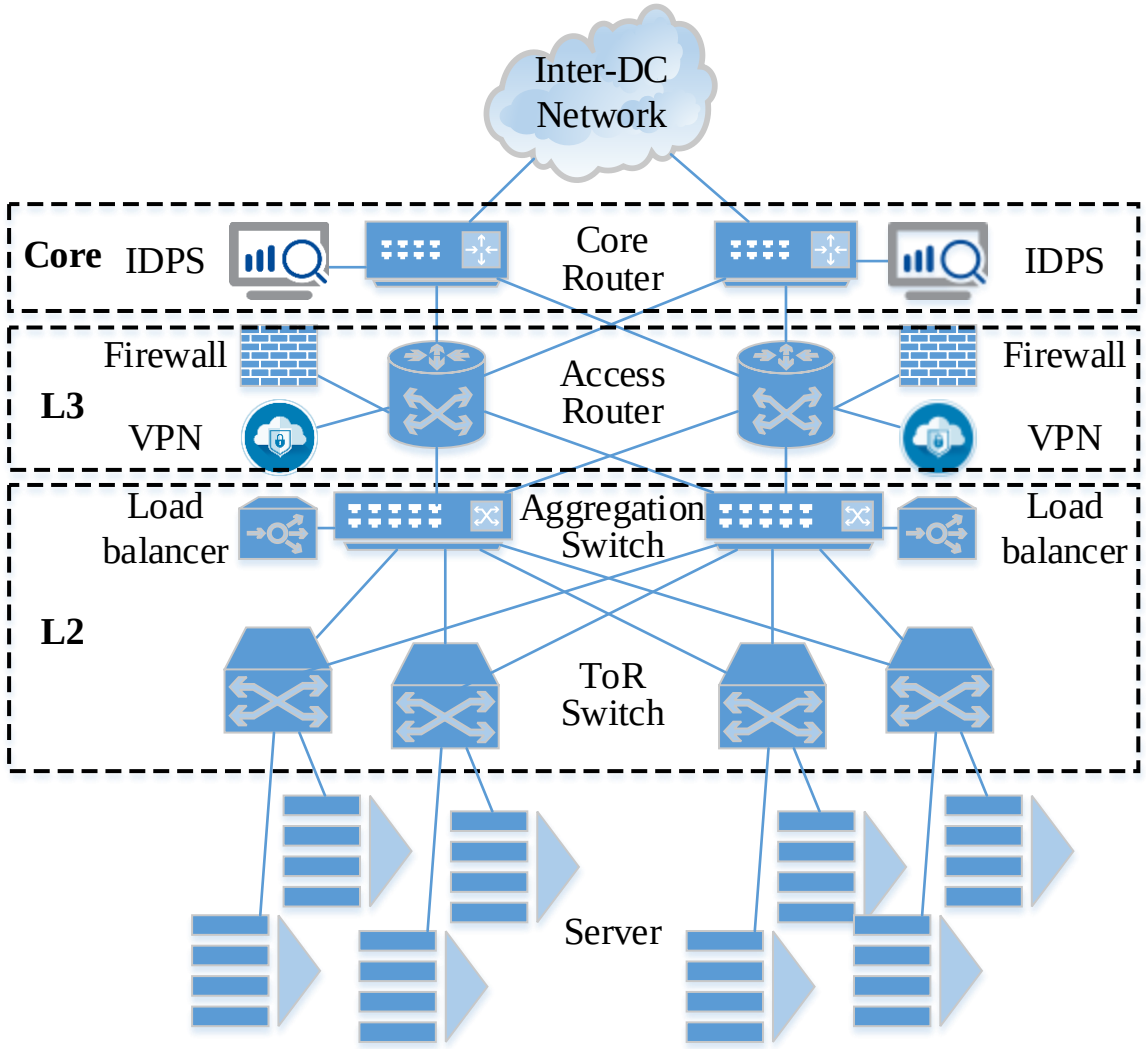
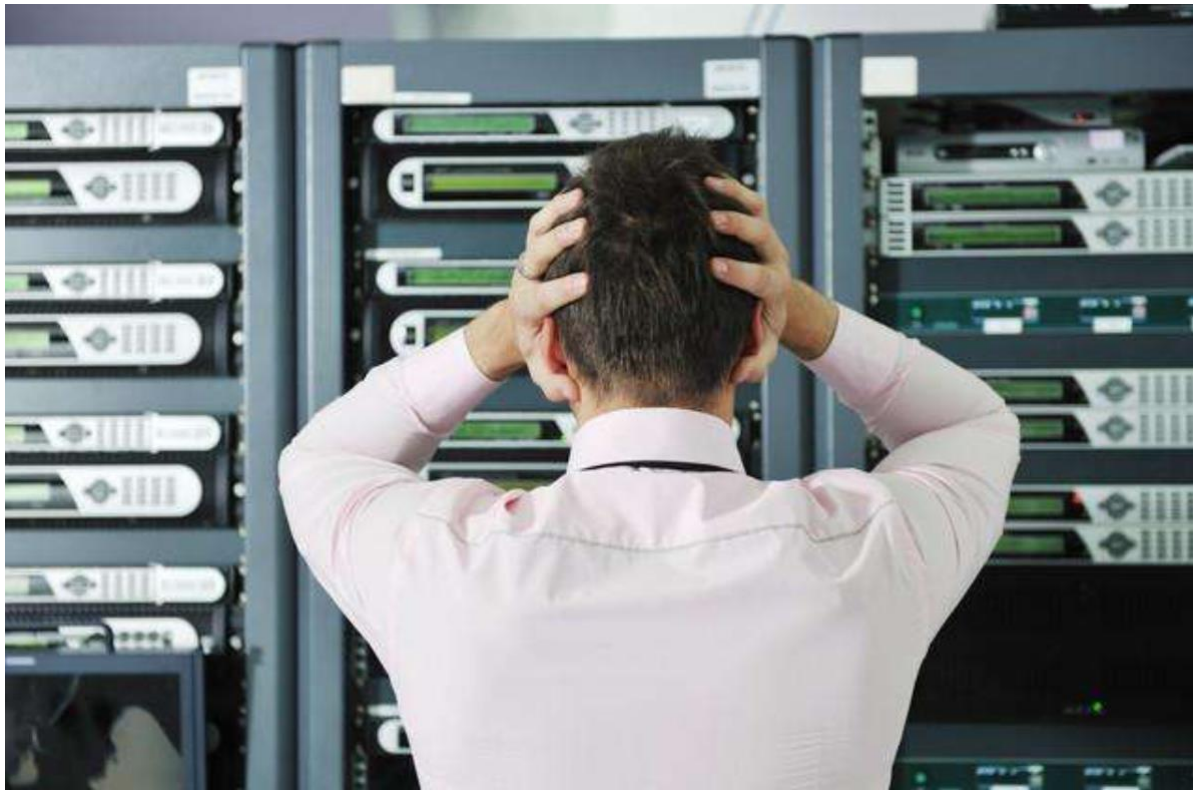
Device-Agnostic Log Anomaly Classification with Partial Labels

Weibin Meng, Ying Liu, Shenglin Zhang, Dan Pei
Hui Dong, Lei Song, Xulong Luo

Motivation

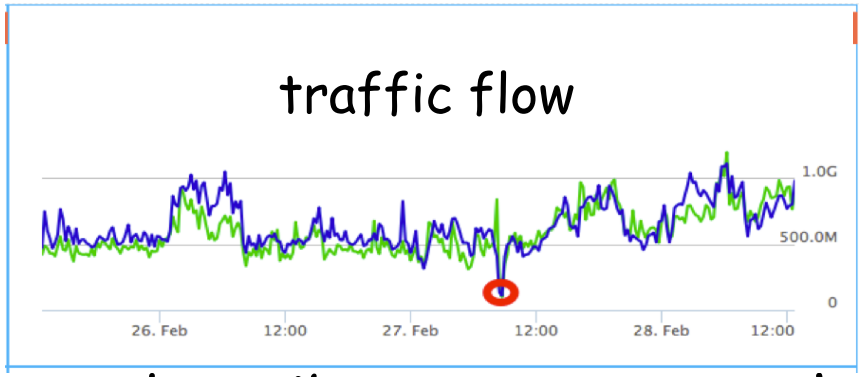
Architecture of Datacenter Networks

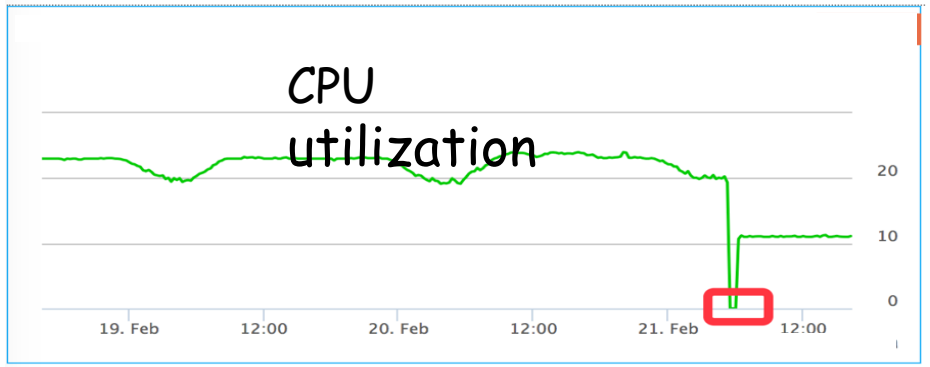




Motivation

- Traditional anomaly detection methods usually monitor KPI curves.
  - KPI need network operators select manually.
  - KPI methods can only find anomalous behaviors





- Logs describe some events that KPI curve can't, such as the root cause.
  - Logs are most valuable data sources for device management.

```
SYSLOG/6/SYSLOG_RESTART: System restarted -- H3C Comware Software.

DEV/2/FAN STATE CHANGE TO FAILURE: Trap 1.3.6.1.4.1.2011.2.23.1.12.1.6(fanfailure): fan ID is 1

P01 OUT_SWITCH 192.168.201.218 2016 %%10DEVM/1/FAN STATE CHANGES TO FAILURE(t): Trap
1.3.6.1.4.1.2011.2.23.1.12.1.6: fan ID is 1

DEV/5/SYSTEM_REBOOT: System is rebooting now.
```

Device logs

Message types are ambiguous for accurate classification

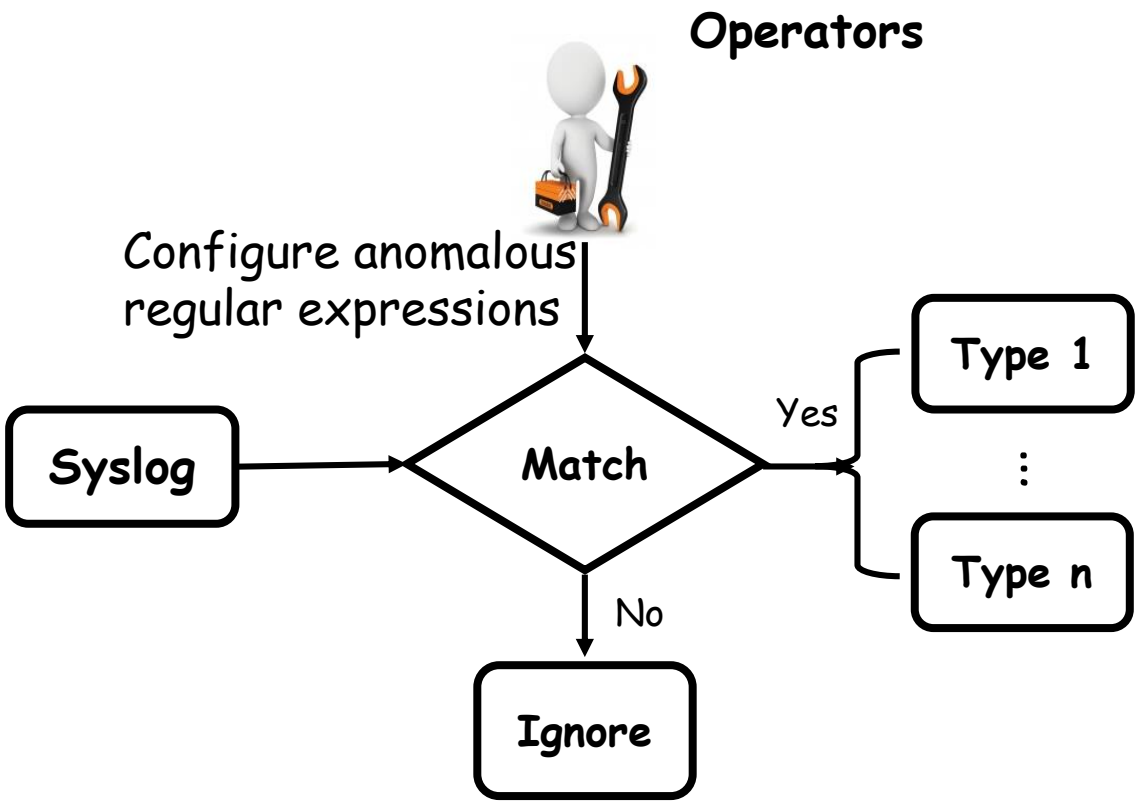
- **Examples of device(switch) log :**

Switch ID	Timestamp	Message Type	Detailed Message
Switch 1	Jun 12 19:03:27 2017	SIF	**Interface** te-1/1/59, changed state **to** down
Switch 2	Jun 13 20:22:03 2017	-	Vlan-interface vlan22, changed state **to** down
Switch 1	Jun 13 20:22:03 2017	SIF	**Interface** te-1/1/17, changed state **to** up
Switch 18	Jun 18 05:21:03 2017	SIF	**Interface** te-1/1/19, changed state **to** up
Switch 22	Jun 15 13:46:43 2017	OSPF	Neighbour vlan23, changed state from Exchange **to** Loading
Switch 28	Jun 15 13:46:43 2017	OSPF	PVID mismatch discovered on Ten-GigabitEthernet 6/0/10 , **to** S12516XAF-38.Int Ten-GigabitEthernet 3/0/17

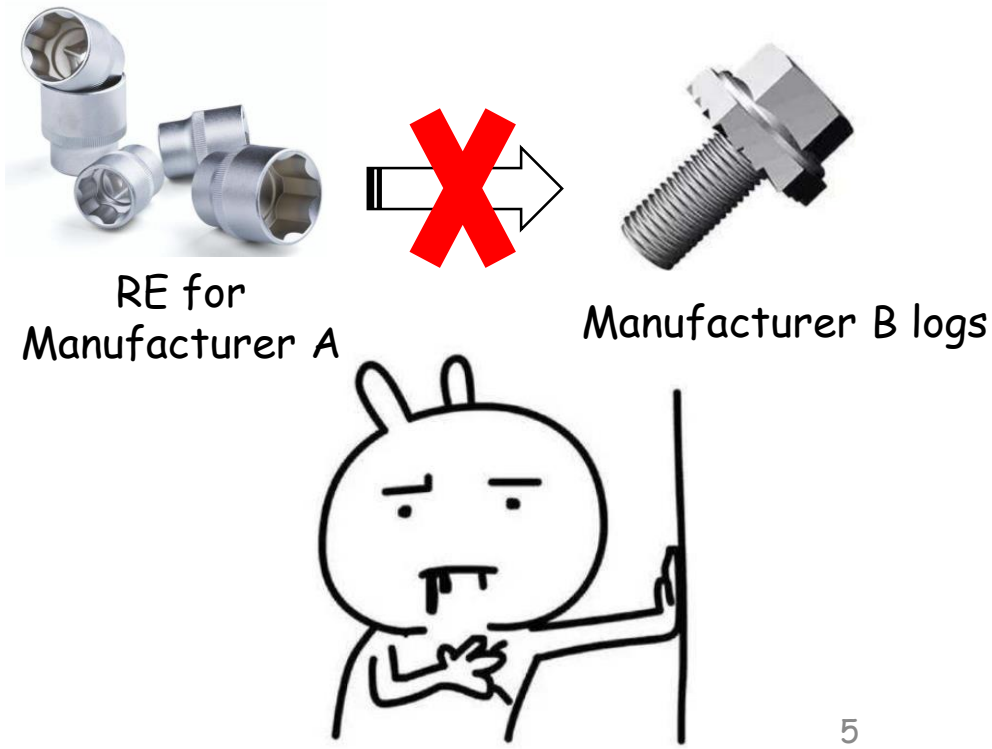
Detailed Messages are **Semi-structured natural languages** provided by device developers

Drawbacks in Regular Expression

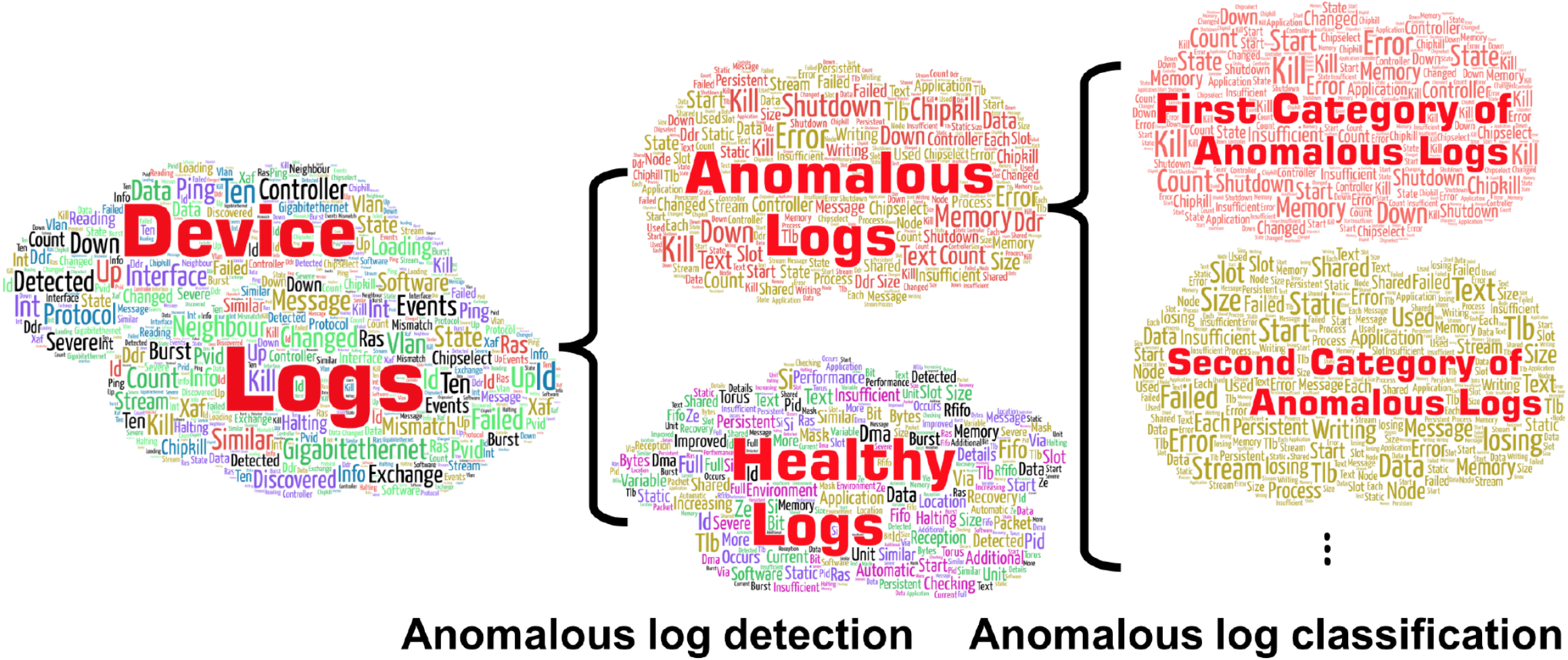
- **Regular Expression** is the popular technique for anomalous log classification.



- **Drawbacks:**
  - Low generality
  - Labor intensity
  - …



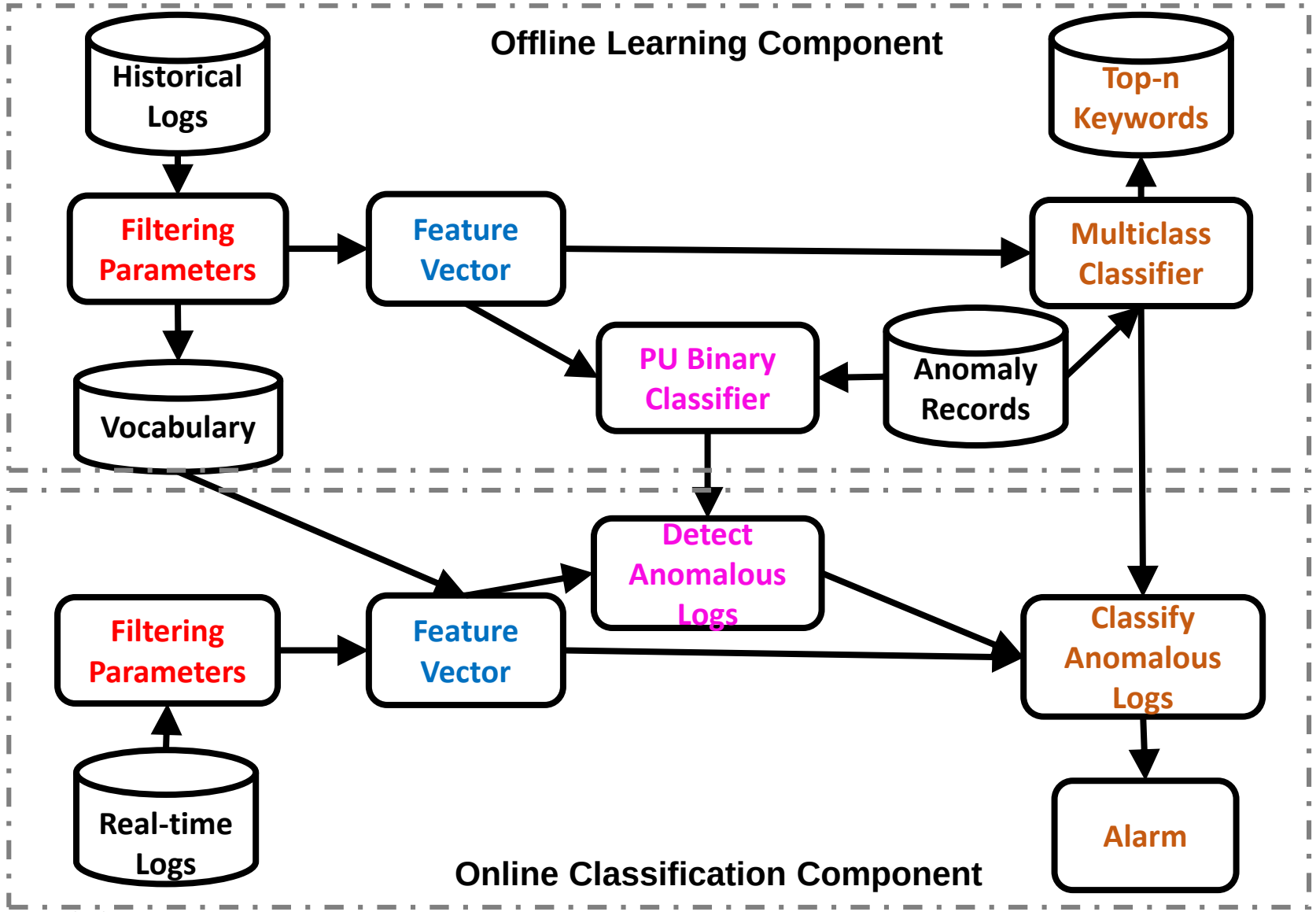
Problem Definitions



Challenges

- **Device-agnostic vocabulary**
  - Device logs are type- specific and manufacturer- specific.
  - It is hard to fit one classification model for all different device types.
- **Partial labels**
  - Network operators only label partial anomalous logs they encountered.
  - Difficult to train a traditional classification model.

LogClass Design Overview



1. Log Preprocessing
2. Feature vector
3. Anomaly detection
4. Anomaly classification

Text feature vector

The universal method to construct a text feature vector is the bag-of-words model.

logs:

L_1	Interface	te-1/1/59	changed	state	to	down	
L_2	VlanInterface	vlan22	changed	state	to	up	
L_3	Neighbour	vlan23	changed	state	from	Exchange	to	Loading

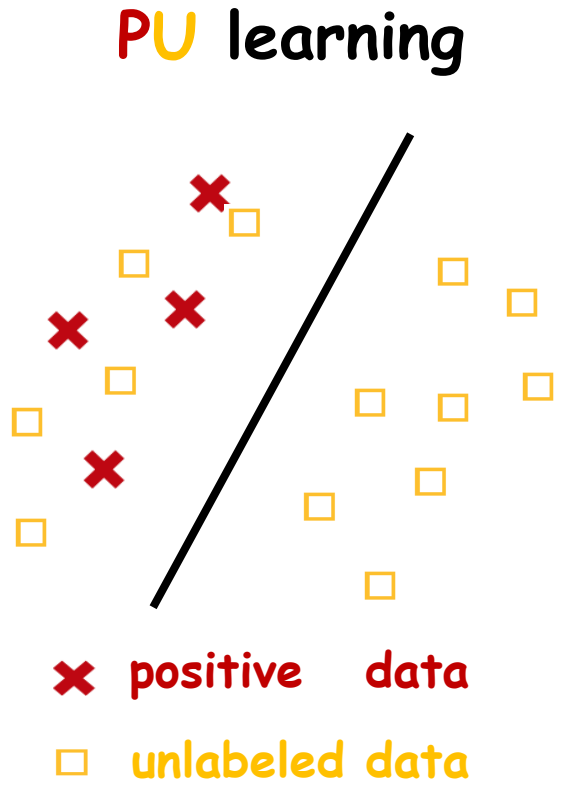
bag-of-words vectors:

Vocabulary	Interface	changed	state	to	down	VlanInterface	Neighbour	from	Exchange	Loading	up
L_1	1	1	1	1	1	0	0	0	0	0	0
L_2	0	1	1	1	0	1	0	0	0	0	1
L_3	0	1	1	1	0	0	1	1	1	1	0

Assign weighting values to each component in vectors. (e.g., TF-IDF)

PU Learning

- Different from tradition classification.
  - In our scnario, labelling all existing anomalous logs is not natural.
- PU Learning input:
  - Positive set **P (Anomalous logs)**
  - Unlabeled set **U (Unlabeled logs)**



(Gang Niu et al. NIPS'16)

Evaluation

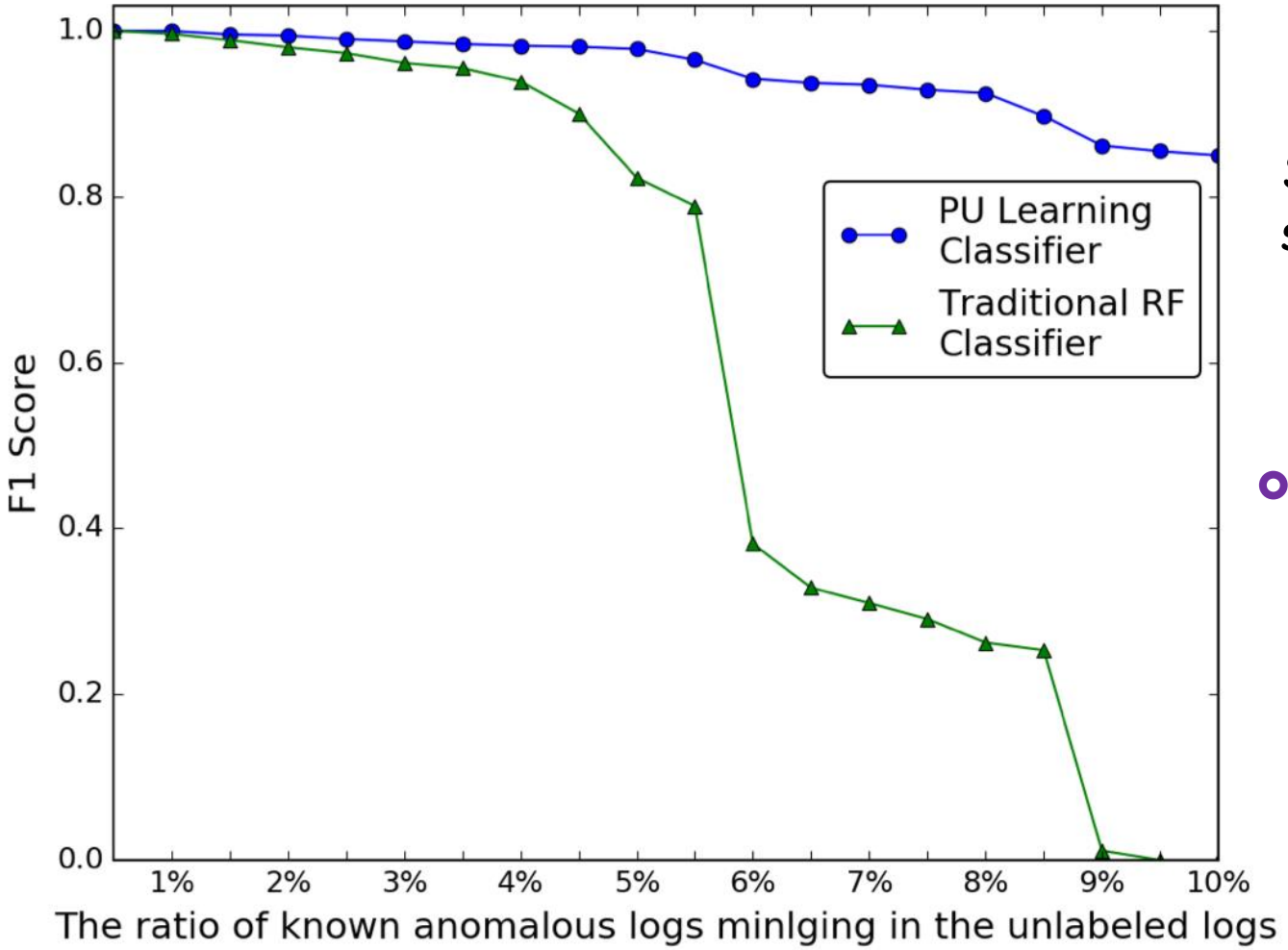
Dataset

- Real-world Switch logs
- 58 switches types
- Two-week period
- 1,758,456 anomalous logs
- 16,702,547 unlabeled logs

Benchmark methods

- Labeled-LDA
- Regular Expression

Evaluation on PU Learning



Sampled anomalous logs randomly cross all switch types and assumed they have no labels.

PU Learning classifier is more stable than traditional classifier.

Evaluation on Anomalous Log Classification

Methods	Macro-F1	Micro-F1	Training Time(s)	Classifying Time(s)
LogClass	95.32%	99.74%	247.73	4.836
L-LDA	89.68%	93.53%	4436.4	28.59
RE	-	-	-	419.47

LogClass is more accurate.

The overheads of L-LDA and RE are larger than LogClass

Conclusion

Challenges

- Device-Agnostic vocabulary
- Partial anomalous logs have labels

LogClass

- PU learning
- Simple NLP techniques

Evaluation

- Real-world switch logs.

Thank you!

mwb16@mails.tsinghua.edu.cn