Graphical Event Model Learning and Verification for Security Assessments

PhD candidate LS2N-GFI Informatique

Dimitri Antakly - Benoît Delahaye - Philippe Leray

IEA/AIE
9th of July 2019

Table of contents

1. Introduction
2. Graphical Event Models (GEM)
   - Introduction of GEMs
   - TGEMs and RTGEMs
3. Motivation and Proposed Strategy
   - Motivation
   - SHD
   - Toy Example
4. Perspectives

- The detection of malicious behaviors in dynamic systems is very complex due to:
  - the big volume of data generated
  - the incompleteness of the data
  - the fast evolution of multi-agent systems
- From a security point of view it is as important to learn a model that best represents reality AND satisfies certain security properties
- In the literature:
  - Machine Learning formalisms
  - Formal verification formalisms

1 Introduction

2 Graphical Event Models (GEM)
- Introduction of GEMs
- TGEMs and RTGEMs

3 Motivation and Proposed Strategy
- Motivation
- SHD
- Toy Example

4 Perspectives

- A *Graphical Event Model* $\mathcal{G} = (\mathcal{L}, E)$ allows to graphically represent an event stream

- Data type:
  - x_{t^*}: An event stream $(t_1, l_1), ..., (t_n, l_n)$, with $0 < t_i < t_{i+1} < t^*$
  - The i_{th} history: $h_i = (t_1, l_1), ..., (t_{i-1}, l_{i-1})$

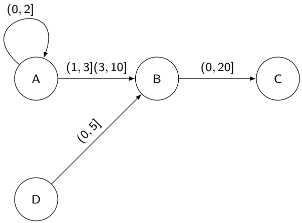
- The data likelihood knowing the model and its parameters:

$$p(x_{t^*} \mid t^*) = \prod_{i=1}^{|x_{t^*}|} \lambda_{l_i}(t_i \mid h_i) \prod_{i=1}^{|x_{t^*}|+1} e^{-\sum_{l \in \mathcal{L}} \int_{t_i-1}^{t_i} \lambda_l(\tau|h_i) d\tau}$$

- Conditional intensity functions in the Markov case:

$$\lambda_l(t \mid h) = \lambda_l(t \mid [h]_{Pa(l)})$$

- A *Timescale* GEM $M = (\mathcal{G}, \mathcal{T})$, a sub-family of GEMs representing more strict temporal dependencies
- Conditional intensity function: $\lambda_l(t \mid h) = \lambda_{l,c_l(h,t)}$
  Example:



For node B: $\lambda_{B,000}$, $\lambda_{B,001}$, $\lambda_{B,010}$, $\lambda_{B,011}$, $\lambda_{B,100}$ $\lambda_{B,101}$, $\lambda_{B,110}$ and $\lambda_{B,111}$

- A *Recursive* TGEM is a structurally and parametrically consistent model that can approximate any TGEM
- The allowed Forward operators are: "add", "split" and "extend"
- It could be learned by a "Greedy Search" (in two steps) based on a BIC score

1 Introduction

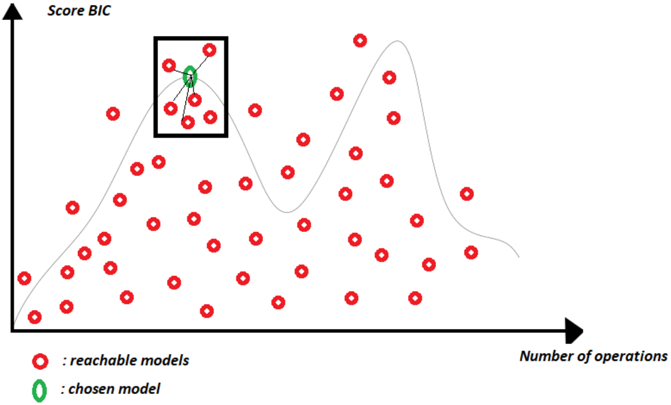
2 Graphical Event Models (GEM)
- Introduction of GEMs
- TGEMs and RTGEMs

3 Motivation and Proposed Strategy
- Motivation
- SHD
- Toy Example

4 Perspectives

- Learn a model that best represents reality and at the same time satisfies *security properties*
- The problem could be defined as:

$$\exists M^*, M^* = argmaxP(D \mid M) \; AND \; P(\phi \mid M) > c$$



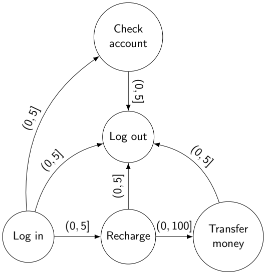
- The Hamming distance is defined as:

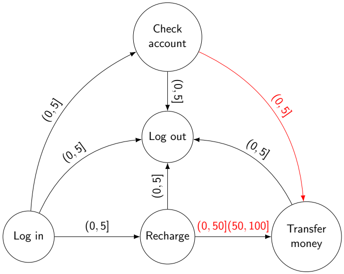
$$\mathrm{SHD}(G_1, G_2) = \sum_{e \in E_{sd}} 1 + \sum_{e \in E_{inter}} d(\mathcal{T}(e, G_1), \mathcal{T}(e, G_2))$$

- The elementary distance is defined as:

$$d(\mathcal{T}(e, G_1), \mathcal{T}(e, G_2)) = \frac{v_{nid}}{v_{nid} + v_{id}}$$

- $\phi = \square^{1000}(\text{Transfer Money} \Rightarrow \text{Recharge}_{(0,20]} \vee \text{Check account}_{(0,5]})$; $P(\phi \mid M^o) > 0.8$





- $\mathrm{SHD}(M^o, M^*) = 1.333$

1 Introduction

2 Graphical Event Models (GEM)
- Introduction of GEMs
- TGEMs and RTGEMs

3 Motivation and Proposed Strategy
- Motivation
- SHD
- Toy Example

4 Perspectives

- The ongoing experiment:
  - The learning phase is finalized, we are looking to improve the global score via a better selection of a starting horizon
  - The neighborhood exploration phase is ongoing
  - The formal verification phase is also under implementation, we are looking to use Statistical Model Checking (SMC) and see if it can be improved
- Apply it on a real world use case with GFI informatique

Thank you !