

Query Log Analysis

Detecting anomalies in DNS traffic

Presentation

- *Goal*
- *Challenges*
- *Implementation*

- *DEMO*

Goal

Design and build a working query log analysis platform using available components and custom development, able to predict, detect and report on common attack and abuse patterns in an open architecture, allowing for future growth and improvement.

Challenges

- Huge data volume
 - efficiency and scalability !
 - easy to stay under the hood
- Wide range of attacks, under constant evolution
- Specific nature of DNS traffic
 - periodicity and trends
 - few (typically two) packets per flow

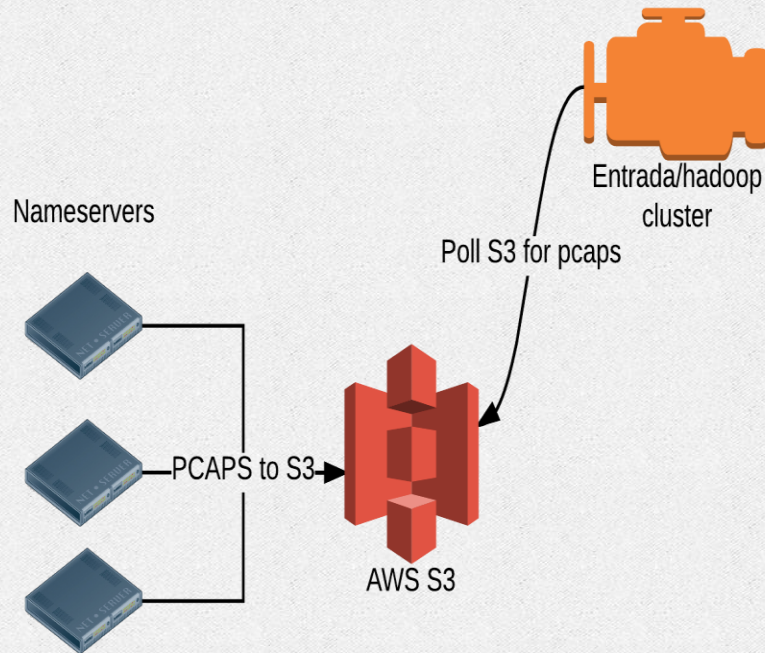
Implementation

- Masters thesis (University of Leuven)
- Main components
 - Entrada
 - QLAD-flow
 - QLAD-global
 - QLAD-UI

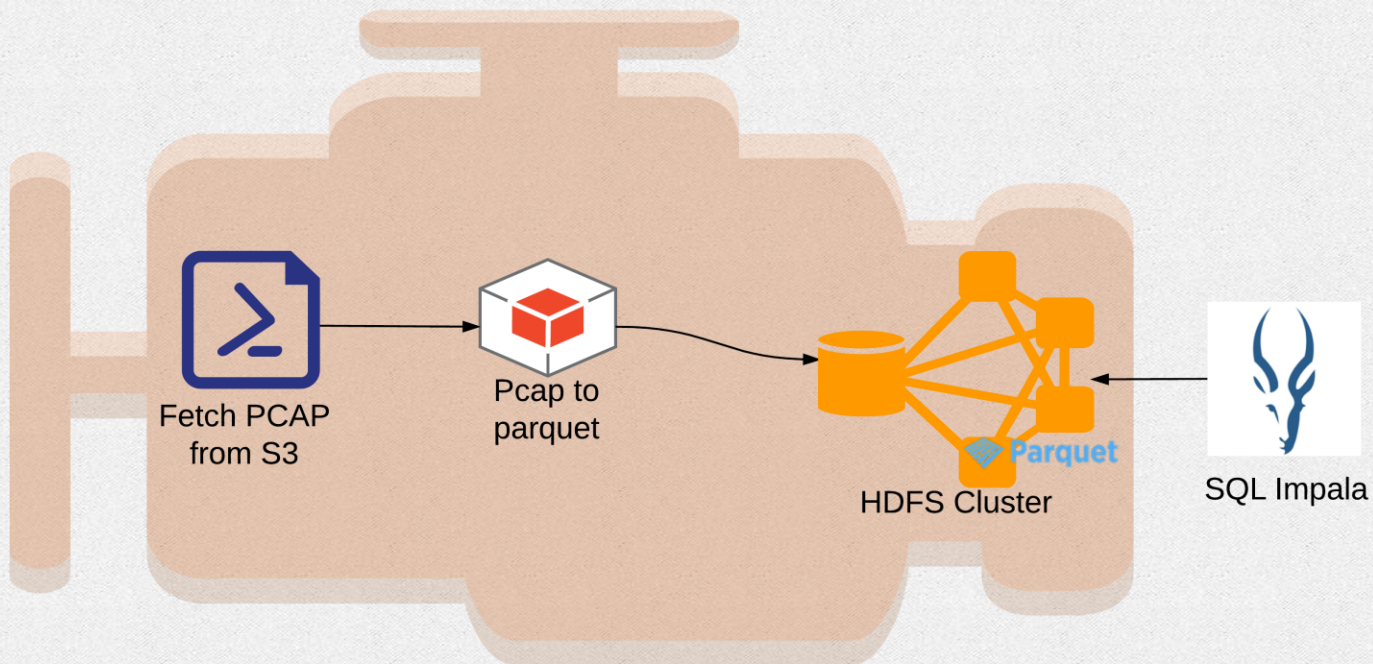
Implementation Entrada: why

- Efficiently store the data
- Efficiently query the data
- Designed to work with DNS

Implementation Entrada: architecture



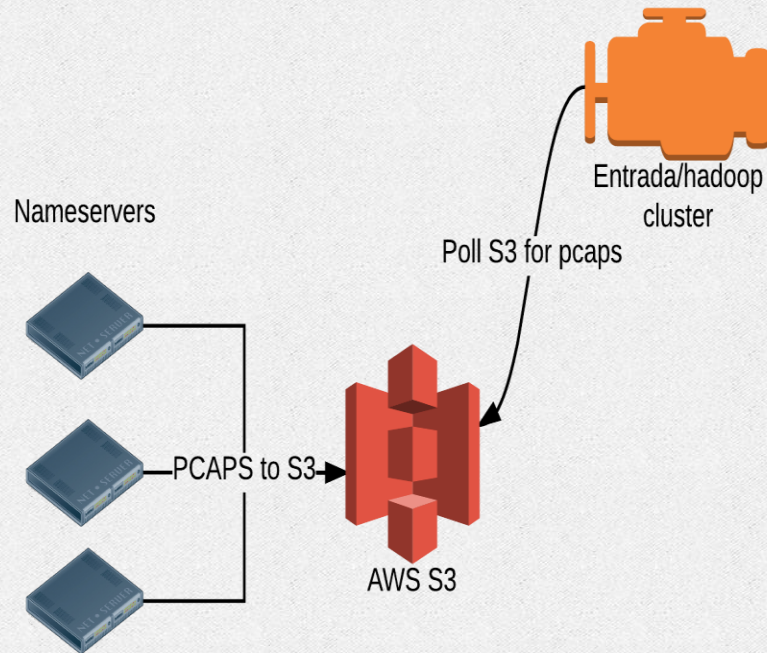
Implementation Entrada: flow



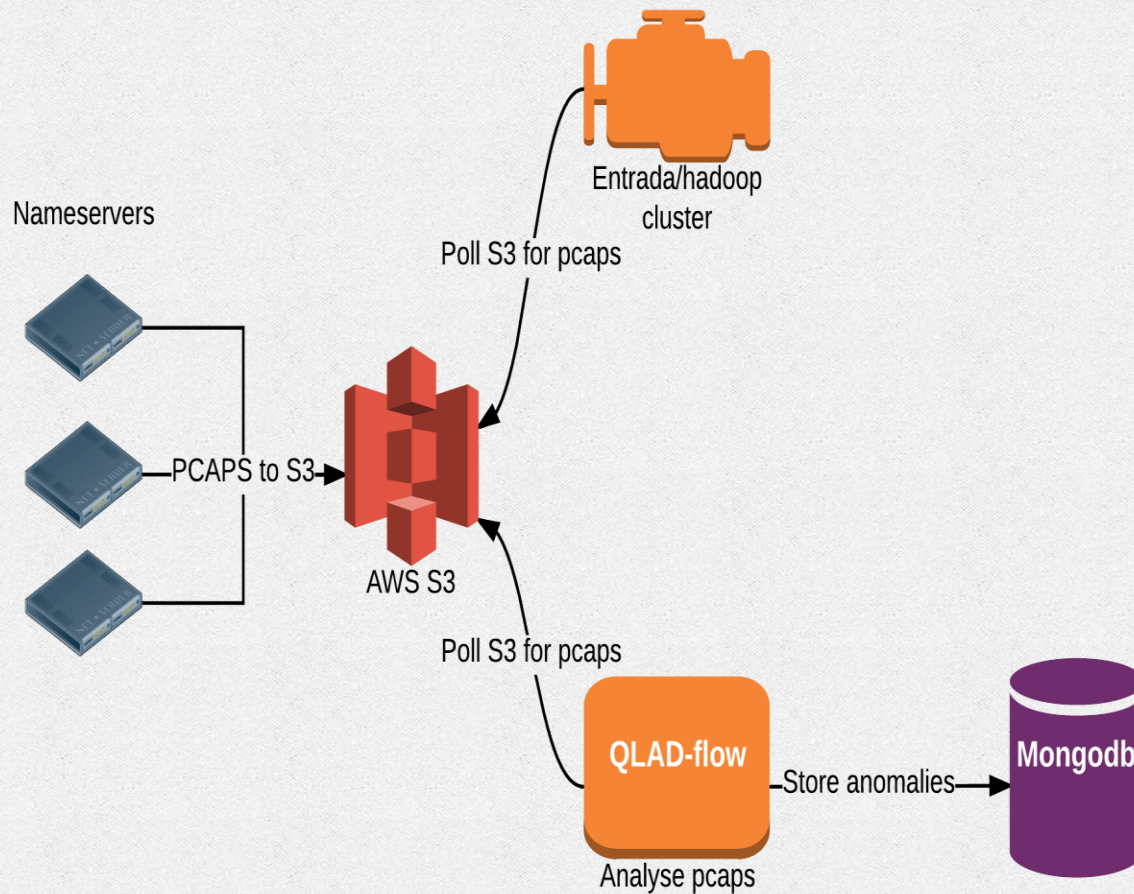
Implementation QLAD-flow: why

- Focus on anomalies (\neq attacks/abuses)
- Inspiration from network anomaly detection
- Able to detect low traffic anomalies
- Uses the algorithm developed by CZ.NIC

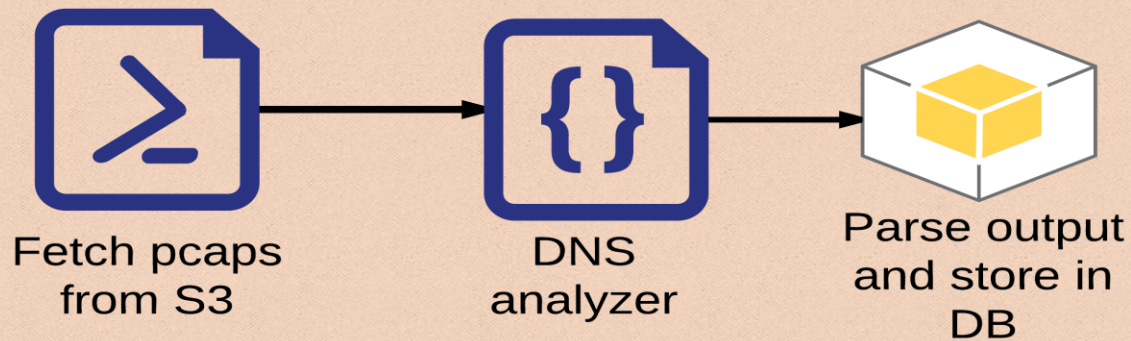
Implementation QLAD-flow: architecture



Implementation QLAD-flow: architecture



Implementation QLAD-flow: flow



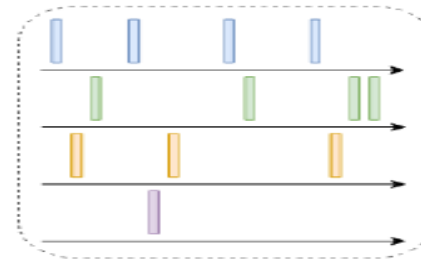
Implementation QLAD-flow: algorithm

① Split traffic into sketches

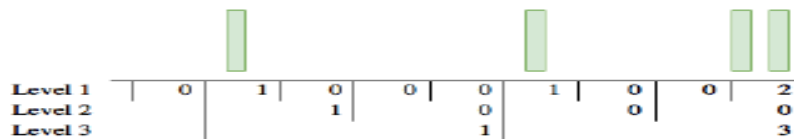


h_1

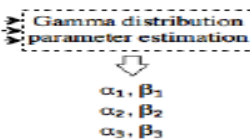
h_2



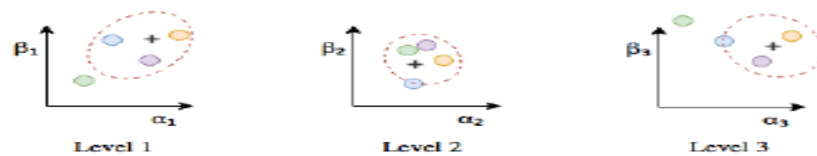
② Create time-aggregated packet count series for each sketch



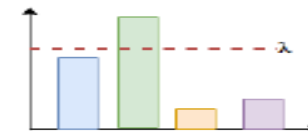
③ Gamma distribution modelling



④ Reference value computation



⑤ Computing the statistical distances



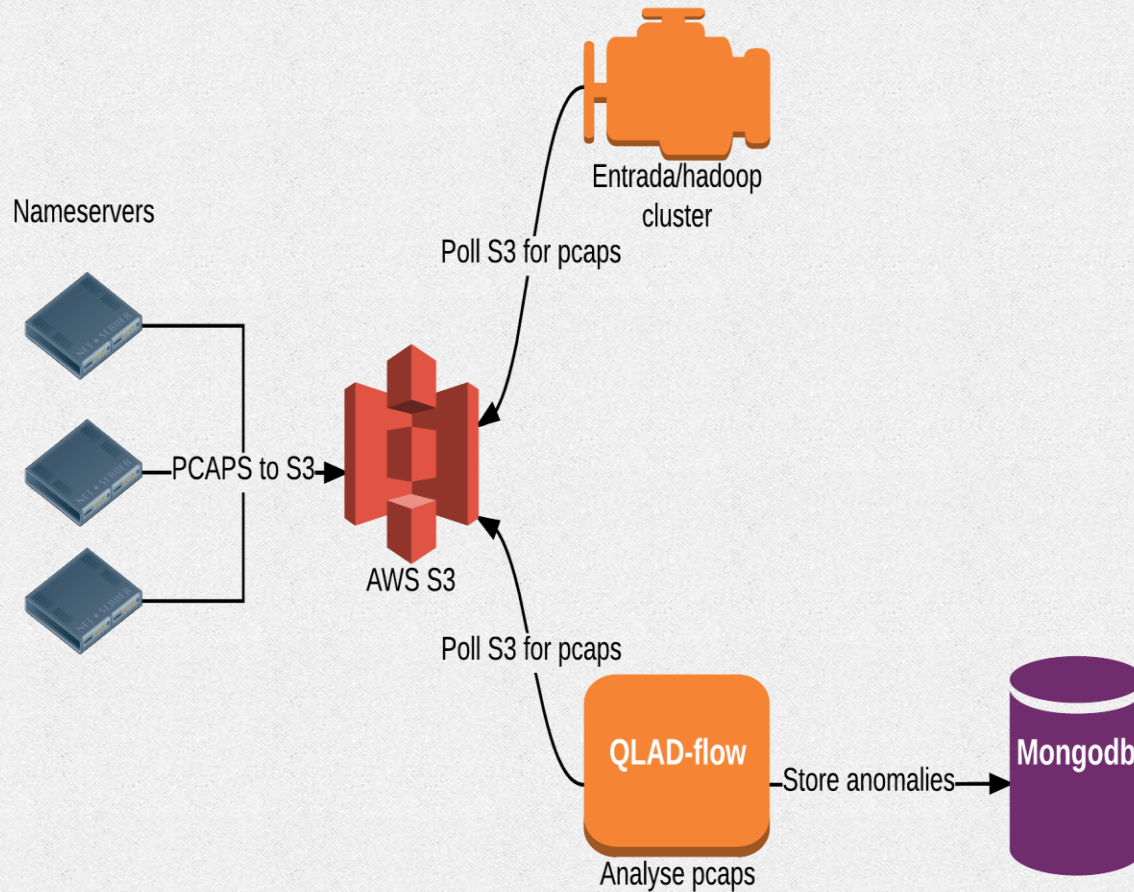
⑥ Anomaly identification by sketch combination



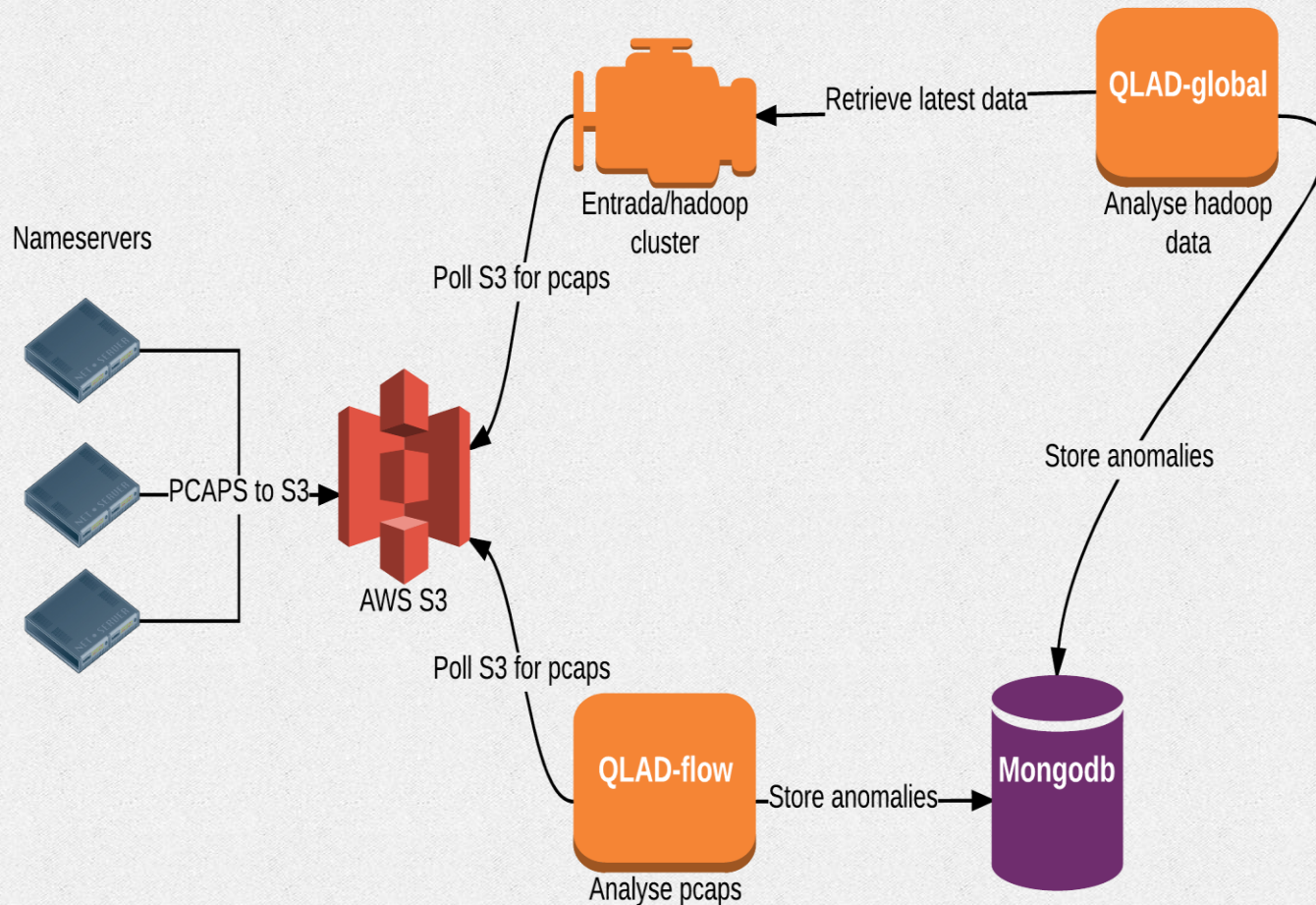
Implementation QLAD-flow: shortcomings

- Some attacks span a lot of flows -> QLAD-flow is unable to detect these (e.g. like DoS with spoofed IP address)
- QLAD-global was implemented to fill the gap

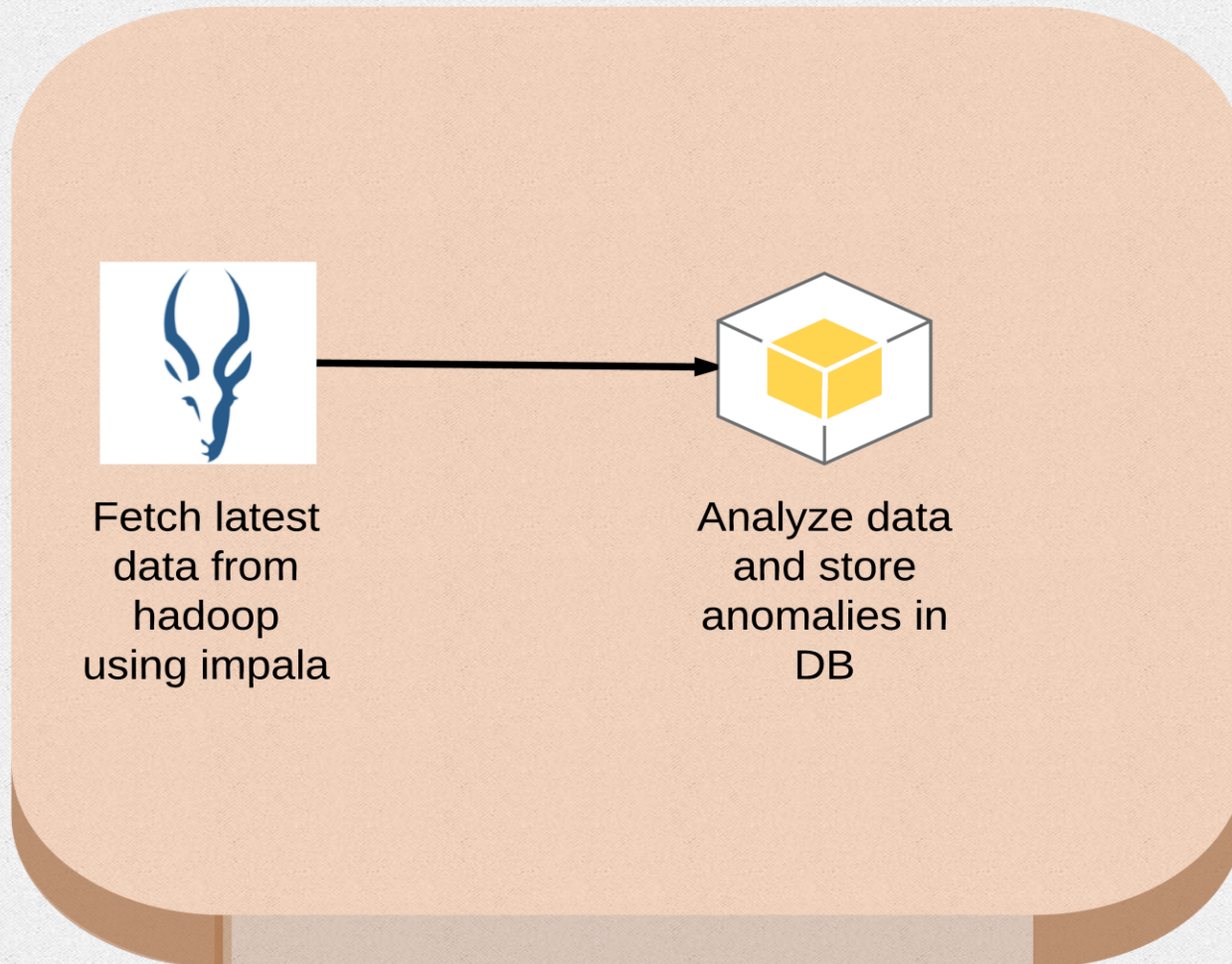
Implementation QLAD-global: architecture



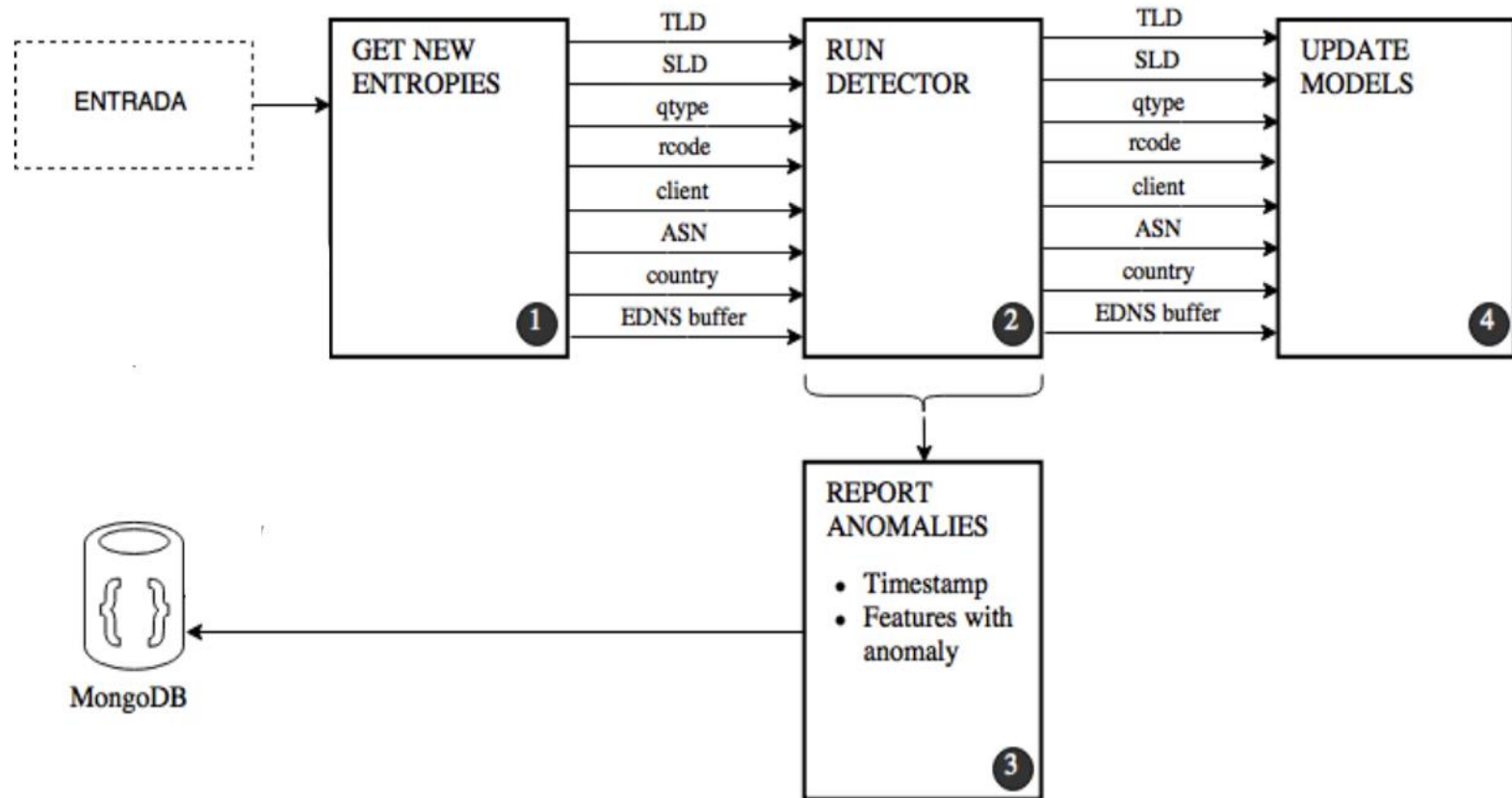
Implementation QLAD-global: architecture



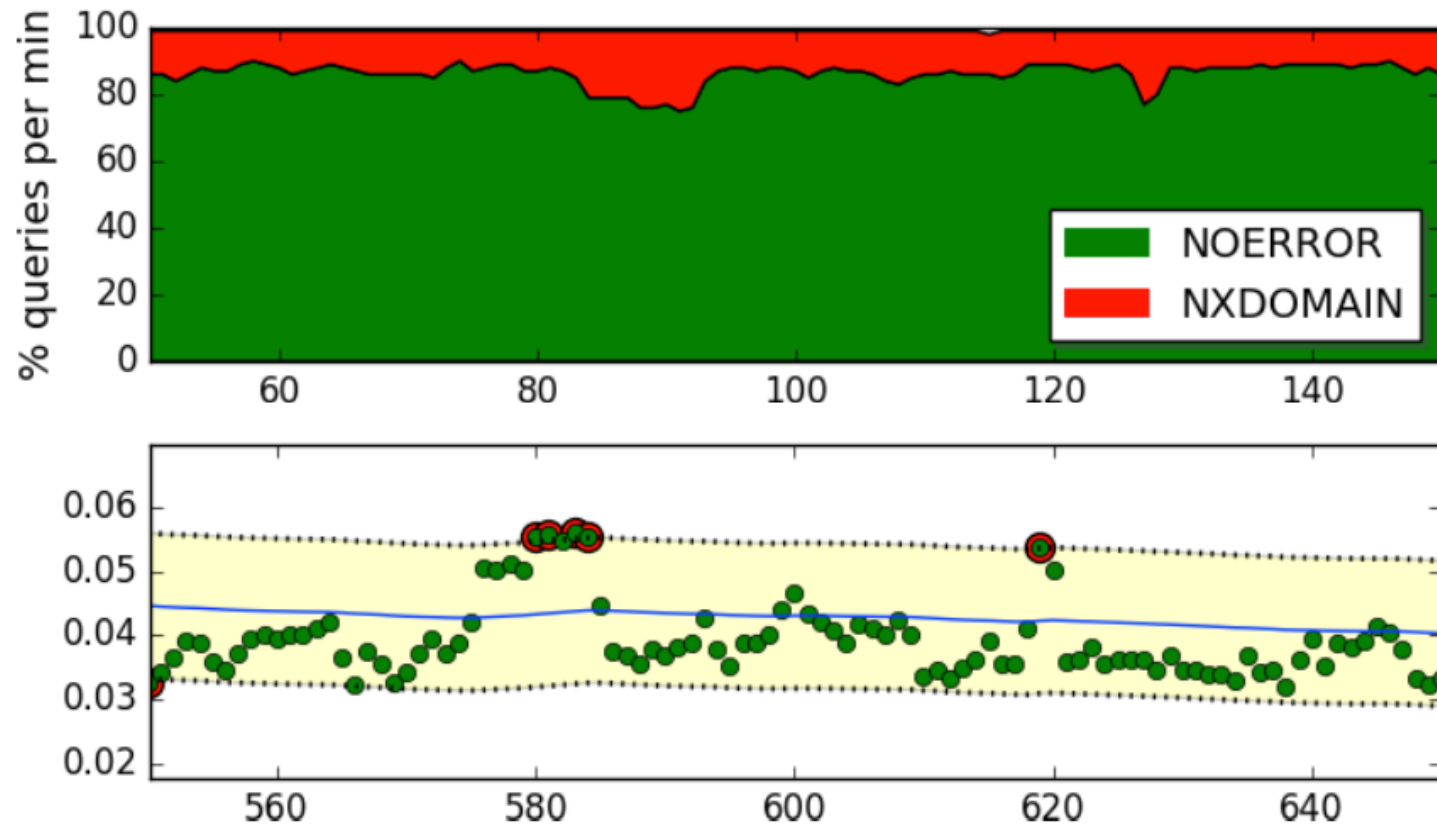
Implementation QLAD-global: flow



Implementation: QLAD-global: algorithm



Implementation: QLAD-global: algorithm result



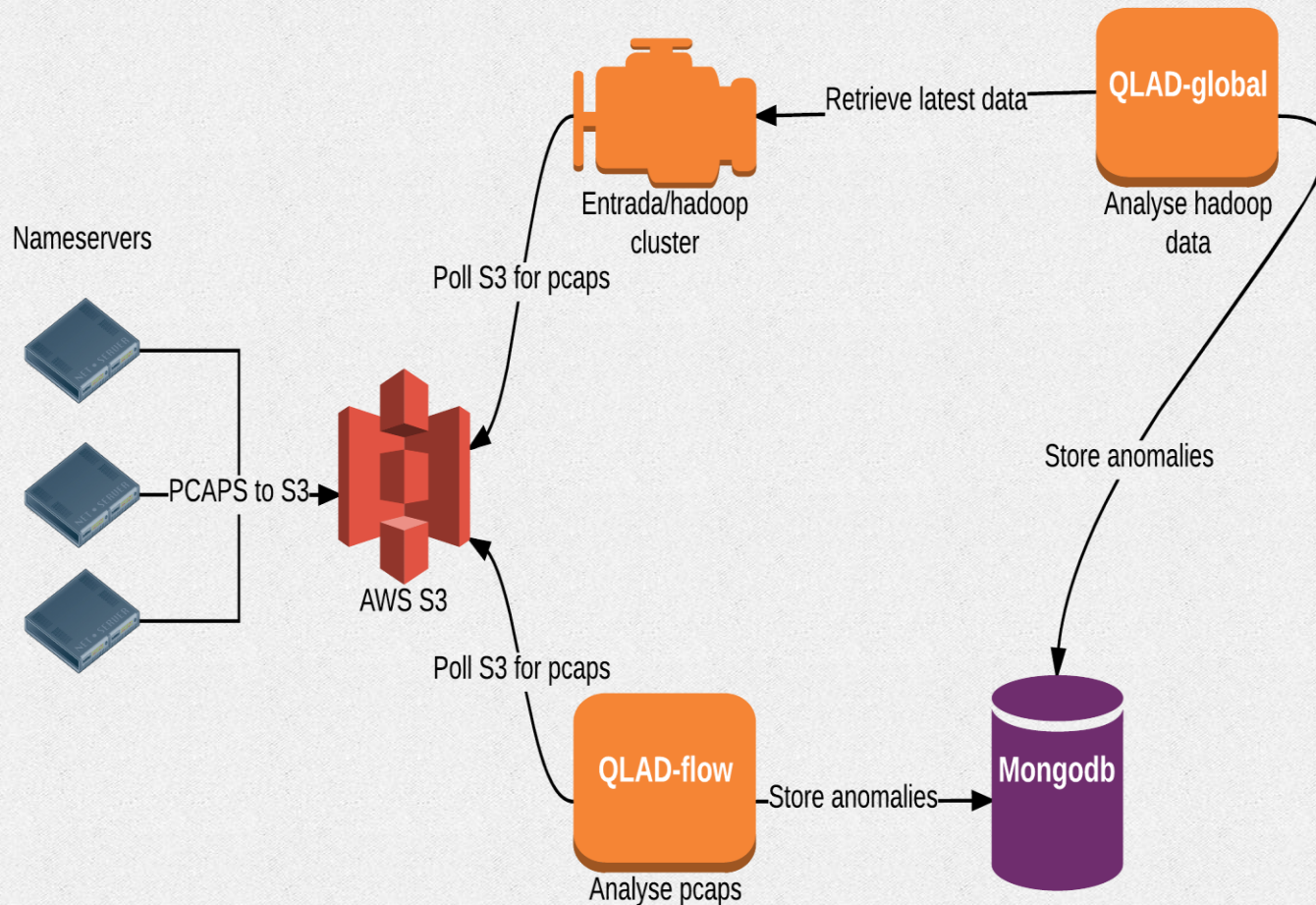
Implementation: QLAD-UI

Automatic classification is challenging

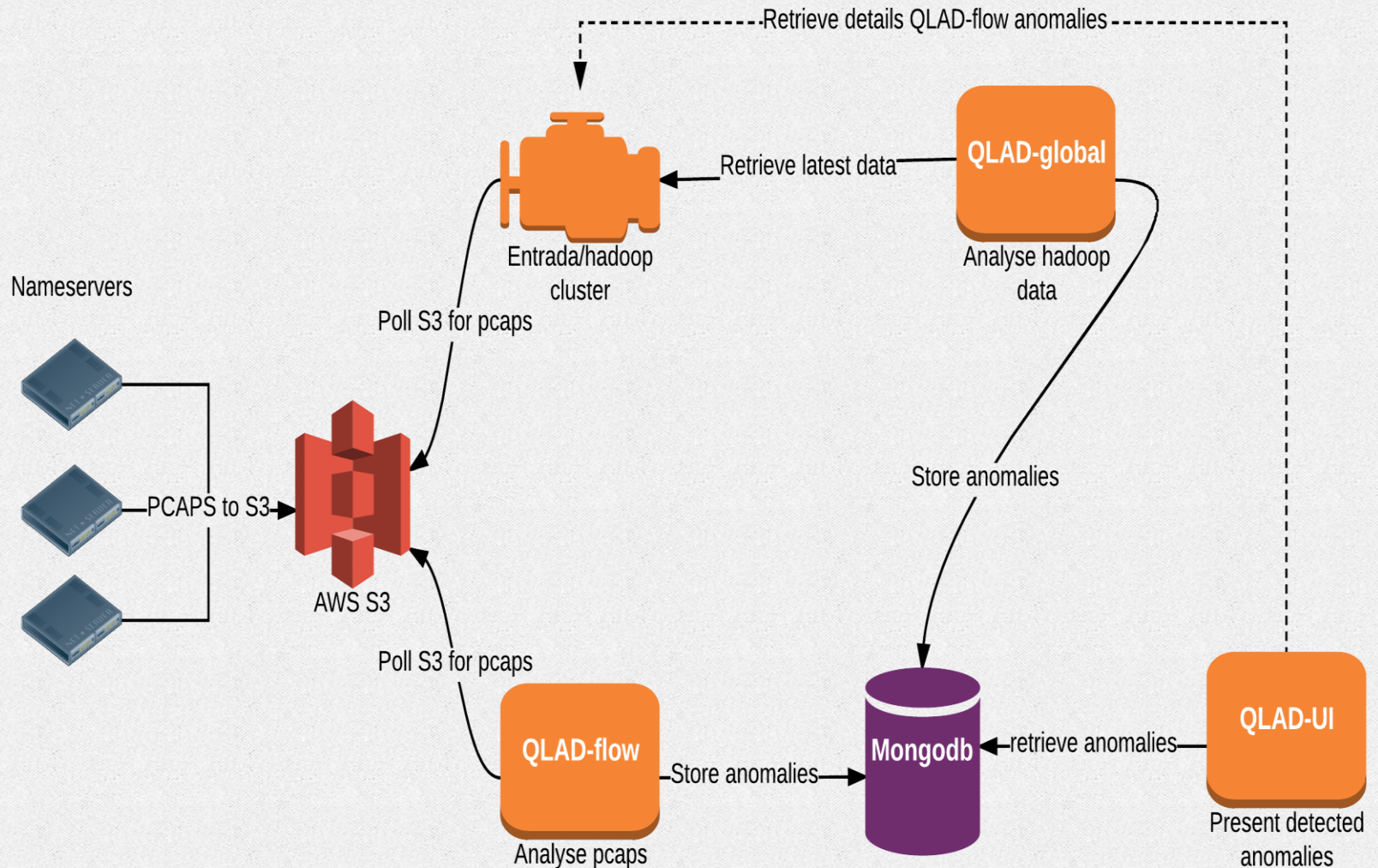
- wide range of anomalies
- subtle differences

=> Rely on human interpretation of the detected anomalies by QLAD-global and QLAD-flow

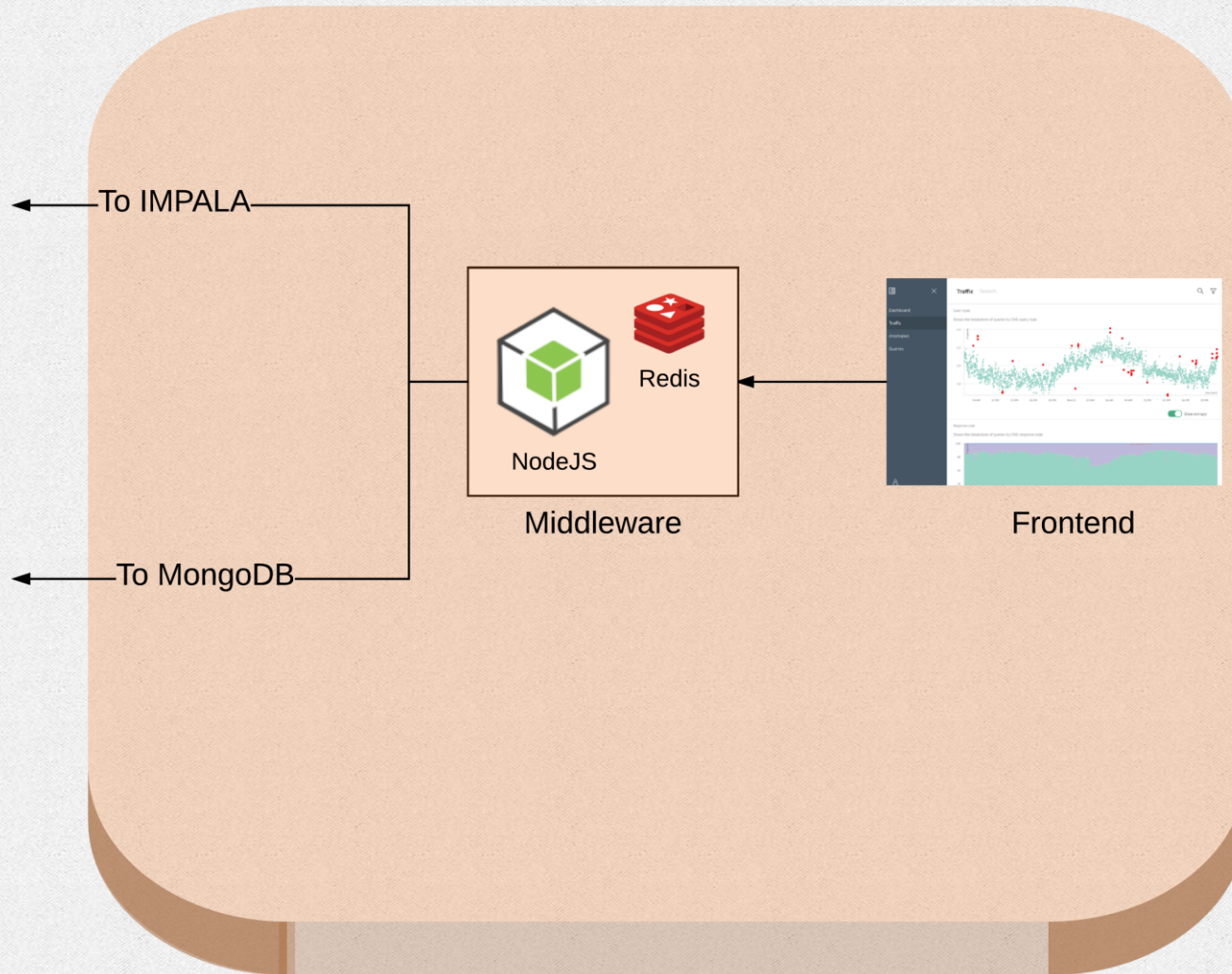
Implementation QLAD-UI: architecture



Implementation QLAD-UI: architecture



Implementation QLAD-UI: flow



Demo

Questions?