BERT를 활용한 로그 시퀀스 내 로그 단위 이상탐지를 위한 프레임워크

권태완 이찬재 정상현 김용신 이태희

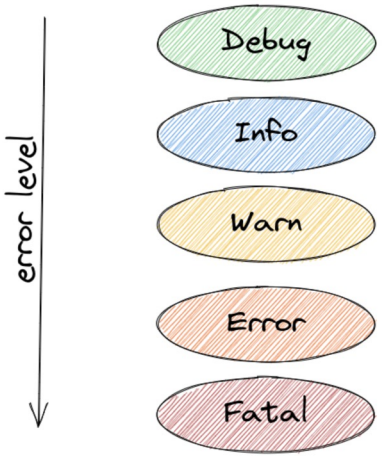
OKESTRO

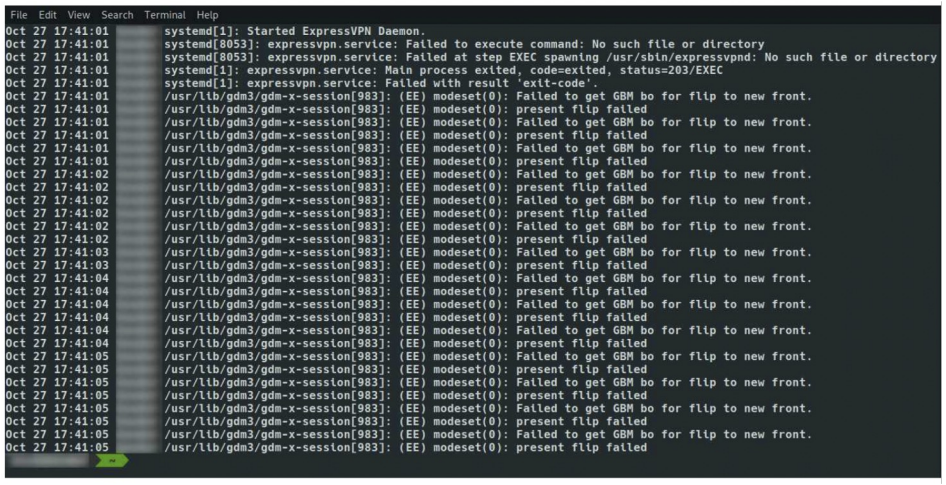
OKESTRO

시스템 로그와 로그 기반 이상 탐지

시스템 로그 : 시스템이 동작하는 동안 발생하는 로그 메시지 또는 로그 메시지들의 집합

시스템 로그 분석의 단계 : log collection → log compression → log parsing → log analysis





```
Oct 27 17:41:01 [redacted] systemd[1]: Started ExpressVPN Daemon.
Oct 27 17:41:01 [redacted] systemd[8053]: expressvpn.service: Failed to execute command: No such file or directory
Oct 27 17:41:01 [redacted] systemd[8053]: expressvpn.service: Failed at step EXEC spawning /usr/sbin/expressvpnd: No such file or directory
Oct 27 17:41:01 [redacted] systemd[1]: expressvpn.service: Main process exited, code=exited, status=203/EXEC
Oct 27 17:41:01 [redacted] systemd[1]: expressvpn.service: Failed with result 'exit-code'.
Oct 27 17:41:01 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): Failed to get GBM bo for flip to new front.
Oct 27 17:41:01 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): present flip failed
Oct 27 17:41:01 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): Failed to get GBM bo for flip to new front.
Oct 27 17:41:01 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): present flip failed
Oct 27 17:41:01 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): Failed to get GBM bo for flip to new front.
Oct 27 17:41:01 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): present flip failed
Oct 27 17:41:02 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): Failed to get GBM bo for flip to new front.
Oct 27 17:41:02 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): present flip failed
Oct 27 17:41:02 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): Failed to get GBM bo for flip to new front.
Oct 27 17:41:02 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): present flip failed
Oct 27 17:41:02 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): Failed to get GBM bo for flip to new front.
Oct 27 17:41:02 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): present flip failed
Oct 27 17:41:03 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): Failed to get GBM bo for flip to new front.
Oct 27 17:41:03 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): present flip failed
Oct 27 17:41:04 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): Failed to get GBM bo for flip to new front.
Oct 27 17:41:04 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): present flip failed
Oct 27 17:41:04 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): Failed to get GBM bo for flip to new front.
Oct 27 17:41:04 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): present flip failed
Oct 27 17:41:04 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): Failed to get GBM bo for flip to new front.
Oct 27 17:41:04 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): present flip failed
Oct 27 17:41:05 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): Failed to get GBM bo for flip to new front.
Oct 27 17:41:05 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): present flip failed
Oct 27 17:41:05 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): Failed to get GBM bo for flip to new front.
Oct 27 17:41:05 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): present flip failed
Oct 27 17:41:05 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): Failed to get GBM bo for flip to new front.
Oct 27 17:41:05 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): present flip failed
Oct 27 17:41:05 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): Failed to get GBM bo for flip to new front.
Oct 27 17:41:05 [redacted] /usr/lib/gdm3/gdm-x-session[983]: (EE) modeset(0): present flip failed
[redacted] ~
```

출처 : *https://www.atatus.com/blog/an-ultimate-guide-to-node-js-logging/*, https://www.systranbox.com/how-to-find-internet-logs-on-linux/

시스템 로그와 로그 기반 이상 탐지

로그 메시지의 구조 : 로그 메시지는 크게 헤더와 메시지의 내용 부분으로 나뉘어진다

로그 파싱 : 로그 메시지에서 헤더와, 내용 부분 중에서 variable들을 제거함으로써, 메시지들을 구조화

1	081109 213908 2549 INFO dfs.DataNode$DataXceiver: 10.251.39.192:50010 Served block blk_-5341992729755584578 to /10.251.39.192
2	081109 214009 2594 INFO dfs.DataNode$DataXceiver: 10.250.5.237:50010 Served block blk_3166960787499091856 to /10.251.43.147
3	081109 214043 2561 WARN dfs.DataNode$DataXceiver: 10.251.30.85:50010 Got exception while serving blk_-2918118818249673980 to /10.251.90.64
…	

Parsing

	Header	Event Template	Parameters
1	[081109, 213908, 2549, INFO, dfs.DataNode$DataXceiver]	* Served block * to *	[10.251.39.192:50010, blk_-5341992729755584578, /10.251.39.192]
2	[081109, 214009, 2594, INFO dfs.DataNode$DataXceiver]	* Served block * to *	[10.250.5.237:50010, blk_3166960787499091856, /10.251.43.147]
3	[081109, 214043, 2561, WARN, dfs.DataNode$DataXceiver]	* Got exception while serving * to *	[10.251.30.85:50010, blk_-2918118818249673980, /10.251.90.64]
…			

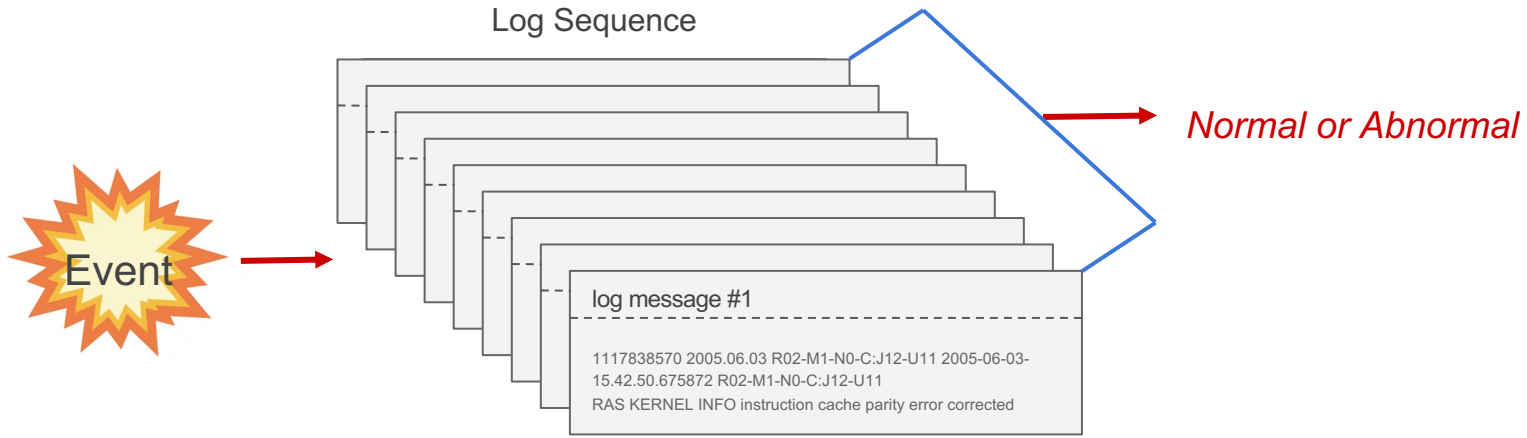
출처 : https://arxiv.org/pdf/2108.01955.pdf

OKESTRO

기존의 방식: 로그 시퀀스 기반의 이상탐지

1개의 로그 메시지가 아닌, 10개 또는 20개로 나열된 로그 메시지의 시퀀스에 대해 정상 / 이상을 탐지하는 방식

로그 메시지는 어느 로그 시퀀스에 포함되는가에 따라 정상일수도, 이상일수도 있다

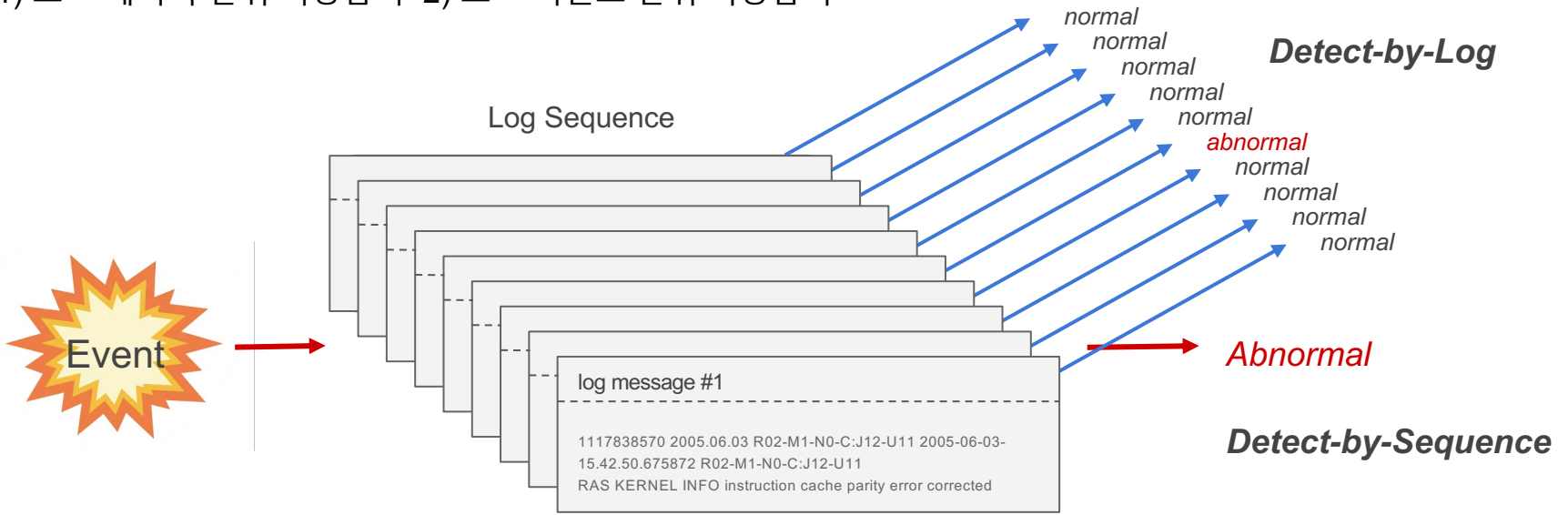


OKESTRO

FineLog: 로그 메시지 단위의 이상탐지

FineLog는 두 단계에 걸쳐 이상탐지를 수행

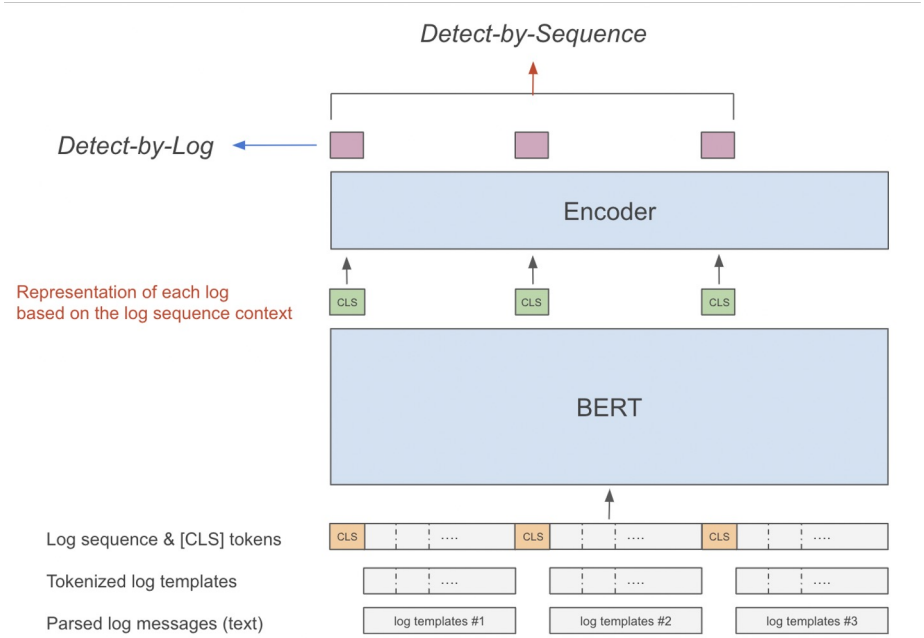
1) 로그 메시지 단위 이상탐지 2) 로그 시퀀스 단위 이상탐지



OKESTRO

FineLog: 로그 메시지 단위의 이상탐지

BERTSUM은 추출 요약 시, 원문에 포함된 각각의 문장들에 대해, 요약문의 포함 여부에 대해 binary classification



OKESTRO

실험 결과

기존의 로그 시퀀스 단위의 예측에서 두 도메인에 걸친 전이학습을 이용한 TransLog 다음으로 높은 성능 기록

로그 메시지 단위의 예측에서도 precision과 recall 모두 높은 성능 기록

imbalance가 심한 시스템 로그 데이터셋에서 로그 시퀀스 단위, 로그 메시지 단위 모두에서 높은 precision 기록

Method	Precision	Recall	F_1 Score
LR	0.13	0.93	0.23
SVM	0.97	0.30	0.46
DeepLog	0.8974	0.8278	0.8612
LogAnomaly	0.7312	0.7609	0.7408
NeuralLog (sup.)	0.61	0.78	0.68
TransLog (sup.)	0.98	**0.98**	**0.98**
LogBERT	0.8940	0.9232	0.9083
LAnoBERT	-	-	0.8749
Ours	**0.9836**	0.9404	**0.9616**

표 1. BGL 데이터셋에서 FineLog의 Detect-by-Sequence 결과

Precision	Recall	F_1 Score
0.9894	0.9654	0.9773

표 2. BGL 데이터셋에서 FineLog의 Detect-by-Log 결과

OKESTRO

Q&A

OKESTRO