amulog: A General Log Analysis Framework for Diverse Template Generation Methods

Satoru Kobayashi[1], Yuya Yamashiro[2], Kazuki Otomo[2], Kensuke Fukuda[1]

1: NII, 2: The University of Tokyo

Mail: sat@nii.ac.jp

Background

- Automated log analysis requires log template generation
  - To classify messages for time-series analysis
- There are too many log template generation methods (> 50)
  - Different assumptions, difficult to compare or combine

-> We need general framework to use the methods uniformly

Contribution

- We find and solve issues for flexible and practical operation
  - Common preprocessing of logs (depending on data)
  - Log template matching for flexible template use
- We designed and implemented a general framework ***amulog***
- We confirmed scalability of amulog with academic network data

Design of *amulog*

Requirements for general framework

(A) Preprocessing logs uniformly

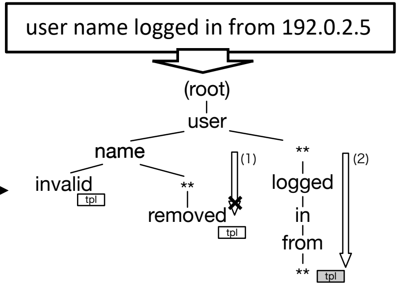
-> Rule-based customizable parser (log2seq)

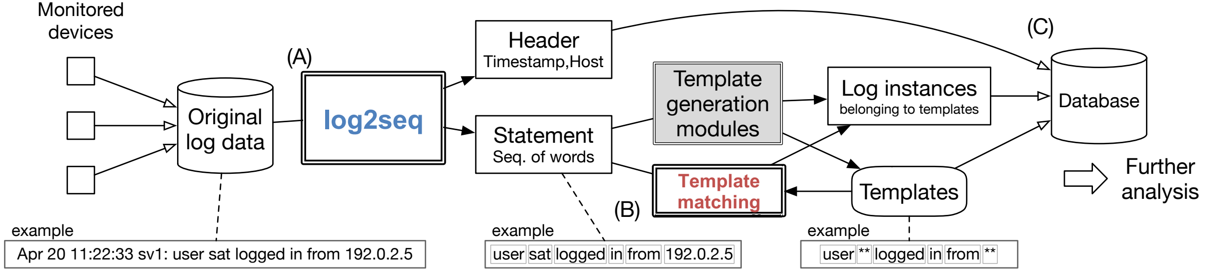
(B) Matching log templates and their instances

-> Tree-based template matching algorithm

(C) Storing parsed data into database

Template matching algorithm





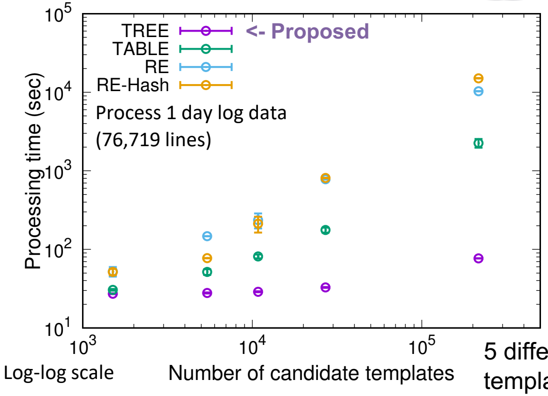
Apr 20 11:22:33 sv1: user sat logged in from 192.0.2.5

user | sat | logged | in | from | 192.0.2.5

user | ** | logged | in | from | **

Evaluation

Using log data of SINET4



5 different template sets

Findings:

- Using Segmentation (TREE, TABLE) is efficient
- TREE is scalable (fast even with 10^5 templates)

Check https://github.com/cpflat/amulog