Detecting inference attacks involving raw sensor data*

Paul Lachat – paul.lachat@insa-lyon.fr
Affiliation: LIRIS, INSA Lyon, France & DIMIS, University of Passau, Germany

Inference involving raw sensor data: A motivating example

Involving personal data

e.g., Age, Sex ⟶ Pr(CVD) < 70% ✔
Cardiovascular disease

Involving raw data

e.g., Raw data ⟶ Activities ✔

Involving both personal & raw data

Age, Sex + Activities ⟶ Pr(CVD) ≥ 70% ✖
Raw data

Yasuhiko Kubota et al. "Physical Activity and Lifetime Risk of Cardiovascular Disease and Cancer". In: *Medicine and science in sports and exercise* 49.8 (Aug. 2017), pp. 1599–1605. pmid: 28350711

Oresti Banos et al. "Design, Implementation and Validation of a Novel Open Framework for Agile Development of Mobile Health Applications". In: *BioMedical Engineering OnLine* 14.2 (Aug. 13, 2015), S6

The user knowledge model: Metadata knowledge unit obtained by issuing queries

User knowledge

Modeled as $\langle a, c \rangle$, where

- a is an attribute identifier, e.g., ■, ■, ■
- c is a selection condition, e.g., (t_b, t_e) a temporal interval

Example queries

Q_1 SELECT ■, ■, ■ WHERE INTERVAL (t_2, t_3)

Q_2 SELECT ■, ■ WHERE INTERVAL (t_1, t_2)

Q_3 SELECT ■, ■ WHERE INTERVAL (t_0, t_1)

Query history log (QHL)

Tracks each user knowledge metadata.

After issuing Q_{1-3}, the QHL is:

Inference detection system (InfDS): For inference channels involving raw sensor data

Overall workflow

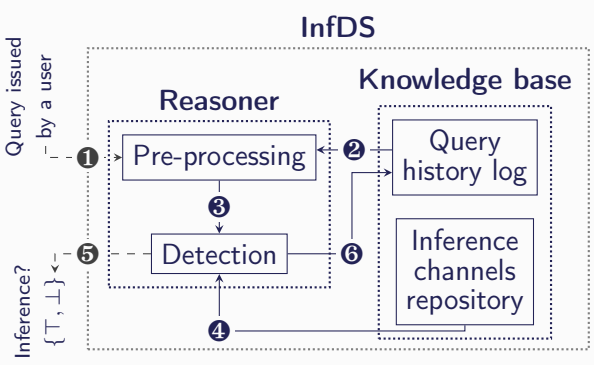
When a new query ❶ is issued by a user u, the system retrieves the previous knowledge of u ❷, i.e., $QHL(u)$, and merges all the metadata into one set ❸. The InfDS then checks if ❸ can exploit one of the registered inference channels ❹ and notifies the detection result ❺. In case no inference is detected, $QHL(u)$ is updated with the new metadata ❻

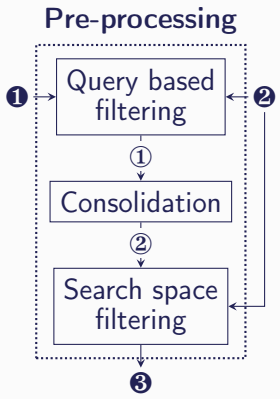
Query based filtering (Qbf)

Based on the query knowledge ❶, extracts from the QHL ❷ only the knowledge that can be consolidated ①

Search space filtering (Ssf)

Extracts from the QHL ❷ the knowledge related to the consolidation results ② that the detection must process ❸

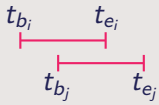




Consolidation

Aims to reduce the QHL size.

Rule example for temporal intervals:



$\Rightarrow \langle ■, (t_{b_i}, t_{e_j}) \rangle$

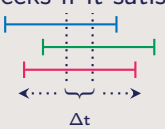
Hence ■ $\in Q_{1\&2} \Rightarrow \langle ■, (t_1, t_3) \rangle$

Detection

Searches knowledge using patterns:

$\bigcup \langle \{■|■|■\}, (t_b, t_e) \rangle$

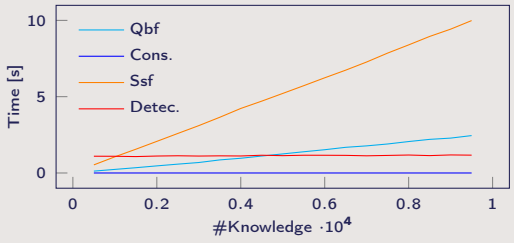
Checks if it satisfy the constraints:



$\exists\ \Delta t \geq 2s$?

Performance evaluation

- Linear complexity based on the size of the user knowledge
- Demonstrates the feasibility on the MHEALTH use case
- The InfDS needs further optimization to be scalable



Additional information

Work supported by the DFH-UFA
*Submitted to DEXA 2021