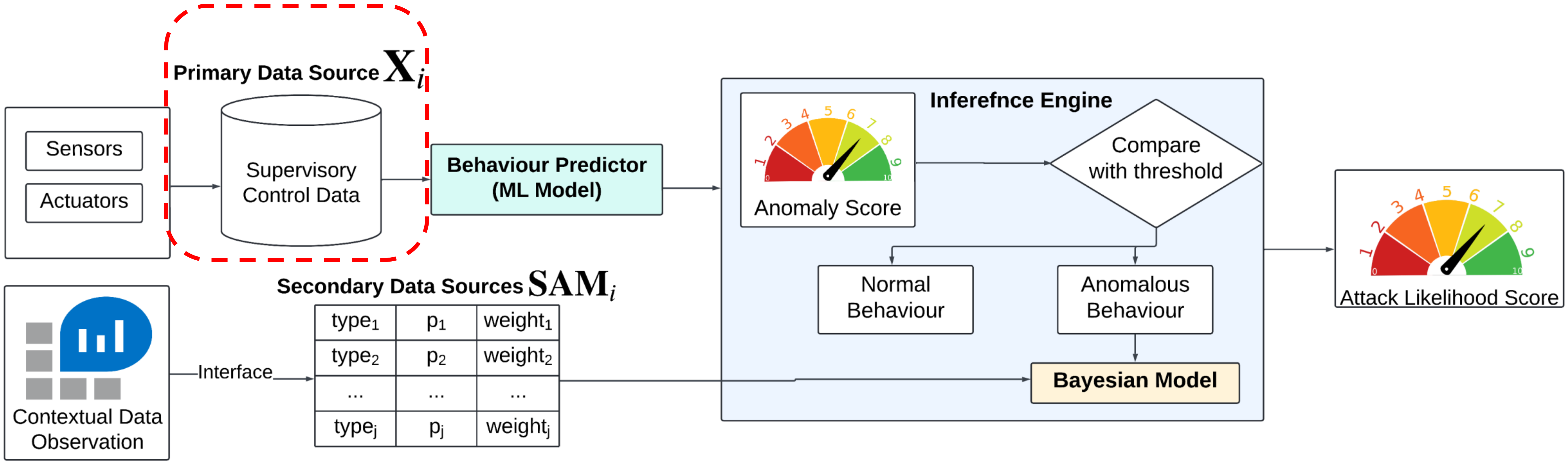
- *Handles uncertainty in threat detection.*
- *Incorporates both control data and Suspicious Activity Markers (SAMs) for more informed decisions.*

AttackTracker Framework

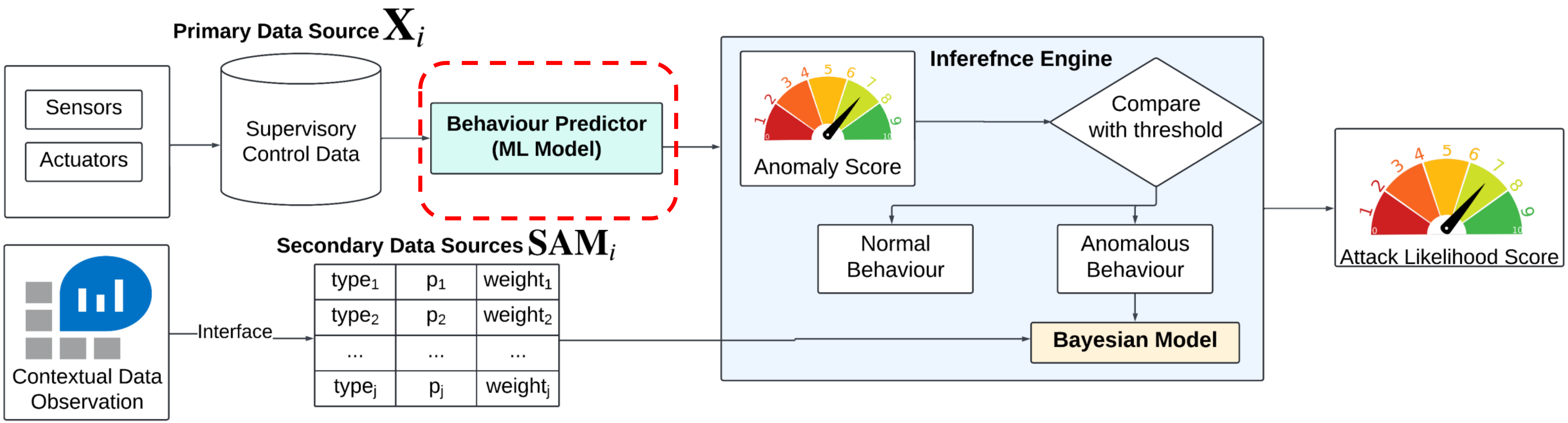
- Hierarchical distributed network of detectors.
  - Local detectors: Behavior Predictor + Inference Engine
  - Higher level detectors: Inference Engine

- Key components:
  - Behavior Predictor: MTCN
  - Inference Engine: Dynamic Scoring , Modified z-score

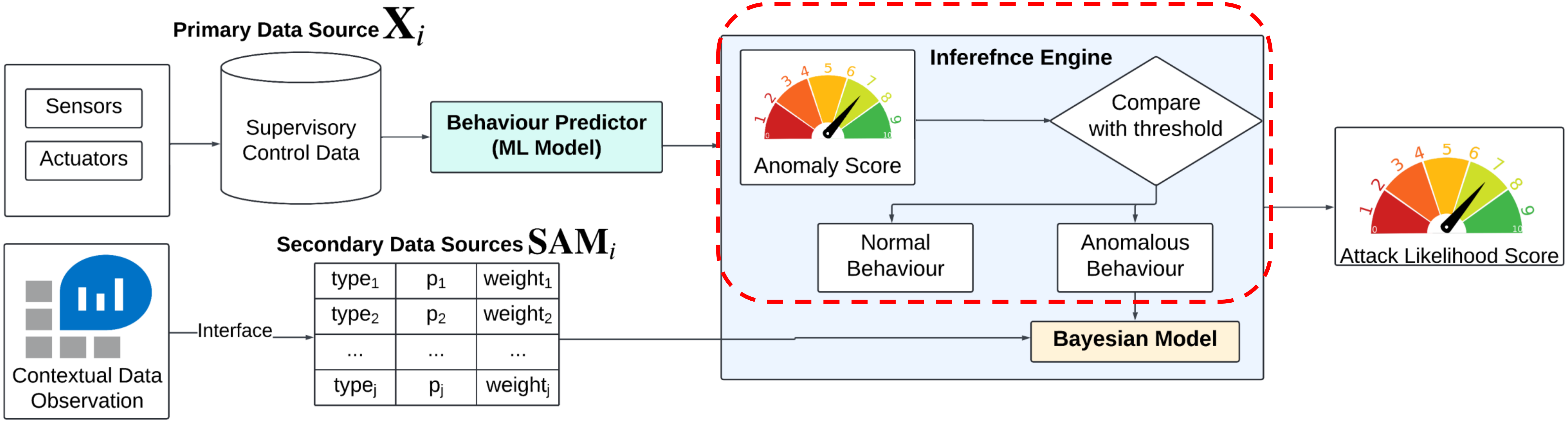
GenAttackTracker Framework



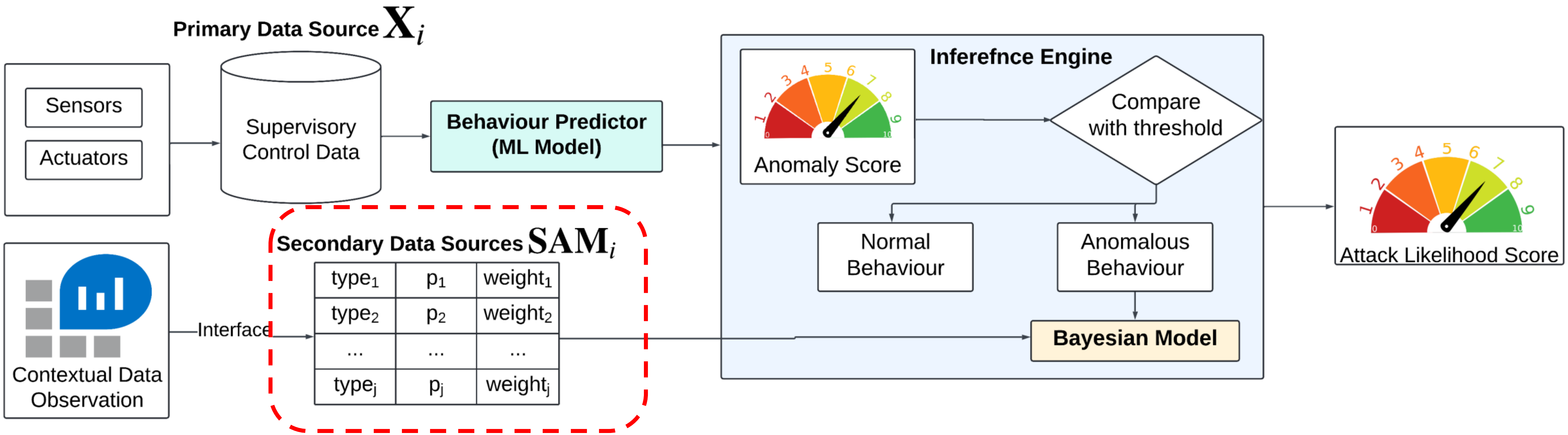
GenAttackTracker Framework



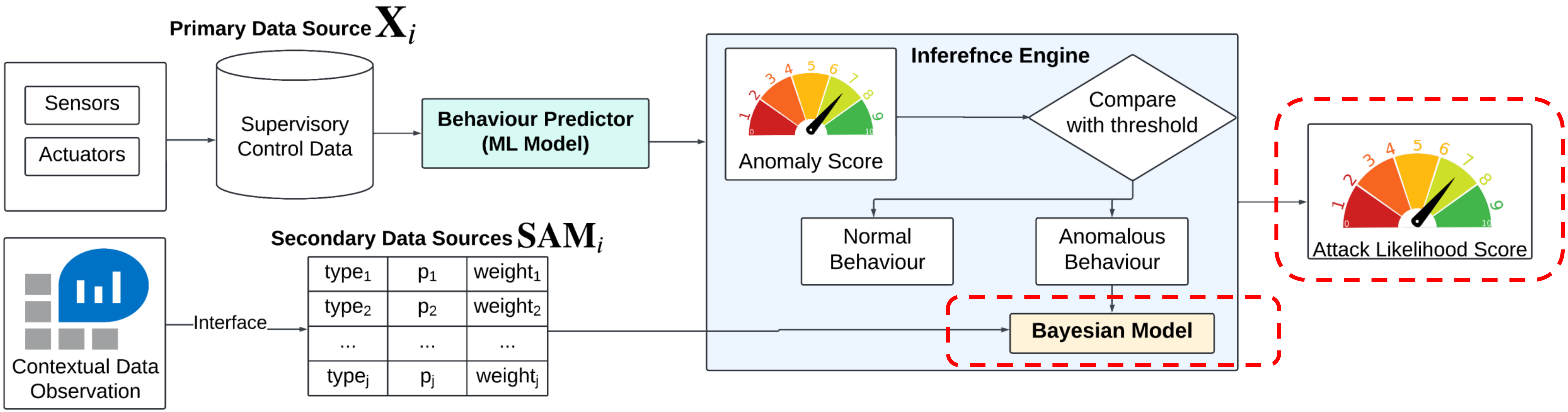
GenAttackTracker Framework



GenAttackTracker Framework



GenAttackTracker Framework



Inference Engine – Bayesian Model

- Hierarchical Model
  - Local Detectors
  - Intermediate Level
  - Global Level

- Key formula:

Prior

$$P(\text{Attack}_i|X_i, \text{SAM}_i) = \frac{P(X_i|\text{Attack}_i) \cdot \left(\prod_{j=1}^{N} (p_{i,j} \times \text{weight}_{i,j})\right) \cdot P(\text{Attack}_i)}{P(X_i) \cdot P(\text{SAM}_i)}$$

Inference Engine – Bayesian Model

- Hierarchical Model
  - Local Detectors
  - Intermediate Level
  - Global Level

- Key formula:

Likelihood

$$P(\text{Attack}_i | X_i, \text{SAM}_i) = \frac{P(X_i|\text{Attack}_i) \cdot \left(\prod_{j=1}^{N} (p_{i,j} \times \text{weight}_{i,j})\right) \cdot P(\text{Attack}_i)}{P(X_i) \cdot P(\text{SAM}_i)}$$

Inference Engine – Bayesian Model

- Hierarchical Model
  - Local Detectors
  - Intermediate Level
  - Global Level

  - Key formula:

$$P(\text{Attack}_i | X_i, \text{SAM}_i) = \frac{P(X_i | \text{Attack}_i) \cdot \left(\prod_{j=1}^{N} (p_{i,j} \times \text{weight}_{i,j}) \right) \cdot P(\text{Attack}_i)}{P(X_i) \cdot P(\text{SAM}_i)}$$

Posterior

Experiments

- Baseline: AttackTracker framwork
- Dataset: SWaT (Secure Water Treatment Testbed)
  - 11 days of operation, including 7 days of normal behavior and 4 days of cyberattacks.
  - 51 variables: Sensors (e.g., flow, pressure) and actuator states (e.g., valve positions, pump statuses).

- Implementation:
  - Toolset: TensorFlow, PyMC3, Scikit
  - Monte Carlo Simulation

Experiments

- Baseline: AttackTracker fram
- Dataset: SWaT (Secure Wate
  - 11 days of operation, includi s of cyberattacks.
  - 51 variables: Sensors (e.g., fl , valve positions, pump statuses).

- Implementation:
  - Toolset: TensorFlow, PyMC3,
  - Monte Carlo Simulation

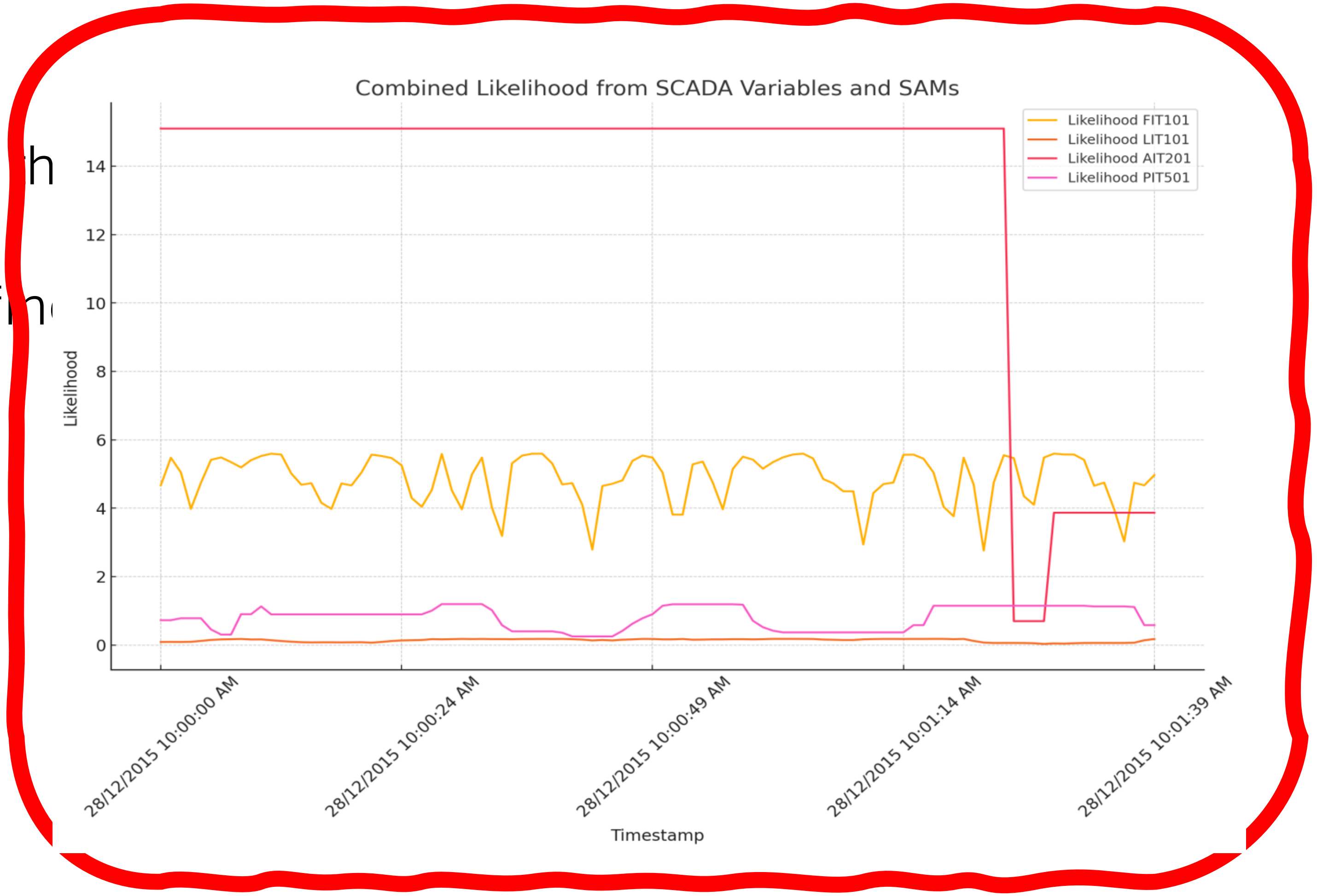
Input: SCADA data X, Suspicious Activity Markers (SAMs) S, anomaly score A
Output: Posterior probability of attack
procedure COMPUTELIKELIHOOD(X, A)
    Compute likelihood L based on SCADA data and anomaly score
    return L
end procedure
procedure CHOOSEPRIORS
    Set prior P_{attack} based on historical SCADA data
    Set prior P_{SAM} from external tools for SAMs
    return P_{attack}, P_{SAM}
end procedure
procedure UPDATEPOSTERIOR(L, P_{attack}, P_{SAM})
    Update posterior $P_{posterior} \leftarrow \frac{L \times P_{attack} \times P_{SAM}}{marginal\_likelihood}$
    return $P_{posterior}$
end procedure
procedure BAYESIANINFERENCE(X, S, A)
    $L \leftarrow$ COMPUTELIKELIHOOD(X, A)
    $P_{attack}, P_{SAM} \leftarrow$ CHOOSEPRIORS
    $P_{posterior} \leftarrow$ UPDATEPOSTERIOR(L, P_{attack}, P_{SAM})
    return $P_{posterior}$
end procedure

Experiments

- Insightful results
  - Provided more reliable threat assessments by continuously updating the posterior probabilities.
  - Incorporating SAMs refined

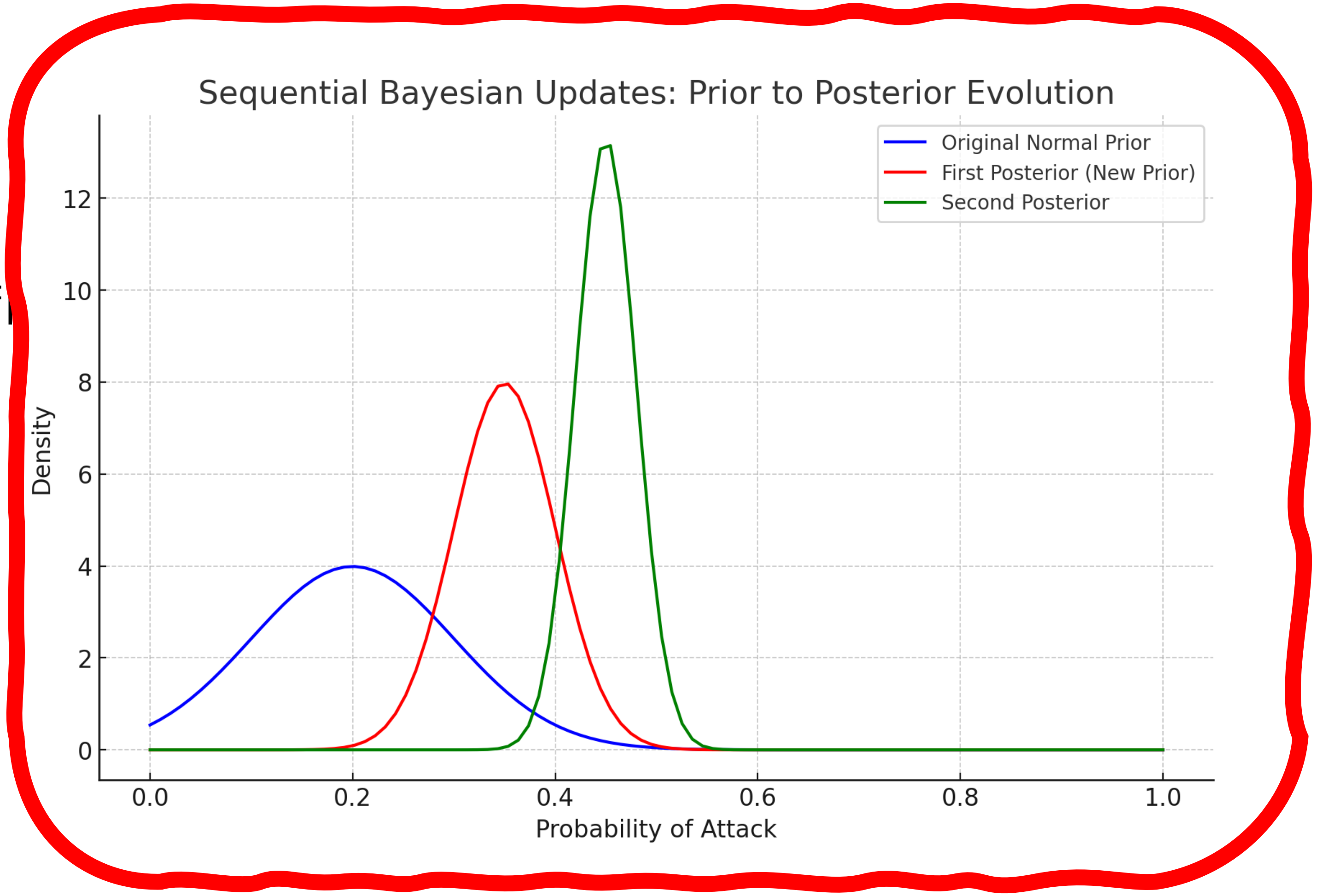
Experiments

- Insightful results
  - Provided more reliable th
    posterior probabilities.
  - Incorporating SAMs refin



Experiments

- Insightful results
  - Provided more reliable posterior probabilities.
  - Incorporating SAMs refi



Conclusion

- GenAttackTracker Contributions:
  - Combined dynamic anomaly scoring with Bayesian inference for enhanced situational awareness.
- Key Achievements:
  - Improved Threat Detection: Increased accuracy in identifying cyber threats with fewer false positives.
  - SAM Integration: Suspicious Activity Markers provided additional context, improving the reliability of threat assessments.
  - Monte Carlo Simulation: Reduced uncertainty in attack likelihood estimation through probabilistic simulations.
- Future Work:
  - Expand the model to analyze interconnected infrastructures.

SIMON FRASER UNIVERSITY

Thank you!

Questions