Hochschule Hannover
University of Applied Sciences and Arts
-
Fakultät III Medien, Information und Design

Hochschule Hannover
University of Applied Sciences and Arts
-
Fakultät IV Wirtschaft und Informatik

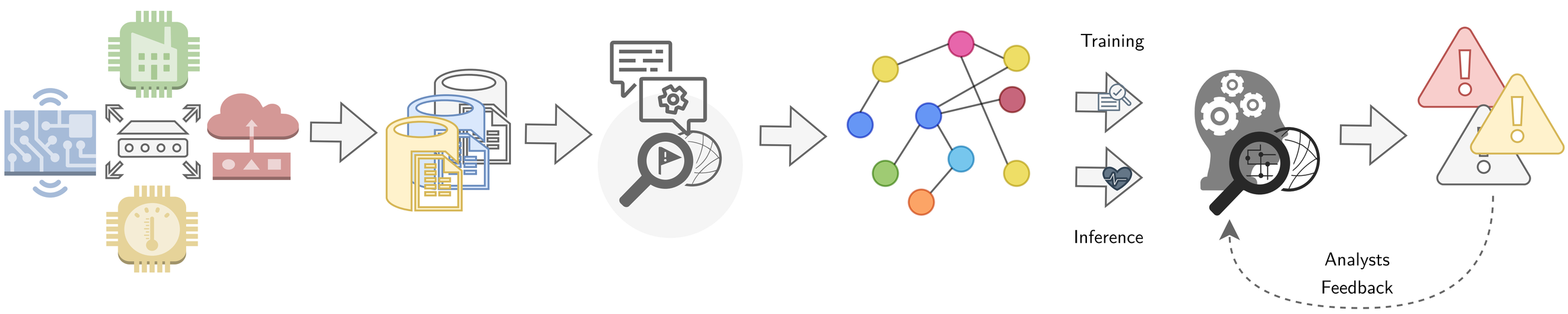
PRO ME TE U S

TRUST at HsH

Data|H
INSTITUTE FOR APPLIED DATA SCIENCE HANNOVER

With a bit of imagination Prometeus stands for: Enhancing cybersecurity with PROtocol MEssage analysis and anomaly detection using TExt UnderStanding.

Concept



From Raw Logs...

Excerpt of HDFS Dataset [1].

Timestamp	Component	Content
1226289780	dfs.DataNode$DataXceiver	10.251.30.134:50010 Served block blk_2039230511363331616 to /10.251.65.203
1226306817	dfs.DataNode$DataXceiver	10.251.107.50:50010 Served block blk_−2285729896739318683 to /10.251.70.5
1226314574	dfs.FSNamesystem	BLOCK* NameSystem.addStoredBlock: blockMap updated: 10.251.65.203:50010 is added to blk_−3949633496933134307 size 67108864
1226342014	dfs.DataNode$DataXceiver	10.251.39.179:50010:Got exception while serving blk_−8558281124275910849 to /10.251.30.134:
1226374998	dfs.FSNamesystem	BLOCK* NameSystem.delete: blk_−765820154946060546 is added to invalidSet of 10.251.65.203:50010

Static Parts

Mined Event Templates by Drain [2].

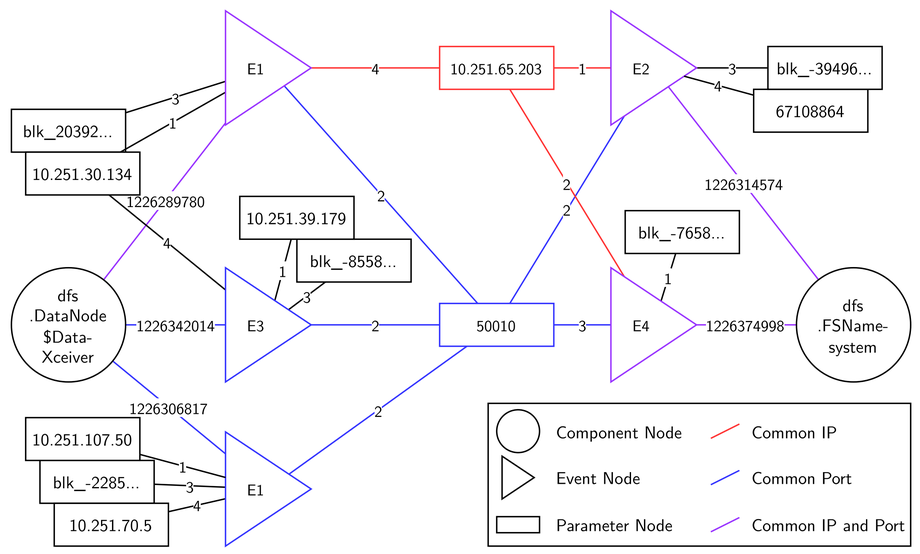
EventId	EventTemplate
E1	<*>:<*> Served block <*> to /<*>
E2	BLOCK* NameSystem.addStoredBlock: blockMap updated: <*>:<*> is added to <*> size <*>
E3	<*>:<*>:Got exception while serving <*> to /<*>:
E4	BLOCK* NameSystem.delete: <*> is added to invalidSet of <*>:<*>

Dynamic Parts

Extracted Parameters.

Timestamp	Component	EventId	Parameters
1226289780	dfs.DataNode$DataXceiver	E1	10.251.30.134, 50010, blk_2039230511363331616, 10.251.65.203
1226306817	dfs.DataNode$DataXceiver	E1	10.251.107.50, 50010, blk_−2285729896739318683, 10.251.70.5
1226314574	dfs.FSNamesystem	E2	10.251.65.203, 50010, blk_−3949633496933134307, 67108864
1226342014	dfs.DataNode$DataXceiver	E3	10.251.39.179, 50010, blk_−8558281124275910849, 10.251.30.134
1226374998	dfs.FSNamesystem	E4	blk_−765820154946060546, 10.251.65.203, 50010

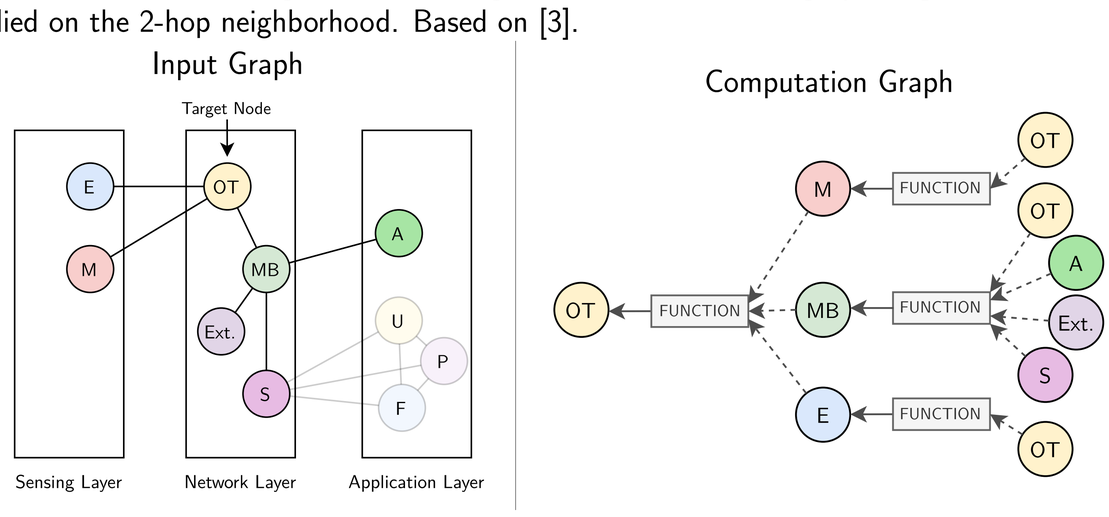
...to System Behavior Graph



Using the HDFS dataset as an example, Drain was able to reduce 11.1 million log lines to 29 event templates.

Message Passing

Simplified representation of a CPS in a three-layer architecture with exemplary characteristics of the nodes and their edges. For the target node *OT* the message passing framework is applied on the 2-hop neighborhood. Based on [3].



E: Event, M: Measurement
OT: OT Component, MB: Message Broker, Ext.: External Service, S: Server
A: Application Data, U: User, P: Process, F: File

Conclusion

Research objectives:

- Generic log data as input for graph transformation algorithms,
- Customized graph construction,
- GNN-based anomaly detection methods.

Main difficulties:

- Finding realistic test environments,
- False positives due to noise,
- Unknown and various log message formats.

References

[1] Wei Xu et al. "Detecting Large-Scale System Problems by Mining Console Logs". In: *Proceedings of the ACM SIGOPS 22nd Symposium on Operating Systems Principles*. SOSP '09. New York, NY, USA: Association for Computing Machinery, 2009, pp. 117–132. doi: 10.1145/1629575.1629587.

[2] Pinjia He et al. "Drain: An Online Log Parsing Approach with Fixed Depth Tree". In: *2017 IEEE International Conference on Web Services (ICWS)*. 2017, pp. 33–40. doi: 10.1109/ICWS.2017.13.

[3] William L. Hamilton. "Graph Representation Learning". In: *Synthesis Lectures on Artificial Intelligence and Machine Learning* 14.3 (2020), pp. 1–159. url: https://www.cs.mcgill.ca/~wlh/grl_book/files/GRL_Book.pdf.

VolkswagenStiftung