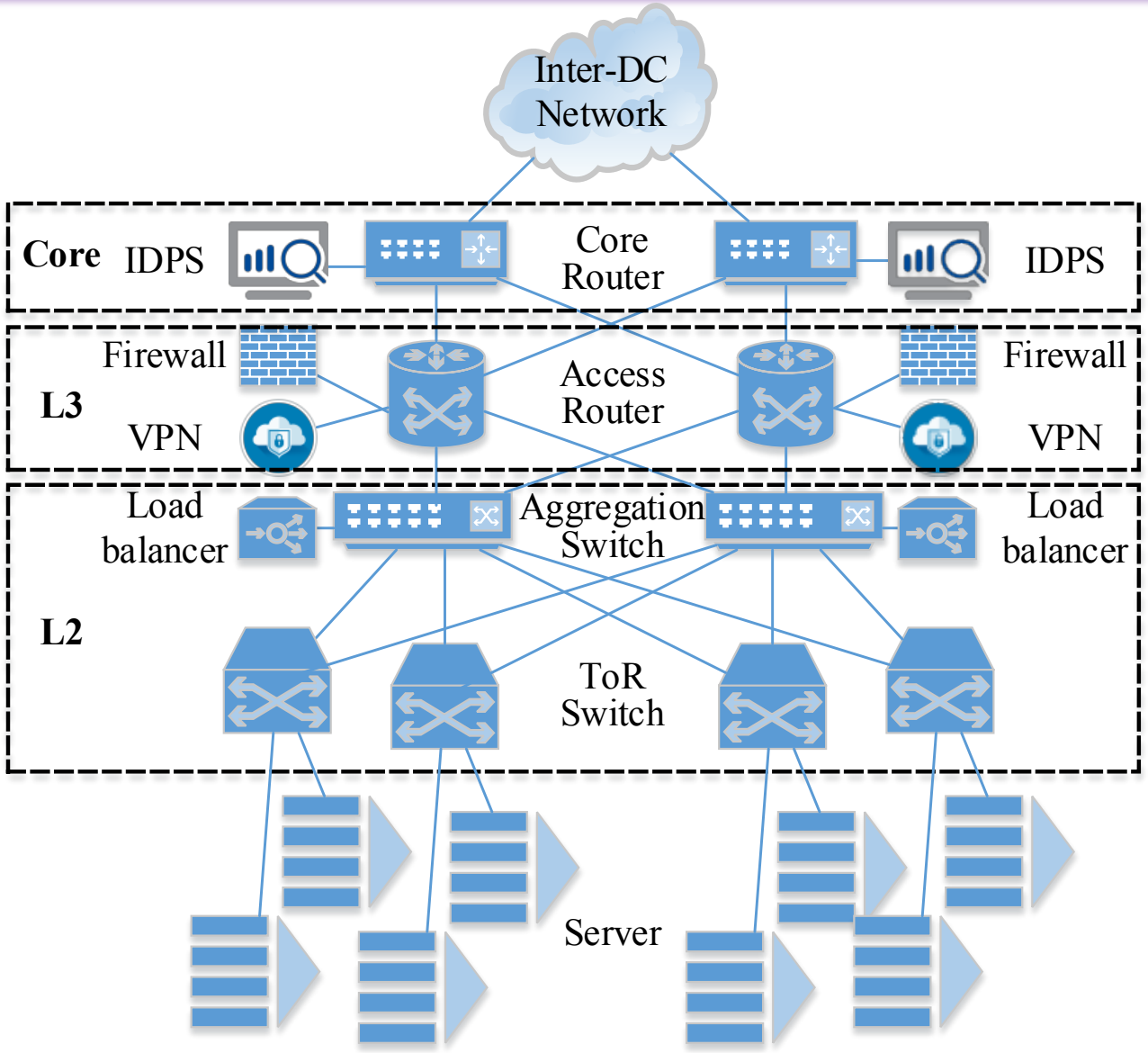
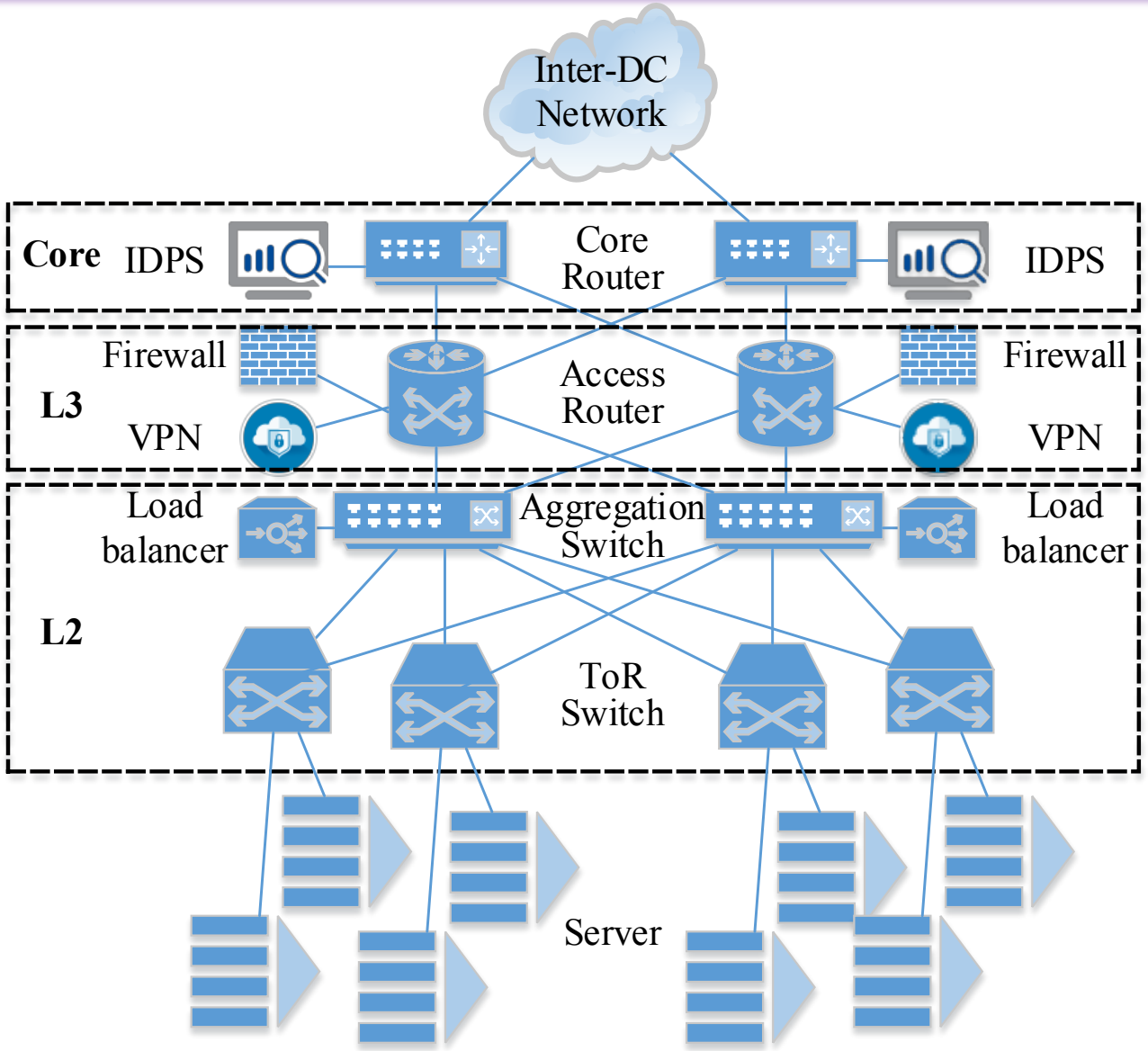
PreFix: Switch Failure Prediction in Datacenter Networks

Shenglin Zhang

Network Devices in Data Center Networks

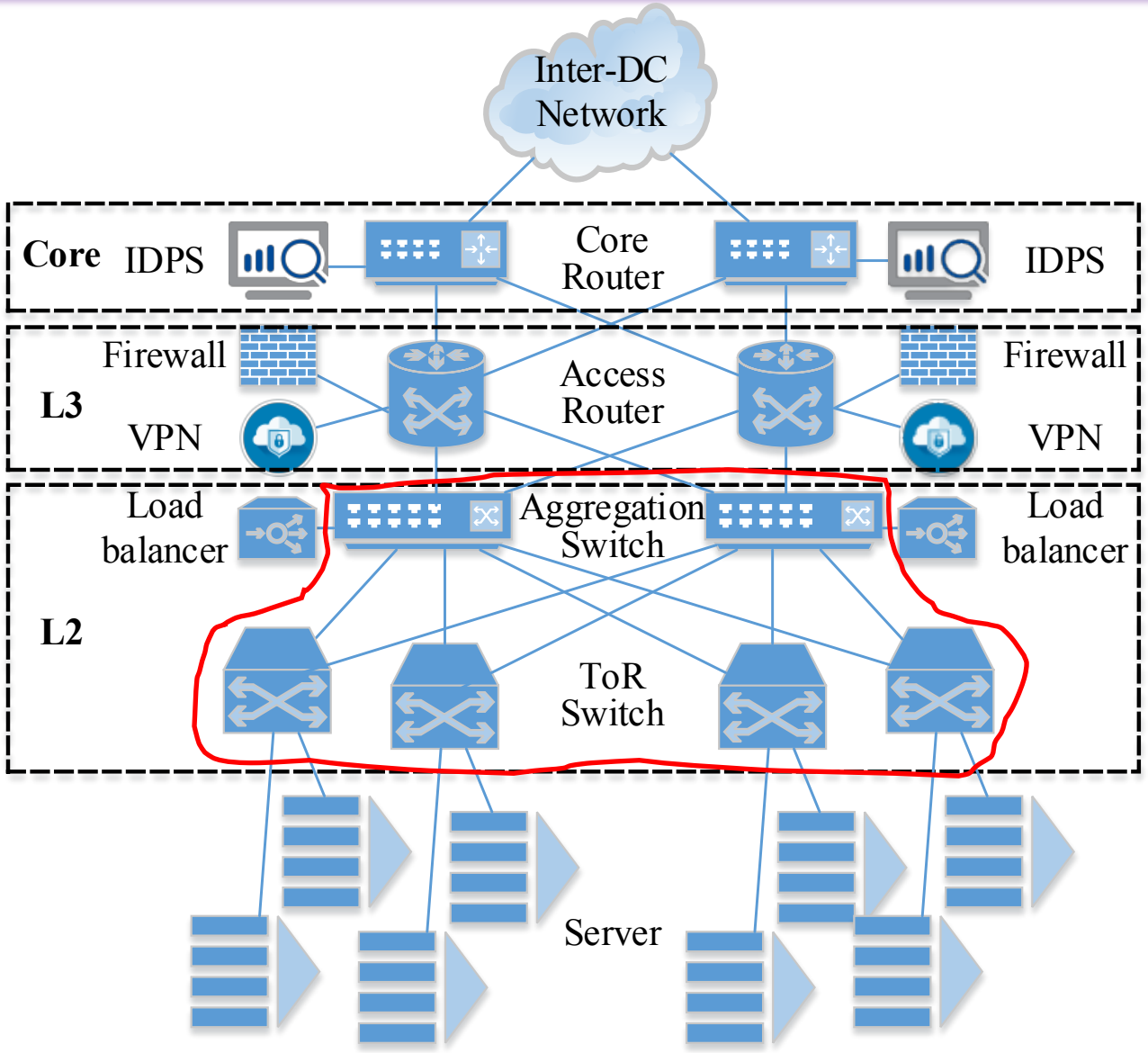


Network Devices in Data Center Networks



- Switch
  - Top-of-rack switch
  - Aggregation switch
- Router
  - Access router
  - Core router
- Middle box
  - Firewall
  - Intrusion detection and prevention system (IDPS)
  - Load balancer
  - VPN

Network Devices in Data Center Networks



- Switch
  - Top-of-rack switch
  - Aggregation switch
- Router
  - Access router
  - Core router
- Middle box
  - Firewall
  - Intrusion detection and prevention system (IDPS)
  - Load balancer
  - VPN

Scale of Network Devices in Datacenter

Microsoft (C. Guo, et al., SIGCOMM'15)

- Hundreds of thousands to millions of servers
- Hundreds of thousands of switches
- Millions of cables and fibers

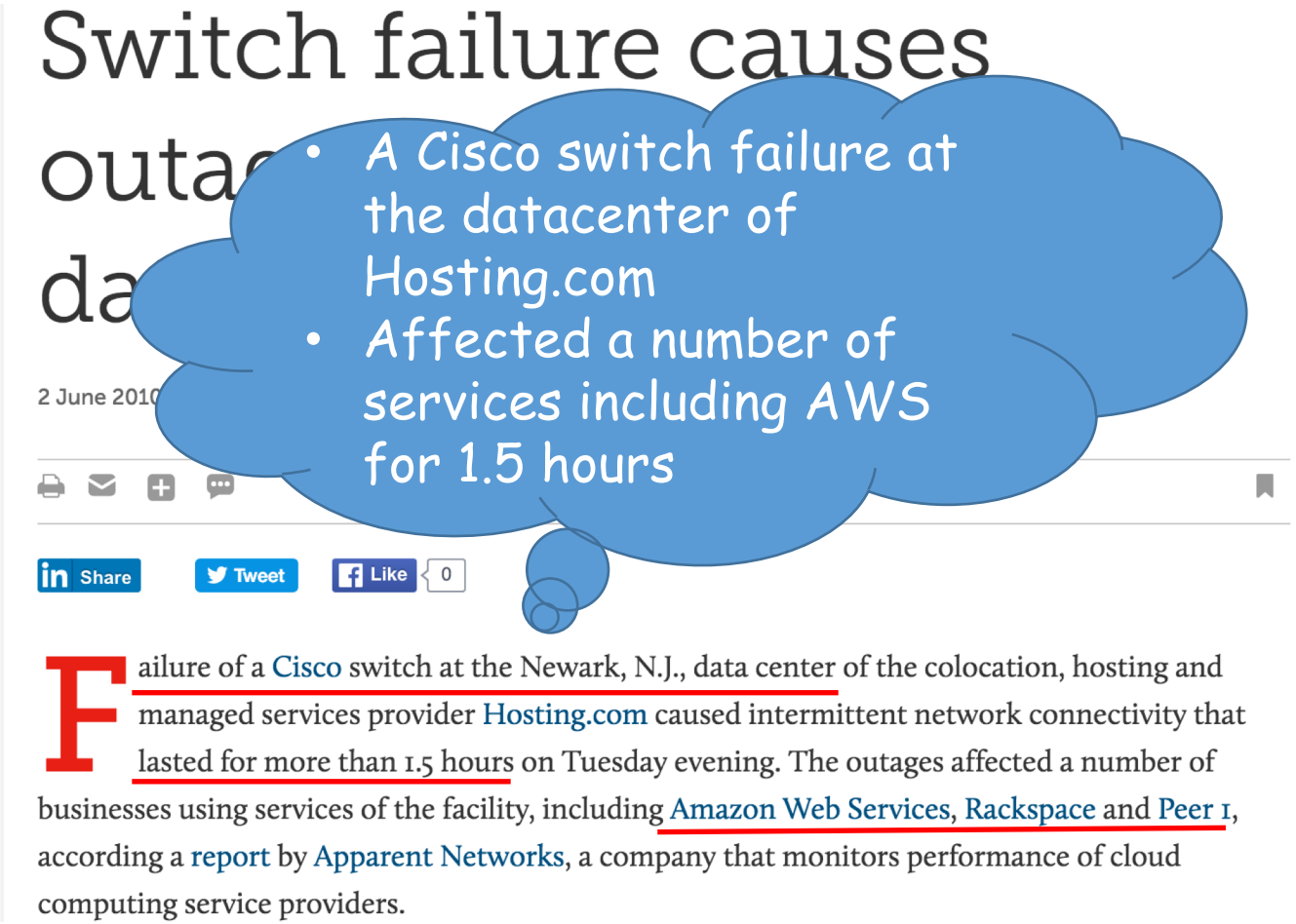
Scale of Network Devices in Datacenter

Microsoft (C. Guo, et al., SIGCOMM'15)	• Hundreds of thousands to millions of servers • Hundreds of thousands of switches • Millions of cables and fibers
Baidu	• Hundreds of thousands of servers • Tens of thousands of switches

Scale of Network Devices in Datacenter

Microsoft (C. Guo, et al., SIGCOMM'15)	• Hundreds of thousands to millions of servers • Hundreds of thousands of switches • Millions of cables and fibers
Baidu	• Hundreds of thousands of servers • Tens of thousands of switches
Swich failures are the norm rather than the exception (P. Gill, et al., SIGCOMM'11)	• More than 400 switch failures per year

Switch Failures Lead to Outages



Switch failure causes outa… da…

2 June 2010

Share Tweet Like 0

Failure of a Cisco switch at the Newark, N.J., data center of the colocation, hosting and managed services provider Hosting.com caused intermittent network connectivity that lasted for more than 1.5 hours on Tuesday evening. The outages affected a number of businesses using services of the facility, including Amazon Web Services, Rackspace and Peer 1, according a report by Apparent Networks, a company that monitors performance of cloud computing service providers.

- A Cisco switch failure at the datacenter of Hosting.com
- Affected a number of services including AWS for 1.5 hours

Switch Failures Lead to Outages

Switch failure causes outa... da...

2 June 2010

- A Cisco switch failure at the datacenter of Hosting.com
- Affected a number of services including AWS for 1.5 hours

Share Tweet Like 0

Failure of a Cisco switch at the Newark, N.J., data center of the colocation, hosting and managed services provider Hosting.com caused intermittent network connectivity that lasted for more than 1.5 hours on Tuesday evening. The outages affected a number of businesses using services of the facility, including Amazon Web Services, Rackspace and Peer 1, according a report by Apparent Networks, a company that monitors performance of cloud computing service providers.

Switch failure shuts down computer network at data center

AP By The Associated Press
May 24, 2016 8:49 am

CHESTER, Va. (AP) — The computer network of a data center in Chester went dark after a switch failure.

The Richmond Times-Dispatch (http://bit.ly/20v8U5T) reports that Saturday's outage at the Commonwealth Enterprise Solutions Center affected access to the network by almost every executive branch agency the center serves, including the Department of Motor Vehicles.

Email, cellphones and agency computer servers in the center went dark, causing outage for inbound and outbound ... the DMV ...

V...

- The datacenter network went dark after a switch failure
- Almost every executive branch agency are affected for a few hours

Hardware Failure of Switches

- "An event that occurs when the switch is not functioning for forwarding traffic because of hardware errors" (P. Gill, et al., Microsoft, SIGCOMM' 11)

Hardware Failure of Switches

Observable

(F. Salfner, ACM Comput. Surv. 2010)

- A human
- A server
- Another network device
- If not result in incorrect output, it is not failure

Failure tickets

- Regular expression match with syslogs
- Feedback by Internet services
- Monitoring results of interfaces

Previous Proposed Solutions

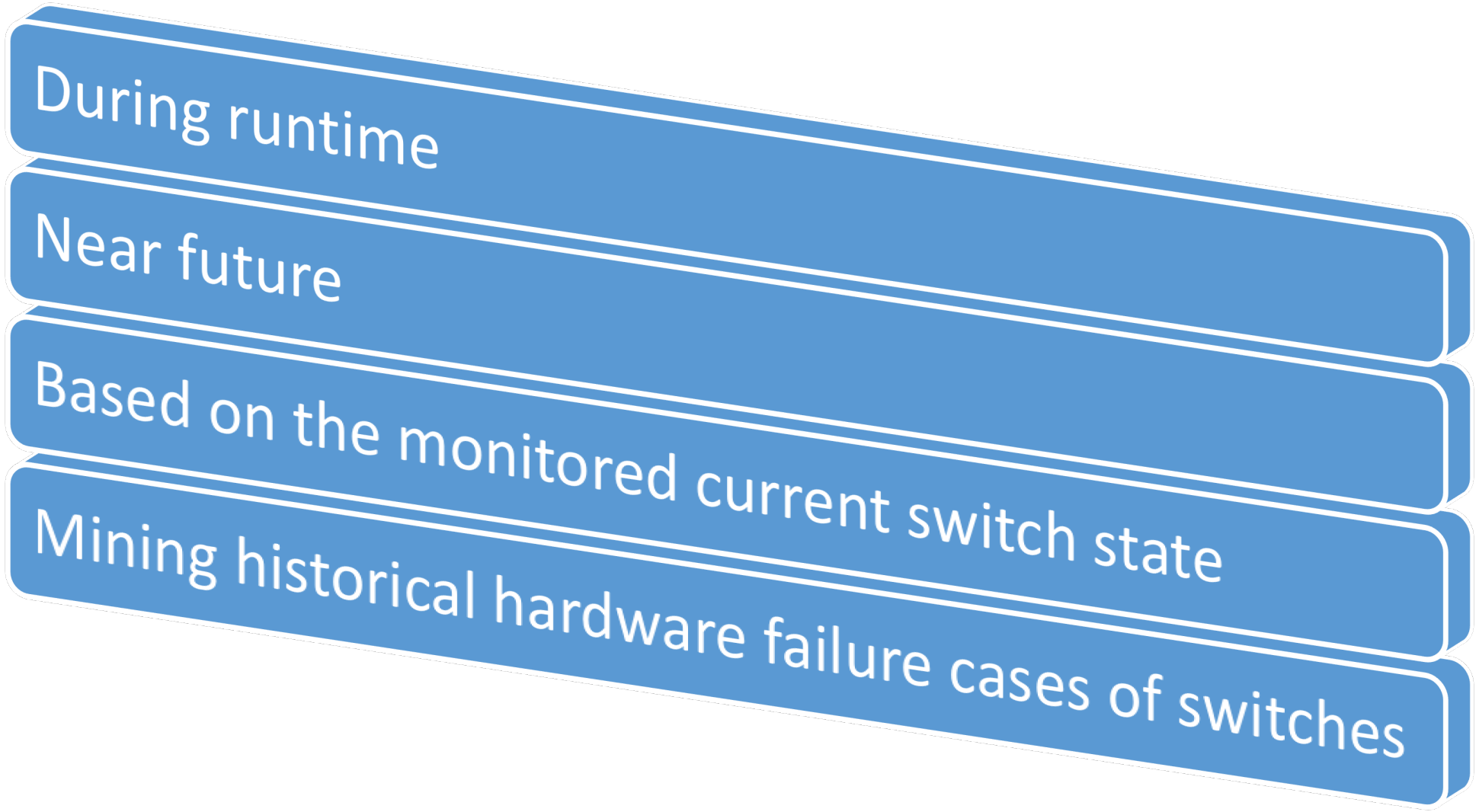
Change the protocols and network topologies

- Aim to automatically failover
- ToR switches do not have hot backups

Locate and diagnose failed switches

- Face deployment challenges
- Take time to locate and fix the failed switches
- Drop packets silently

Hardware Failure Prediction for Switches



Hardware Failure Prediction for Switches Based on Syslogs

- Sep 8 15:44:30 192.168.191.85 192.168.191.85 : [SIF]Interface ae3, changed state to down
- Sep 8 15:45:51 192.168.191.85 192.168.191.85 : [SIF]Vlan-interface vlan22, changed state to down
- Sep 8 15:46:59 192.168.191.85 192.168.191.85 : [SIF]Interface ae3, changed state to up
- Sep 8 15:47:21 192.168.191.85 192.168.191.85 : [SIF]Vlan-interface vlan22, changed state to up
- Sep 8 15:48:30 192.168.191.85 192.168.191.85 : [OSPF]Neighbour(rid:10.231.0.42, addr:10.231.38.85) on vlan22, changed state from Full to Down
- Sep 8 15:49:35 192.168.191.85 192.168.191.85 : [SIF]Interface ae3, changed state to down
- Sep 8 15:49:45 192.168.191.85 192.168.191.85 : [SIF]Vlan-interface vlan22, changed state to down
- Sep 8 15:50:42 192.168.191.85 192.168.191.85 : [SIF]Interface ae3, changed state to up
- Sep 8 15:50:59 192.168.191.85 192.168.191.85 : [SIF]Vlan-interface vlan22, changed state to up
- Sep 8 15:51:22 192.168.191.85 192.168.191.85 : [OSPF]A single neighbour should be configured
- Sep 8 15:51:52 192.168.191.85 192.168.191.85 : [OSPF]A single neighbour should be configured
- Sep 8 15:52:46 192.168.191.85 192.168.191.85 : [SIF]Interface ae1, changed state to down
- Sep 8 15:53:24 192.168.191.85 192.168.191.85 : [SIF]Vlan-interface vlan20, changed state to down
- Sep 8 15:54:31 192.168.191.85 192.168.191.85 : [OSPF]Neighbour(rid:10.231.0.40, addr:10.231.36.85) on vlan20, changed state from Full to Down
- Sep 8 15:55:12 192.168.191.85 192.168.191.85 : [SIF]Interface ae1, changed state to up
- Sep 8 15:56:47 192.168.191.85 192.168.191.85 : [SIF]Vlan-interface vlan20, changed state to up
- Sep 8 15:59:01 192.168.191.85 192.168.191.85 : [OSPF]A single neighbour should be configured
- Sep 8 16:31:20 whole machine failure (labelled by the operators)

Challenges

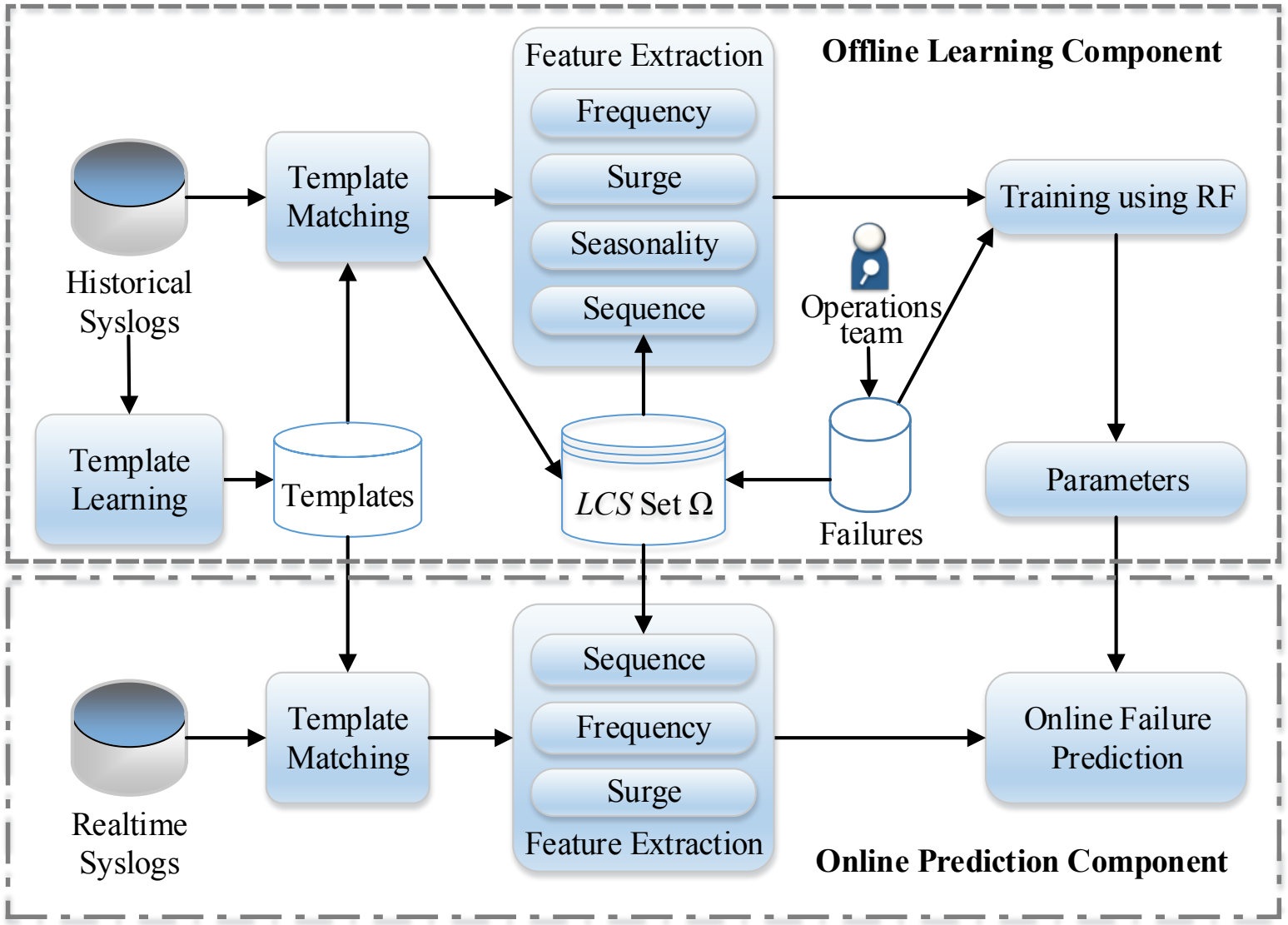
Noisy signals in syslog data

- Syslogs are highly diverse
  - Across several geographical locations, network layers, protocols, services
  - Normal login events of operators
  - Interface up/downs
  - Fail to send/receive packets
- Rarely contain failure omens

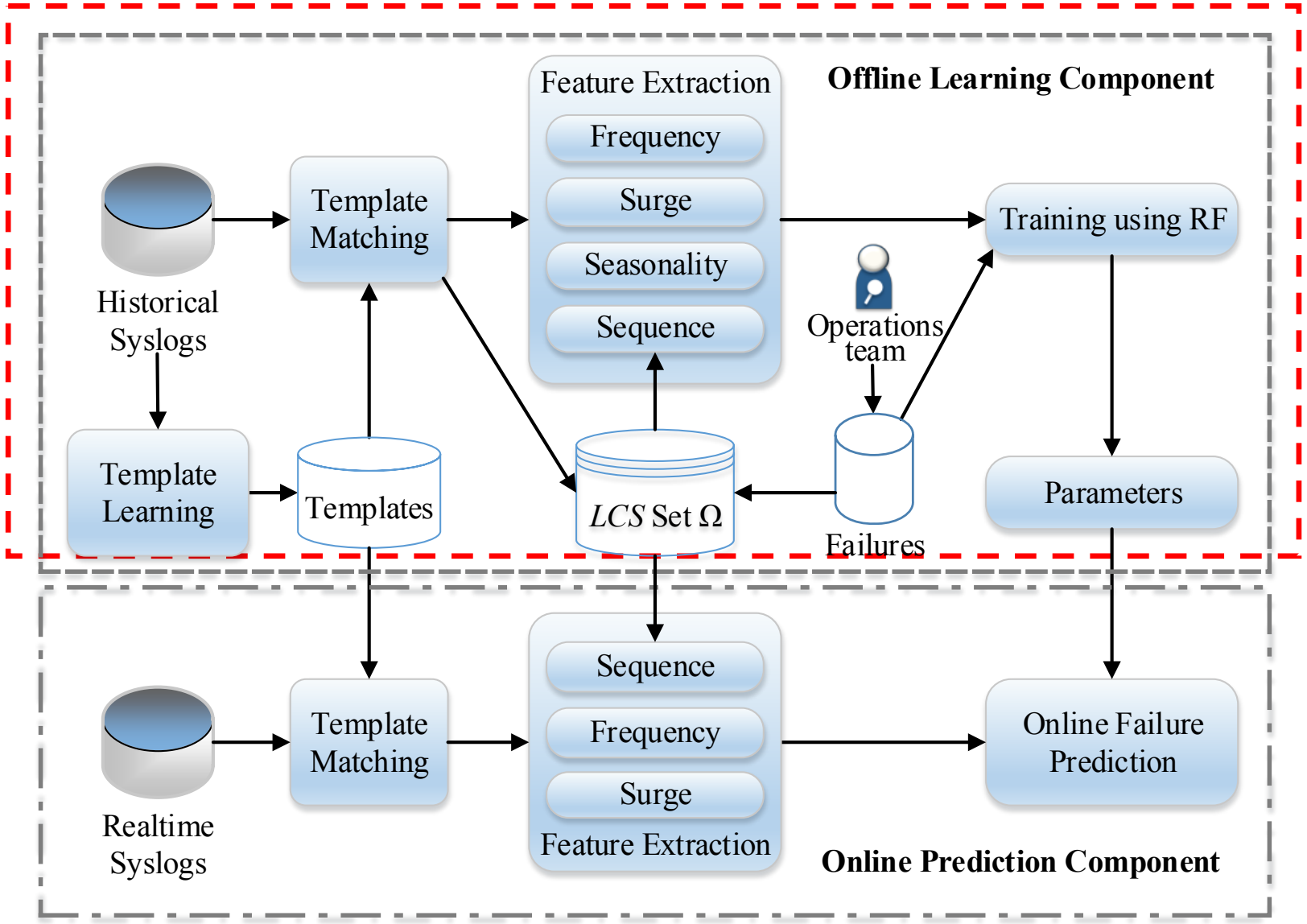
Sample imbalance

- Low failure possibility for a single switch
- Failure omen time bins: failure non-omen time bins =1:72500
- Low false alarms and high recall at the same time

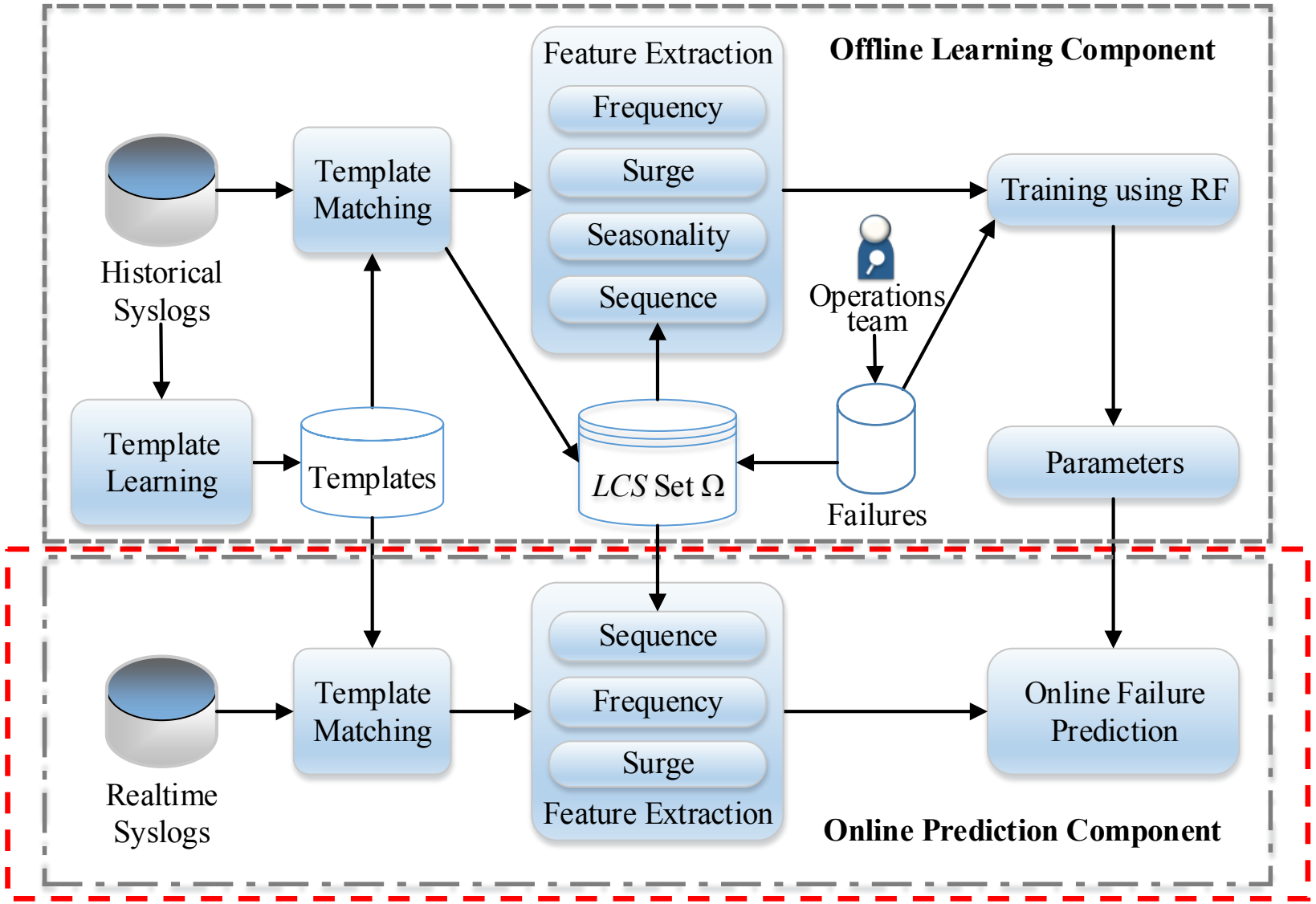
Design Overview



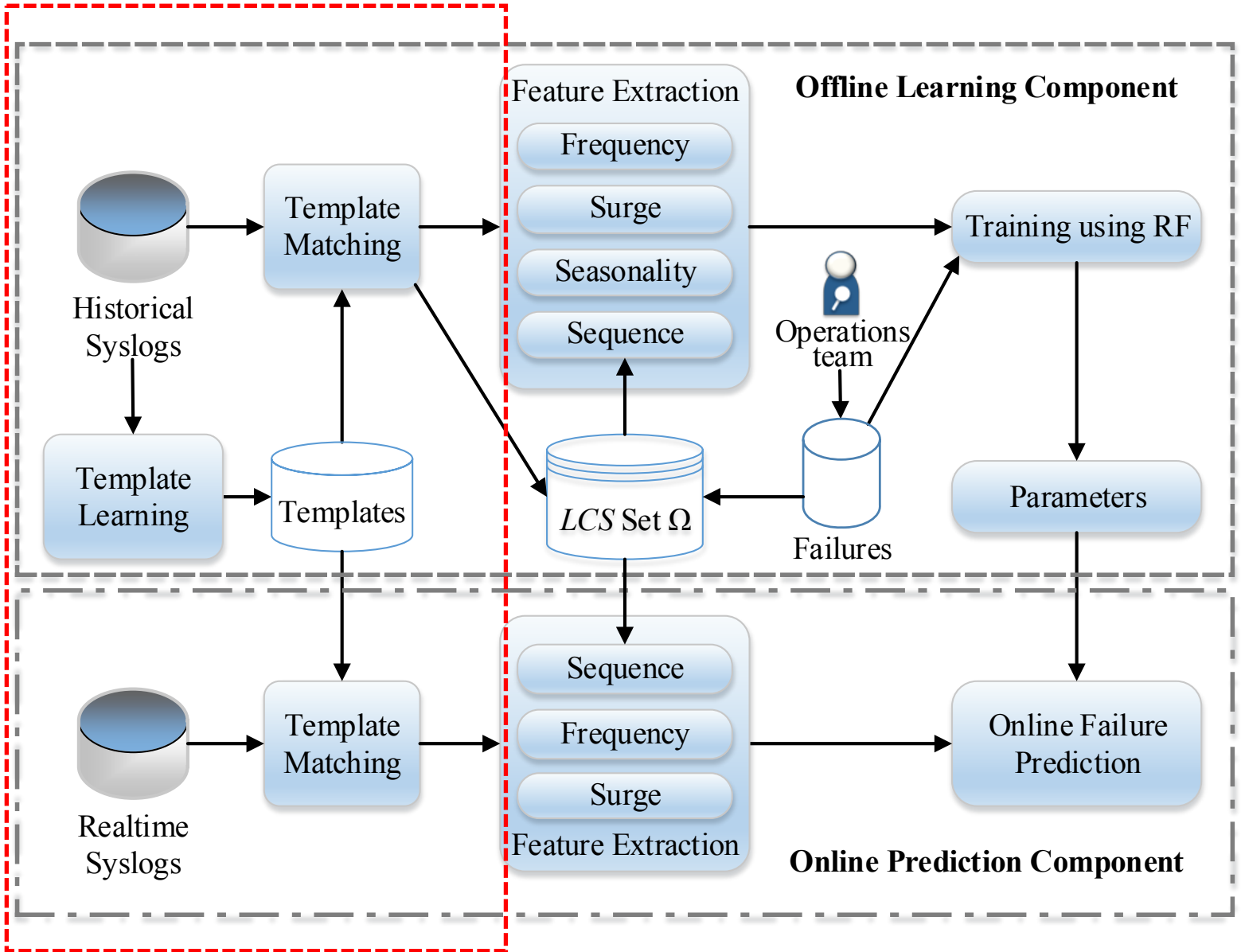
Design Overview



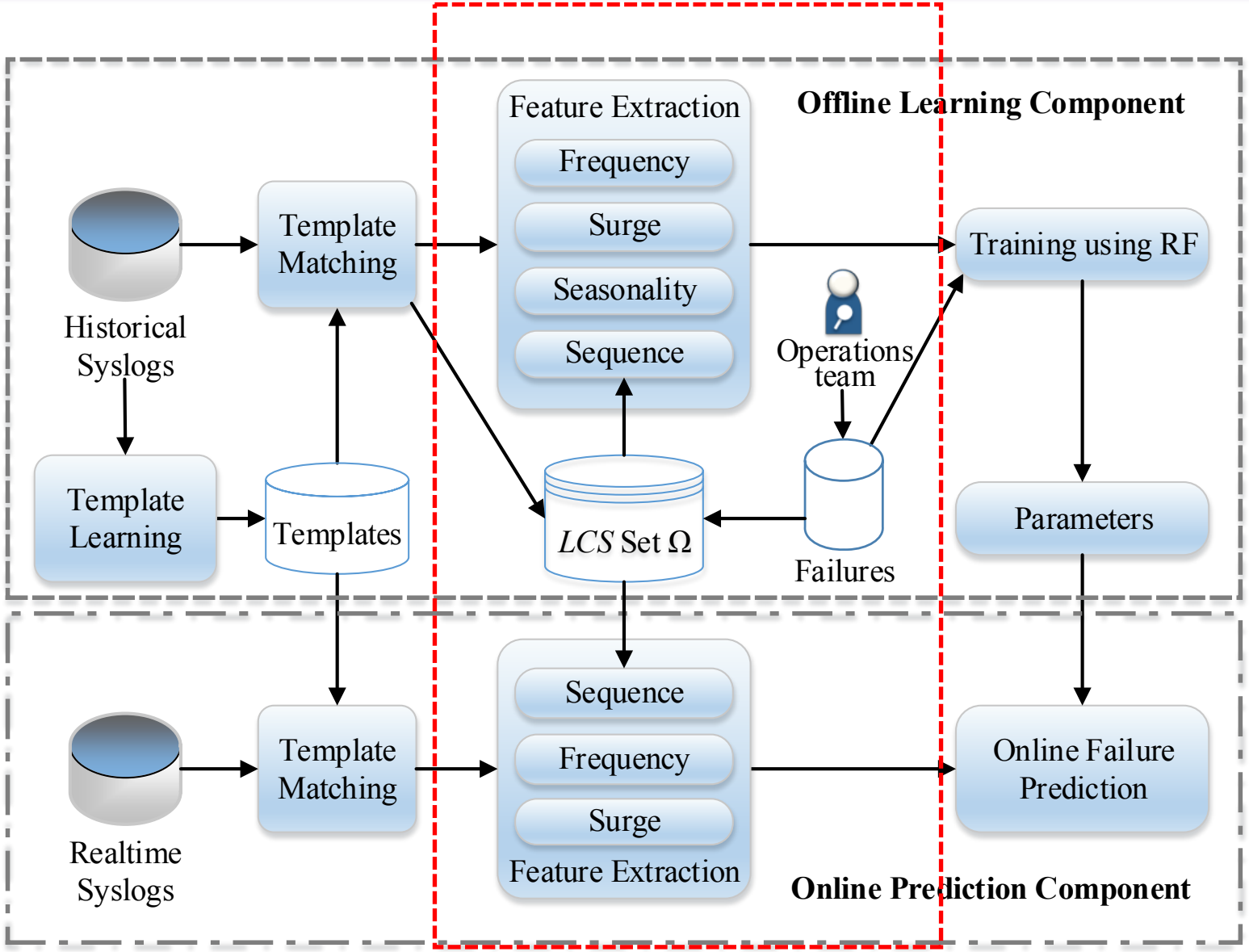
Design Overview



Design Overview



Design Overview



Feature Extraction

Sequence

Several failures share common syslog sequences

Surge

Some syslogs are indicative of failures when they occur in a sudden burst
E.g., interface up/down

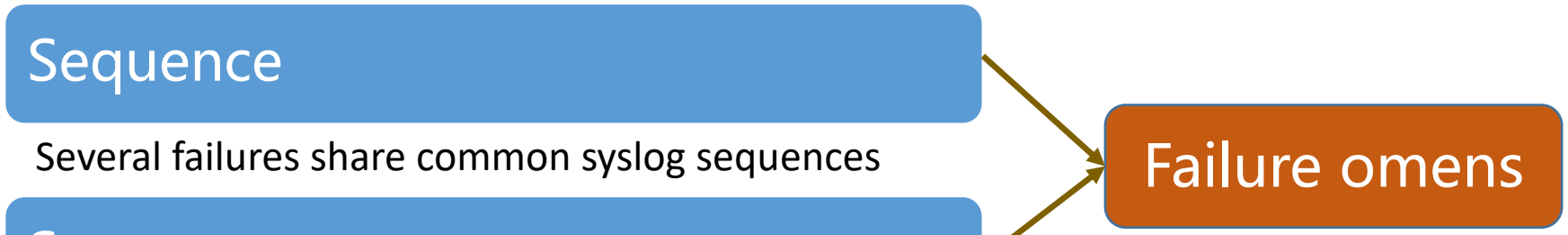
Frequency

Frequent syslogs can be ignored
E.g., package loss ratio of PING session

Seasonality

Some syslogs are periodic and irrelevant to failures
E.g., daily maintenance operations

Feature Extraction



Several failures share common syslog sequences

Some syslogs are indicative of failures when they occur in a sudden burst
E.g., interface up/down

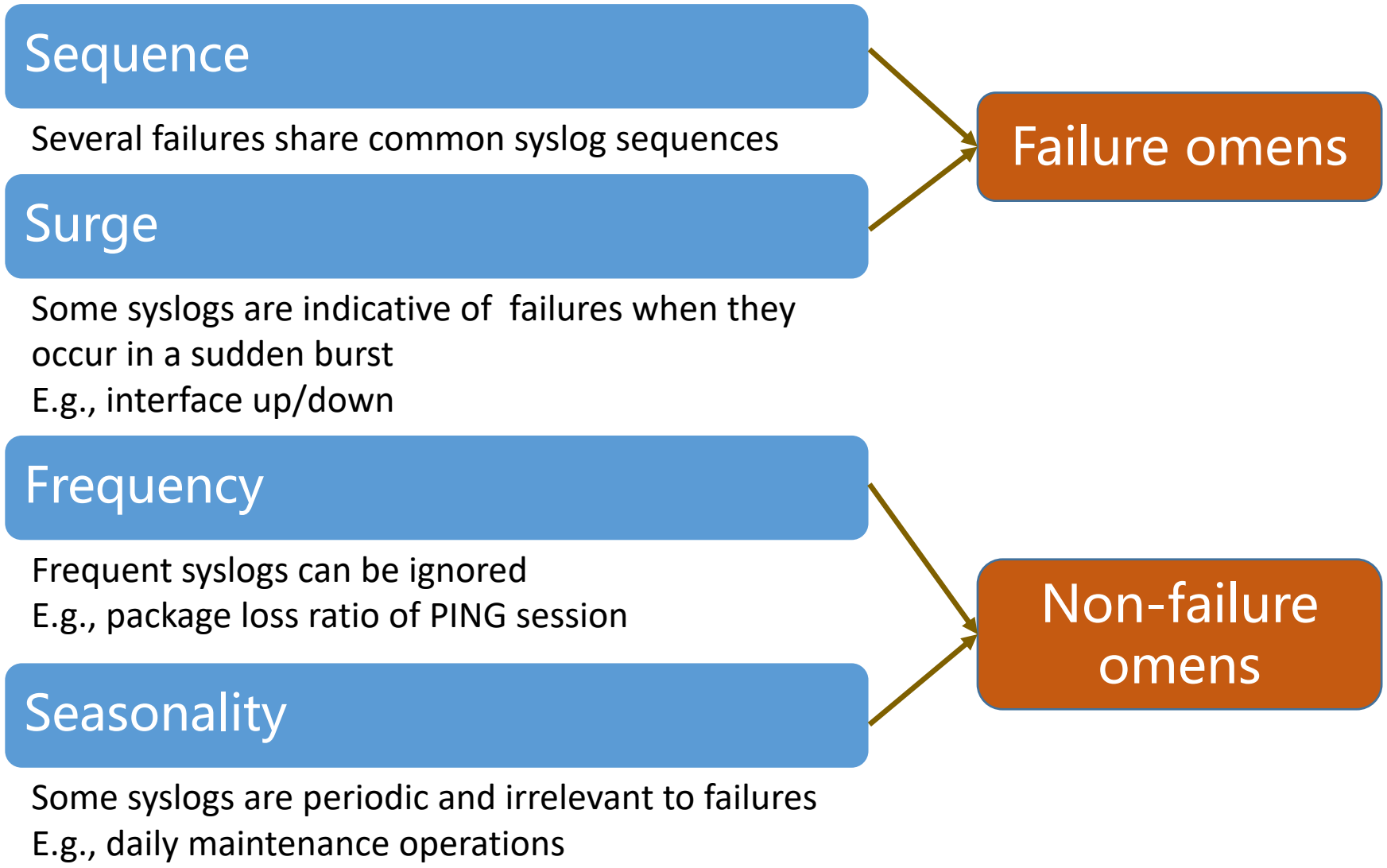
Frequency

Frequent syslogs can be ignored
E.g., package loss ratio of PING session

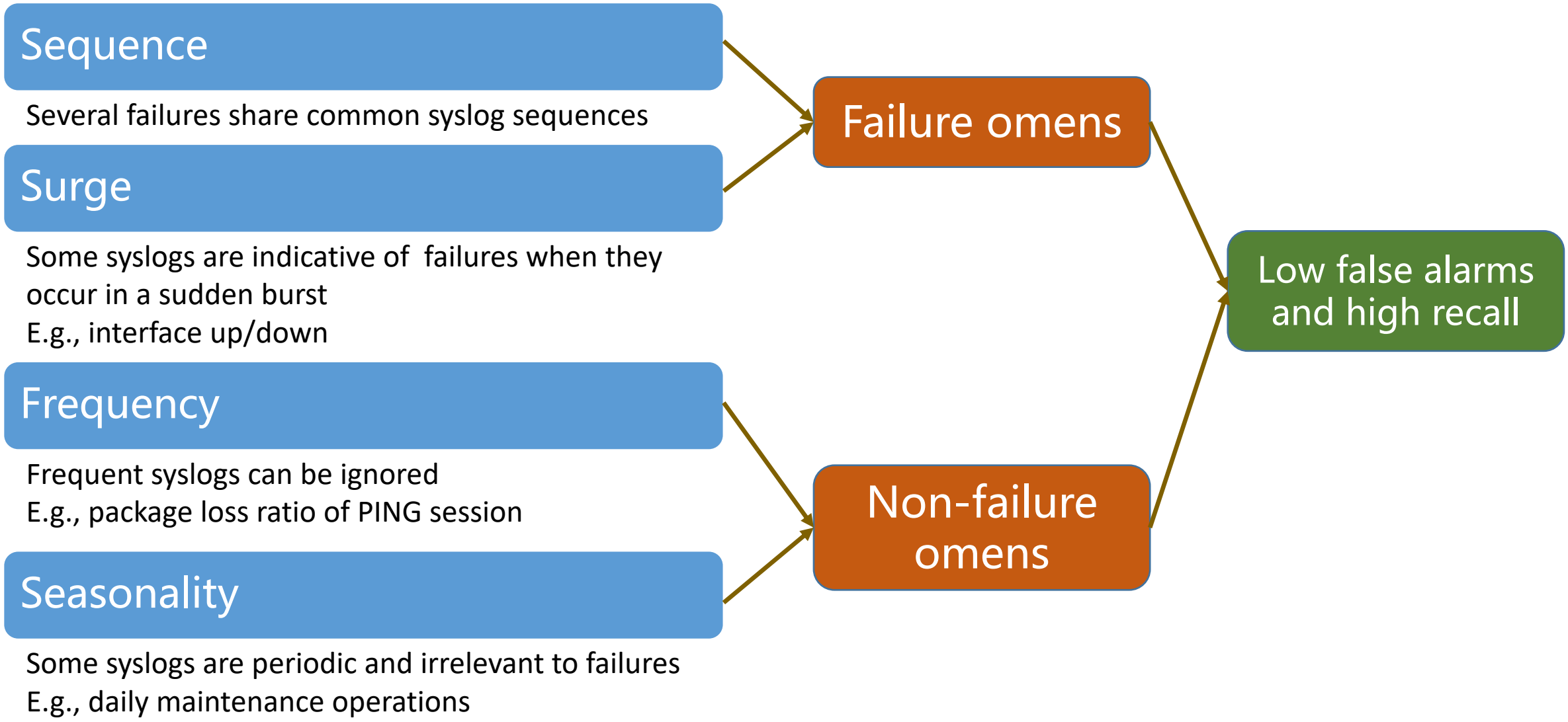
Seasonality

Some syslogs are periodic and irrelevant to failures
E.g., daily maintenance operations

Feature Extraction



Feature Extraction



Feature Extraction

Sequence

Several failures share common syslog sequences

Surge

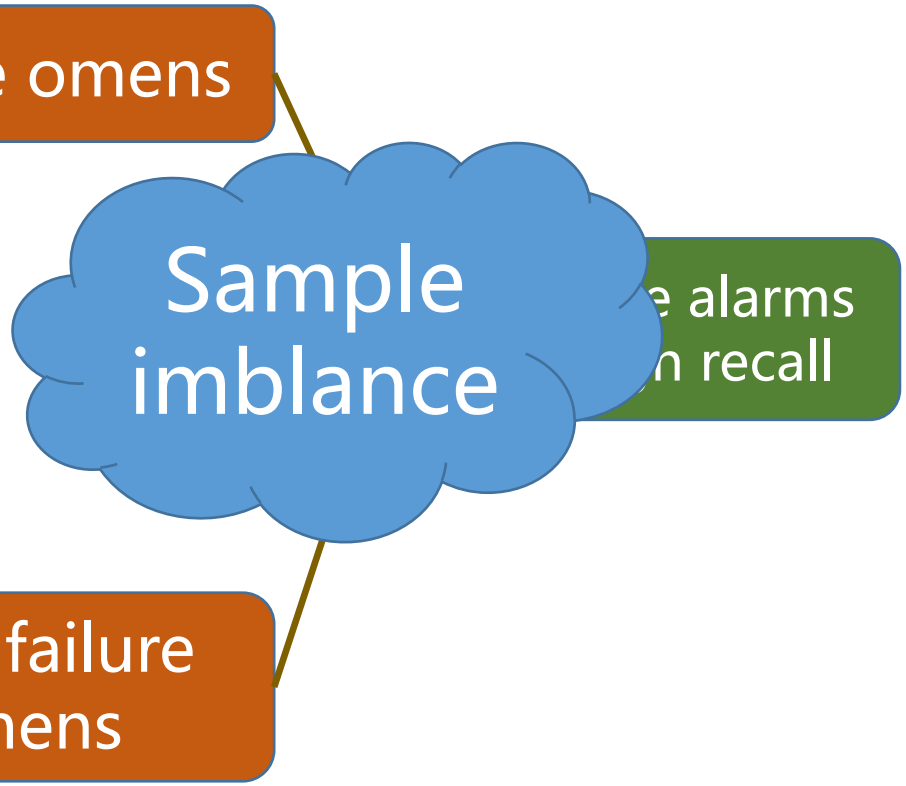
Some syslogs are indicative of failures when they occur in a sudden burst
E.g., interface up/down

Frequency

Frequent syslogs can be ignored
E.g., package loss ratio of PING session

Seasonality

Some syslogs are periodic and irrelevant to failures
E.g., daily maintenance operations



Syslogs Before a Failure (Within 2 Hours)

- Sep 8 15:44:30 192.168.191.85 192.168.191.85 : [SIF]Interface ae3, changed state to down
- Sep 8 15:45:51 192.168.191.85 192.168.191.85 : [SIF]Vlan-interface vlan22, changed state to down
- Sep 8 15:46:59 192.168.191.85 192.168.191.85 : [SIF]Interface ae3, changed state to up
- Sep 8 15:47:21 192.168.191.85 192.168.191.85 : [SIF]Vlan-interface vlan22, changed state to up
- Sep 8 15:48:30 192.168.191.85 192.168.191.85 : [OSPF]Neighbour(rid:10.231.0.42, addr:10.231.38.85) on vlan22, changed state from Full to Down
- Sep 8 15:49:35 192.168.191.85 192.168.191.85 : [SIF]Interface ae3, changed state to down
- Sep 8 15:49:45 192.168.191.85 192.168.191.85 : [SIF]Vlan-interface vlan22, changed state to down
- Sep 8 15:50:42 192.168.191.85 192.168.191.85 : [SIF]Interface ae3, changed state to up
- Sep 8 15:50:59 192.168.191.85 192.168.191.85 : [SIF]Vlan-interface vlan22, changed state to up
- Sep 8 15:51:22 192.168.191.85 192.168.191.85 : [OSPF]A single neighbour should be configured
- Sep 8 15:51:52 192.168.191.85 192.168.191.85 : [OSPF]A single neighbour should be configured
- Sep 8 15:52:46 192.168.191.85 192.168.191.85 : [SIF]Interface ae1, changed state to down
- Sep 8 15:53:24 192.168.191.85 192.168.191.85 : [SIF]Vlan-interface vlan20, changed state to down
- Sep 8 15:54:31 192.168.191.85 192.168.191.85 : [OSPF]Neighbour(rid:10.231.0.40, addr:10.231.36.85) on vlan20, changed state from Full to Down
- Sep 8 15:55:12 192.168.191.85 192.168.191.85 : [SIF]Interface ae1, changed state to up
- Sep 8 15:56:47 192.168.191.85 192.168.191.85 : [SIF]Vlan-interface vlan20, changed state to up
- Sep 8 15:59:01 192.168.191.85 192.168.191.85 : [OSPF]A single neighbour should be configured
- Sep 8 16:31:20 whole machine failure (labelled by the operators)

Transfer to Template Tag Sequence

- The syslogs before failure 1 (2h)
  - 48 49 46 47 63 48 49 46 47 62 62 48 49 63 46 47 62

Transfer to Template Tag Sequence

- The syslogs before failure 1 (2h)
  - 48 49 46 47 63 48 49 46 47 62 62 48 49 63 46 47 62
- The syslogs before failure 2 (2h)
  - 0 48 48 48 48 48 46 46 46 46 46 48 48 46 46 48 46 48 48 46 46 48 48 46 46 48 46 48 49 63 51 50 46 47 62 48 48 46 46 51 50 51 50 48 49 48 49 63 51 46 47 50 63 46 47 48 49 62 62 46 47 62 48 49 46 47 62 48 49 63 51 50 46 47 62 56 57 58 59 44

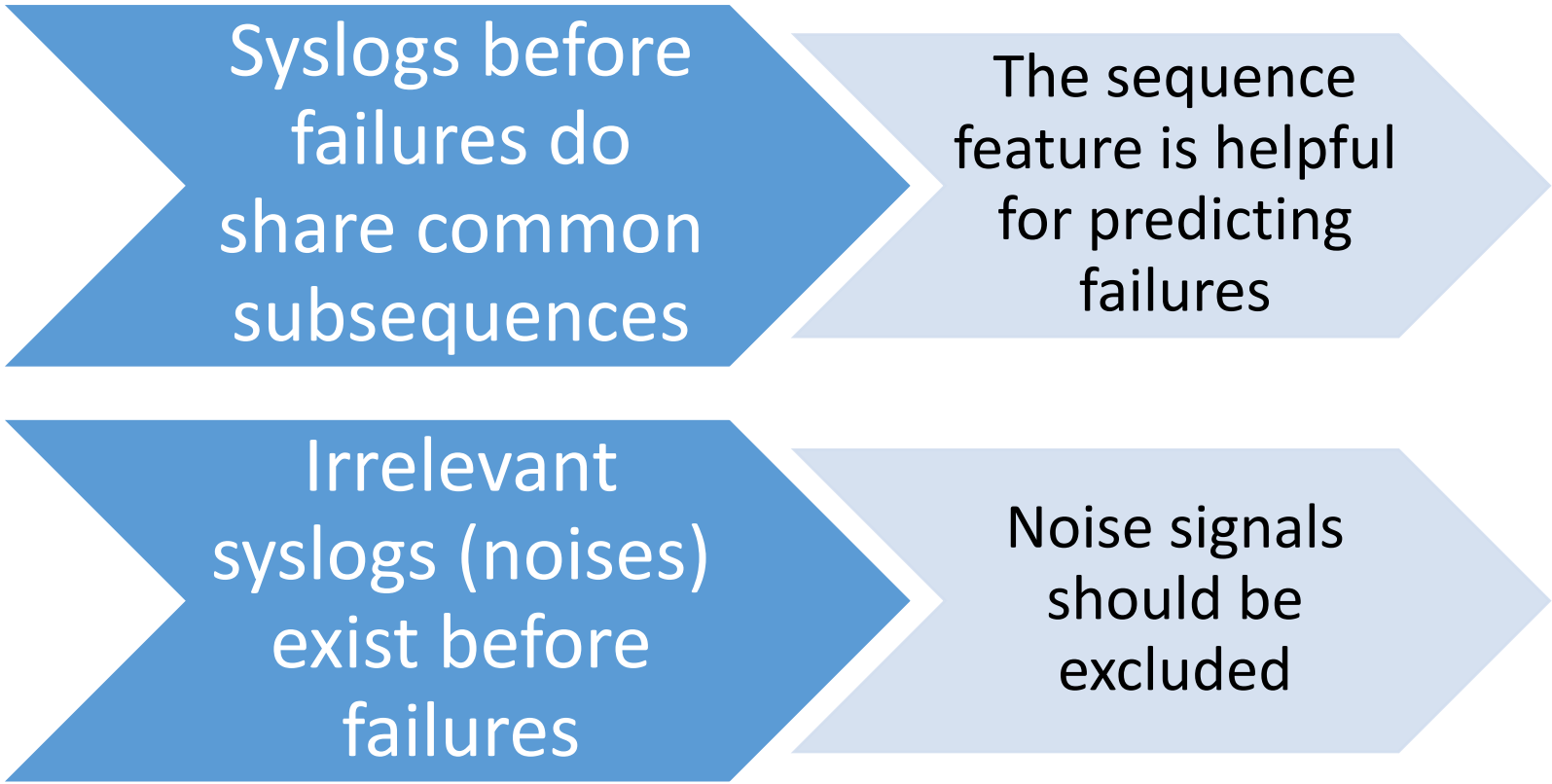
Transfer to Template Tag Sequence

- The syslogs before failure 1 (2h)
  - 48 49 46 47 63 48 49 46 47 62 62 48 49 63 46 47 62
- The syslogs before failure 2 (2h)
  - 0 48 48 48 48 48 46 46 46 46 46 48 48 46 46 48 46 48 48 46 46 48 48 46 46 48 46 48 49 63 51 50 46 47 62 48 48 46 46 51 50 51 50 48 49 48 49 63 51 46 47 50 63 46 47 48 49 62 62 46 47 62 48 49 46 47 62 48 49 63 51 50 46 47 62 56 57 58 59 44
- The syslogs before failure 3 (2h)
  - 48 48 49 49 63 63 46 46 47 47 62 62 56 56 57 57 58 58 59 59

Transfer to Template Tag Sequence

- The syslogs before failure 1 (2h)
  - 48 49 46 47 63 48 49 46 47 62 62 48 49 63 46 47 62
- The syslogs before failure 2 (2h)
  - 0 48 48 48 48 48 46 46 46 46 46 48 48 46 46 48 46 48 48 46 46 48 48 46 46 48 46 48 49 63 51 50 46 47 62 48 48 46 46 51 50 51 50 48 49 48 49 63 51 46 47 50 63 46 47 48 49 62 62 46 47 62 48 49 46 47 62 48 49 63 51 50 46 47 62 56 57 58 59 44
- The syslogs before failure 3 (2h)
  - 48 48 49 49 63 63 46 46 47 47 62 62 56 56 57 57 58 58 59 59
- The syslogs before failure 4 (2h)
  - 51 50 48 49 63 46 47 62 48 49 46 47 62 51 51 50 50 51 50 48 49 63 51 46 47 50 62 48 49 46 47 62 48 49 63 46 47 62 56 57 58 59 48 49 63 46 47 62 48 49 46 47 48 49 63 51 46 47 50 62 62

Insights of the above examples



The LCS^2 method

- LCS^2
  - First step: filter noises and get longest common subsequences (LCSes)
  - Second step: measure the similarity

The LCS^2 method

- LCS^2
  - First step: filter noises and get longest common subsequences (LCSes)
  - Second step: measure the similarity
- Filter noises and get LCSes
  - Seq 1: 48 49 46 47 63 48 49 46 47 62 62 48 49 63 46 47 62
  - Seq 2: 48 48 49 49 63 63 46 46 47 47 62 62 56 56 57 57 58 58 59 59
  - Seq 3: 50 62 48 49 46 47 62 48 49 63 46 47 62 56 57 58 59 48 49 63 46 47 62 48 49 46 47 48 49 63 51 46 47 50 62 62
  - Seq 1 ∩ Seq 2: 48 48 49 49 63 46 47 62
  - Seq 1 ∩ Seq 3: 48 49 46 47 63 48 49 46 47 62 48 49 63 46 47 62
  - Seq 2 ∩ Seq 3: 48 48 49 49 63 46 46 47 47 62 62

Ω

The LCS^2 method

- LCS^2
  - First step: filter noises and get longest common subsequences (LCSes)
  - Second step: measure the similarity
- Measure the similarity
  - Measure the similarity between current sequence and omen sequences
  - For each LCS_i in Ω
    - LCS_{ci} is the LCS between the current sequence and LCS_i
    - Calculate the ratio of the length of LCS_{ci} to that of LCS_i, R_i
    - Apply $\max(R_i)$ as the sequential feature score of the current sequence

The LCS^2 method

- LCS^2
  - First step: filter noises and get longest common subsequences (LCSes)
  - Second step: measure the similarity
- Advantages
  - Computationally efficient
  - Filter noises for sequences before failures

Evaluation Experimens

- Dataset
  - Three switch models
  - 9397 switches
  - 2-year period
  - 20+ data centers

Switch model	Method	Precision	Recall	$F1$	FPR
M1	PreFix	87.35%	74.36%	80.33%	2.49×10^{-5}
	SKSVM	8.25%	76.09%	14.89%	1.96×10^{-3}
	HSMM	32.27%	95.3%	48.21%	4.63×10^{-4}
M2	PreFix	59.79%	58.59%	59.18%	5.43×10^{-6}
	SKSVM	4.47%	8.72%	5.91%	2.57×10^{-5}
	HSMM	0.28%	60.58%	0.56%	2.94×10^{-3}
M3	PreFix	84.00%	52.50%	64.61%	2.48×10^{-5}
	SKSVM	0.79%	91.91%	1.58%	2.85×10^{-2}
	HSMM	26.32%	11.11%	15.63%	7.72×10^{-5}

Evaluation Experimens

- Dataset
  - Three switch models
  - 9397 switches
  - 2-year period
  - 20+ data centers

Switch model	Method	Precision	Recall	$F1$	FPR
M1	PreFix	87.35%	74.36%	80.33%	2.49×10^{-5}
	SKSVM	8.25%	76.09%	14.89%	1.96×10^{-3}
	HSMM	32.27%	95.3%	48.21%	4.63×10^{-4}
M2	PreFix	59.79%	58.59%	59.18%	5.43×10^{-6}
	SKSVM	4.47%	8.72%	5.91%	2.57×10^{-5}
					10^{-3}
					10^{-5}
					10^{-2}
					10^{-5}

Average recall: 61.81% , mean time between false alarms: 8494 days(23.3 years)

Conclusion

Challenges

- Noisy signals in syslog data
- Sample imblance

Four features

- Sequence, seasonality, surge and frequency
- LCS^2 method

Evaluation

- Real-world data

Future work

- Switch failure prediction across different switch models

Thank you!

Q&A

zhangsl@nankai.edu.cn