Event Pattern Identification in Anonymized System Logs

Siavash Ghiassvand\(^\ddagger\) and Florina M. Ciocar\(^*\)

\(\ddagger\)Technische Universität Dresden, Germany \(\star\)University of Basel, Switzerland

1. Motivation and Challenges

- Improving the performance of pattern detection algorithms
- Filtering system logs based on document frequency

2. The Method

3. Pattern Detection

4. Results

- Data filtering significantly speeds-up the identification process
- The frequency of events is reduced by 50%

5. Conclusion

- A small portion of system logs contains the most alarming information.
- Filtering system logs based on document frequency can significantly reduce the required storage capacity.
- The data can be fully anonymized but still useful for some statistical analysis.
- System log filtering significantly increases the performance of pattern detection algorithms.
- Reducing the volume of non-informative data and combining other sources of information increase the accuracy of analysis results, without requiring additional computational power.

Footnotes

- Event 1: Each event is a log, in which all variables are replaced by constant sample values.
- Event 2: Part of the system logs is responsible for more than 90% of syslog network traffic.
- Event 3: The system logs related to a single event and can safely be ignored.
- Event 4: The most alarming events are among the 1% of all system logs.
- Event 5: Events related to several errors, including "file system" failures, are located among the 74% of all system logs.

References


Acknowledgments

This work was partly supported by the German Research Foundation (DFG) in the DFG Research Center for Advanced Information Systems Engineering (CIS 17).