FUJITSU

shaping tomorrow with you

A Resource Contention Analysis Framework for Diagnosis of Application Performance Anomalies in Consolidated Cloud Environments

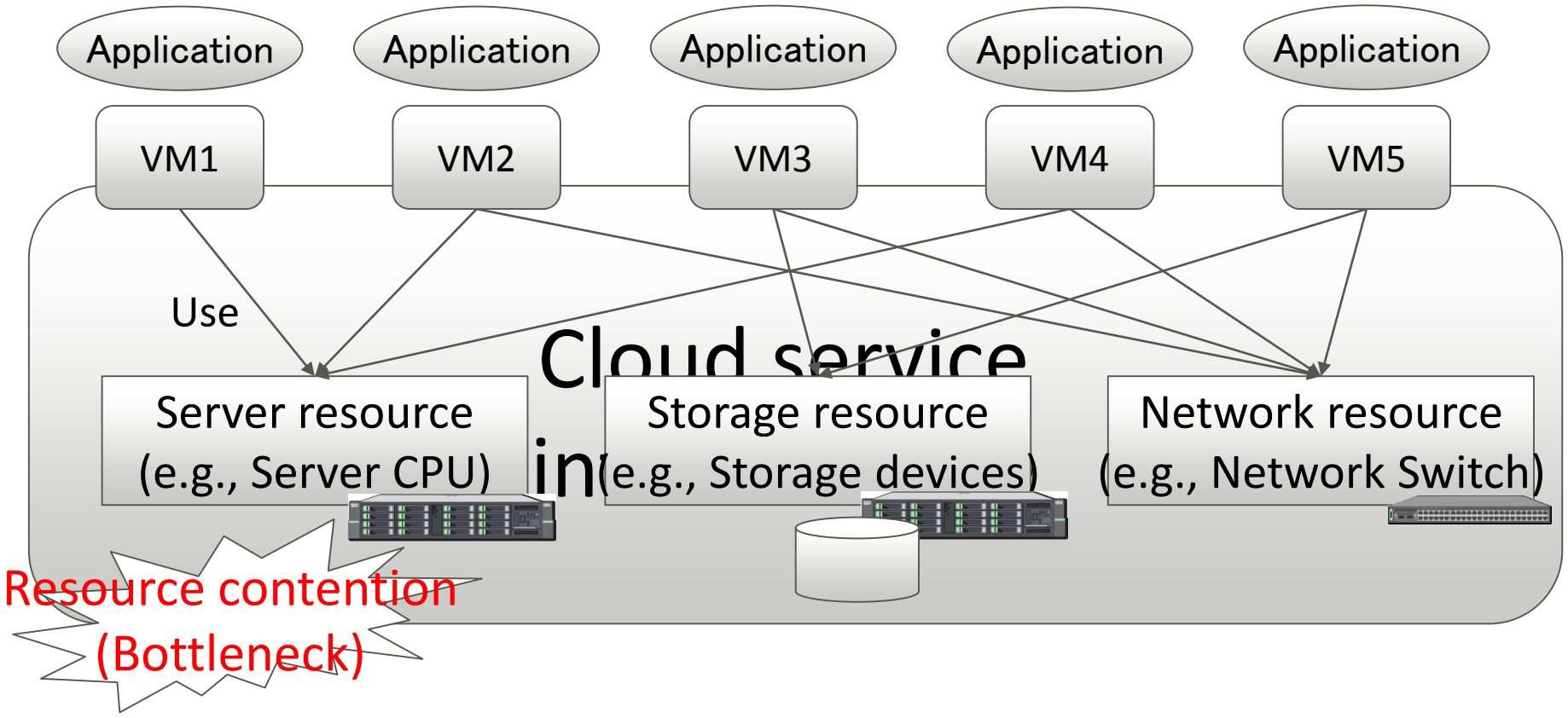
Tatsuma Matsuki, Naoki Matsuoka
Fujitsu Laboratories LTD.

ICPE 2016.3.12-16

Performance Problems in Cloud

FUJITSU

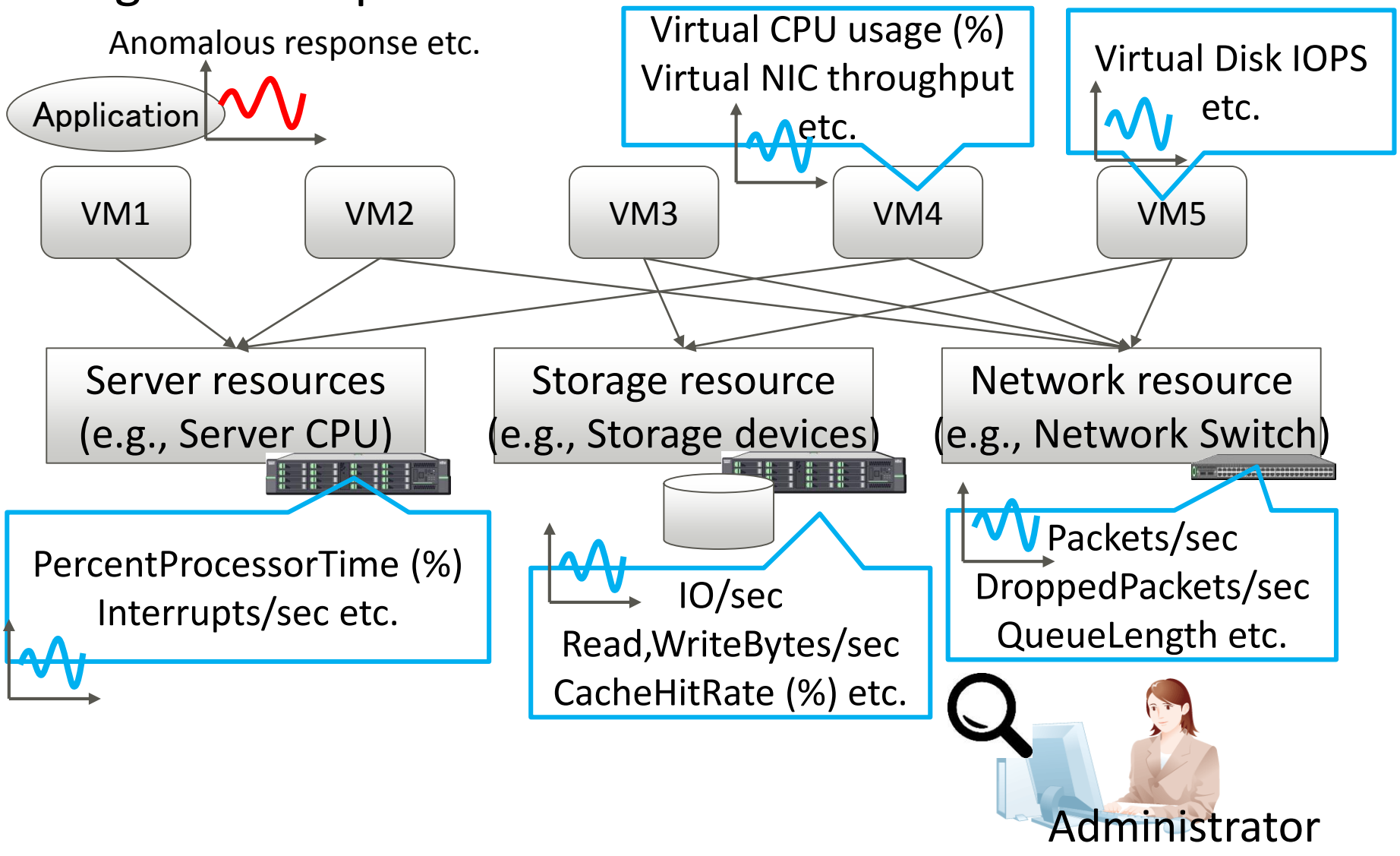
- Physical resources in cloud are shared within VMs and applications.
- Anomalous workloads of applications create bottlenecks in cloud



Performance Diagnosis

FUJITSU

- Diagnose the performance anomalies with metrics

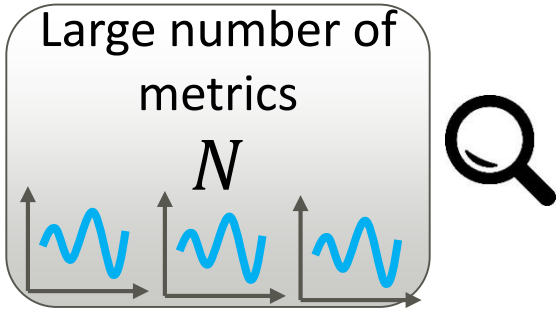


How much does the diagnosis cost?

FUJITSU

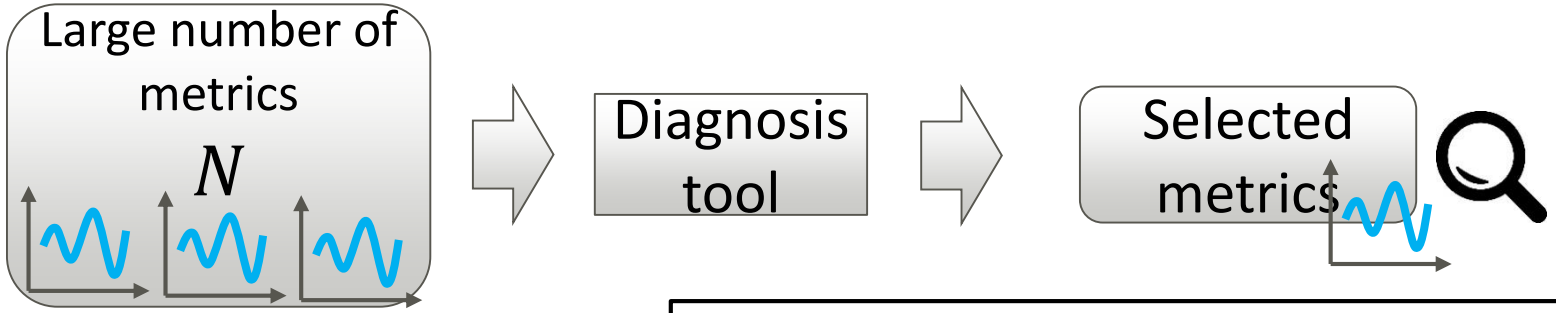
- Without any diagnosis tools
  - All available N metrics have to be investigated

Administrator



Diagnosis cost: $\mathrm{X} = \alpha N$

- With a diagnosis tool
  - Diagnosis tool can automatically select causal metrics

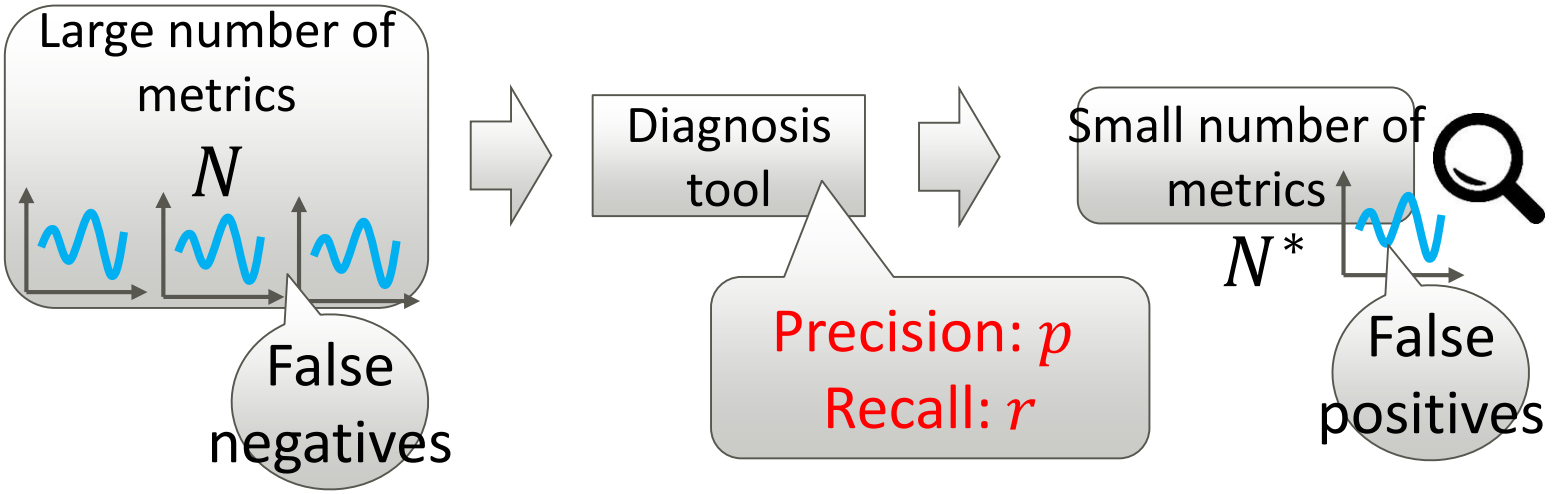


Diagnosis cost: $\mathrm{X}^* = 0$?

The Remaining Diagnosis Cost

FUJITSU

- A diagnosis tool is not perfect !
  - The diagnosis tool leaves some diagnosis costs for users.



- *The Remaining Diagnosis Cost* is estimated as:

$$X^* = \alpha N^*(1-p) + \alpha(N-N^*)(1-r)$$

Selected metrics include false positives with probability, $1-p$

Additional investigation is required with probability, $1-r$

The Remaining Diagnosis Cost

FUJITSU

- Transform the formula as follows

$$X^* = \alpha N^*(1-p) + \alpha(N-N^*)(1-r)$$

$$= X - \underbrace{\alpha N\left(\frac{d-1}{d}r + \frac{1}{d}p\right)}_{\text{Gain of the diagnosis tool}} \quad (N = dN^*, d \geq 1)$$

Recall is more important than *precision* !

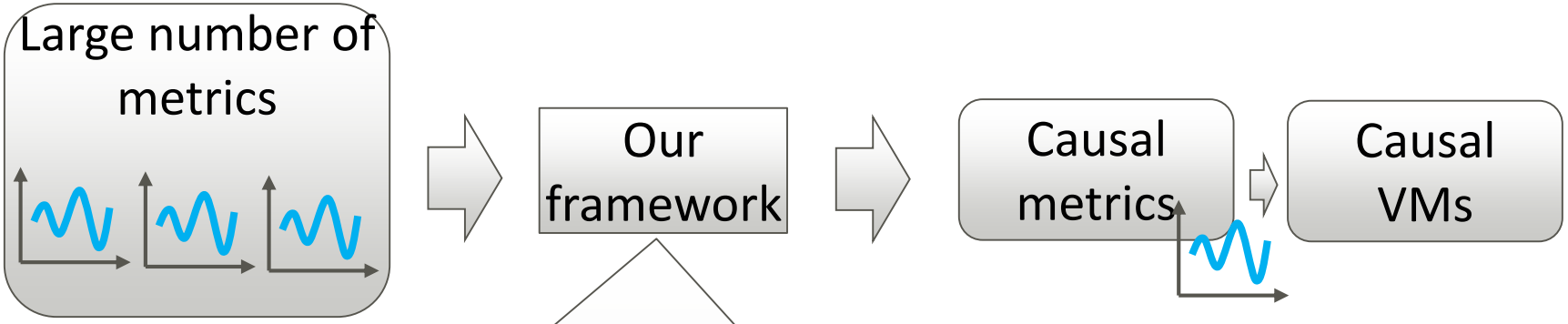
When $d \to \infty$, the gain is approaching $\alpha N r$
and p makes no contributions.

Our Framework Approach

FUJITSU

Objective of our framework

- Select causal metrics and VMs with better recall



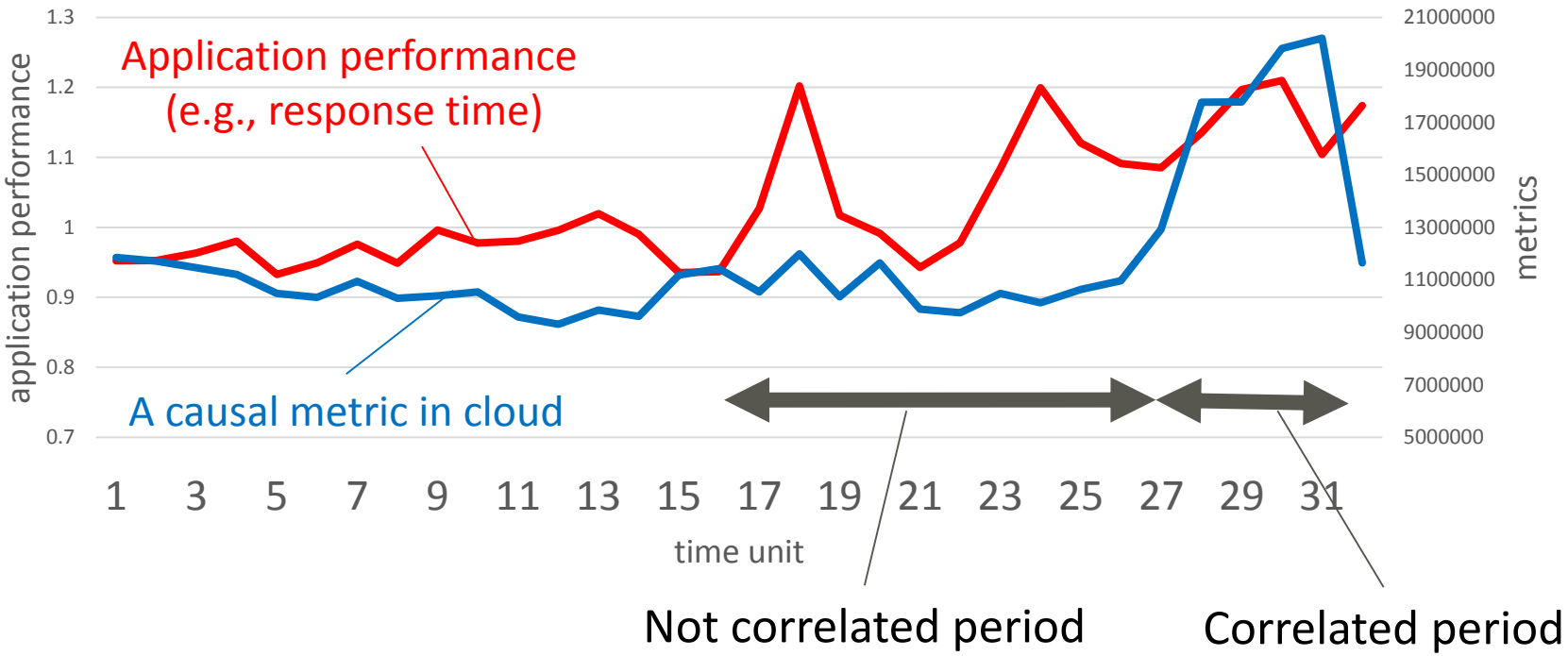
- Statistical correlation analysis
  - Evaluate Pearson correlation between time-series
  - Mining *Association Rules* in time-series data

    => reduce false negatives and achieve better recall!

Why the false negatives occur?

FUJITSU

- Temporary correlated metrics lead to false negatives



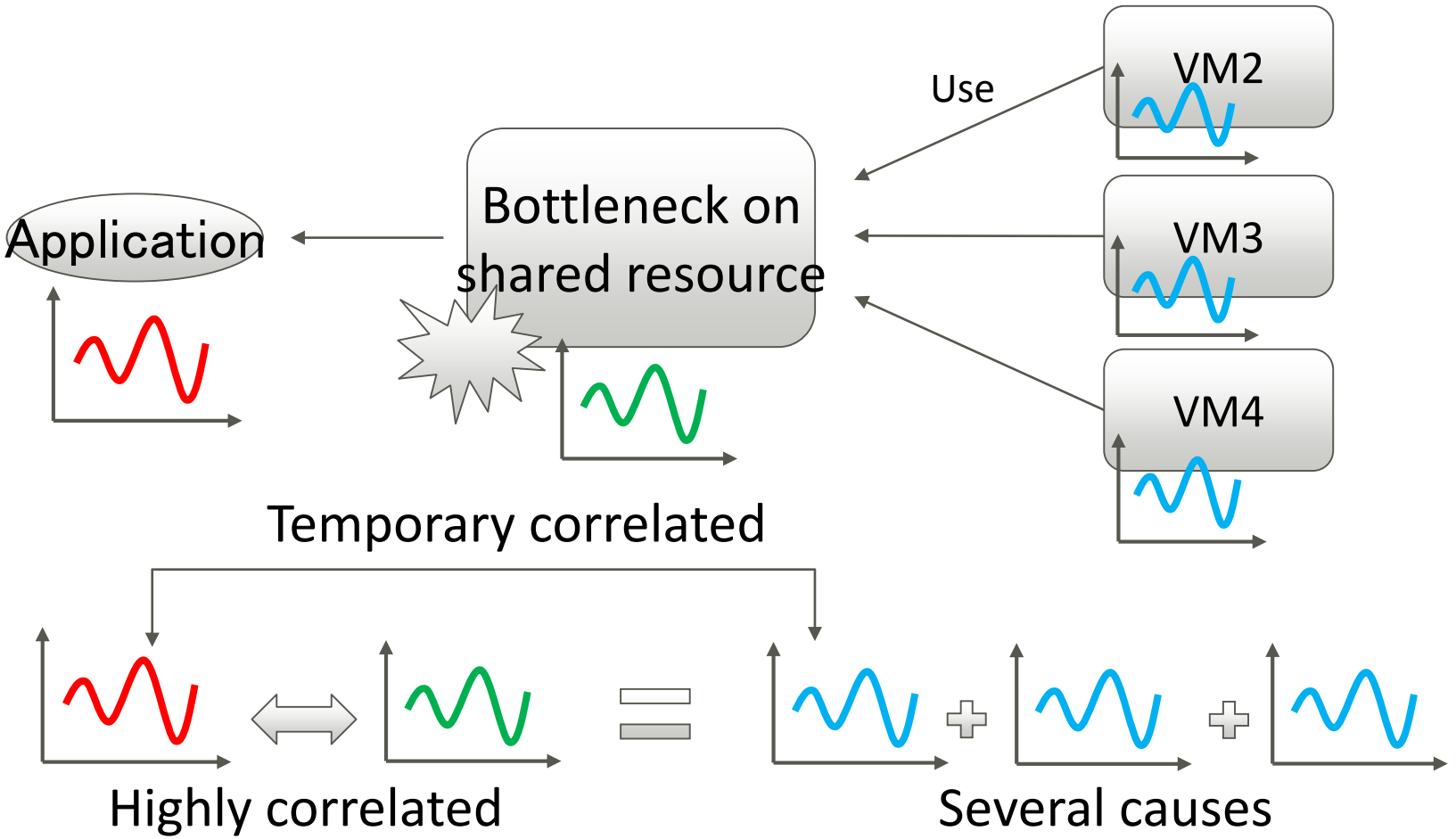
Statistical Pearson correlation is not sufficient!

Correlation coefficient is 0.559 in this case

Temporary Correlated Metrics

FUJITSU

Why does the temporal correlation occur?



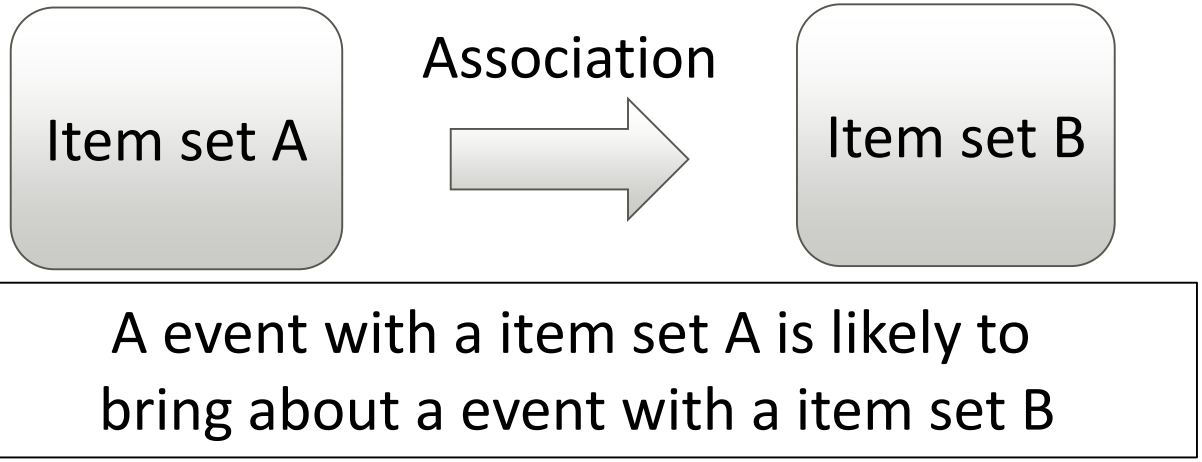
When several causes exist, the temporary correlations occur!

Association Rule Analysis

FUJITSU

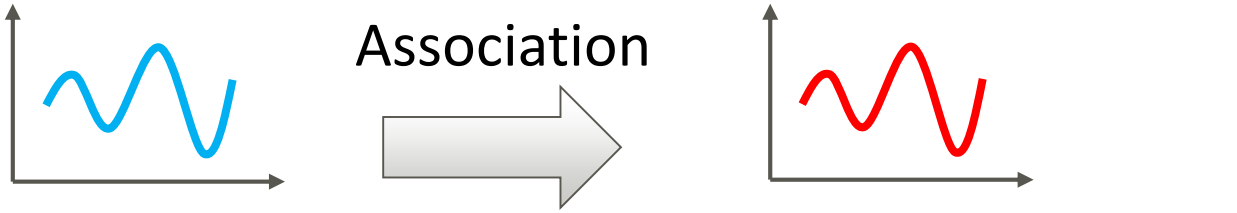
- Widely used in data-mining area
- Discovering relations between variables in databases

Original



A event with a item set A is likely to
bring about a event with a item set B

Our framework

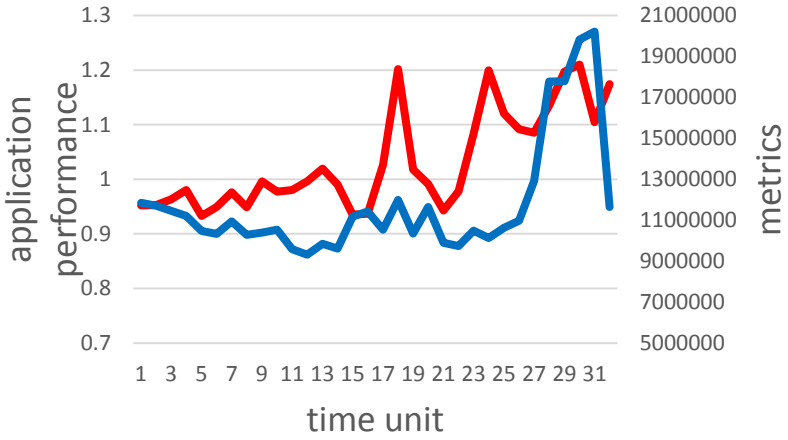


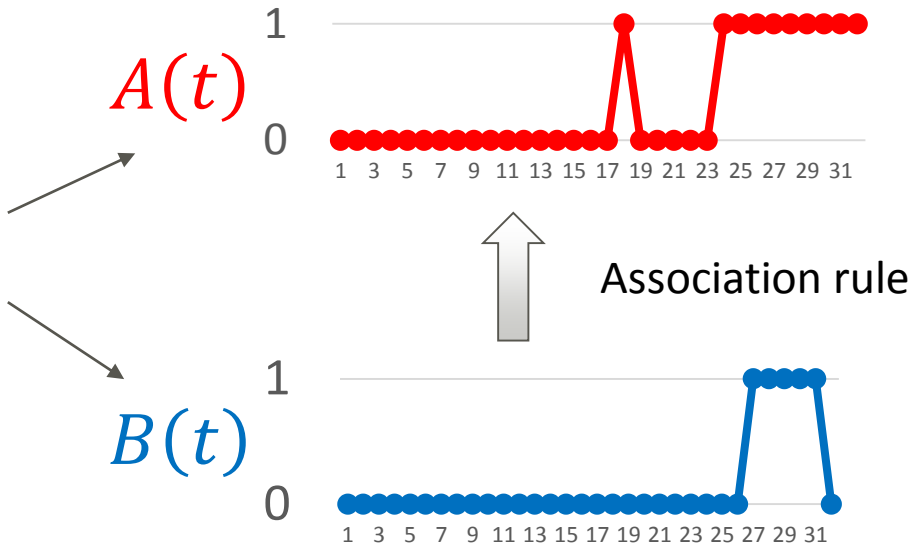
A anomalous metric behavior is likely to
bring about a anomalous application performance

Association Rule in time-series

FUJITSU

- Calculate the *confidence* and *support* measures





$$\text{confidence}(\boldsymbol{B} \rightarrow \boldsymbol{A}) = \frac{\sum_t A(t) \cdot B(t)}{\sum_t B(t)} = \frac{5}{5} = 1$$

A valid association rule exits?

$$\text{support}(\boldsymbol{B} \rightarrow \boldsymbol{A}) = \frac{\sum_t A(t) \cdot B(t)}{\sum_t A(t)} = \frac{5}{10} = 0.5$$

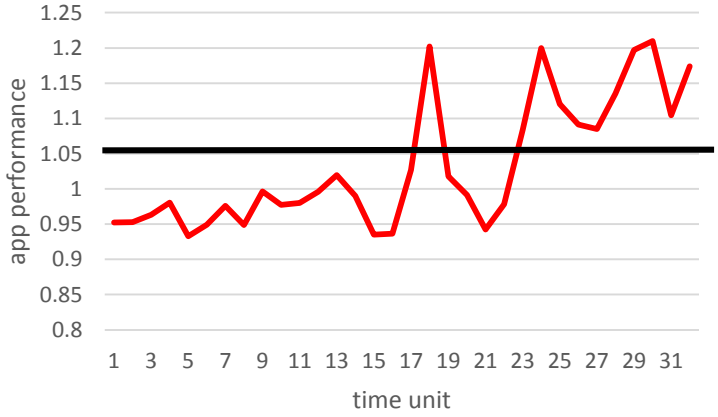
How strong the association?

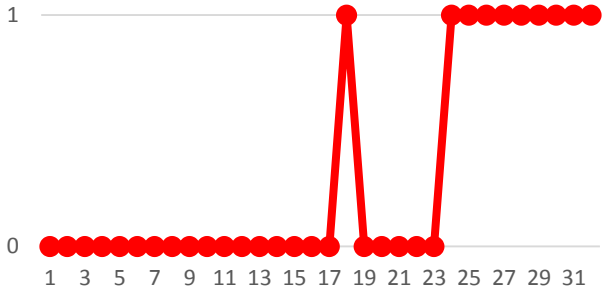
Association Rule Calculation

FUJITSU

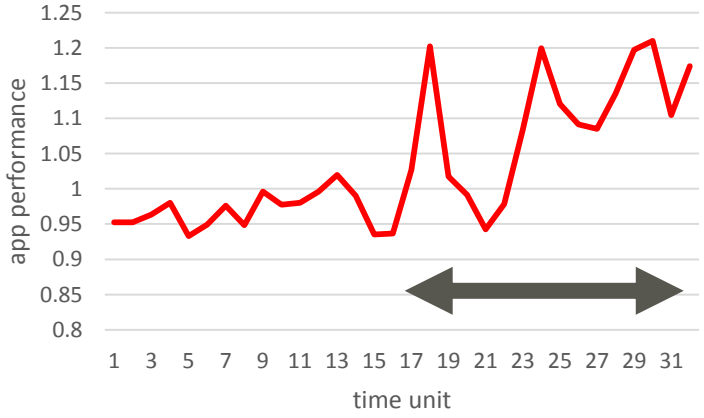
- Application performance discretizing

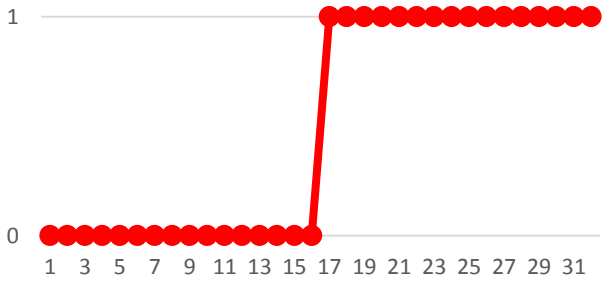
Example1: User-defined threshold





Example2: User-defined anomalous period



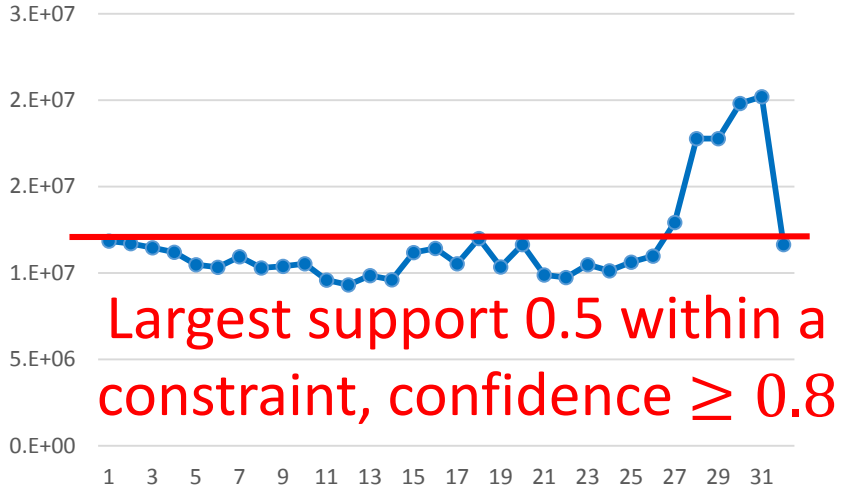


Association Rule Calculation

FUJITSU

Metrics discretizing

Investigate the optimal discretizing threshold for each metric



time unit t	$m(t)$
1	11851874
2	11708705
3	11464851
4	11207176
5	10479006
6	10336769
7	10941646
8	10293431
..	...
30	19815091
31	20206165
32	11647891

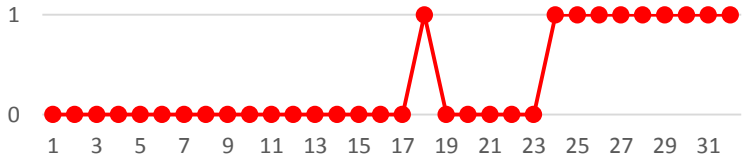
Candidates of the optimal threshold

Optimal threshold

$$= \arg\max_{m(t)} \text{support}\left(\boldsymbol{B}_{m(t)} \rightarrow \boldsymbol{A}\right) \quad s.t.\ \text{confidence}\left(\boldsymbol{B}_{m(t)} \rightarrow \boldsymbol{A}\right) \geq 0.8$$

Discretized time series with threshold $m(t)$

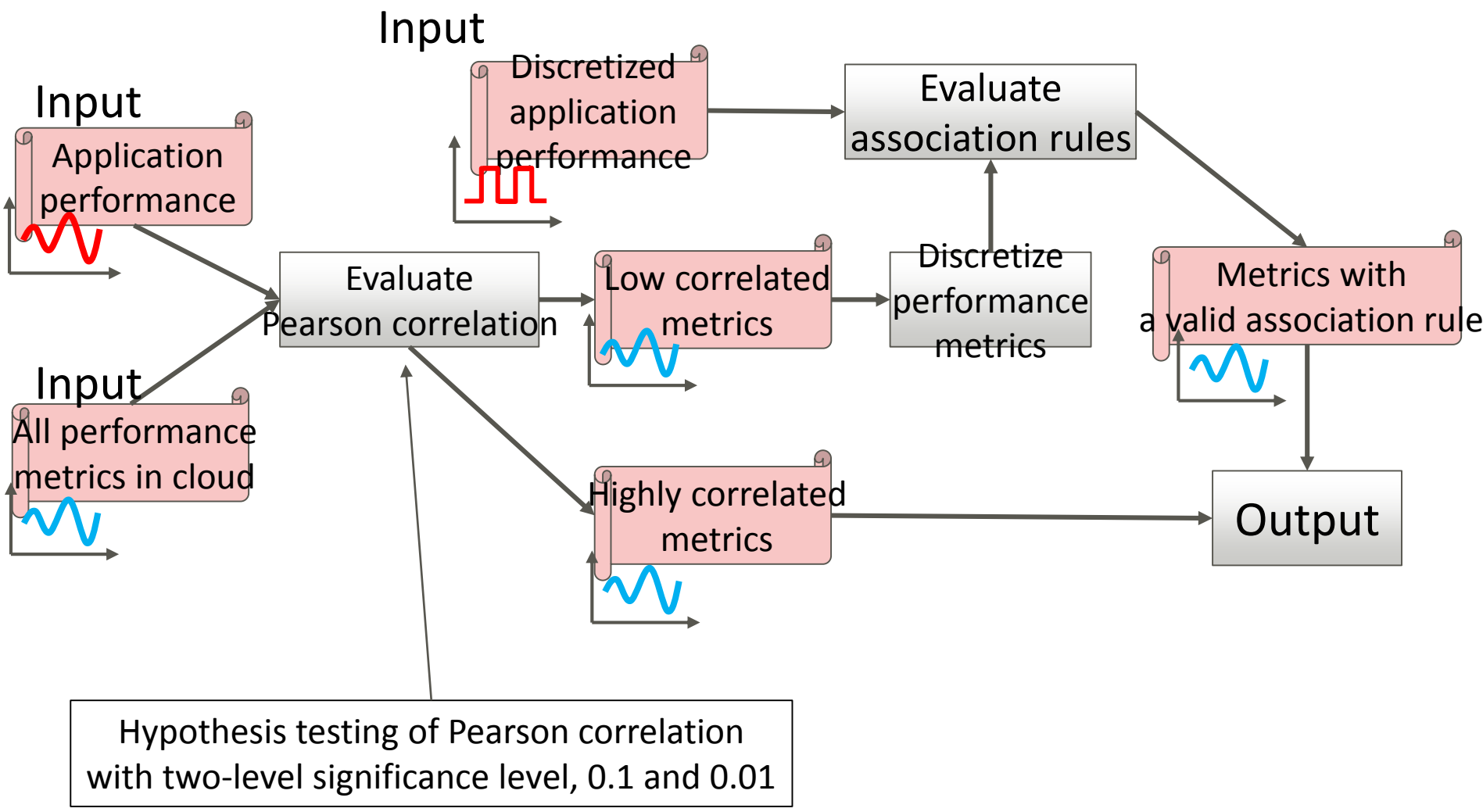
Given the discretized application performance



Our Framework Overview

FUJITSU

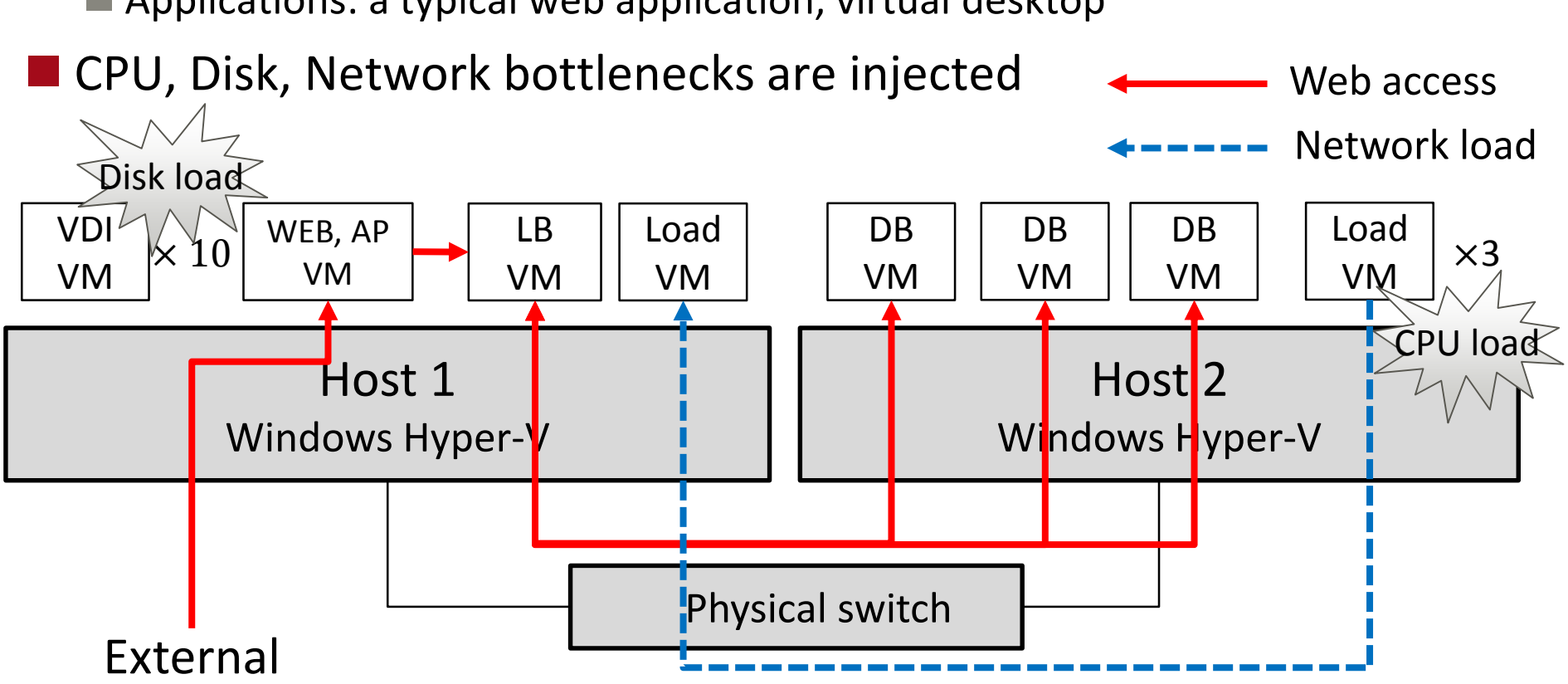
- Applying the association rule mining



Experimental Evaluation

FUJITSU

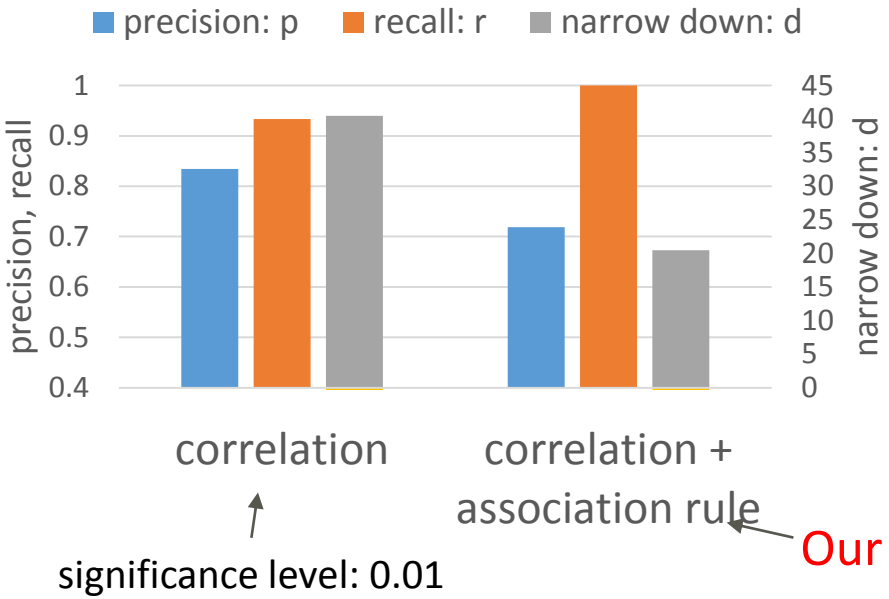
- Experimental setup
  - 2 host servers, a physical switch
  - Windows performance counters: 6,110 metrics
  - Applications: a typical web application, virtual desktop
- CPU, Disk, Network bottlenecks are injected



Experimental Results

FUJITSU

One of the main evaluation results

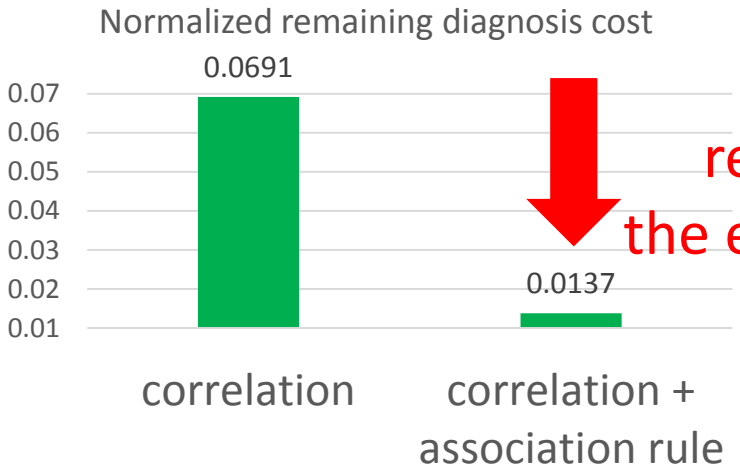


significance level: 0.01

Our framework

① Tradeoffs between precision and recall

② Larger number of metrics are selected by our framework



reduce the expected cost

① Quantify the tradeoffs

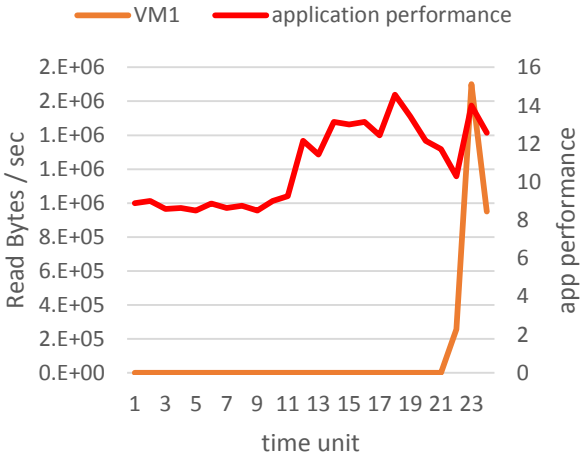
② We expected that our framework reduces the diagnosis cost by > 80% in this case.

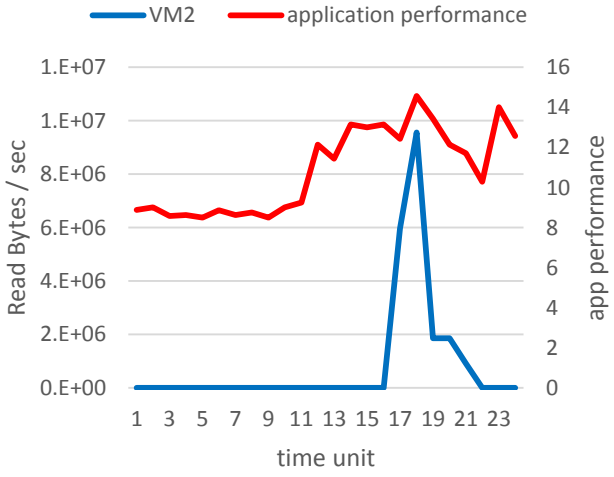
Testing on a Real Data Set

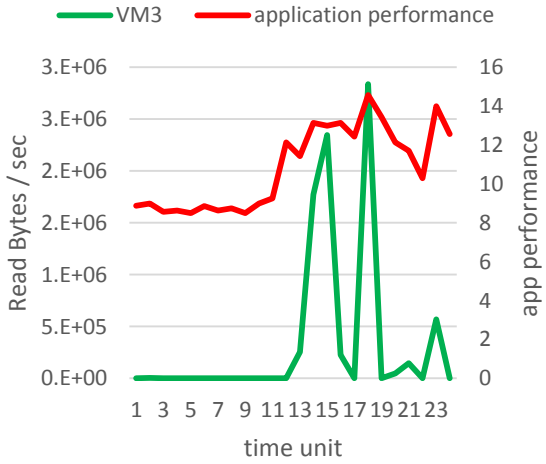
FUJITSU

- Real data set from a Virtual Desktop Infrastructure (VDI)
  - About 300 VMs are running on Windows Hyper-V
  - Windows performance counters: about 35,000 metrics
  - Application performance is obtained from benchmark results
  - Performance problems caused by storage resource bottlenecks.
- Examples of the selected causal metrics
  - Support threshold is set to 0.2

Our framework can capture these metrics!







Conclusion

FUJITSU

- We proposed a performance diagnosis framework
  - Our framework selects metrics that cause application performance problem
  - Our framework can capture temporary correlated metrics with application performance

    => **use association rule mining technique**

    => reduce false negatives
- We evaluated our framework
  - From the perspective of **the remaining diagnosis cost**
  - **Verification of the accuracy of the remaining diagnosis cost is future work**.
- We tested our framework on a real data set
  - Temporary correlated metrics actually exit
  - Our framework can capture those metrics

Questions?