Towards Autonomous IT Operations through Artificial Intelligence

Dan Pei

清華大學 | NetMan Tsinghua

About myself

- Tenured Associate Professor in Computer Science @ Tsinghua University
- Homepage: http://netman.aiops.org/~peidan
- Email: peidan@tsinghua.edu.cn Wechat: peidanwechat
- Research direction: AI for IT Operations; Autonomous IT Operations
- UCLA Ph.D. Best Ph.D. Thesis Award in UCLA CS in 2005.
- Joined Tsinghua CS Department in December 2012, with Government Endorsement (“ Recruitment Program of Global Talents”)
- Previously a Principal Researcher at AT&T Research, a co-founder and founding CEO of a mobile health company in Beijing, before joining Tsinghua.
- ACM/IEEE Senior Member
- During AT&T days, supervised interns from CMU, Cornell, Princeton, UCLA, GaTech, Michigan, Northwestern etc. Now @ Google, MSR, IBM, Purdue, Northeastern, HKUST

My Research Group @ Tsinghua: NetMan

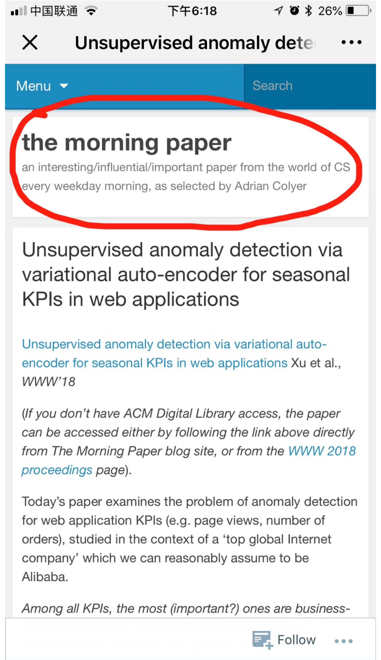
- Currently advising~15 of Ph.D. and M.S. students at Tsinghua.

- Two affiliated assistant professors and two post-docs

- Graduated 10 PhDs (3 went to MSRA, two went to Nankai University, one becomes a CEO, one goes to Alibaba)

Industry Collaborators

Research results are covered by technology media such as MIT technology Review, Hacker News, Mother Board, Morning paper, and many Chinese media.



中国联通 下午6:18 26%

Unsupervised anomaly dete

Menu Search

the morning paper

an interesting/influential/important paper from the world of CS every weekday morning, as selected by Adrian Colyer

Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications

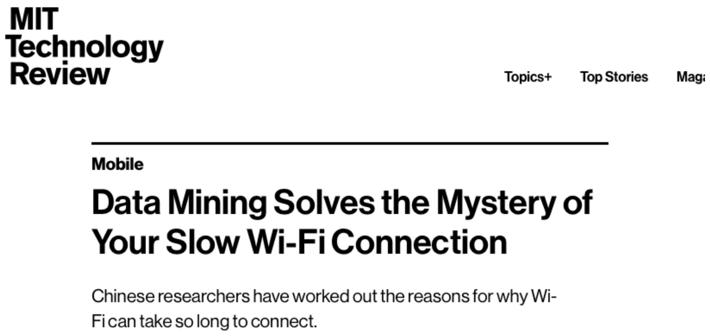
Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications Xu et al., *WWW'18*

(If you don't have ACM Digital Library access, the paper can be accessed either by following the link above directly from The Morning Paper blog site, or from the WWW 2018 proceedings *page).*

Today's paper examines the problem of anomaly detection for web application KPIs (e.g. page views, number of orders), studied in the context of a 'top global Internet company' which we can reasonably assume to be Alibaba.

Among all KPIs, the most (important?) ones are business-

Follow



MIT Technology Review

Topics+ Top Stories Mag

Mobile

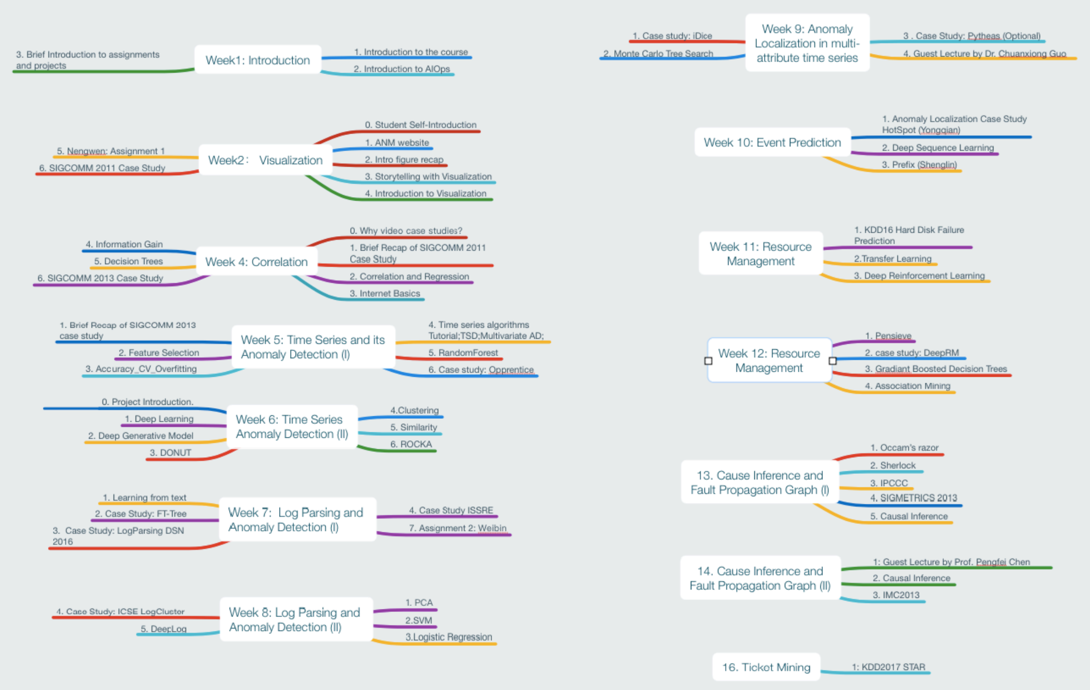
Data Mining Solves the Mystery of Your Slow Wi-Fi Connection

Chinese researchers have worked out the reasons for why Wi-Fi can take so long to connect.

Publications:

100+ AIOps papers and 20+ issued US Patents. Published in SIGCOMM、WWW、SIGMETRICS、TON、INFOCOM、IMC、CoNEXT etc.

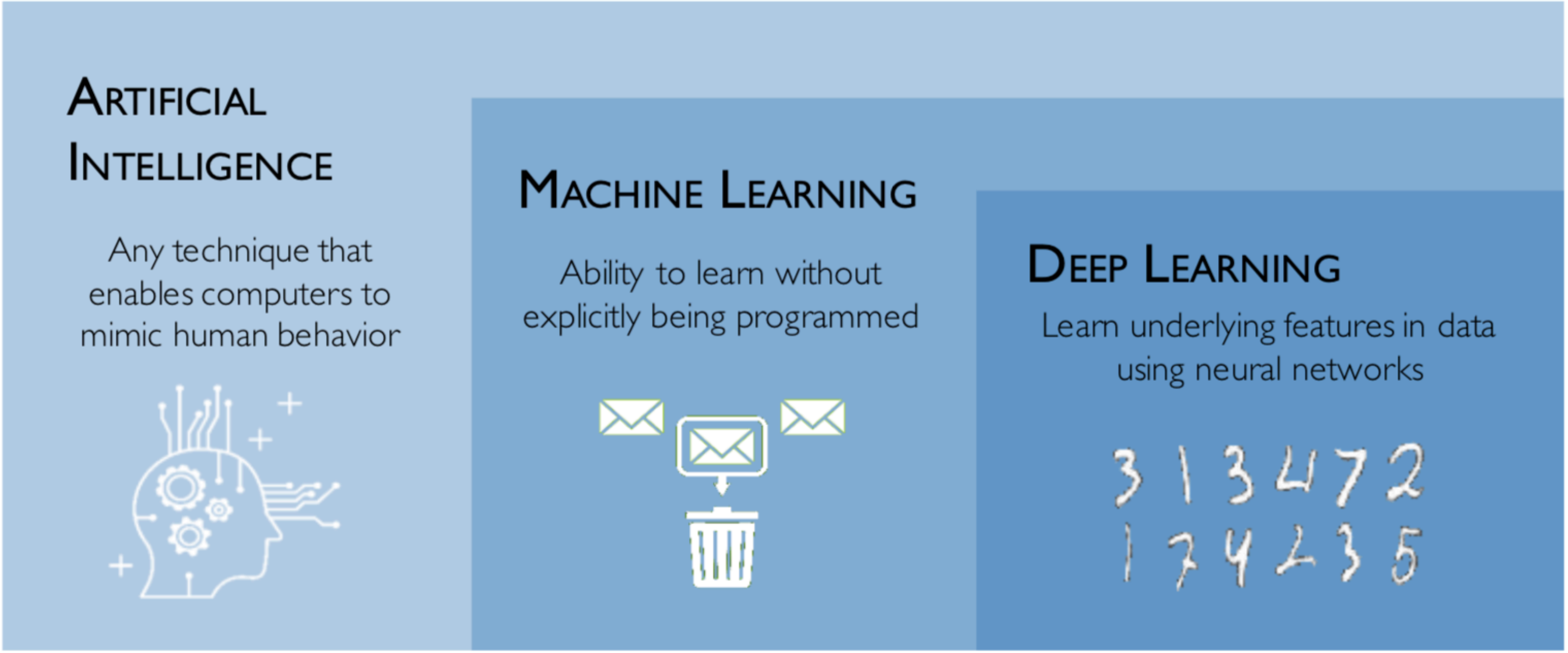
AIOps Course (in English) at Tsinghua: http://course.aiops.org



Outline

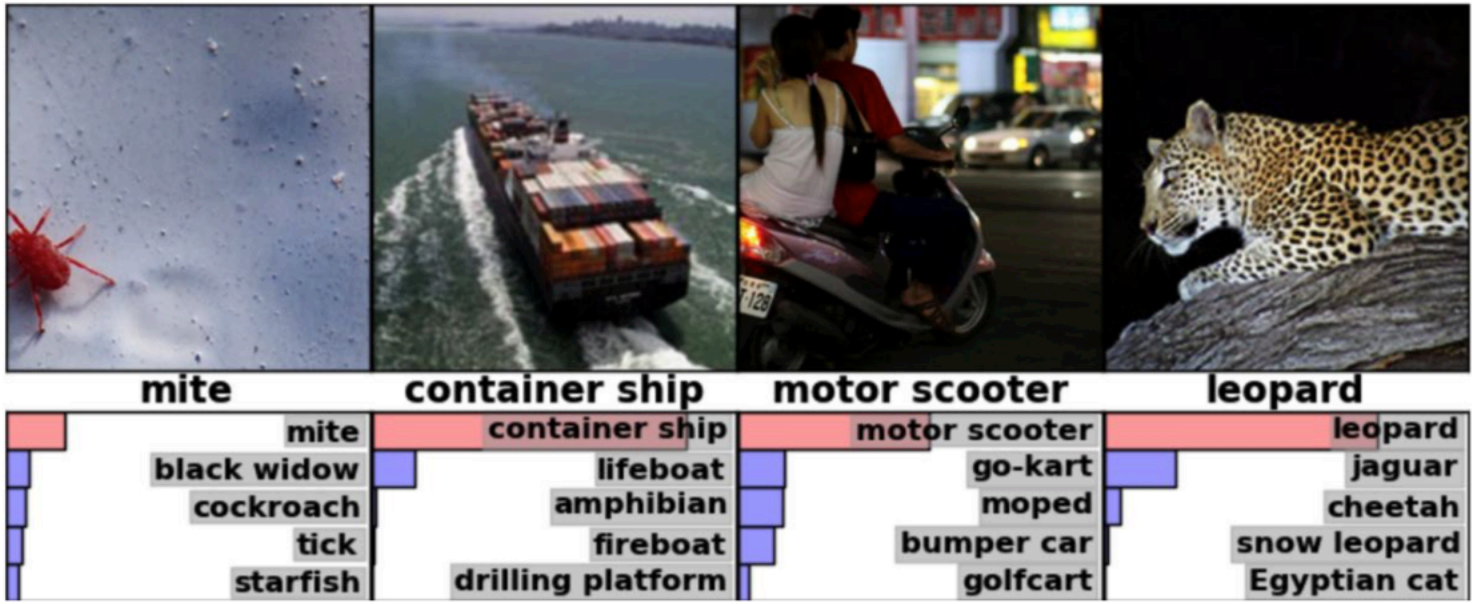
- *AI is changing the world*
- AI for IT Operations
- Operations center tour

What are AI, Machine Learning and Deep Learning?



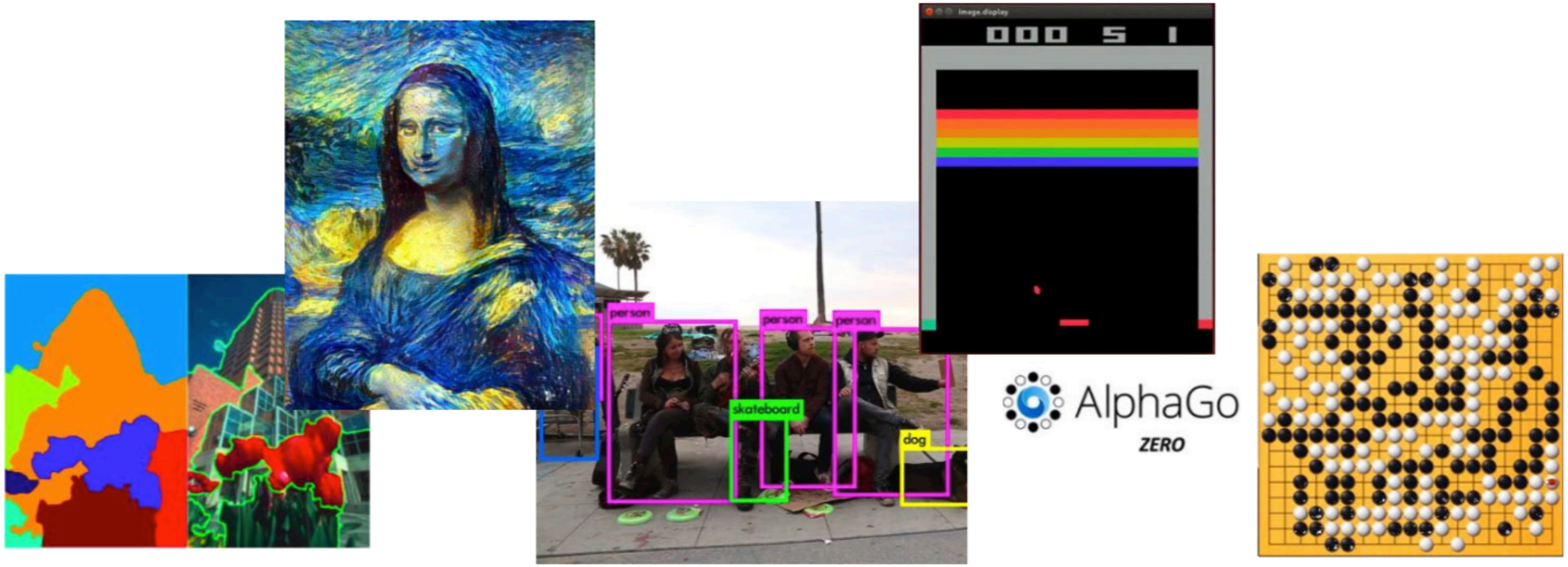
Deep Learning Success: Vision

Image Recognition



Deep Learning Success

And so many more…



Why Now?

1952 Stochastic Gradient Descent

1958 Perceptron
- Learnable Weights

⋮

1986 Backpropagation
- Multi-Layer Perceptron

1995 Deep Convolutional NN
- Digit Recognition

⋮

Neural Networks date back decades, so why the resurgence?

1. Big Data
- Larger Datasets
- Easier Collection & Storage

IMAGENET

WIKIPEDIA
The Free Encyclopedia

2. Hardware
- Graphics Processing Units (GPUs)
- Massively Parallelizable

3. Software
- Improved Techniques
- New Models
- Toolboxes

TensorFlow

Industries being changed by AI

- Finance
- Education
- **TMT**
- **Medical & Health**
- **Automobile**
- **Manufacturing**

Deep Learning Success: Audio

Other sequences-model applications:

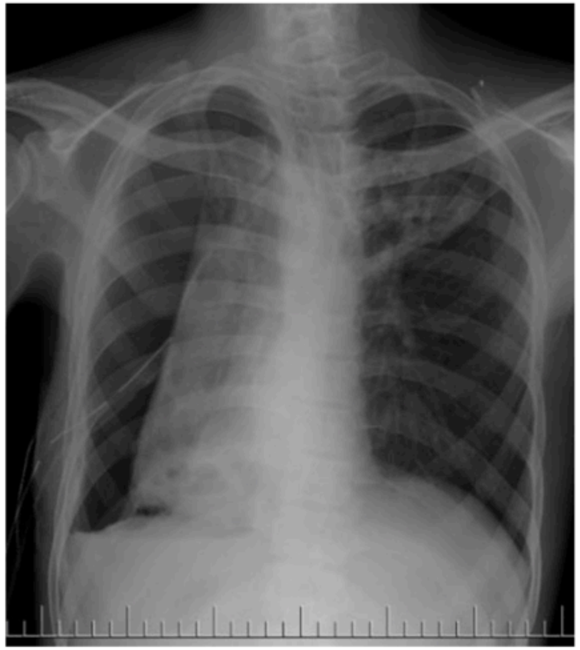
- predict stock price
- machine translation
- …

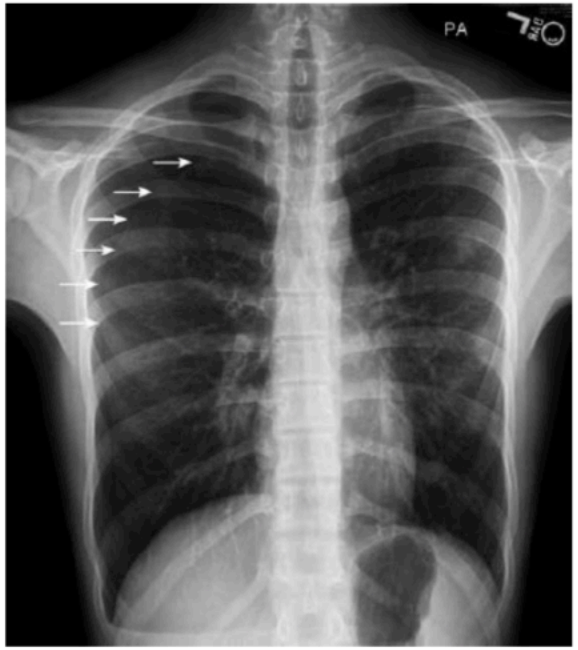
Music Generation

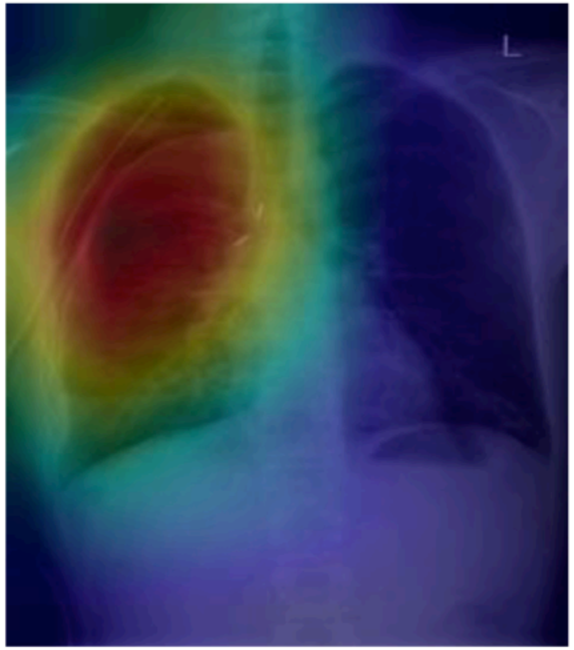
Temporal dependence

Deep Learning Success: Vision

Detect pneumothorax in real X-Ray scans







5 Applications Of AI In The Automotive Industry

1

Driving Features

AI lends itself perfectly to powering advanced safety features for connected vehicles.

2

Cloud Services

The application of artificial intelligence cloud platforms ensure that data is available when needed.

3

Automotive Insurance

AI speeds up the process of filing claims when accidents do occur.

4

Car Manufacturing

Robots are driving optimisation and the rethinking of processes and production in innovative new ways.

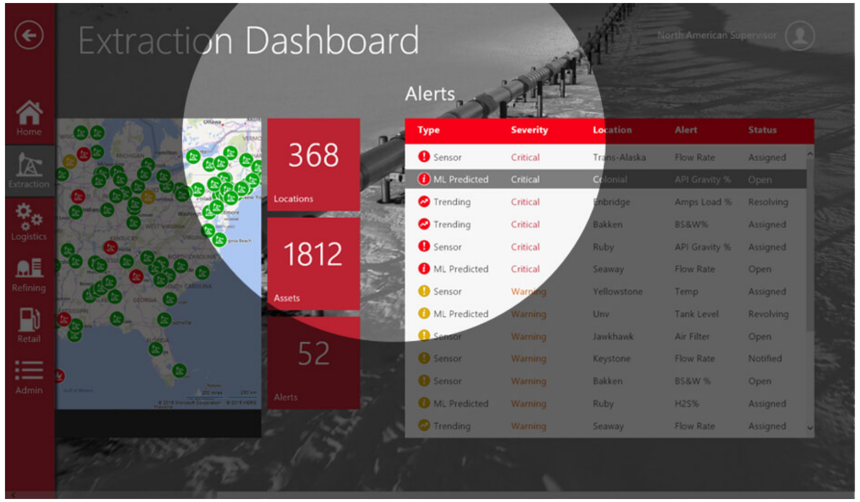
5

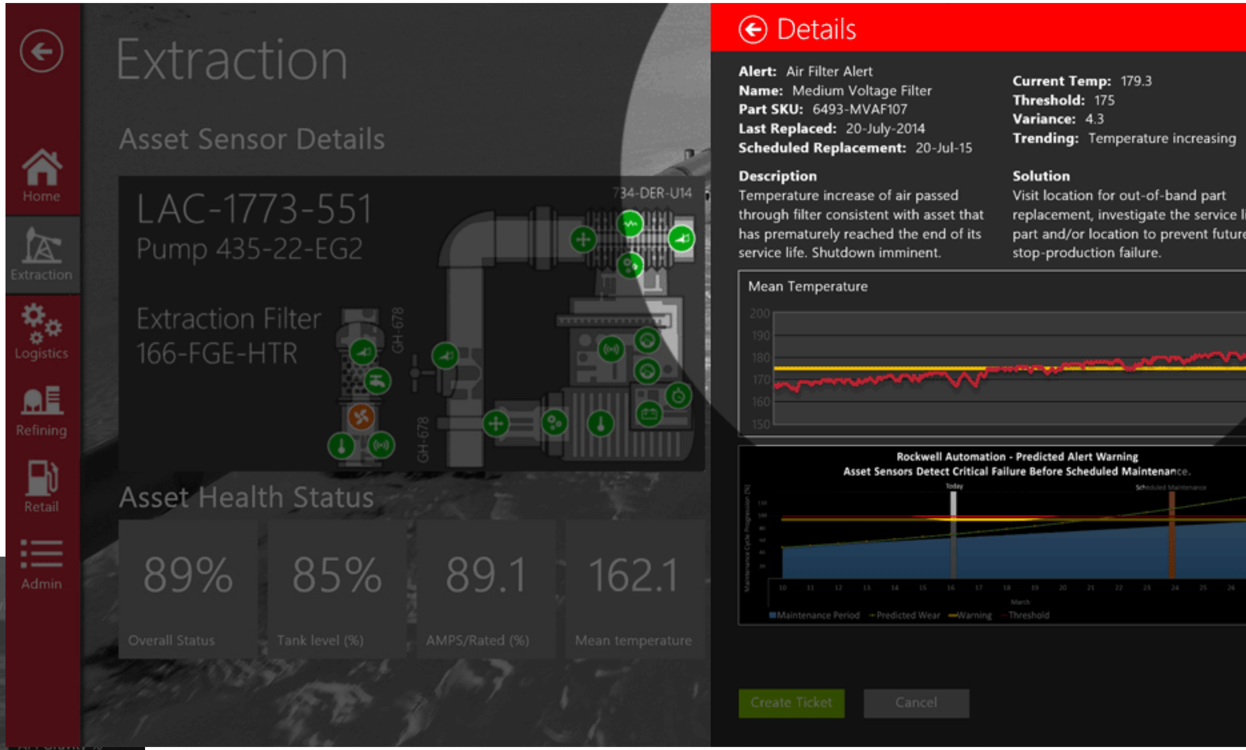
Driver Monitoring

AI software detects driver behavior in four key areas: driver identification, recognition, monitoring and infotainment control.

https://youtu.be/nBs3K0bsxyc

Predictive Maintenance





Machine Learning is a high-level programming language

Success in specific application scenario in specific area in specific industry:
quality assurance in manufacturing industry

Wood Floor

(Play video)

Tobacco Leaf

Steel Industry

8K video monitoring of the production line

Traditional programming language:

hard-coded logic

Machine learning as a programming language

hard-coded logic + fuzzy logic learned from data

The capability boundary of current AI technologies

AI is good at solving problems that satisfy the following five conditions simultaneously:

(1) With abundant data or knowledge

(2) With deterministic Information

(3) With complete Information

(4) Well-defined

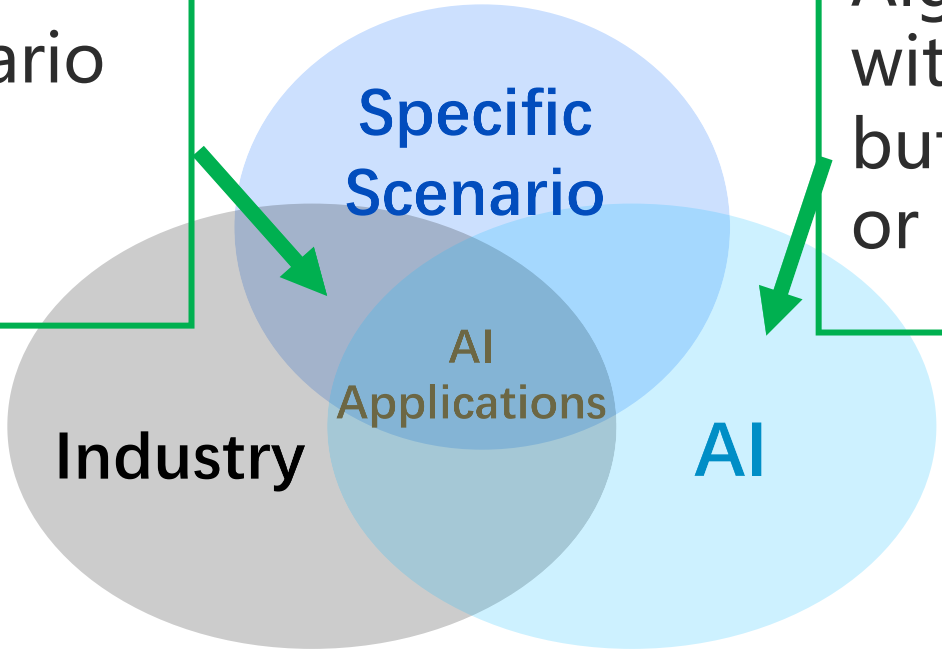
(5) Single-domain or limited-domain

——CAS Fellow, Prof Bo Zhang

Why success only in specific application scenario in specific area in specific industry?

Industry people familiar with scenario and industry

Algorithm people familiar with general AI,
but not specific industry or specific scenario



Traditional programming language:

hard-coded logic

Machine learning as a programming language

hard-coded logic + fuzzy logic learned from data

Pitfalls: use ML algorithms as Blackbox to tackle a specific scenario in a specific industry

a specific scenario in a specific industry

Huge Gap

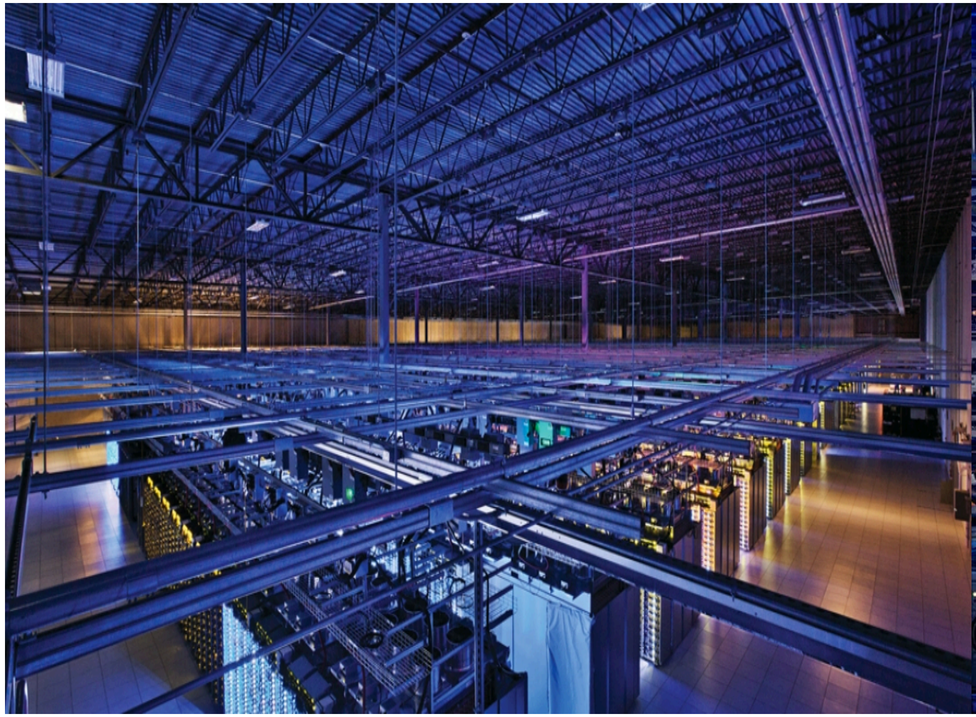
General Machine Learning Algorithms

ARIMA, Time Series Decomposition, Holt-Winters, CUSUM, SST,DiD,DBSCAN, Pearson Correlation, J-Measure, Two-sample test, Apriori, FP-Growth, K-medoids, CLARIONS, Granger Causality, Logistic Regression, Correlation analysis (event-event, event-time series, time series-time series) , hierarchical clustering、Decision tree, Random forest, support vector machine, Monte Carlo Tree search, Marcovian Chain, multi-instance learning, transfer learning, CNN, RNN ,VAE, GAN, NLP

Outline

- AI is changing the world
- AIOps: AI for IT Operations and Autonomous IT Operations
  - *What is AIOps*
  - Value of AIOps: brief case studies
  - Industry Leader’s Opinion
  - Is AIOps necessary?
  - Is AIOps feasible?
  - An in-depth case study
- Operations center tour

IT Operations is one of the technology foundations of the increasingly digitalized world.



IT operations are responsible for ensuring the digitalized businesses and societies run reliably, efficiently and safely, despite the inevitable failures of the imperfect underlying hardware and software.

But IT Operations are currently labor-intensive, heavily relied on human experience, very stressful, and ineffective.

Failure Discovery → Failure Mitigation → Failure Repair → Failure Avoidance

IT Operations Companies

servicenow

Valued at 44 Billion USD

splunk

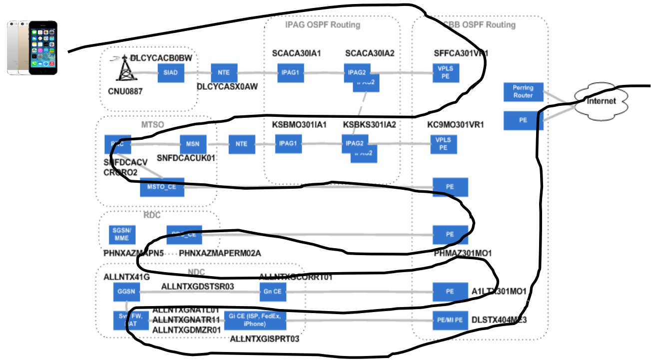
Valued at 20 Billion USD

elastic

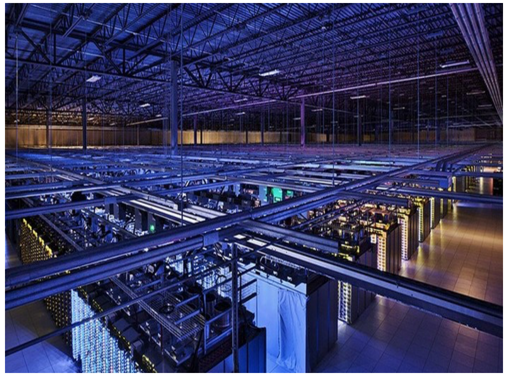
Valued at 7 Billion USD

AIOps： Autonomous IT Operations through Machine Learning

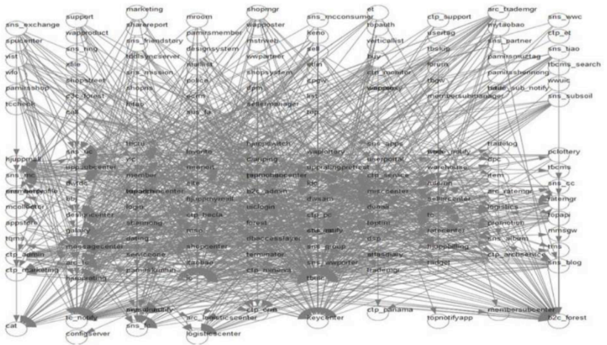
Large & complex
access network



Large & complex
data center



Large & complex
application software



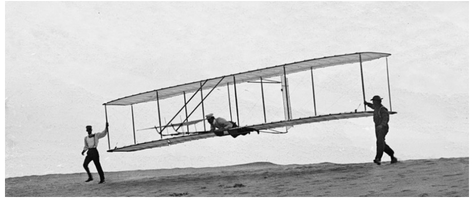
2012 淘宝核心链路应用拓扑图

- Imagine that you are running an Internet-based service with hundreds of thousands of servers and many software modules, a large, complex, cross-layer, and rapidly evolving distributed system.

- You want to achieve 99.999% service reliability, but the terabytes of machine-generated monitoring data and hundreds of operators (IT operation engineers) alone won’t get you there, because of the high complexity and sheer scale of the software/hardware system and the vast amount of machine-generated data.

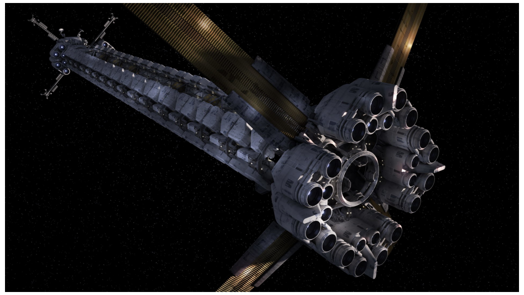
- Machine learning is the direction to enable Autonomous IT Operations autonomous.

Towards Autonomous IT Operations

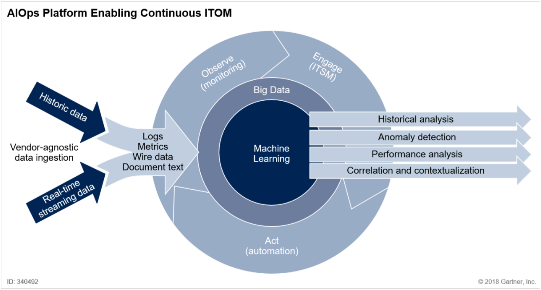


Manual-Driven

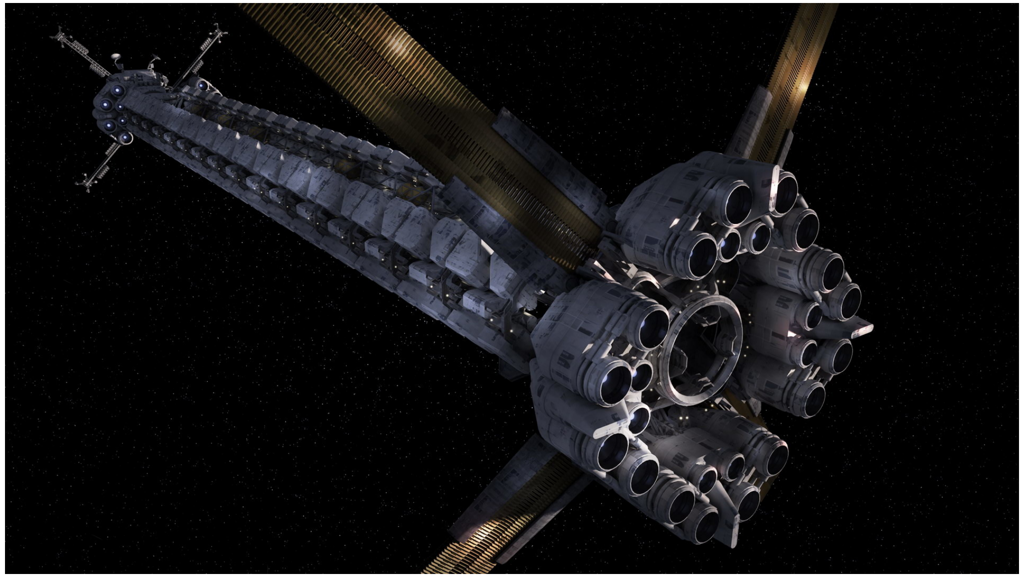
Automated but with Manual Decision



Autonomous



Ultimate Goal: Autonomous IT Operations



Spaceship Covenant: 2000 passengers and 15 crew members all in hibernation. Flying towards Planet Origae-6。Only one awaken android crew.

Spaceship Avalon: 5000 passengers and 258 crew members. Flying towards Planet Homestead II, 120-year trip.

Autonomous IT Operations: Automatically deal with all four causes of changes to IT systems

- Software & hardware failures--> Automatic Healing

- Software changes --> Autonomous software deployment

- Change of user request amount & Pattern --> Elastic Resource Allocation

- Malicious Attacks-->Autonomous Defense

“Most people overestimate what they can do in one year and underestimate what they can do in ten years.”

-- Bill Gates

Outline

- AI is changing the world

- AIOps: AI for IT Operations and Autonomous IT Operations
  - What is AIOps
  - *Value of AIOps: brief case studies*
  - Industry Leader’ s Opinion
  - Is AIOps necessary?
  - Is AIOps feasible?
  - An in-depth case study

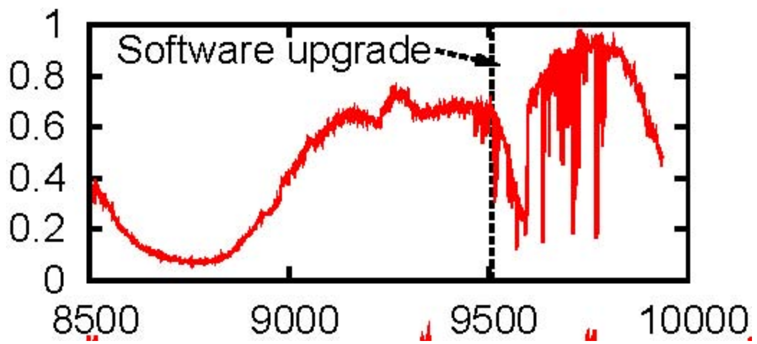
- Operations center tour

Reduced Business Loss: Rapid Assessment of Software Changes

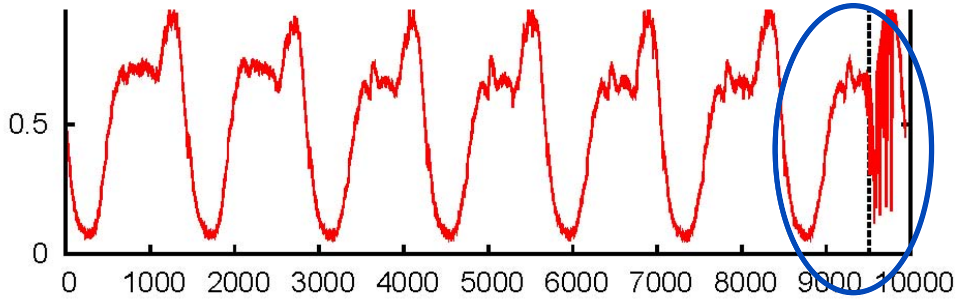
- A buggy deployment causes significant revenue Loss
- Manual trouble shooting takes 1.5 hours

Customer complaints → Inspecting KPIs → Troubleshooting

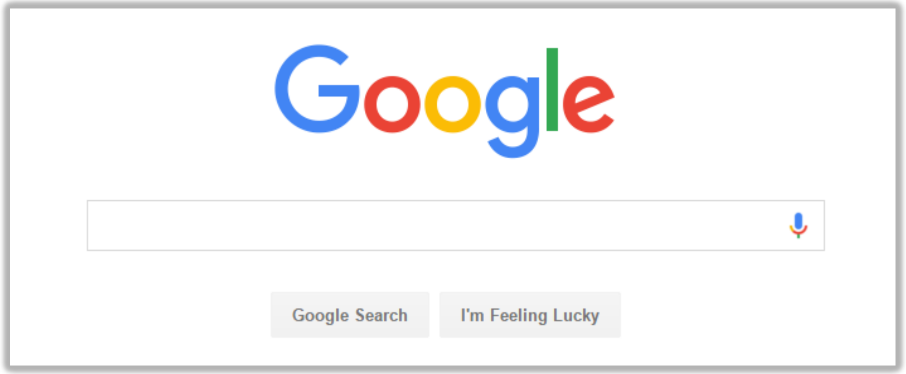
- AIOps solution takes less than 10 minutes



Joint Work with Baidu
Published in ACM CoNext 2015



Web Search Engines

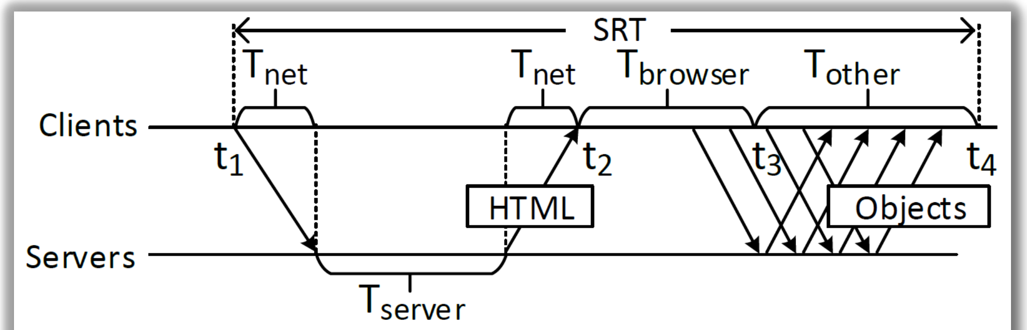


Search Response Time (SRT)

t_1 A search query is submitted

t_4 The result page Is rendered

$$SRT = t_4 - t_1$$



Search Response Time Matters

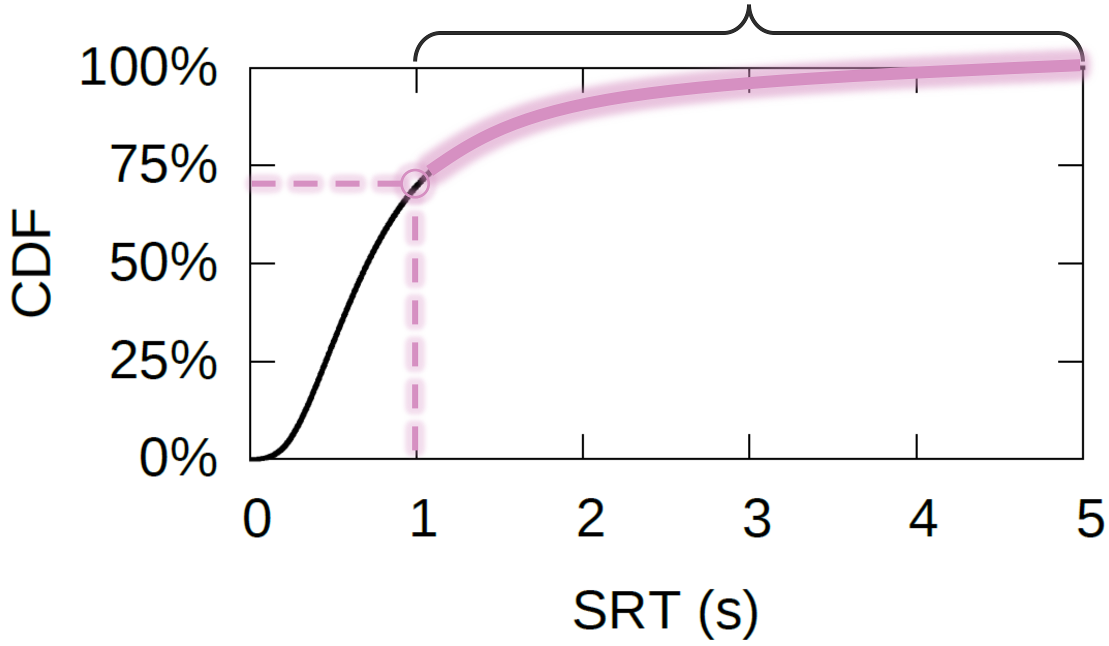
+500ms revenue ↓ 1.2%
[Eric Schurman, Bing]

+100ms~400ms queries ↓ 0.2%~0.6%
[Jake Brutlag, Google]

Given two content-wise identical search result pages,
users are more likely to perform clicks on the fast page
[SIGIR 2014]

Search Response Time in the Wild

User's flow of thought is interrupted
if pages take **longer than 1s** to load



Why?

https://www.nngroup.com/articles/response-times-3-important-limits/

Monitoring SRT: Search Logs

$\underbrace{\hspace{20em}}$ **Measurable attributes that can potentially impact SRT**

SRT	User's ISP	Browser engine	# of Images	Ads	Server Load	…
800ms (Low SRT)	China Unicom	WebKit	10	Yes	1000 queries/s	…
1200ms (High SRT)	China Telecom	Trident 5.0	5	No	500 queries/s	…
……						

Improved Revenue: Reduced Page Response Time

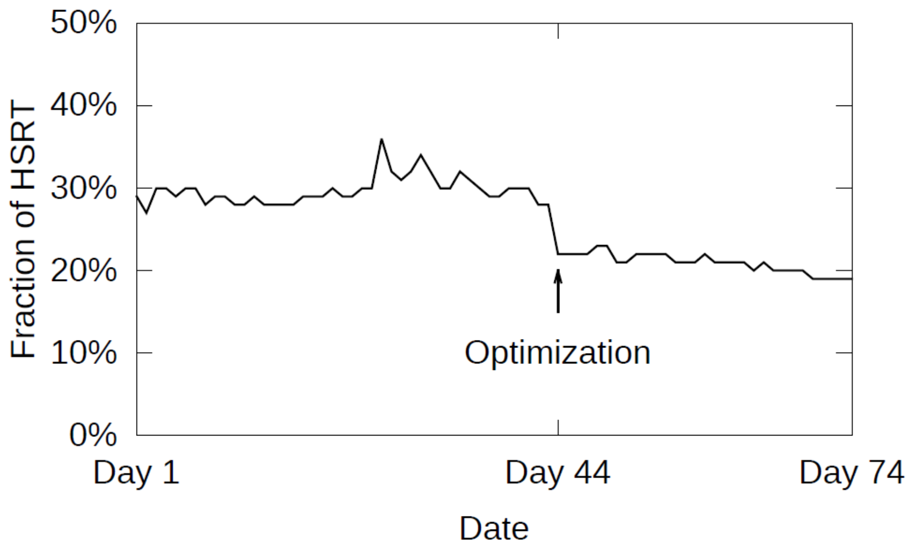
amazon -100ms ->Sales 1%
[Greg Linden, Amazon]

Google -100ms~400ms -> Revenue 0.2%~0.6%
[Jake Brutlag, Google]

After deploying the solutions suggested by AIOps：

Slow responses (>1s) are reduced from 30% to 20%

80th-percentile response time is reduced by 253 ms



(a) Fraction of HSRT each day

Saves 30 man-months (estimated) of manual analysis

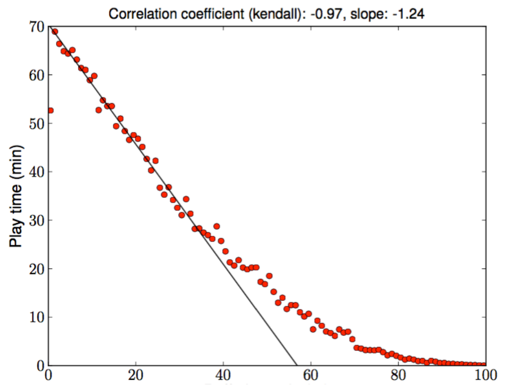
Joint Work with Baidu
Published in IEEE INFOCOM 2016

AIOps Leads to Better User Experience -> Longer Engagement -> More Revenue

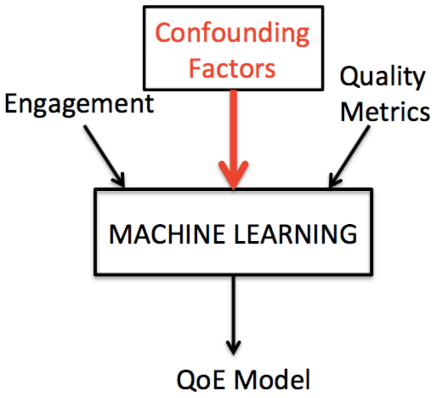
Linear Regression
SIGCOMM 2011

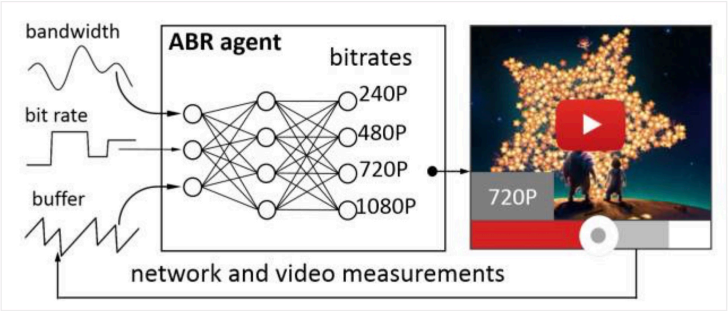
Decision Trees
SIGCOMM 2013

Reinforcement Learning
SIGCOMM 2017



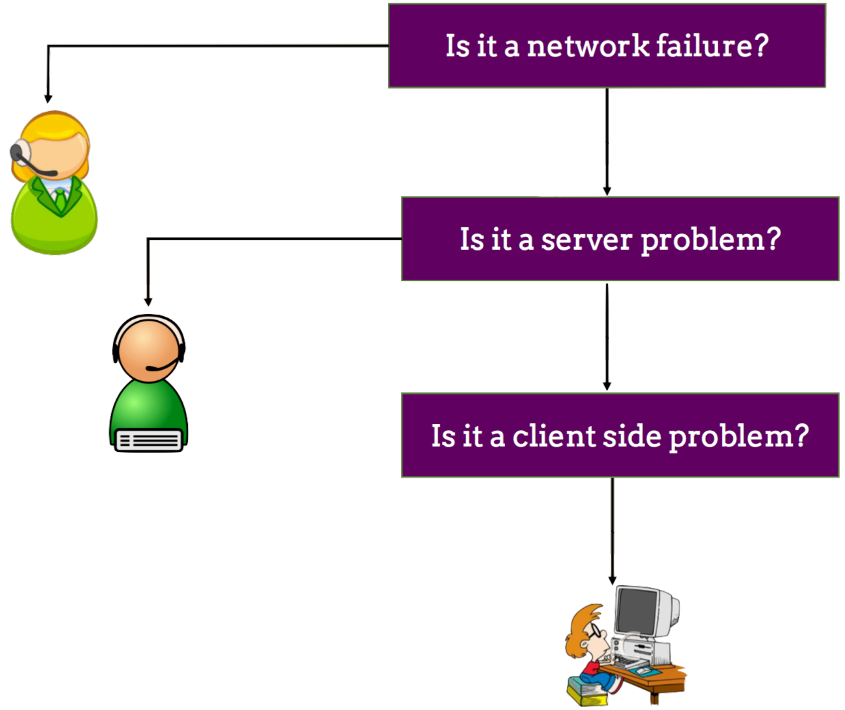
Adding as a feature

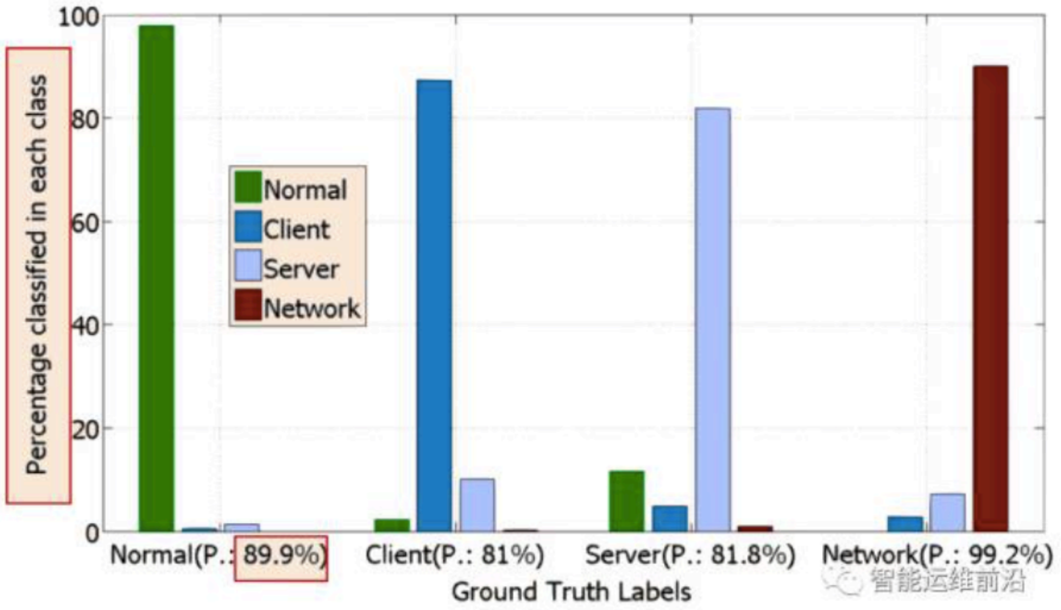




Conviva/CMU/MIT work

AIOps Quickly Decides the Responsibility Boundary: Reduced loss

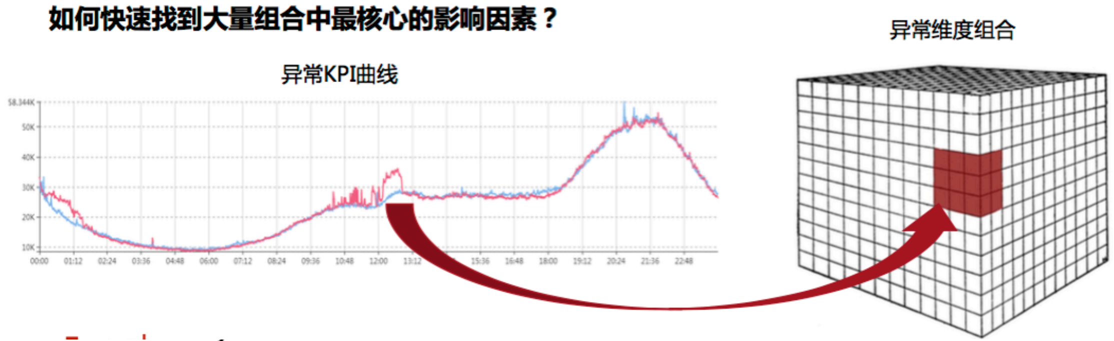


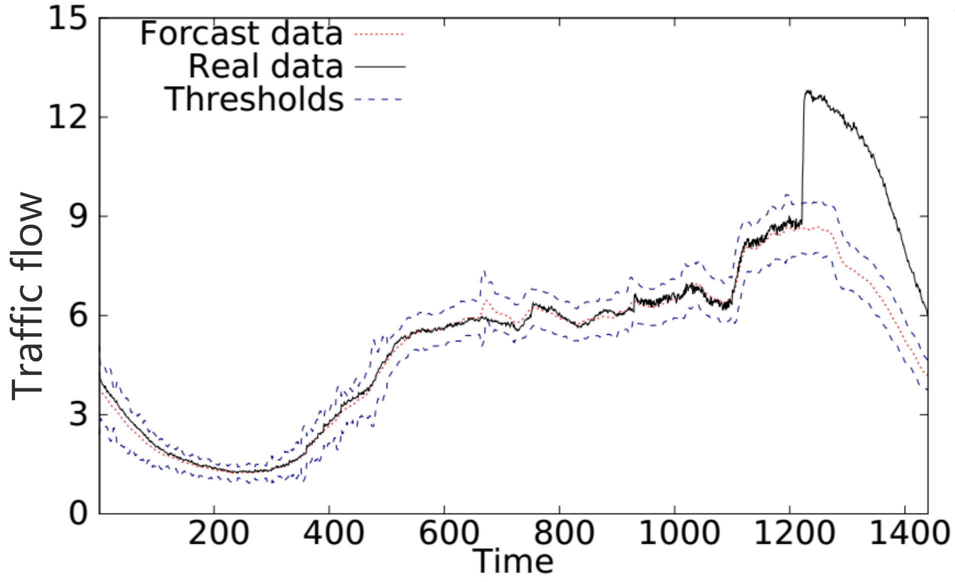


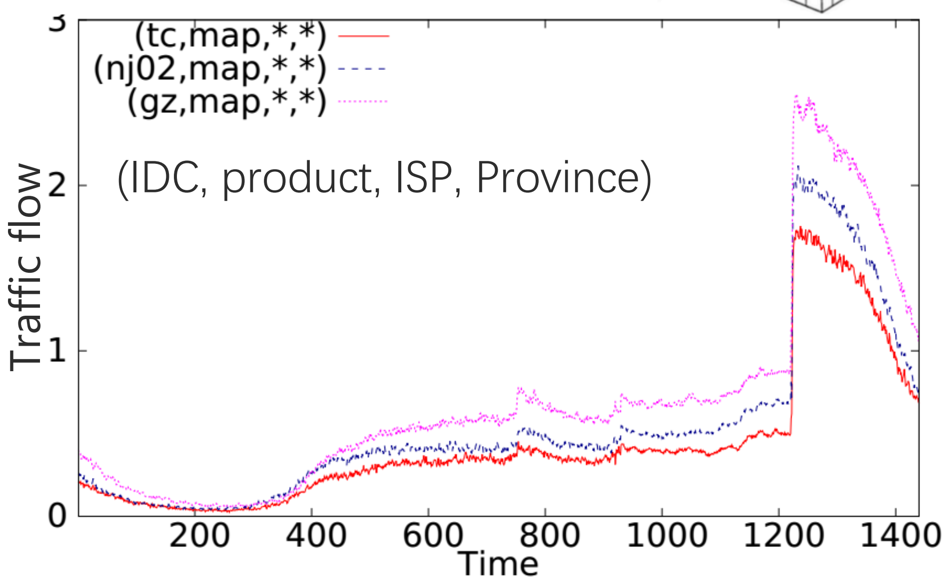
Microsoft Azure Work.
Published in SIGCOMM 2016

Localizing the Anomalous Regions: Reduced Loss

Manual localization: 90 minutes
AIOps: 30s





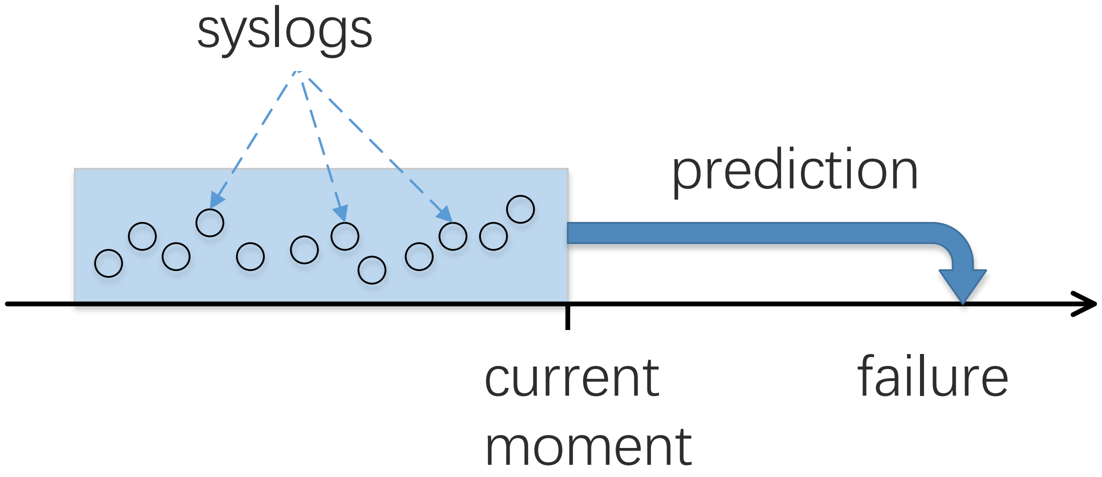


Collaboration with Baidu. Tencent implemented a variant to improve its video streaming service

DC Switch Failure Prediction->Preventive Replacement->Avoided Loss

Problem: Baidu-customized switches intermittently drop/delay packets, causing QoE drop at the application layer.

Reboot stops the problem for some while.
Question: Can we predict the this problem 2 hours before it happens again? Then just switch the traffic away from this switch and reboot it.



- Precision: 82.15%
- Recall: 74.74%
- FPR: 3.75×10^{-5}

Table 2: One Example of Benign Request.

Original Request	POST http://localhost:8080/tienda1/publico/autenticar.jsp modo=entrar&login=caria&pwd=egipciaca&remember=off&B1=Entrar		
Token Sequence	tienda1 publico autenticar jsp modo entrar login _OTHER_ pwd _OTHER_ remember off b1 entrar		
Recovered Token Sequence	tienda1 publico autenticar jsp modo entrar login _OTHER_ pwd _OTHER_ remember on b1 entrar		
BLEU	0.8091	**Malicious Score**	0.1909

Table 3: One Example of Malicious Request.

Original Request	POST http://m.thepaper.cn/admin_UploadDataHandler.ashx ------WebKitFormBoundaryRvkd1dbq3x1OJhUH\x0D\x0AContent-Disposition: form-data; name=\x22uploadify\x22; filename=\x2220170215180046.jpg\x22\x0D\x0A *Content-Type: image/jpeg*\x0D\x0A\x0D\x0A **<%eval request(\x22T\x22)%>**\x0D\x0A------WebKitFormBoundaryRvkd1dbq3x1OJhUH\x0D\x0AContent-Disposition: form-data; name=\x22saveFile\x22\x0D\x0A\x0D\x0At.asp\x0D\x0A------WebKitFormBoundaryRvkd1dbq3x1OJhUH\x0D\x0AContent-Disposition: form-data; name=\x22Upload\x22\x0D\x0A\x0D\x0ASubmit Query\x0D\x0A------WebKitFormBoundaryRvkd1dbq3x1OJhUH--		
Token Sequence	_OTHER_ ashx _OTHER_ content disposition form data name uploadify filename _pnum_0_ jpg content type image jpeg eval request onechr _OTHER_ content disposition form data name _OTHER_ onechr asp _OTHER_ content disposition form data name upload submit query _OTHER_		
Recovered Token Sequence	_OTHER_ _OTHER_ do php _OTHER_ eval get_magic_quotes_gpc stripslashes _post chr _pnum_0_ chr _pnum_1_ _post chr _pnum_2_ chr _pnum_3+_ z0 _pnum_3+_ ini_set display_errors _pnum_3+_ set_time_limit _pnum_3+_ set_magic_quotes_runtime _pnum_3+_ echo onechr dirname _server script_filename if onechr onechr dirname _server path_translated		
BLEU	0	**Malicious Score**	1.0

Detecting previously unseen attacks: 99% accuracy --> more secure

Based on self-translation

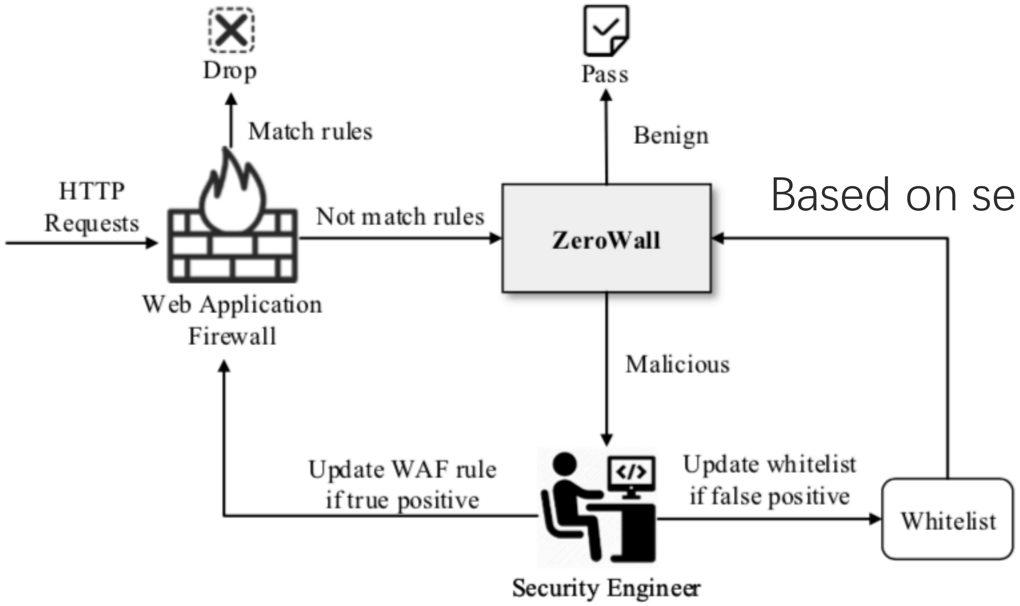
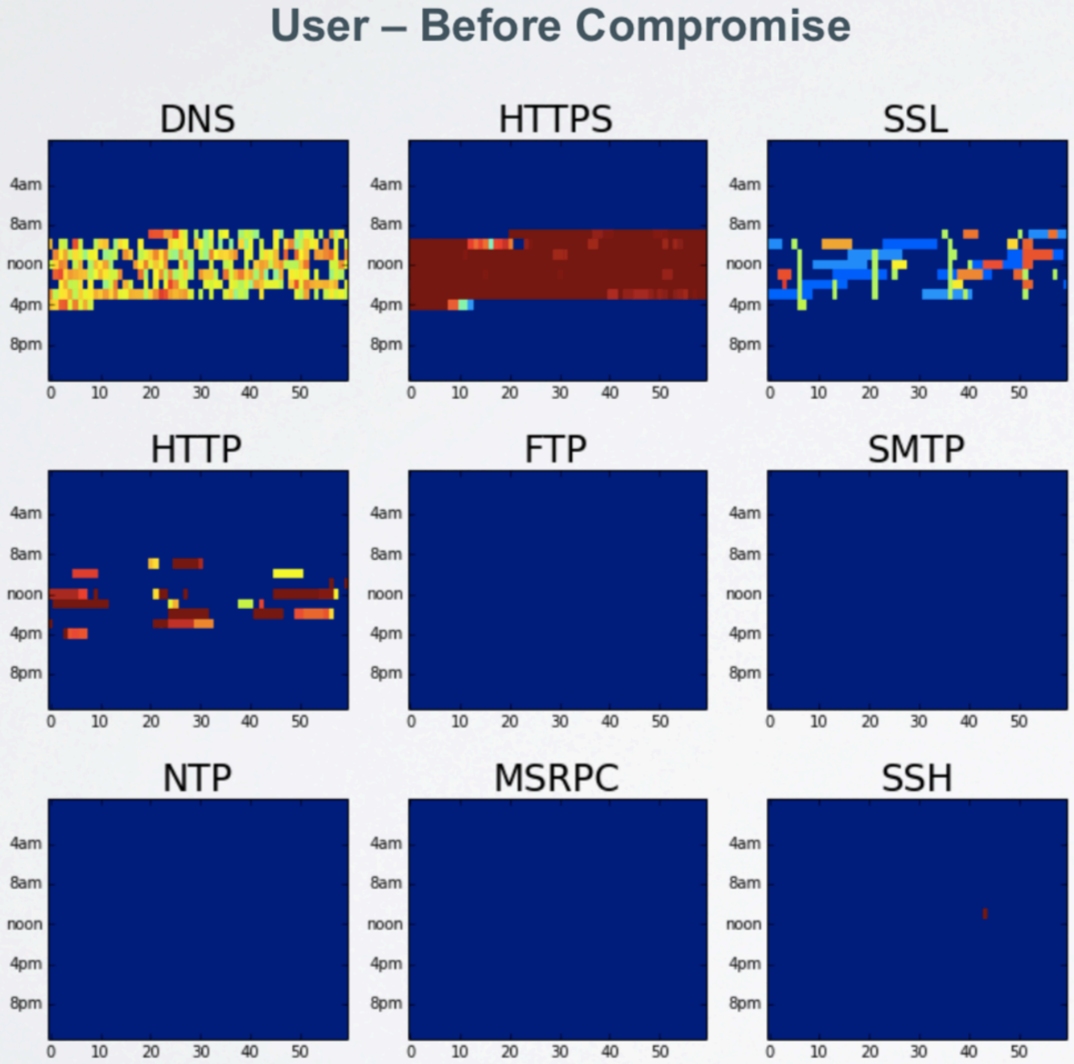
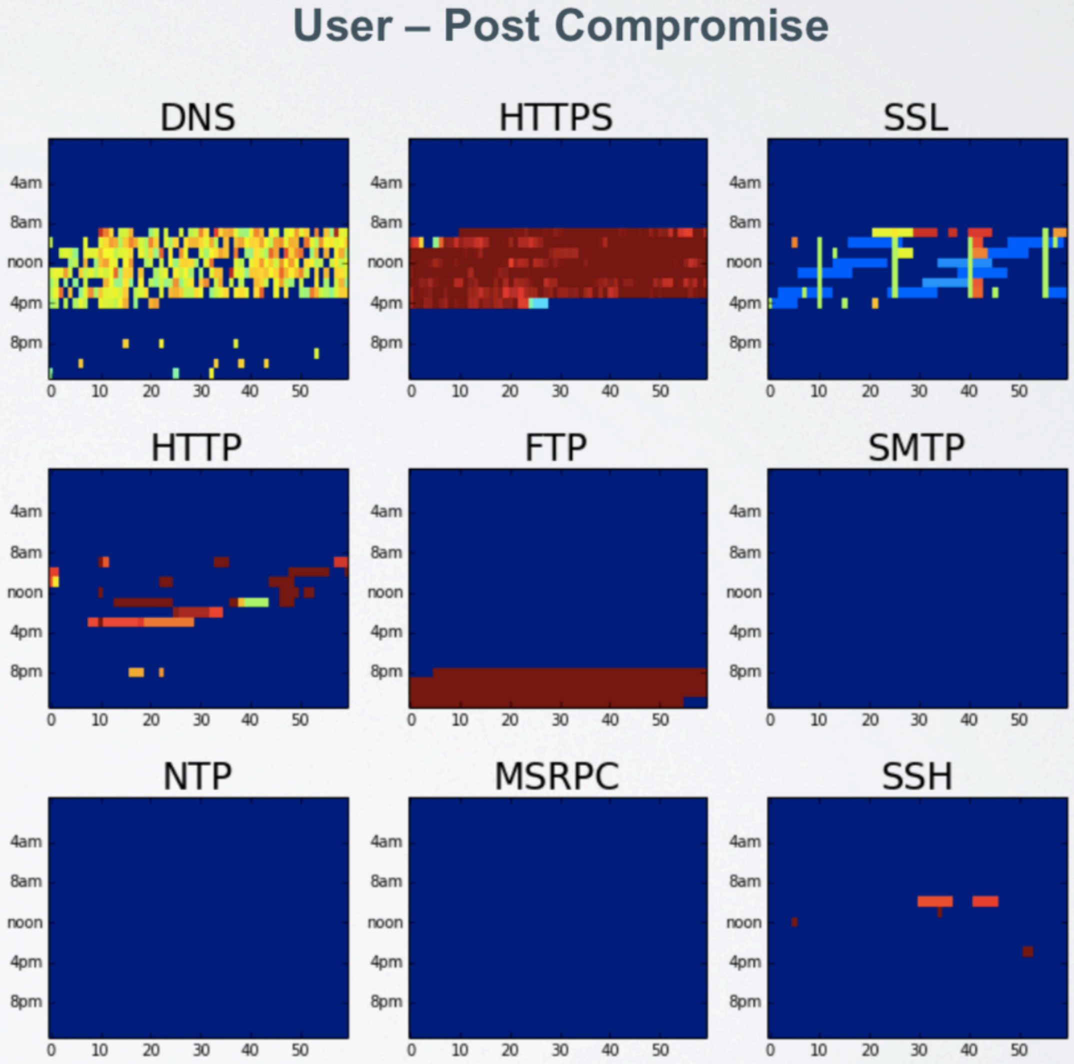


Figure 1: The workflow of *ZeroWall*.

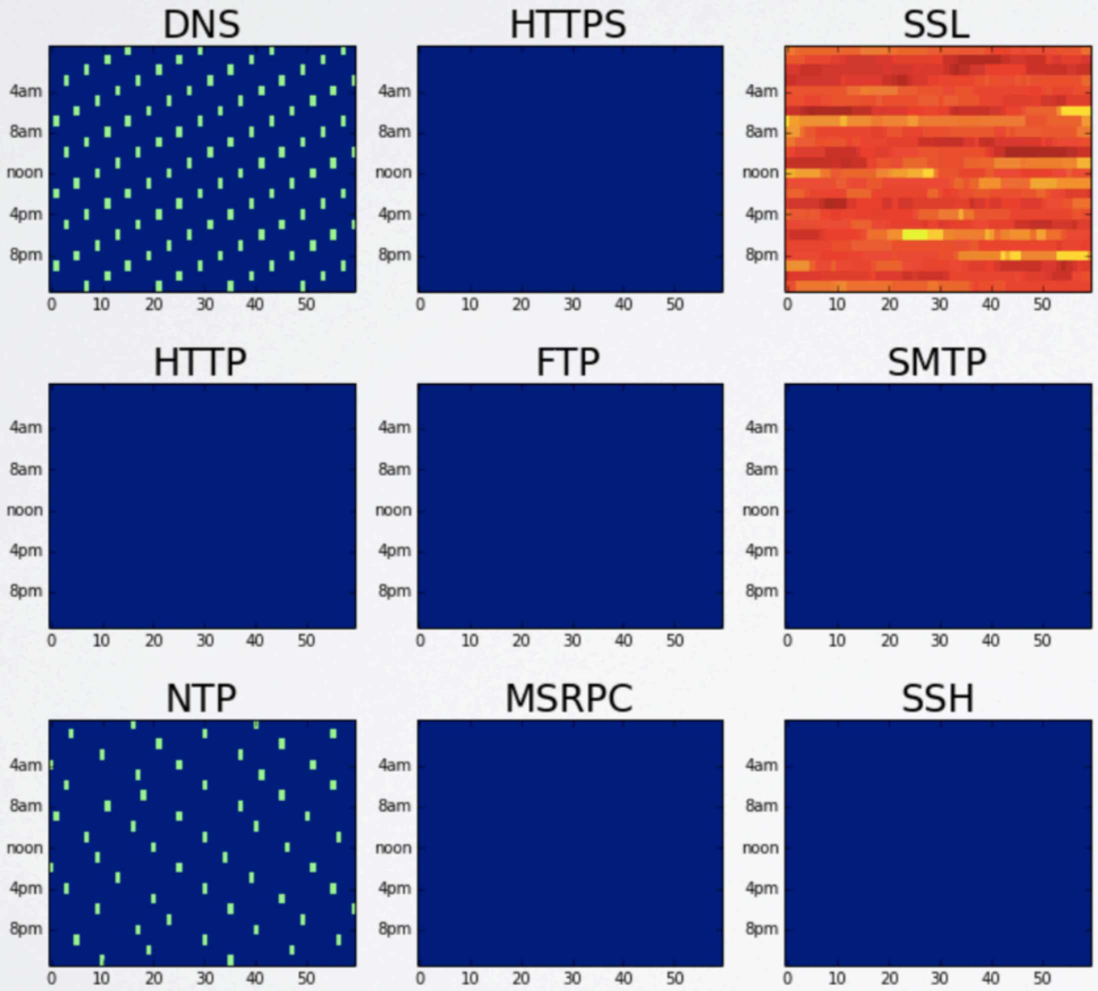
BEHAVIOR ANOMALY USER | EXFILTRATION



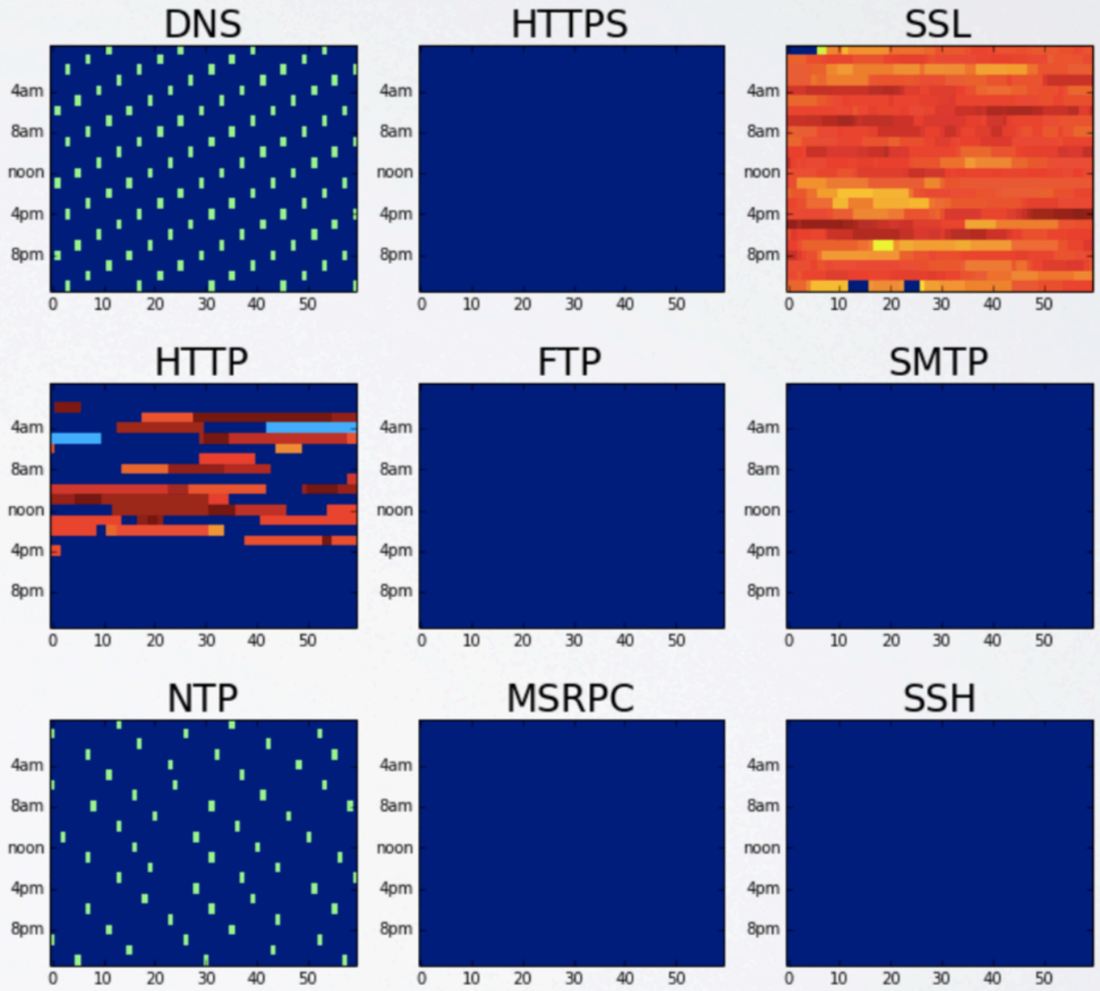


BEHAVIOR ANOMALY IOT DEVICE | DATA DOWNLOAD

Dropcam – Before Compromise



Dropcam – Post Compromise



Outline

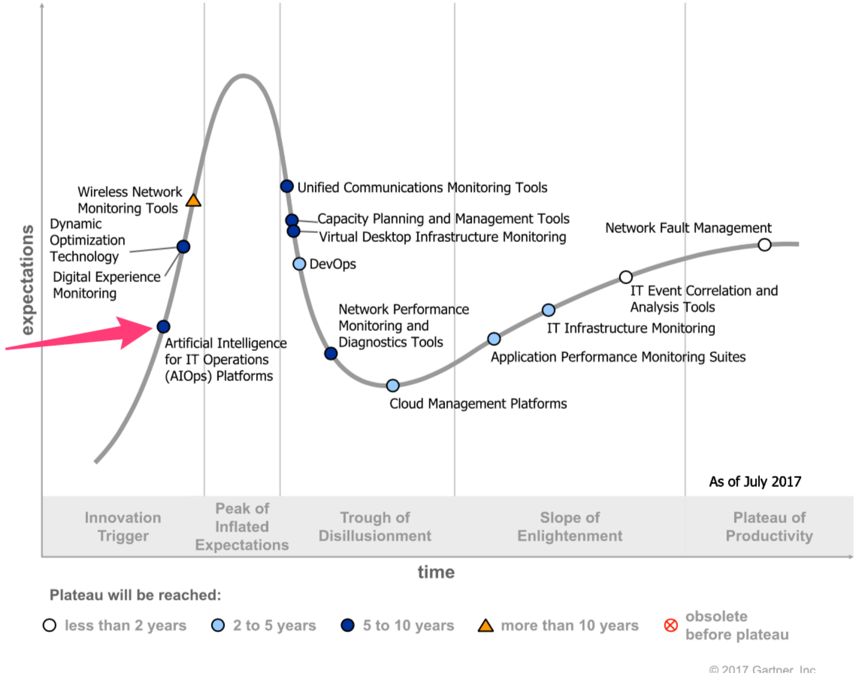
- AI is changing the world

- AIOps: AI for IT Operations and Autonomous IT Operations
  - What is AIOps
  - Value of AIOps: brief case studies
  - *Industry Leader's Opinion*
  - Is AIOps necessary?
  - Is AIOps feasible?
  - An in-depth case study

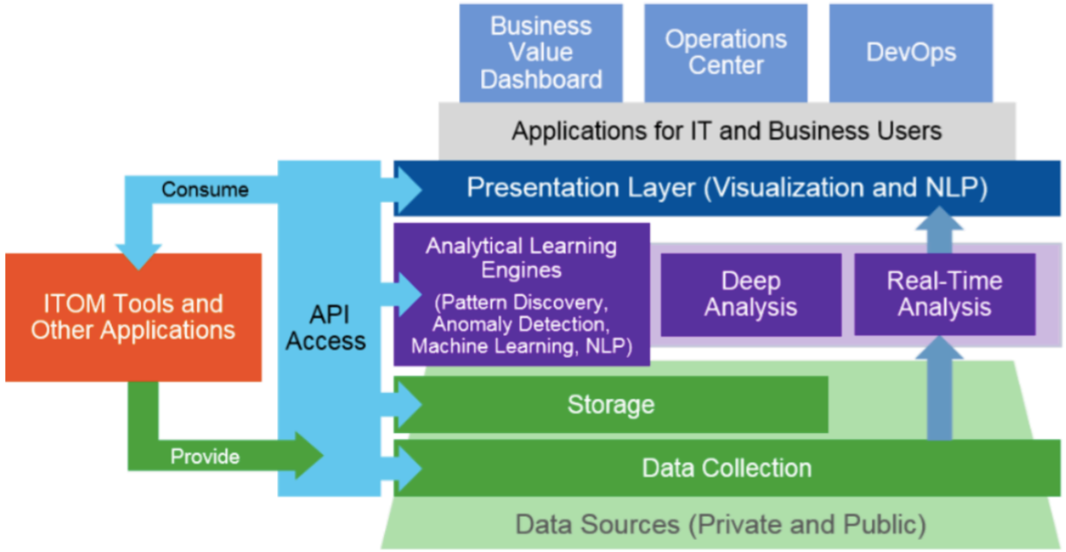
- Operations center tour

AIOps is rising

- According to Gartner Report：
- AIOps global deployment ratio: 10% （2017） → **50% （2020）**



Source: Gartner (July 2017)



Source: Gartner (March 2016)

“In addition to control plane and data plane, Internet needs an AI-based knowledge plane”
- Dave Clark, the Architect of the Internet, in his SIGCOMM 2003 paper.

A Knowledge Plane for the Internet

David D. Clark*, Craig Partridge◆, J. Christopher Ramming† and John T.

*M.I.T Lab for Computer Science
200 Technology Square
Cambridge, MA 02139
{ddc,jtw}@lcs.mit.edu

◆BBN Technologies
10 Moulton St
Cambridge, MA 02138
craig@bbn.com

†SRI
333 Rav
Menlo Pa
chrisramm

ABSTRACT

We propose a new objective for network research: to build a fundamentally different sort of network that can assemble itself given high level instructions, reassemble itself as requirements change, automatically discover when something goes wrong, and automatically fix a detected problem or explain why it cannot do so.

We further argue that to achieve this goal, it is not sufficient to improve incrementally on the techniques and algorithms we know today. Instead, we propose a new construct, the Knowledge Plane, a pervasive system within the network that builds and maintains high-level models of what the network is supposed to do, in order to provide services and advice to other elements of the network. The knowledge plane is novel in its reliance on the tools of AI and cognitive systems. We argue that cognitive techniques, rather than traditional algorithmic approaches, are best suited to meeting the uncertainties and complexity of our objective.

transparent network with rich end-sy
deeply embedded assumption of
administrative structure are critical stre
users when something fails, and high
much manual configuration, diagnosis a

Both user and operator frustrations aris
design principle of the Internet—the
with intelligence at the edges [1,2].
without knowing what that data is, or
combination of events is keeping dat
edge may recognize that there is a prob
that something is wrong, because the c
be happening. The edge understands
expected behavior is; the core only de
network operator interacts with the core
as per-router configuration of routes a
for the operator to express, or the netw

Leaders' opinions about AIOps

Huawei CEO Ren Zhengfei:

"AI is the most important tools for managing the networks.

一、巨大的存量网络是人工智能最好的舞台

为什么要聚焦GTS、把人工智能的能力在服务领域先做好呢？对于越来越庞大、越来越复杂的网络，人工智能是我们建设和管理网络的最重要的工具，人工智能也要聚焦在服务主航道上，这样发展人工智能就是发展主航道业务，我们要放到这个高度来看。如果人工智能支持GTS把服务做好，五年以后我们自己的问题解决了，我们的人工智能又是世界一流。

首先，是解决我们在全球巨大的网络存量的网络维护、故障诊断与处理的能力的提升。我们在全球网络存量有一万亿美元，而且每年上千亿的增加。容量越来越大，流量越来越快，技术越来越复杂，维护人员的水平要求越来越高，经验要求越来越丰富，越来越没有这样多的人才，人工智能，大有前途。

Jeff Dean Head of AI, Google

"We can improve everywhere in a system that have tunable parameters or heuristics"

Anywhere We've Punted to a User-Tunable Performance Option!

Many programs have huge numbers of tunable command-line flags, usually not changed from their defaults

```
--eventmanager_threads=16
--bigtable_scheduler_batch_size=8
--mapreduce_merge_memory=134217728
--lexicon_cache_size=1048576
--storage_server_rpc_freelist_size=128
...
```

Anywhere We're Using Heuristics To Make a Decision!

Compilers: instruction scheduling, register allocation, loop nest parallelization strategies, ...

Networking: TCP window size decisions, backoff for retransmits, data compression, ...

Operating systems: process scheduling, buffer cache insertion/replacement, file system prefetching, ...

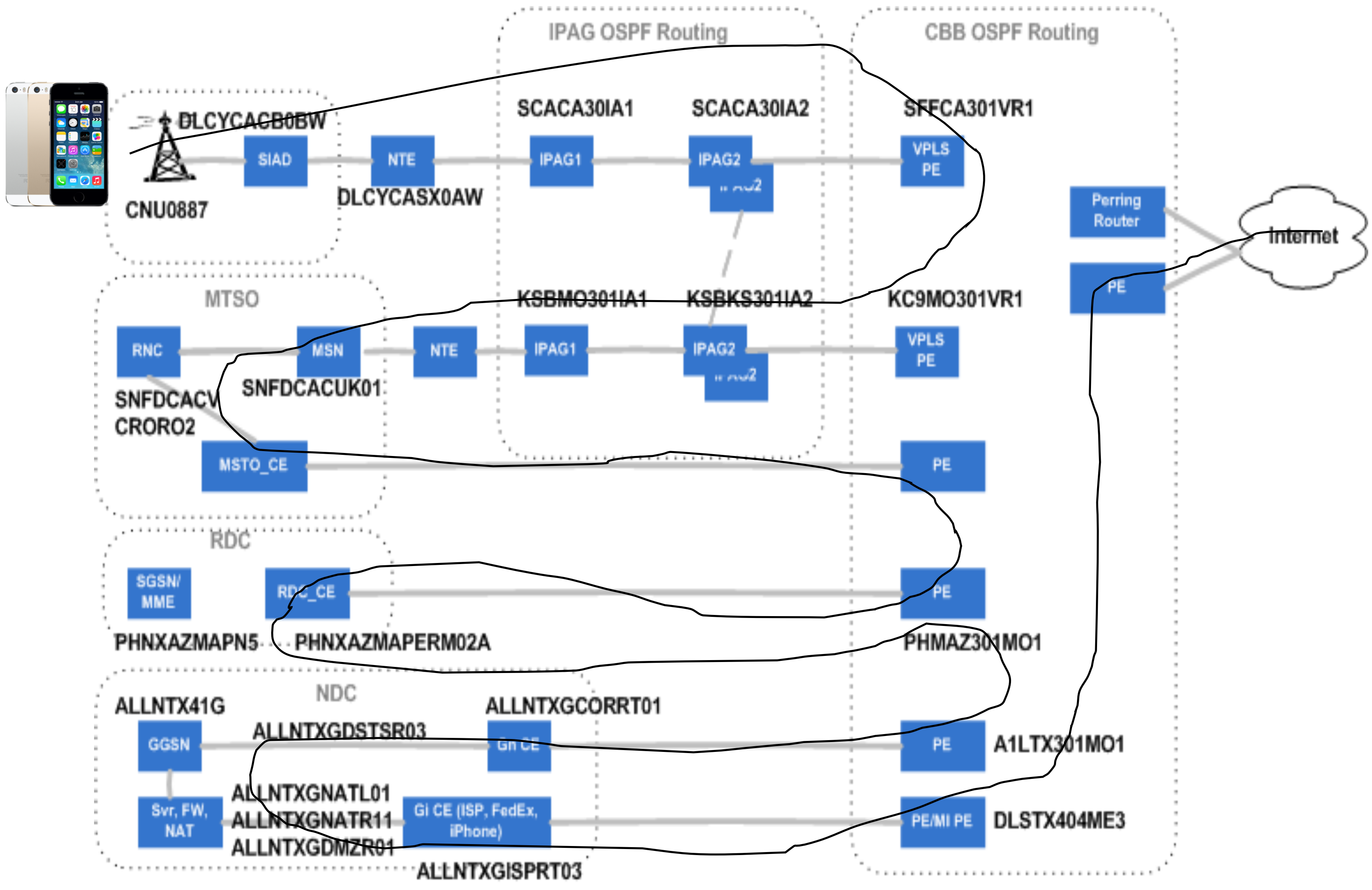
Job scheduling systems: which tasks/VMs to co-locate on same machine, which tasks to pre-empt, ...

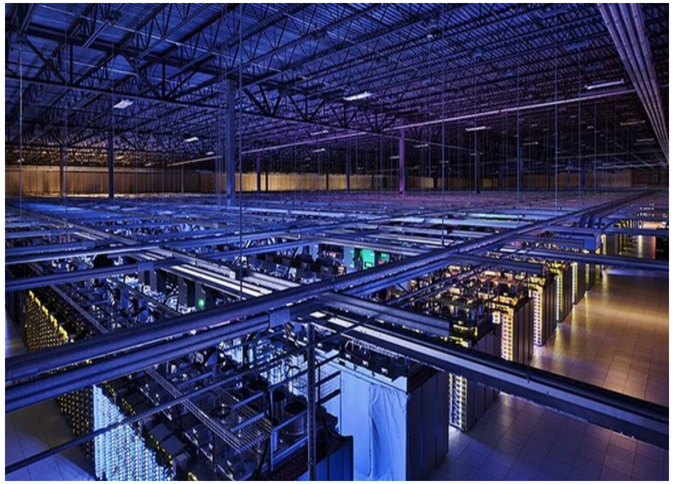
ASIC design: physical circuit layout, test case selection, ...

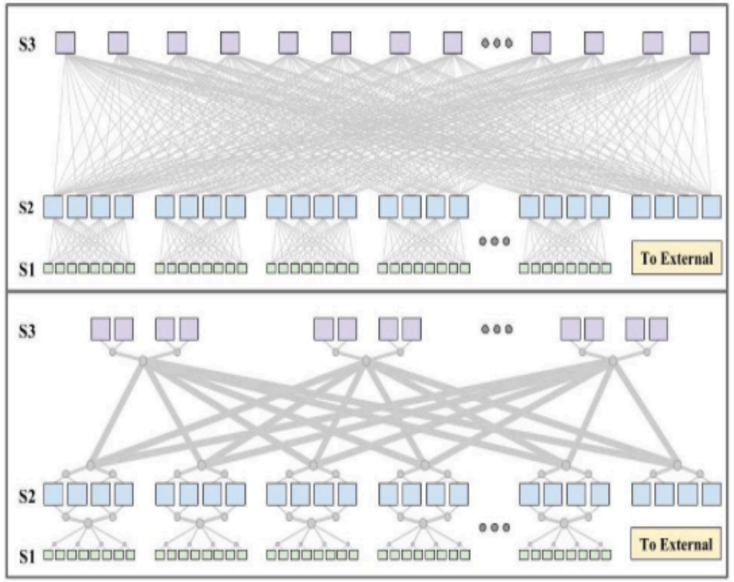
Outline

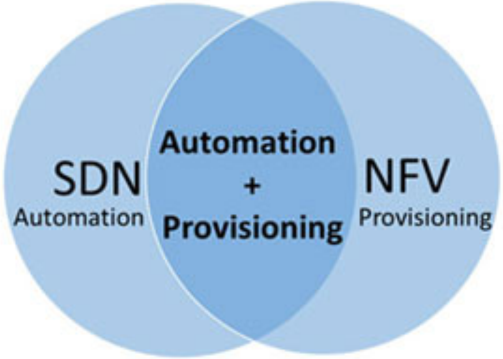
- AI is changing the world
- AIOps: AI for IT Operations and Autonomous IT Operations
  - What is AIOps
  - Value of AIOps: brief case studies
  - Industry Leader’ s Opinion
  - *Is AIOps necessary?*
  - Is AIOps feasible?
  - An in-depth case study
- Operations center tour

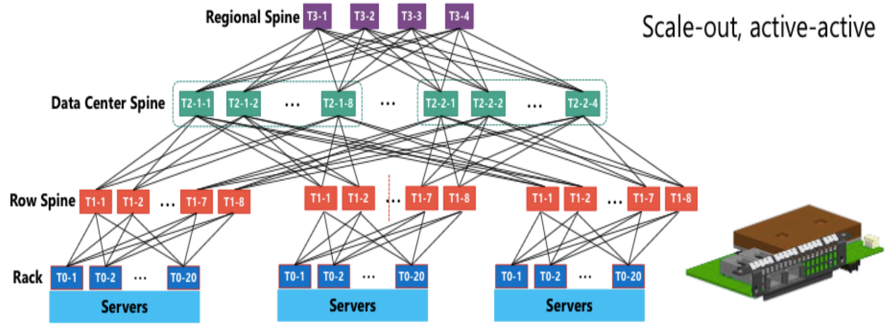
Complex Access Networks

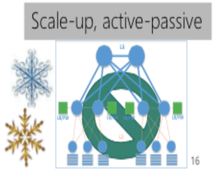






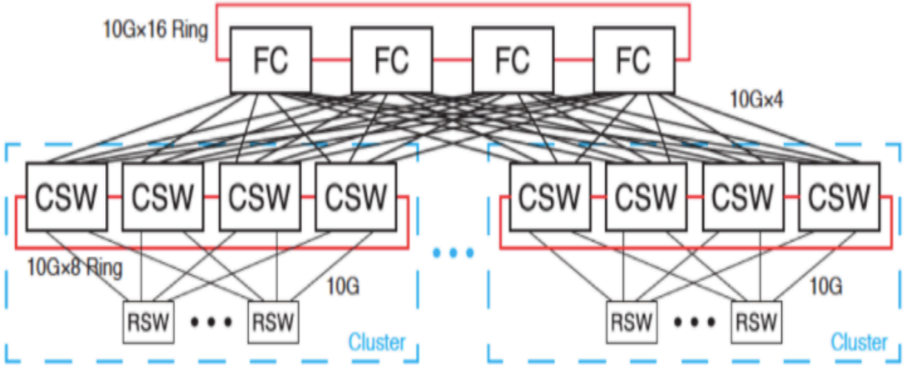




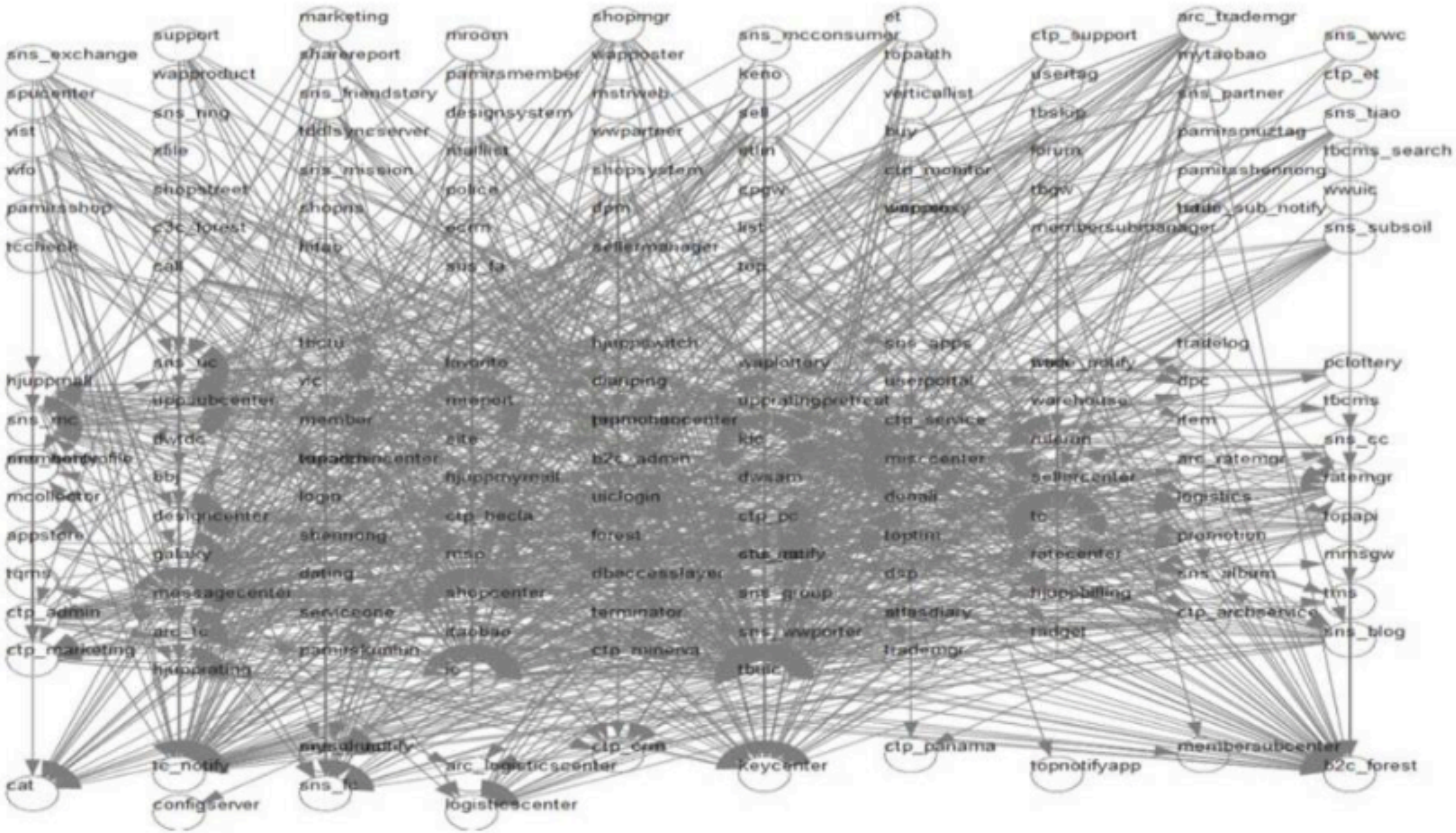


Outcome of >10 years of history, with major revisions every six months

Microsoft

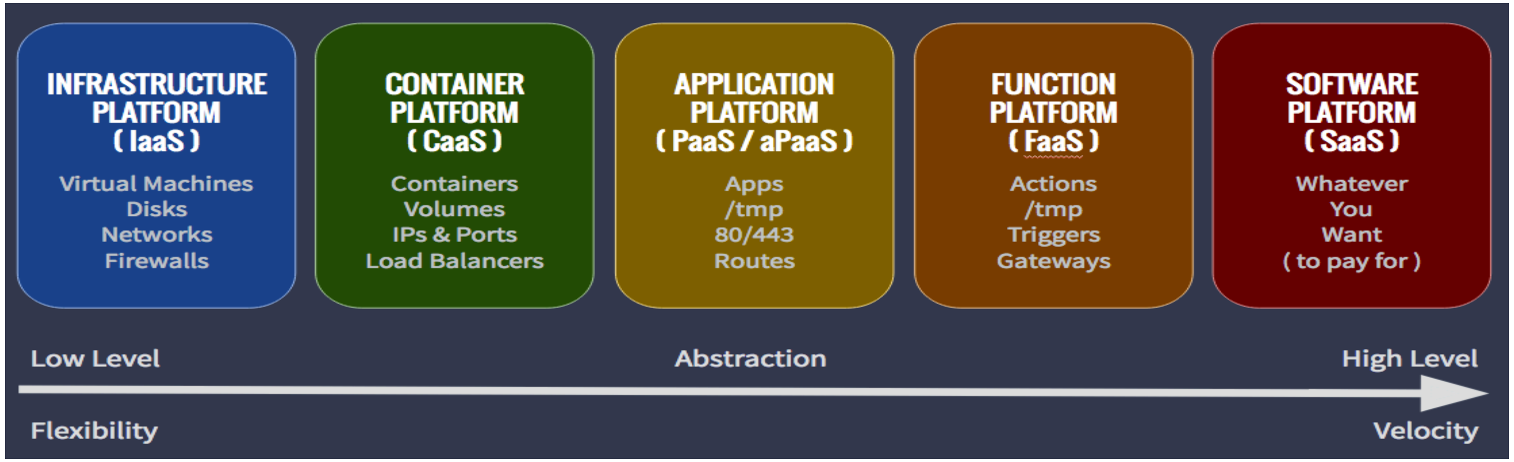


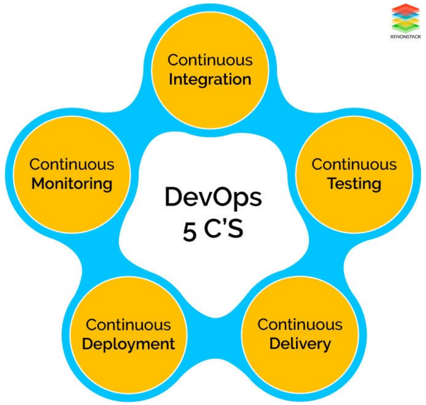
Taobao’ s application dependency in 2012

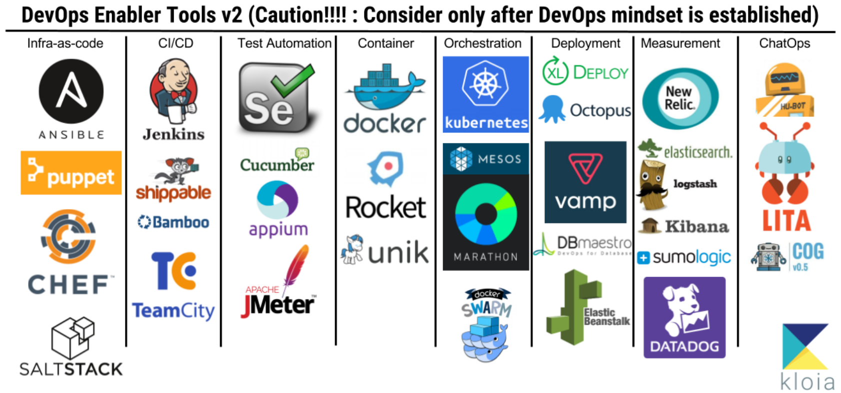


2012 淘宝核心链路应用拓扑图

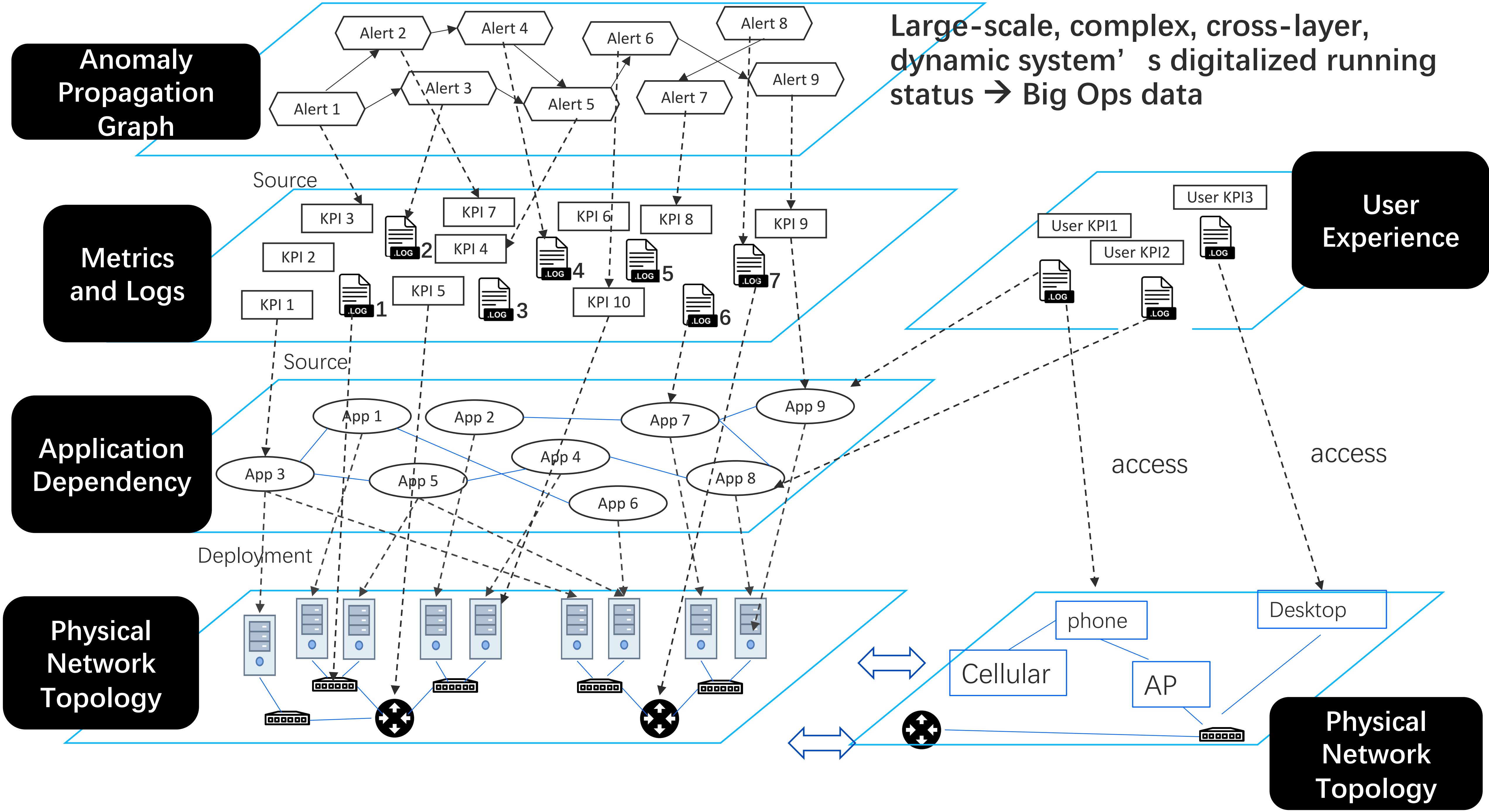
Evolving Techniques Enable Frequent Software Changes



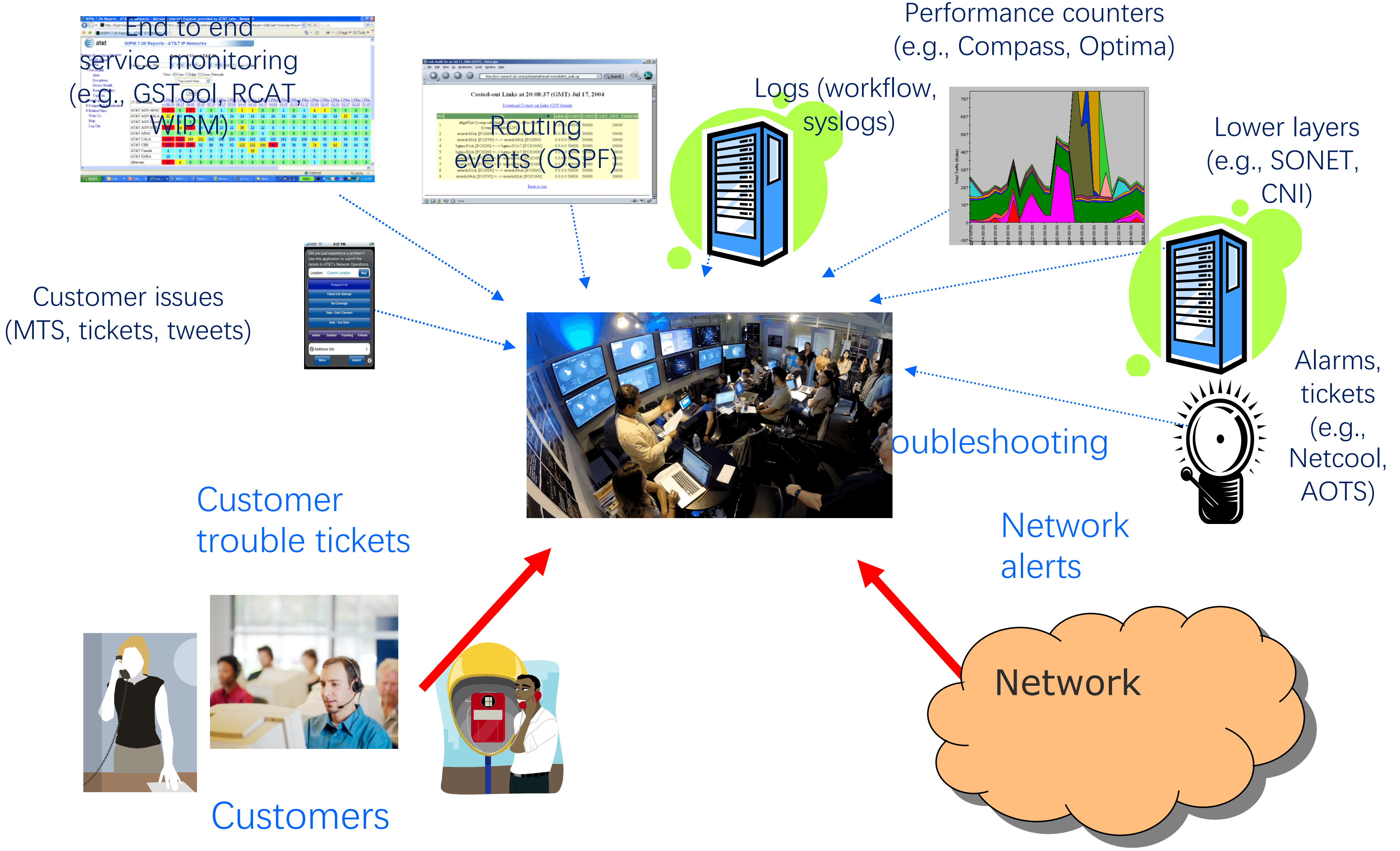




Large-scale, complex, cross-layer, dynamic system’ s digitalized running status → Big Ops data

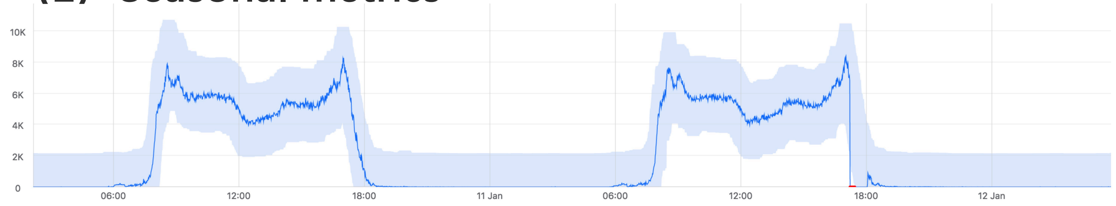


There are a sheer volume of device-generated monitoring data during daily operations

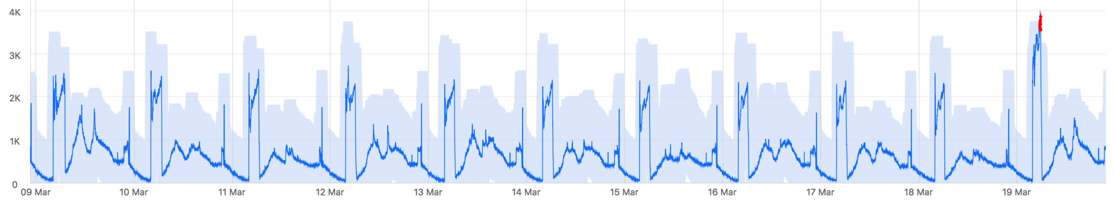


Diverse Metrics and Their Diverse Anomalies

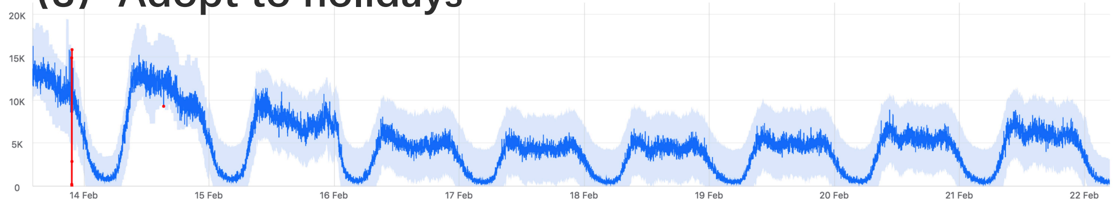
(1) Seasonal metrics



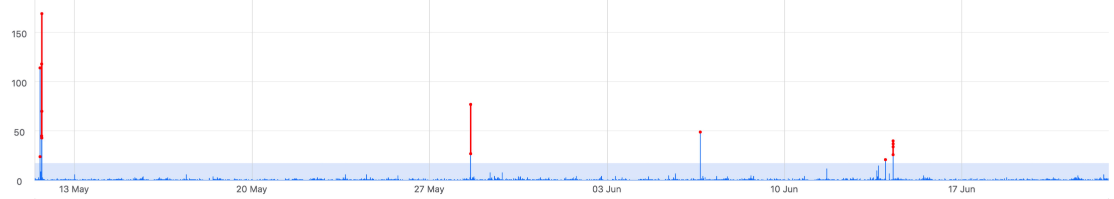
(2) Periodicity shift



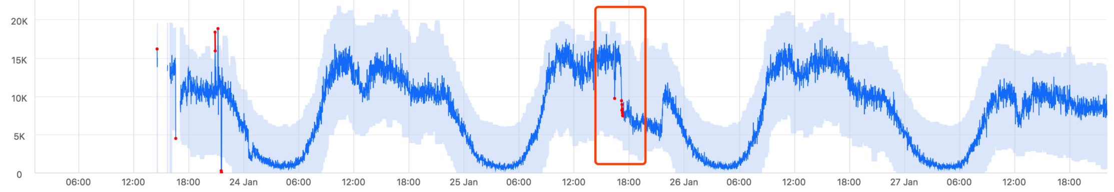
(3) Adopt to holidays



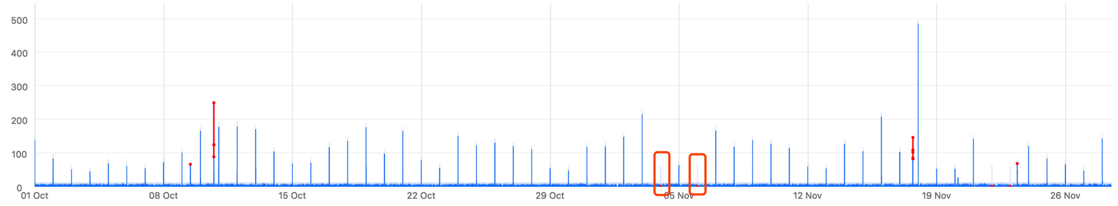
(4) Identify variable metrics and obtain extreme threshold



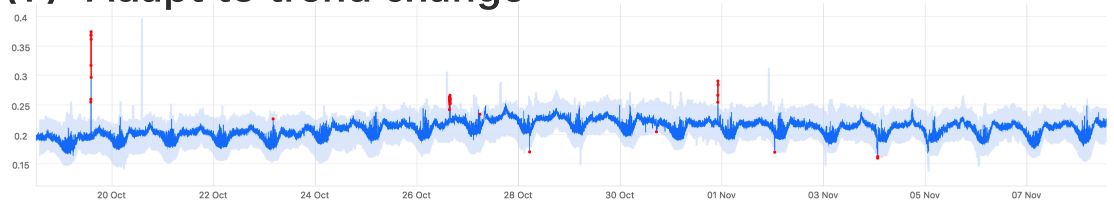
(5) Detect too rapid a change



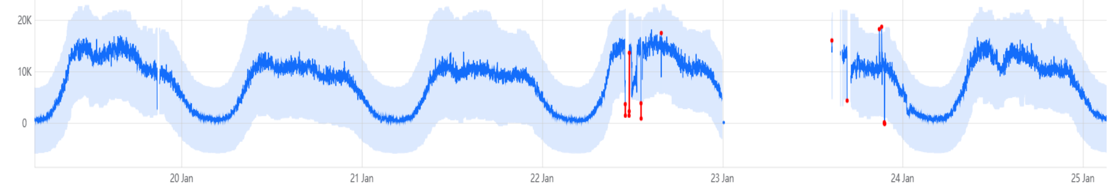
(6) Detect the lack of seasonality.



(7) Adapt to trend change



(8) Robust against data loss or interruption



There are more than one thousand types of logs in top 20 banks in China

App logs

Network Device Logs

- 交换机日志
- 路由器日志
- 防火墙日志
- F5日志
- …

OS logs

- UNIX日志
- Linux日志
- Windows日志
- …

Environment Logs

- 电力日志
- …

DB logs

- Oracle日志
- DB2日志
- Informix日志
- SQLServer日志
- MySQL日志
- …

Middlge-ware Logs

- MQ日志
- Tuxedo日志
- Weblogic日志
- Tomcat日志
- Apache日志
- …

```
2018-10-10 20:53:51,194 [JAgentSocketServer.cpp:121] WARN  agent 9995 - Listening Port : 20510
2018-10-10 20:53:51,194 [RequestHandlerService.cpp:189] WARN  agent 9995 - RequestHandlerService::handle_input(ACE_HANDLE=38)
2018-10-10 20:53:51,195 [ResponseCOUNT.cpp:159] INFO  agent 9995 - IO: Command (1) INITIALISE_PROCESS
2018-10-10 20:53:51,195 [ResponseCOUNT.cpp:302] INFO  agent 9995 - ResponseCOUNT: rc=0
2018-10-10 20:53:51,199 [ResponseCOUNT.cpp:159] INFO  agent 9995 - IO: Command (2) INITIALISE_ROOT
2018-10-10 20:53:51,199 [ResponseCOUNT.cpp:302] INFO  agent 9995 - ResponseCOUNT: rc=0
2018-10-10 20:53:51,204 [ResponseCOUNT.cpp:159] INFO  agent 9995 - IO: Command (3) INITIALISE_THREAD
```

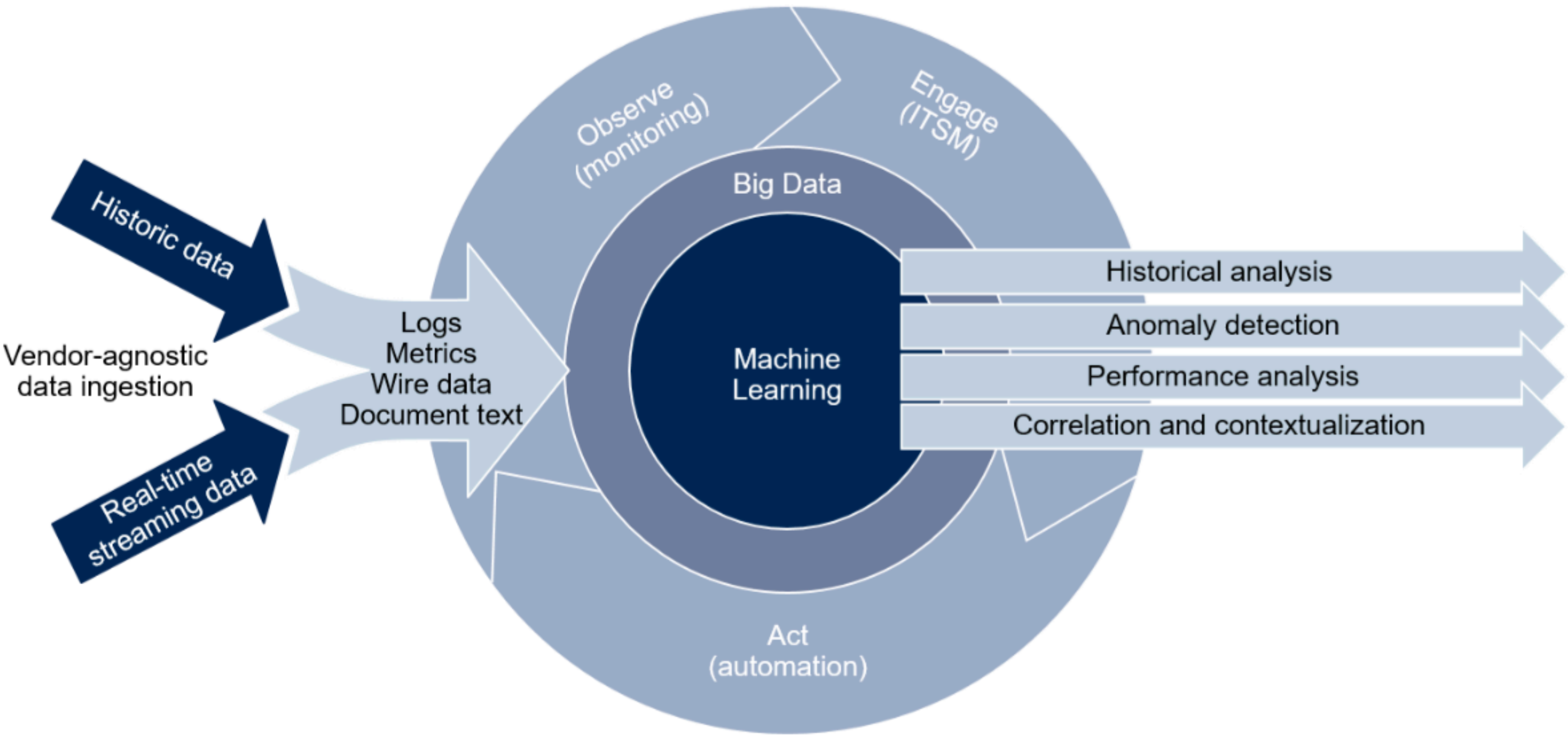
```
 INFO [WebContainer : 15] - queryForList:IDA_TEMPLATE.LISTDATA_MOST_CLICK
 INFO [WebContainer : 8] - queryForList:IDA_NOTICE.LISTDATA_BY_USER
com.teradata.ida.auth.dto.SysUserVO@2c3d3e1d
[8/10/18 8:29:31:581 CST] 00000032 SystemOut     O  INFO [WebContainer : 1] - queryForList:IDA_TEMPLATE_AUTH.findTemplateByRoleId
DEBUG [WebContainer : 7] - 2018-08-10 08:29:32 DEBUG |CsParamSetAction|showAtomsBygid|Start||start=0|limit=25|page=1|fromIndex=0|toInd
 INFO [WebContainer : 7] - queryForList:SEG_BIZ_ATOM_DEF.findAtomByRoleAndShowArea
```

```
EXPLANATION:
Channel program 'CS_EDI_S' ended abnormally.
ACTION:
Look at previous error messages for channel program 'CS_EDI_S' in the error
files to determine the cause of the failure.
----- amqrmrsa.c : 487 -----------------------------------------------
08/07/2018 10:14:54 AM - Process(29670.329016) User(mqm) Program(amqrmppa)
AMQ9513: Maximum number of channels reached.
```

We have no choice but relying on AI to take advantage of the Big Data from Ops

- Volume
- Velocity
- Variety
- Value

AIOps Platform Enabling Continuous ITOM



ID: 340492

Outline

- AI is changing the world

- AIOps: AI for IT Operations and Autonomous IT Operations
    - What is AIOps
    - Value of AIOps: brief case studies
    - Industry Leader’s Opinion
    - Is AIOps necessary?
    - *Is AIOps feasible?*
    - Levels of AIOps
    - An in-depth case study

- Operations center tour

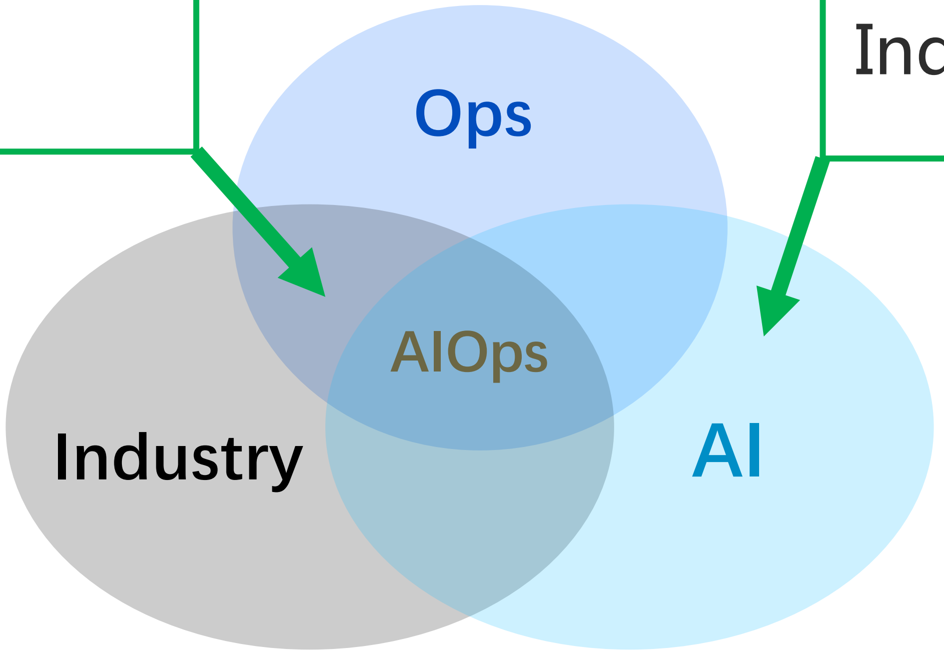
AIOps has *the necessities required for successful ML applications*

- Machine learning tools (algorithms and systems)
- *Applications that show the value*
- *Large amount of data*
- *Labels and the experts who can label*

AIOps is still challenging because its interdisciplinary nature

Ops people familiar with Ops and Industry,
but not AI

Algorithm people familiar with general AI,
but not Ops and Industry



Pitfalls: use ML algorithms as Blackbox to tackle Ops challenges

Failure Discovery → Failure Mitigation → Failure Repair → Failure Avoidance

↑

Huge Gap

↑

General Machine Learning Algorithms

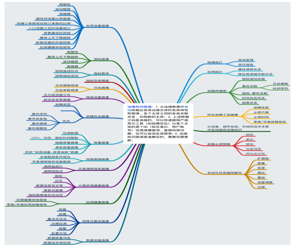
ARIMA, Time Series Decomposition, Holt-Winters, CUSUM, SST,DiD,DBSCAN, Pearson Correlation, J-Measure, Two-sample test, Apriori, FP-Growth, K-medoids, CLARIONS, Granger Causality, Logistic Regression, Correlation analysis (event-event, event-time series, time series-time series) , hierarchical clustering、Decision tree, Random forest, support vector machine, Monte Carlo Tree search, Marcovian Chain, multi-instance learning, transfer learning, CNN, RNN ,VAE, GAN, NLP

AIOps Architecture： Divide the complex task and Conquer

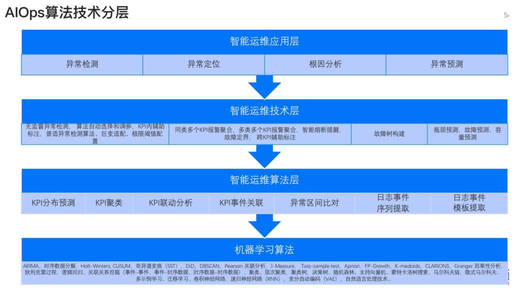
(1) Abundant data
(2) Deterministic information
(3) Complete information
(4) Well defined
(5) Single domain

These two types of modules must be solvable by existing ML algorithms

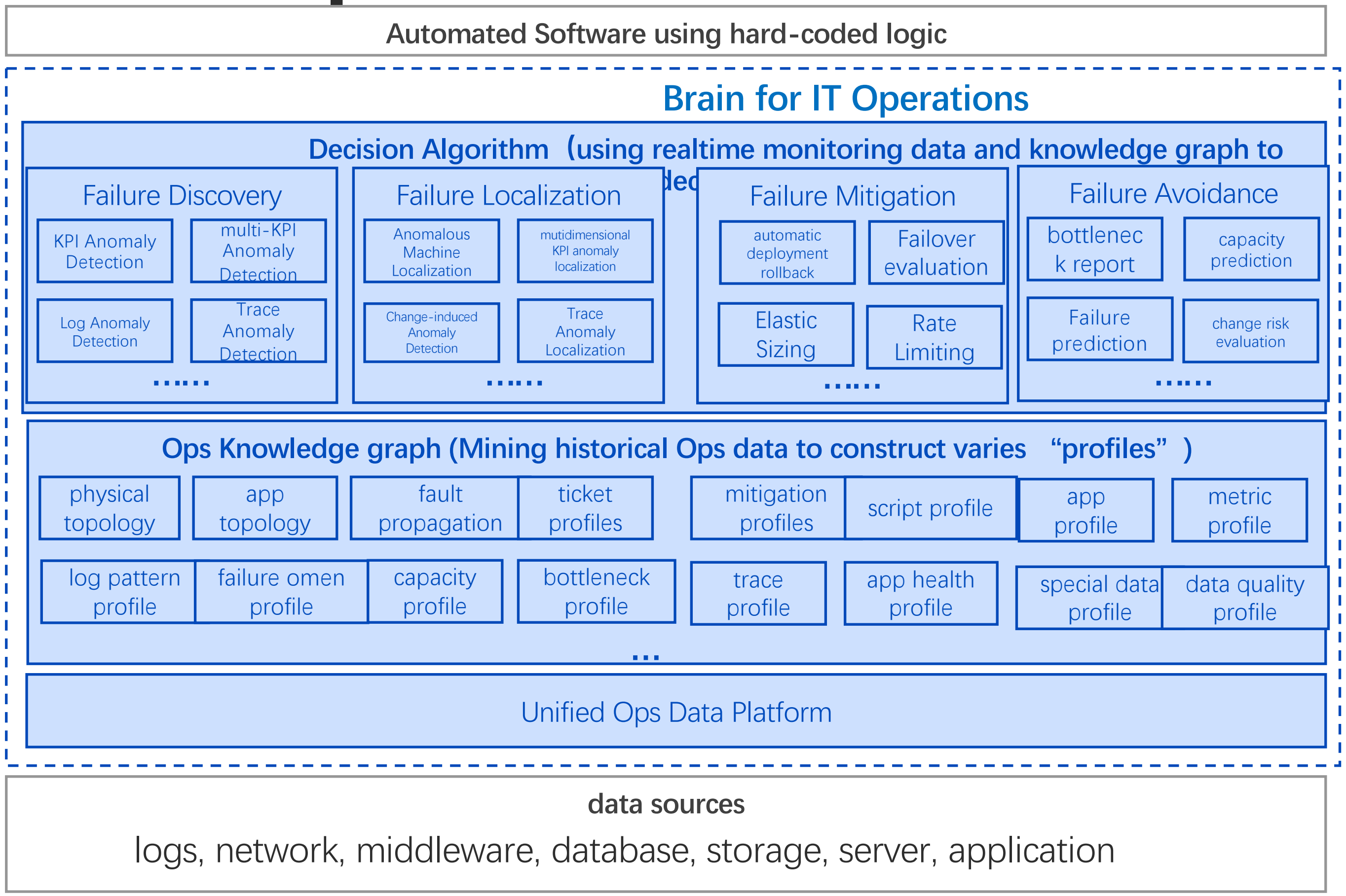
Brain: Knowledge Graph



Brain：

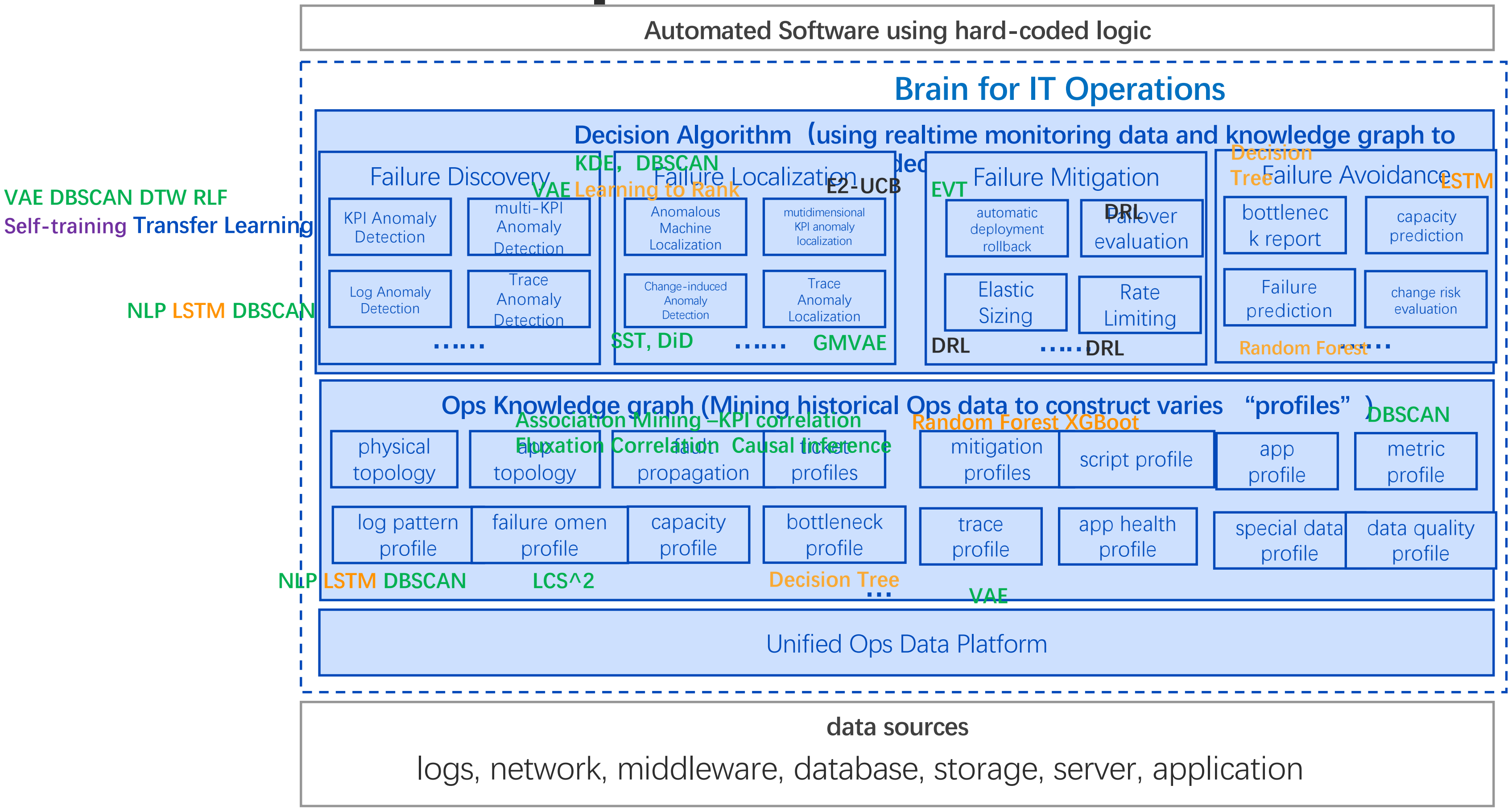


Brain for IT Operations



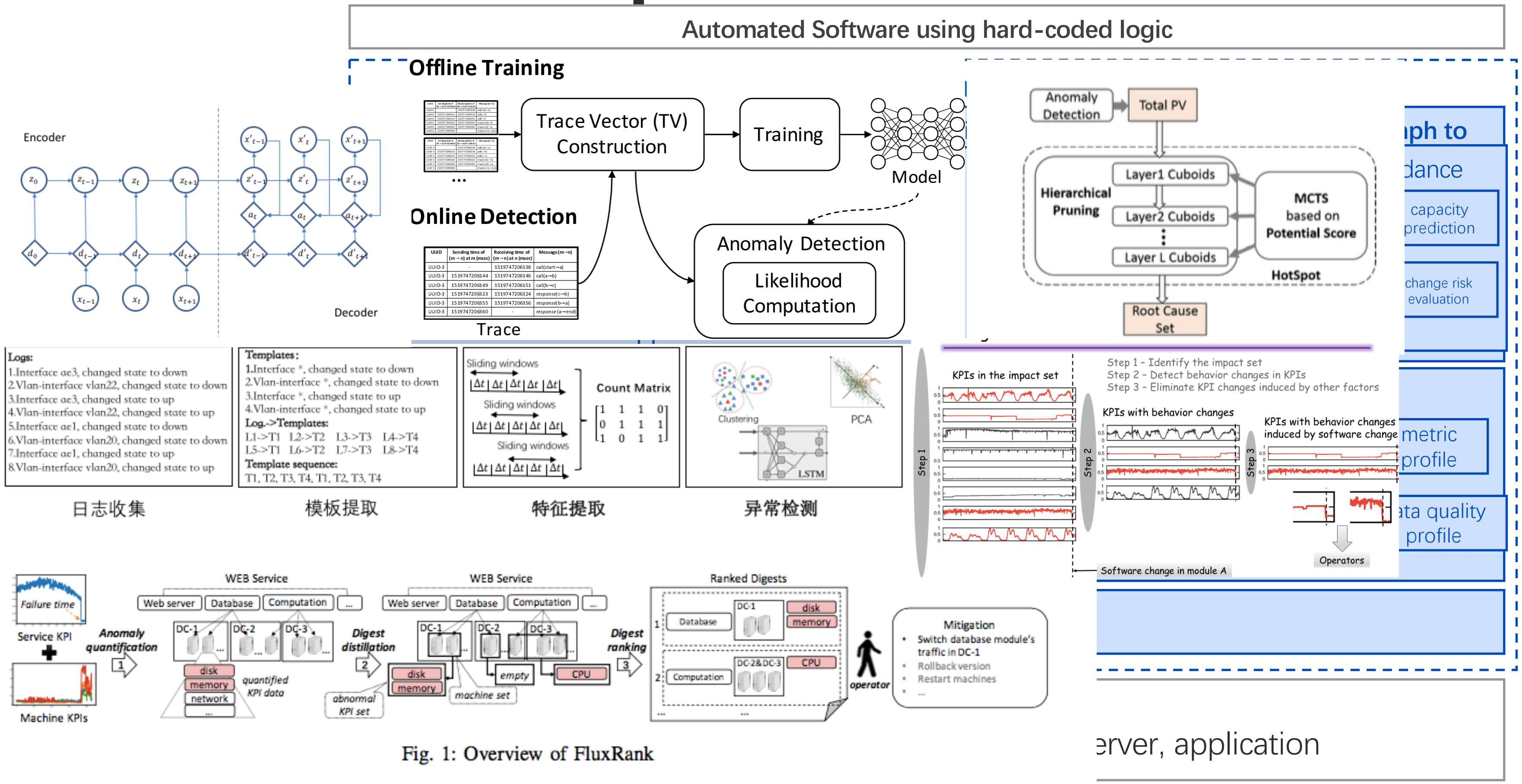
Unsupervised Reinforcement Learning Supervised but with labels Semi-supervised Learning Transfer Learning

Brain for IT Operations



Unsupervised Reinforcement Learning Supervised but with labels Semi-supervised Learning Transfer Learning

Brain for IT Operations



Unsupervised Reinforcement Learning Supervised but with labels Semi-supervised Learning Transfer Learning

Outline

- AI is changing the world

- AIOps: AI for IT Operations and Autonomous IT Operations
  - What is AIOps
  - Value of AIOps: brief case studies
  - Industry Leader’ s Opinion
  - Is AIOps necessary?
  - Is AIOps feasible?
  - *Levels of AIOps*
  - An in-depth case study

- Operations center tour

Levels of Autonomous IT Operations

- Cores Per Op (CPO): The average number of x86bCPU cores managed by an Op (40hours/week)
- Assumption: Organization tries their best to achieve certain reliability.
- Try to decoupled with the following factors:
    - Business sectors, scale, architecture, technology, part-time
- Count operators of server, storage, network, middleware, database, application
- Count the hours of operators for triggering scripts, monitoring the big screen, browsing the monitoring data, deal with alerts, troubleshooting, planning, idle time while on duty.
- Do not count the hours of operators for developing IT operations tools.

Level=⌊ Log (CPO/100) ⌋	Cores Per Op (CPO)	Typical Enterprises
Level 0	O(100)	Finance
Level 1	O(1K)	Medium Internet companies running on public clouds
Level 2	O(10K)	Large Internet companies
Level 3	O(100K)	
Level 4	O(1M)	
Level 5	O(10M)	

Example1：Internet Company A

- All x86 servers: 500K with 12 cores each, 500K with 24 cores each。In total there are 13M cores.

- Labor: (200*0.5+200*0.8)*60/40=390 Op
  - 200 operators for server, storage, database, and network
    - 60 hours/week; 50% of working time is for manual operations, and 50% of working time is for tool development.
  - 200 operators for applications and middleware
    - 60 hours/week; 80% of working time is for manual operations

- CPO=13M cores/390 Op=33K cores/Op

- **Level =⌊ Log (CPO/100) ⌋=2**

Example2 : Internet Company B

- All x86 servers: 500K with 12 cores each, 500K with 24 cores each。In total there are 13M cores.

- Labor: (200*0.5+200*0.8)=130 Op
  - 100 operators for server, storage, database, and network
    - 40 hours/week; 50% of working time is for manual operations, and 50% of working time is for tool development.
  - 100 operators for applications and middleware
    - 40 hours/week; 80% of working time is for manual operations

- CPO=13M cores/130 Op=100K cores/Op

- **Level =⌊ Log (CPO/100) ⌋=3**

Example 3： Bank C

- 10K x86 servers with 12 cores each. 500 small computers, each equivalent to 100 cores. 5 Mainframe computers, each equivalent to 2K cores. 180K cores in total

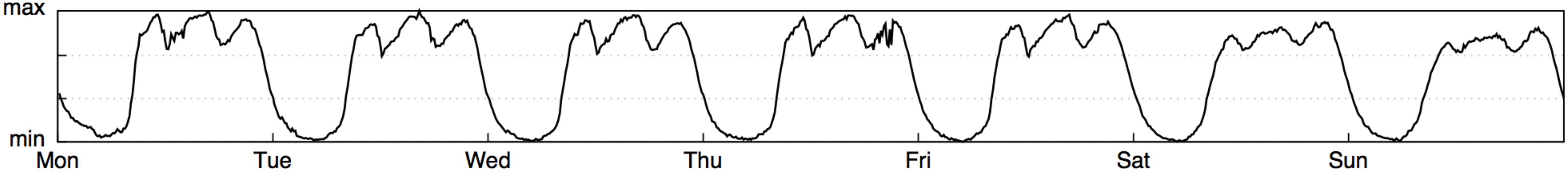
- Labor (100*0.5+100*0.8+200)*60/40=495 Op
  - 100 operators for server, storage, database, and network
    - 60 hours/week; 50% of working time is for manual operations, and 50% of working time is for tool development.
  - 100 operators for applications and middleware
    - 60 hours/week; 80% of working time is for manual operations
  - 200 Outsourced Operators
    - 60 hours/week; full time on manual operations

- CPO=180K Cores/495 Op=363/Op
- **Level =⌊ Log (CPO/100) ⌋=0**
- plan to have 100K x86 servers, and the number of cores increases to 1.26M
  - Keep the CPO, and increase the #Ops to to 1.26M/263=3360, or
  - Keep the #Op=495, but increase the **CPO=1.26M/495=3545 cores/Op; Level=1**

Outline

- AI is changing the world
- AIOps: AI for IT Operations and Autonomous IT Operations
  - What is AIOps
  - Value of AIOps: brief case studies
  - Industry Leader’ s Opinion
  - Is AIOps necessary?
  - Is AIOps feasible?
  - Levels of AIOps
  - *An in-depth case study*
- Operations center tour

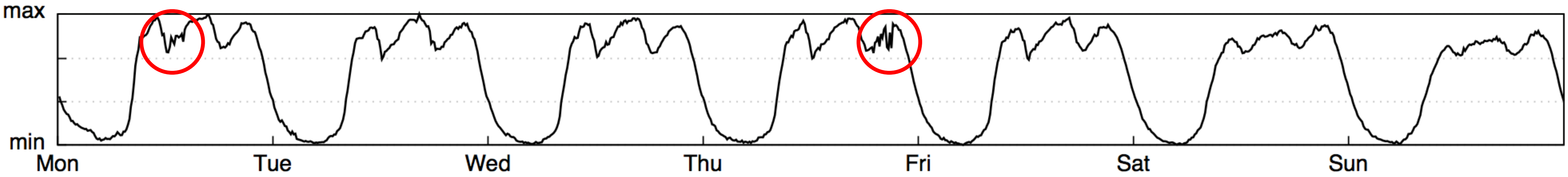
KPIs and Anomaly Detection



Page views (PV) of Baidu

KPIs (Key Performance Indicators): A set of performance measures that evaluate the service quality

KPIs and Anomaly Detection

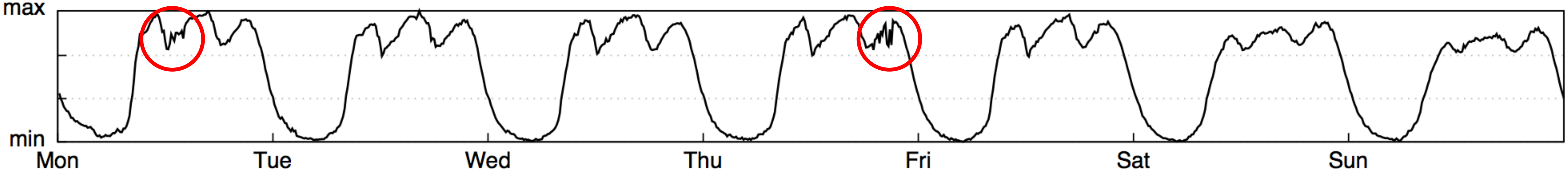


Page views (PV) of Baidu

KPIs (Key Performance Indicators): A set of performance measures that evaluate the service quality

KPI anomalous (unexpected) behaviors → Potential failures, bugs, attacks…

KPIs and Anomaly Detection



Page views (PV) of Baidu

KPIs (Key Performance Indicators): A set of performance measures that evaluate the service quality

KPI anomalous (unexpected) behaviors → Potential failures, bugs, attacks…

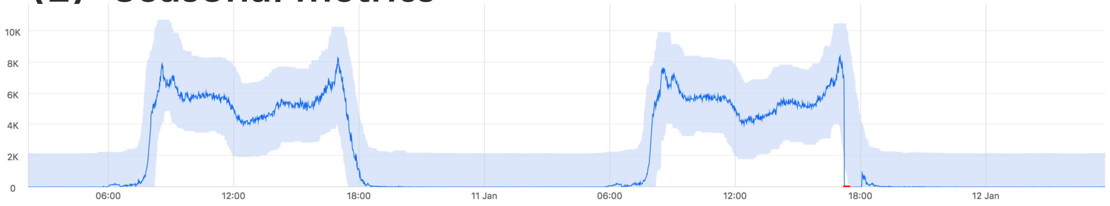
Anomaly detection matters: Find anomalous behaviors of the KPI curve

→ Diagnose and fix it

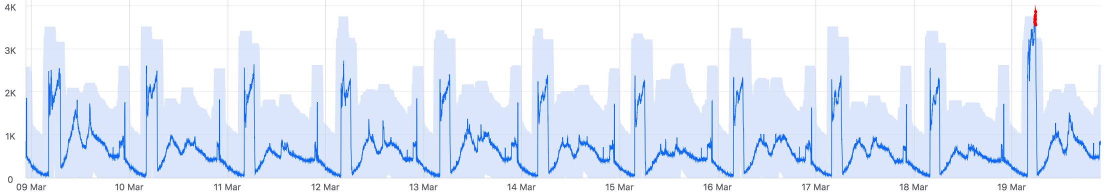
→ Avoid further influences and revenue losses

Diverse Metrics and Their Diverse Anomalies

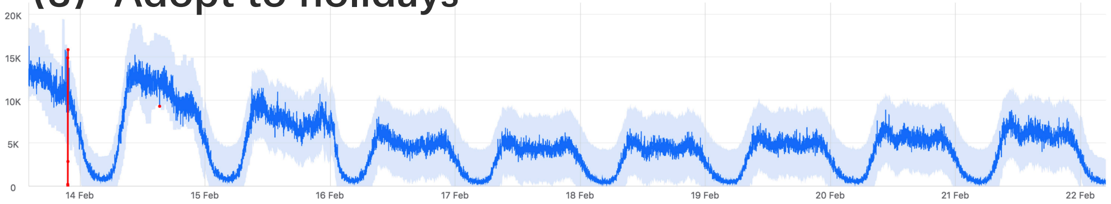
(1) Seasonal metrics



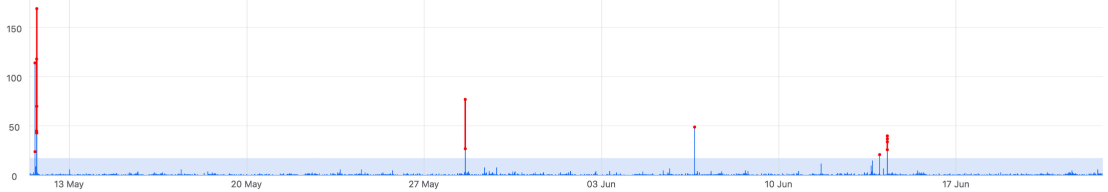
(2) Periodicity shift



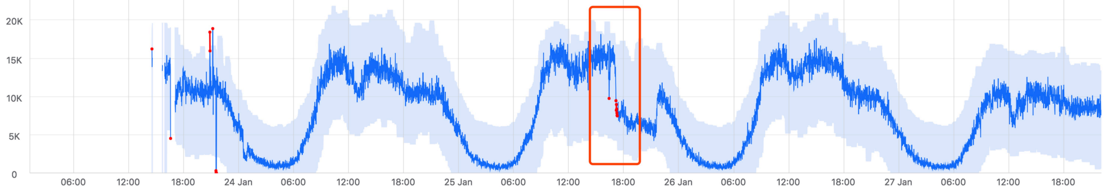
(3) Adopt to holidays



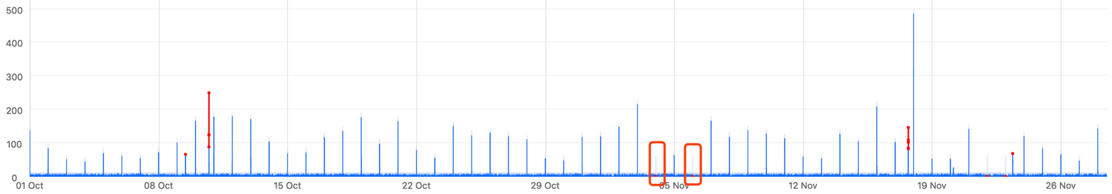
(4) Identify variable metrics and obtain extreme threshold



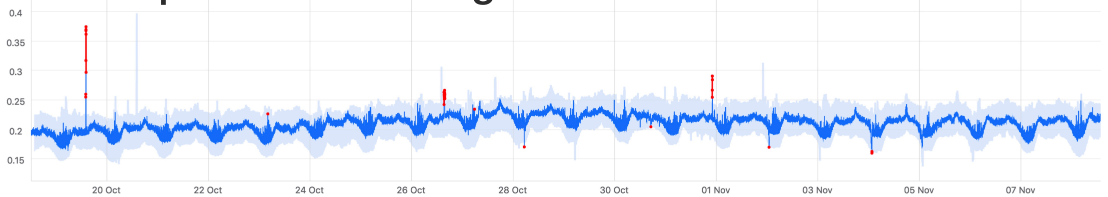
(5) Detect too rapid a change



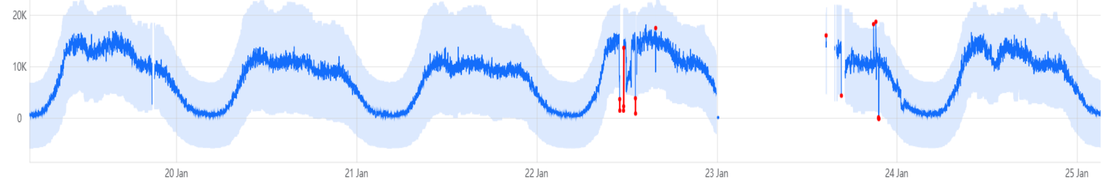
(6) Detect the lack of seasonality.



(7) Adapt to trend change



(8) Robust against data loss or interruption

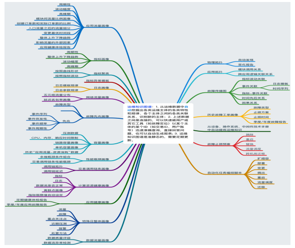


AIOps Architecture： Divide the complex task and Conquer

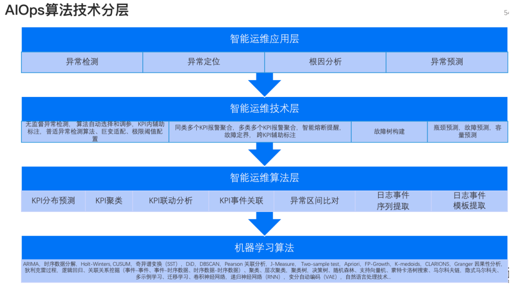
（1） Abundant data
(2) Deterministic information
（3） Complete information
（4） Well defined
（5） Single domain

These two types of modules must be solvable by existing ML algorithms

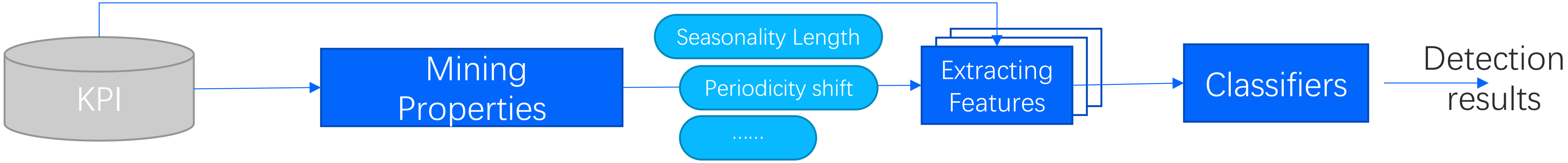
Brain: Knowledge Graph

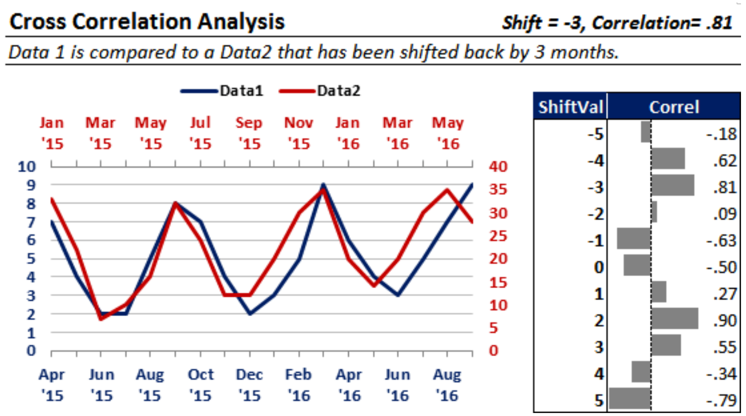


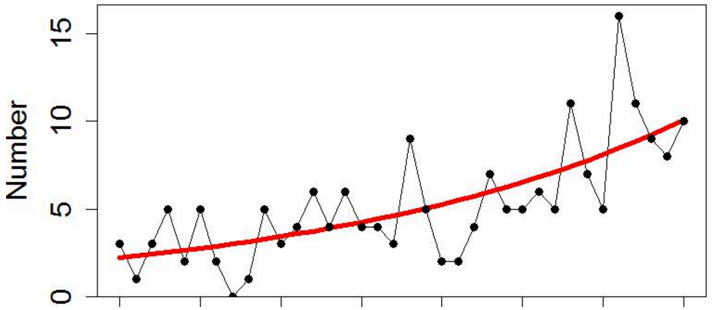
Brain：



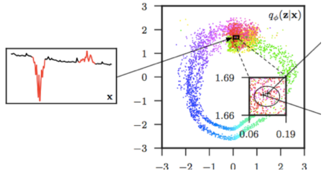
Architecture



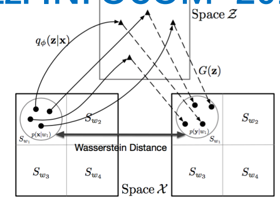




Donut: WWW2018



Buzz: INFOCOM 2019



Label-Less: INFOCOM 2019

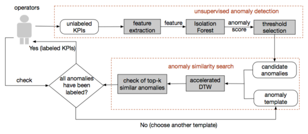
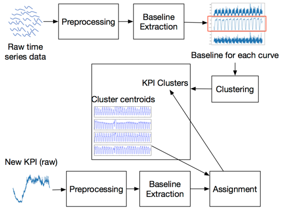
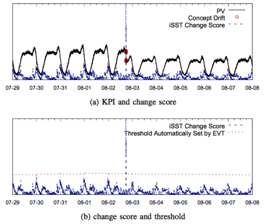


Figure 2: The overall framework of *Label-Less*.

ROCKA: IWQOS 2018



StepWise: ISSRE 2018 Best Paper



Donut: supervised->unsupervised: smooth KPIs

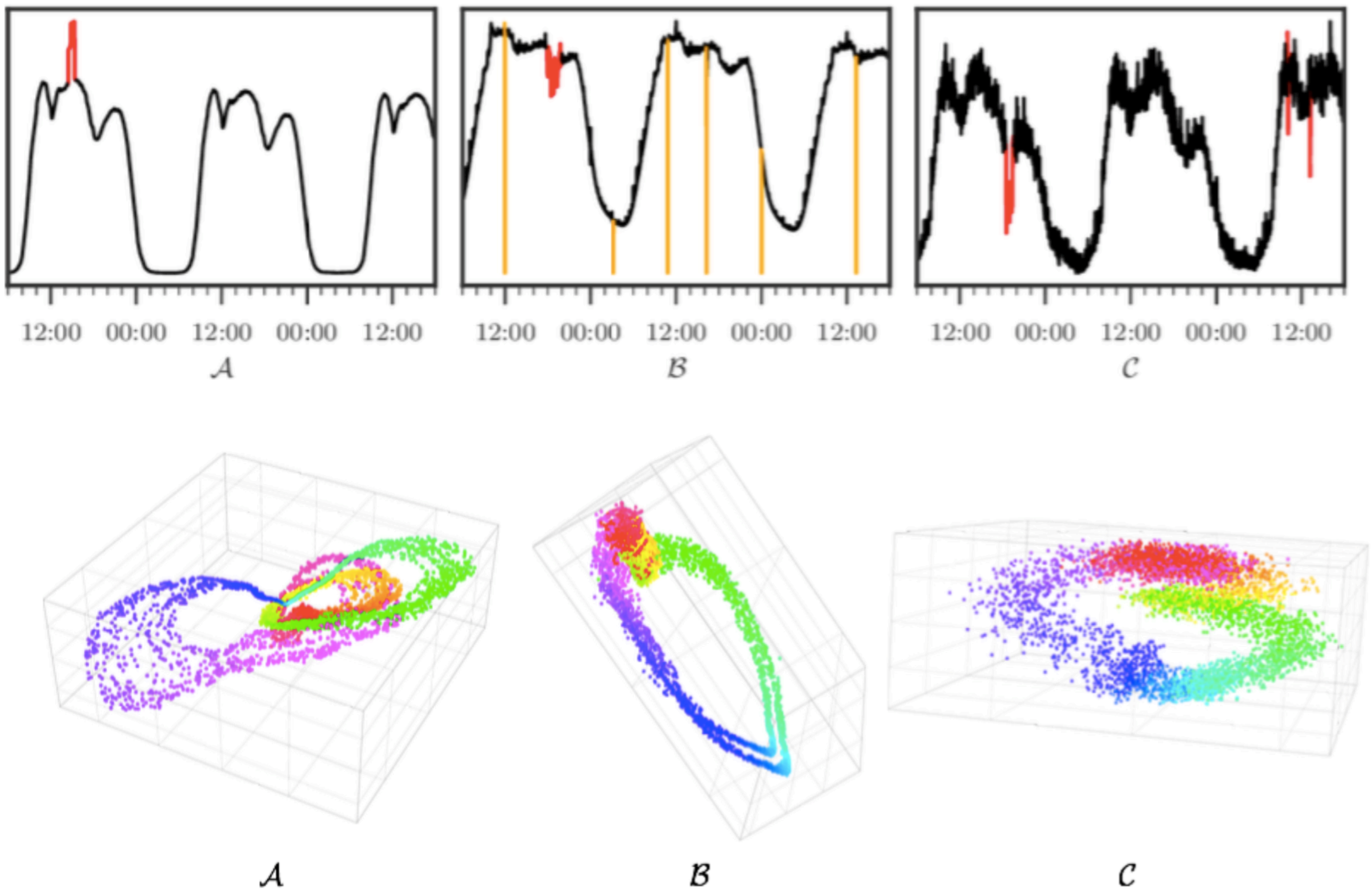
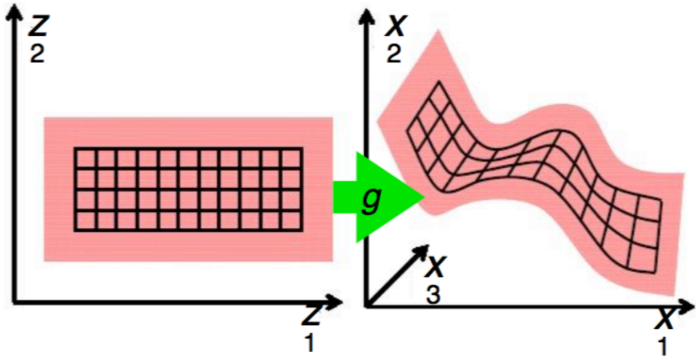
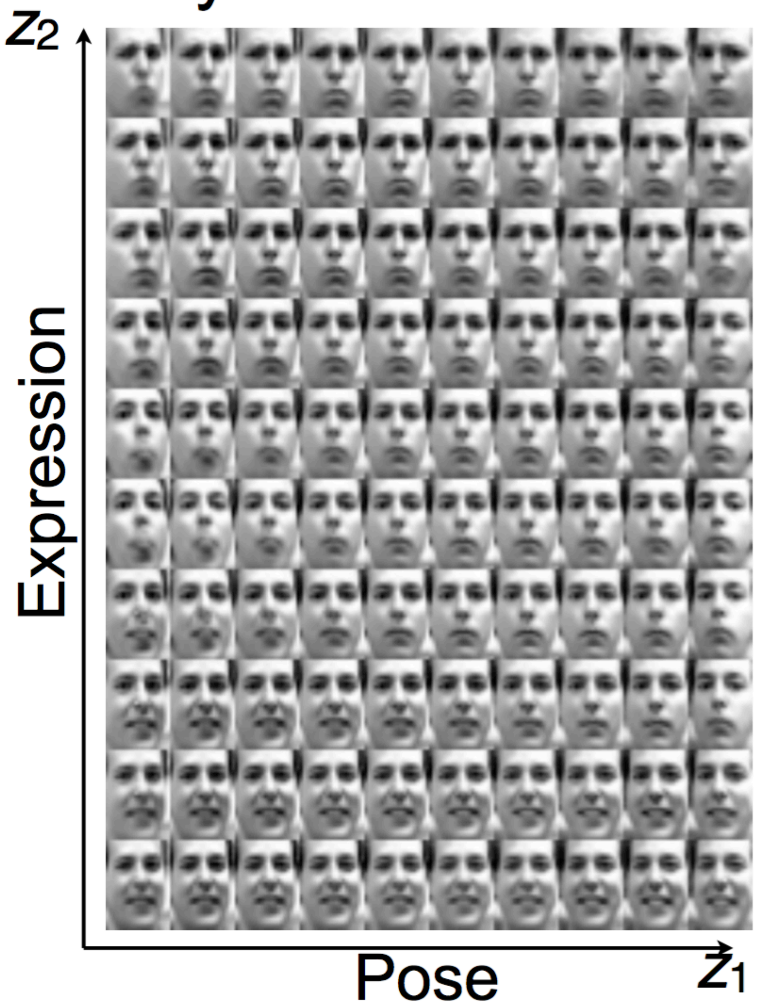


Figure 12: 3-d latent space of all three datasets.

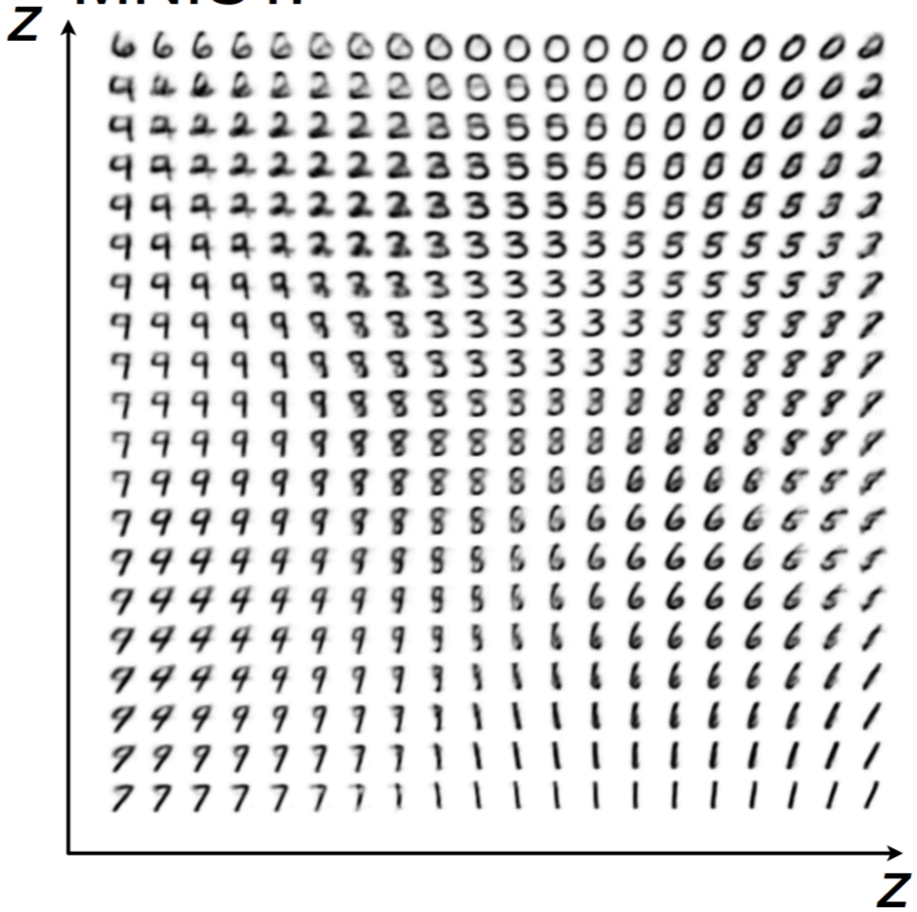
Latent Variable Models



Frey Faces:



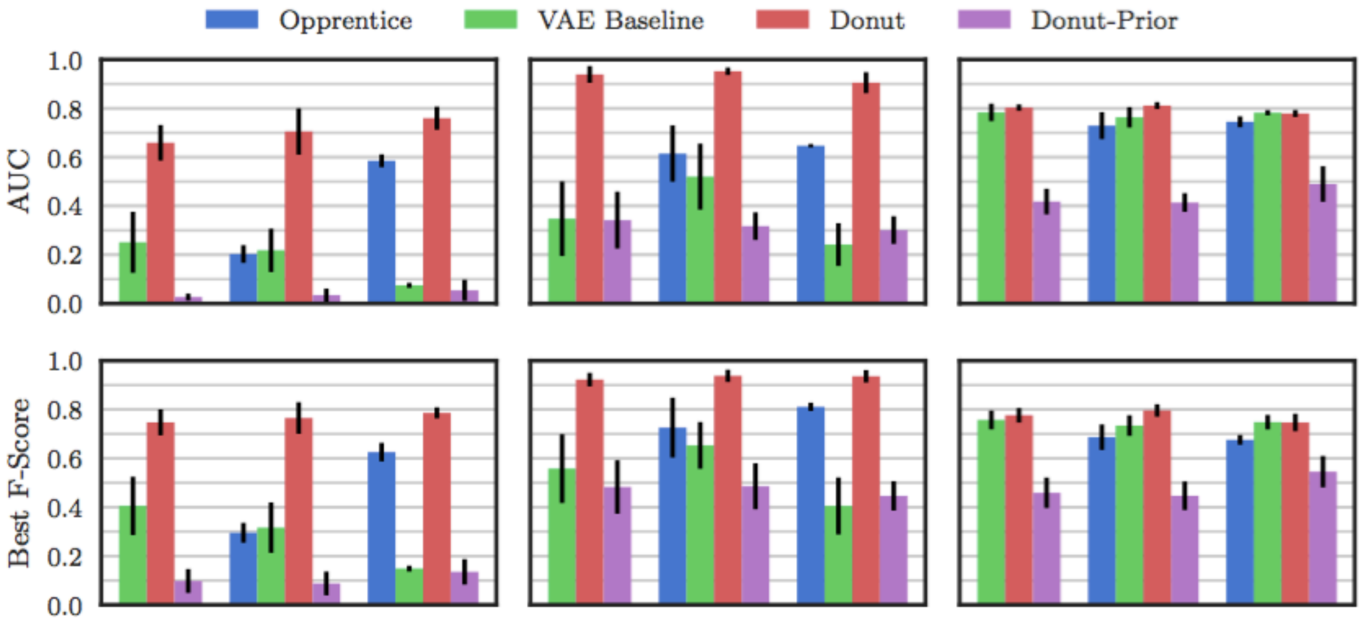
MNIST:

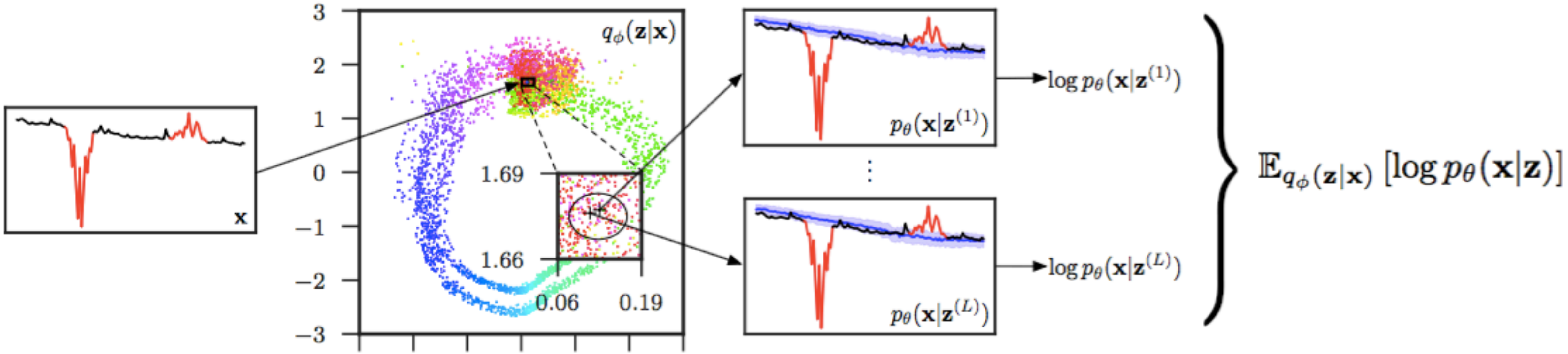


Unsupervised KPI Anomaly Detection Through Variational Auto-Encoder

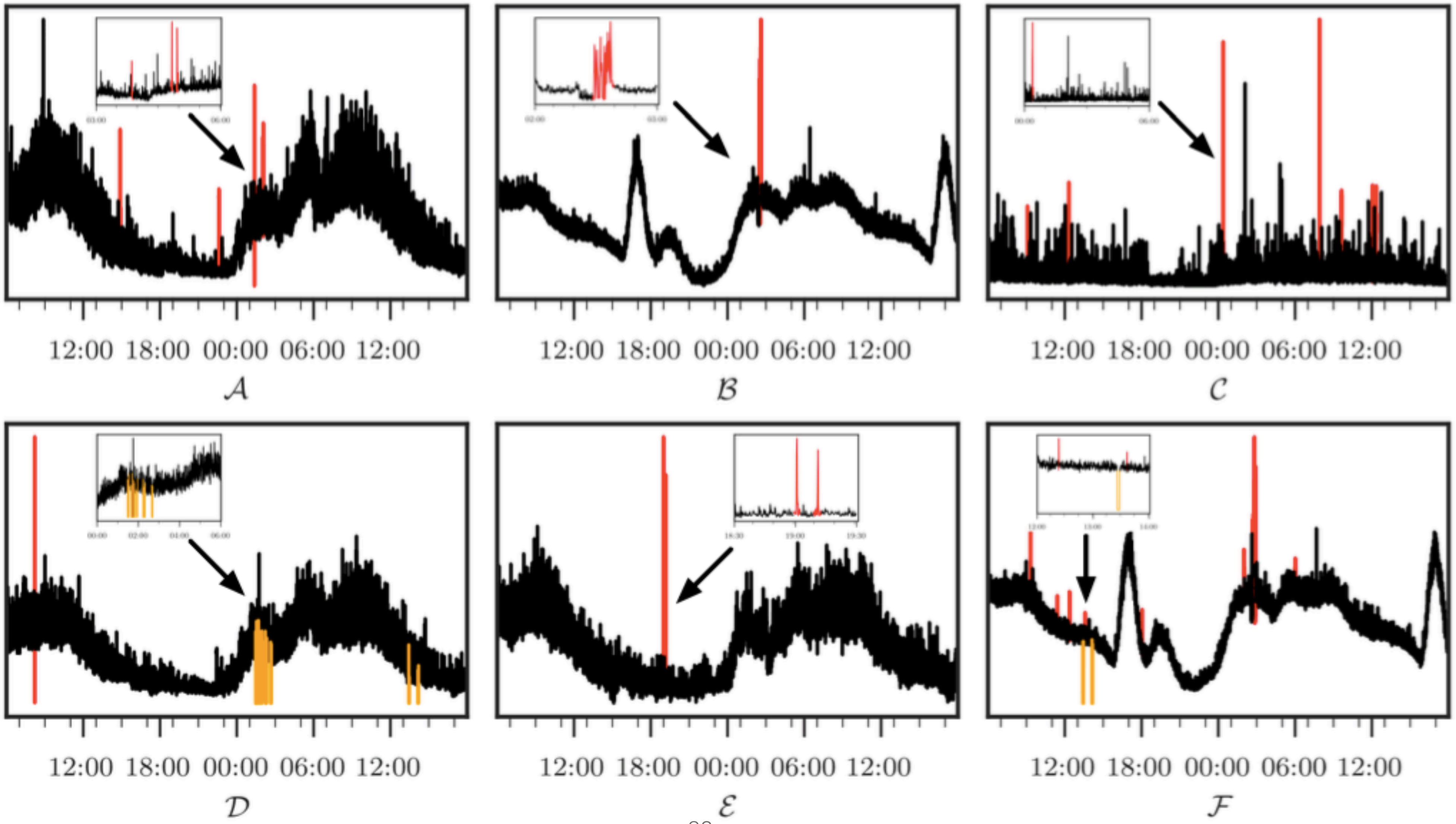
WWW2018

Accuracy of 0.8~0.9，even better than supervised approach.

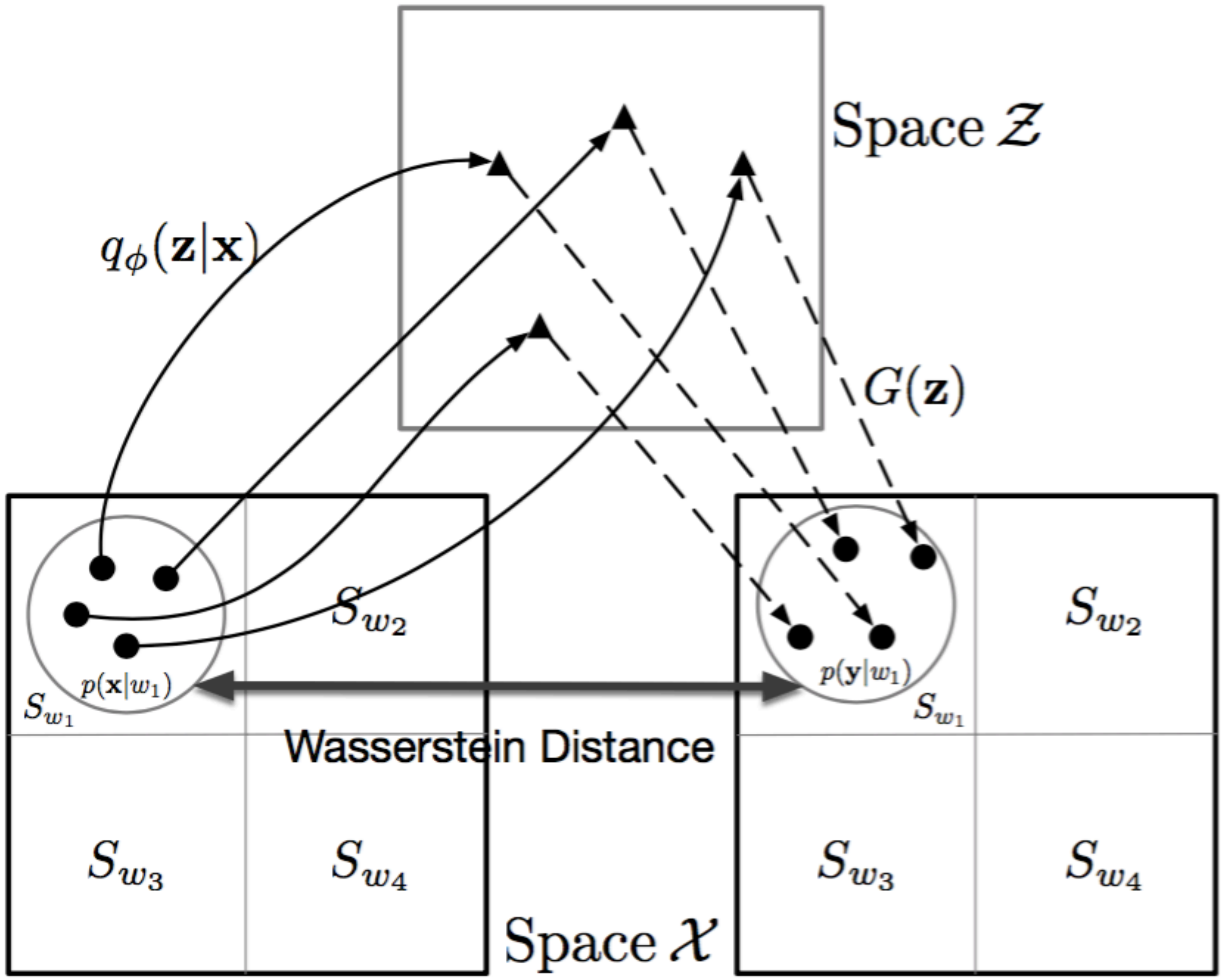




Buzz: Apply Adversarial Training for non-Gaussian noise



Unsupervised Anomaly Detection for Intricate KPIs via Adversarial Training of VAE



INFOCOM 2019

We use two major ideas in Buzz:

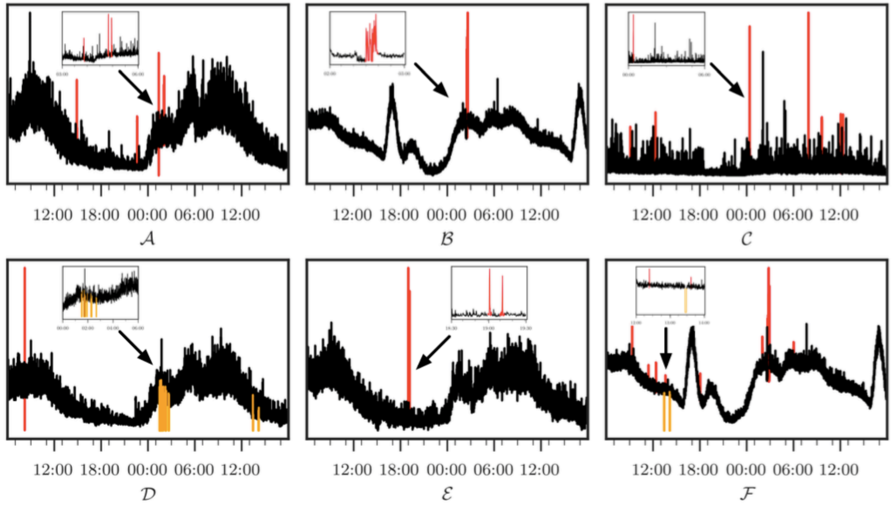
- Wasserstein distance: the distance between the two probability distributions

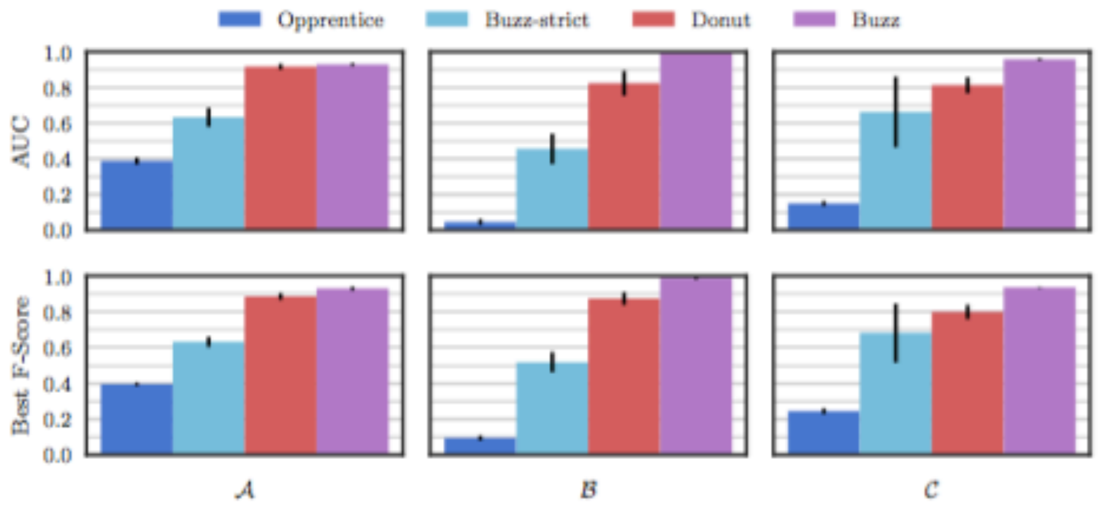
- Partitioning from measure theory.

a powerful and commonly used analysis method for distribution in measure theory.

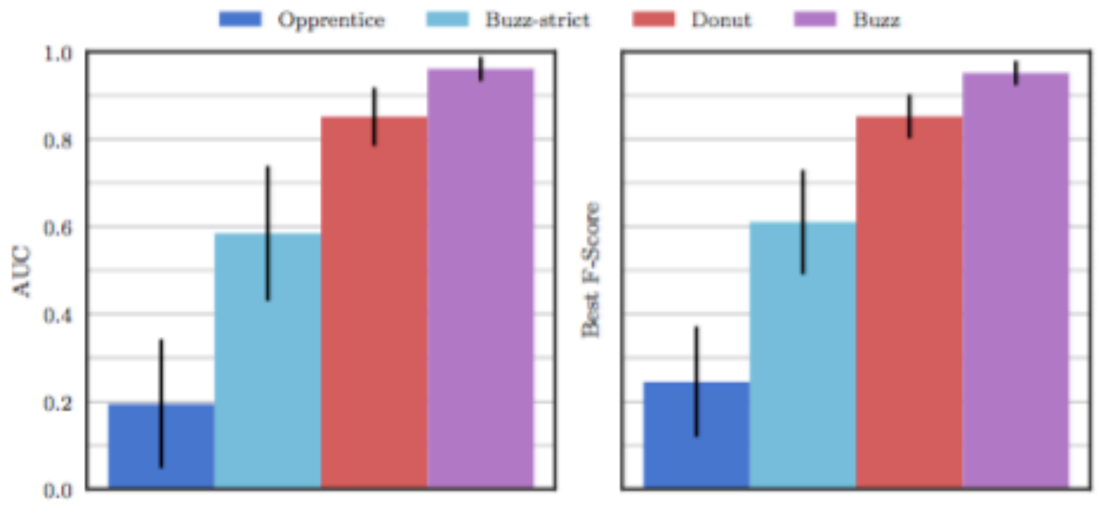
Experiment Results

Best F-Score outperforms Donut by up to 0.15





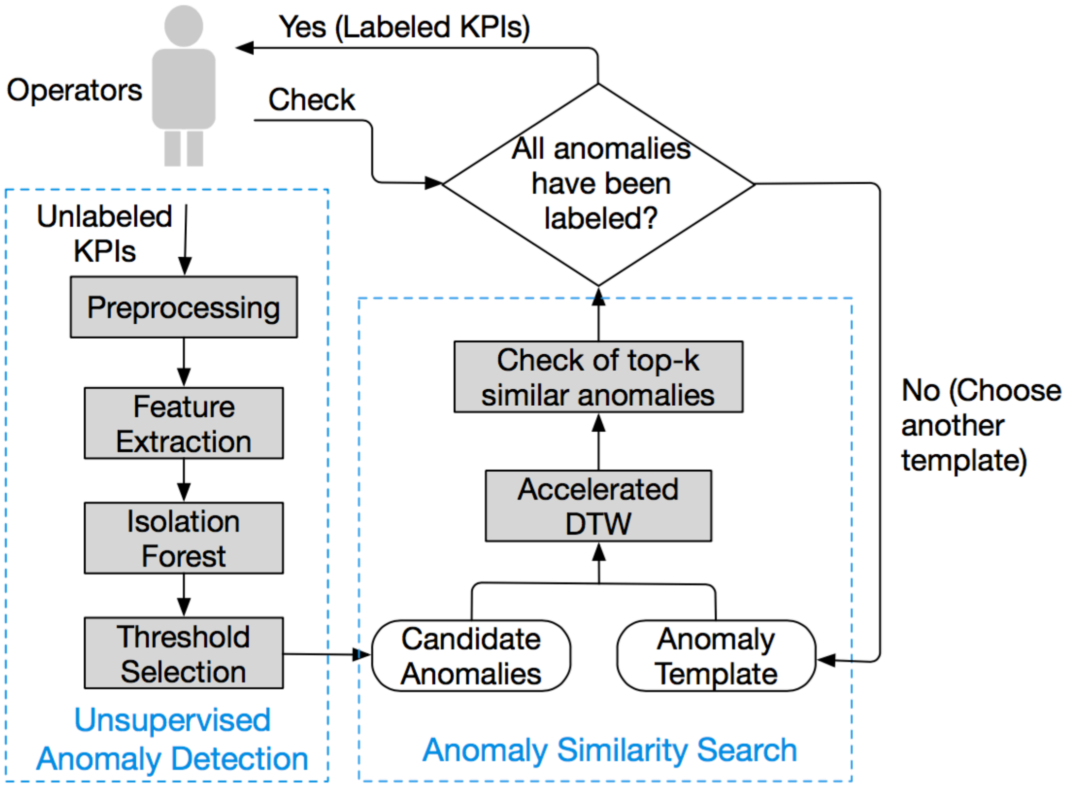
(a) Dateset A, B, C

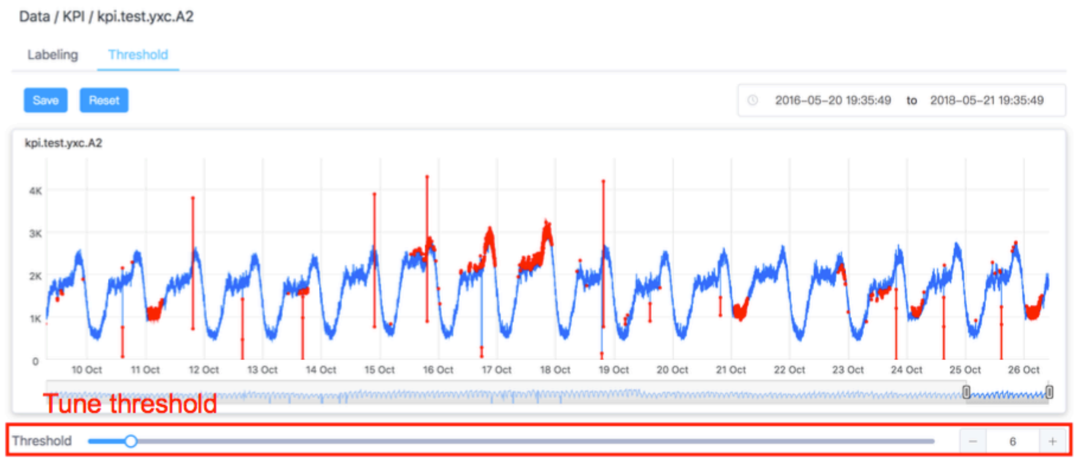


(b) Average of 11 KPIs

Label-Less: A Semi-automatic Labeling Tool for KPI Anomalies

- Best F-score ： 0.95
- Real-time response time：less than 0.5 second
- Reduce operators' labeling overhead by more than 95%

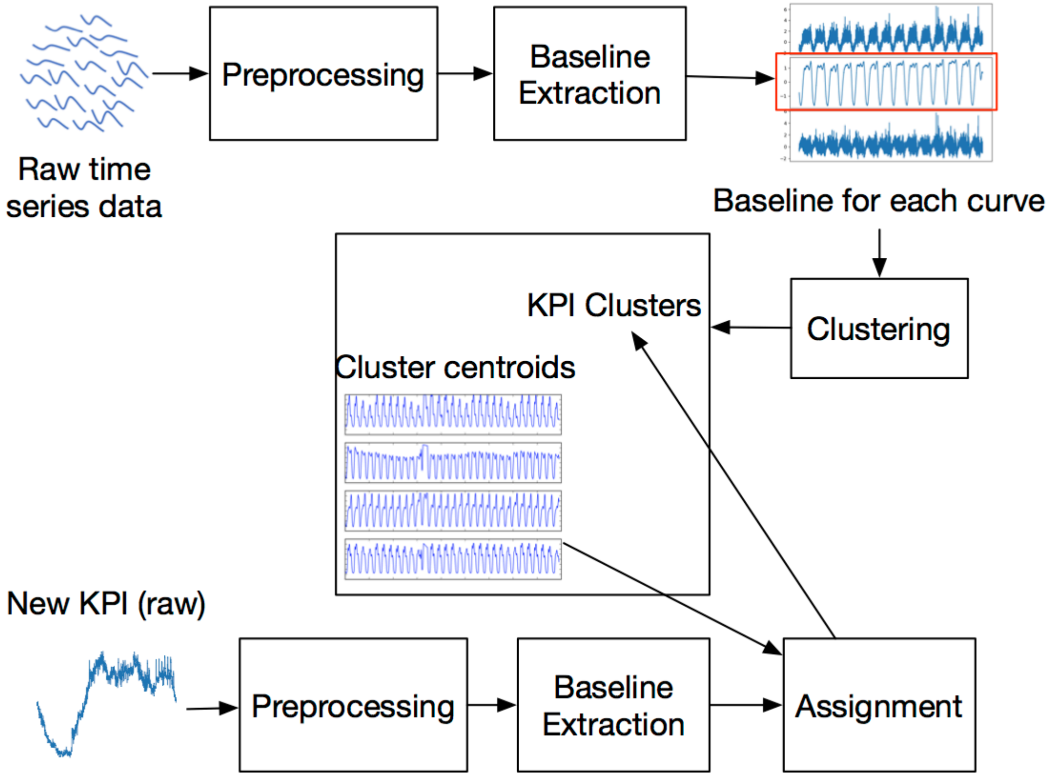




(a) Interface of candidate potential anomalies (labeled in red) given by unsupervised anomaly detection.

(b) Interface of anomaly similarity search. On the left is the anomaly template labeled in pink band; on the right is the similar anomalies given by *Label-Less* sorted by similarity.

Clustering + Transfer Learning to reduce training overhead



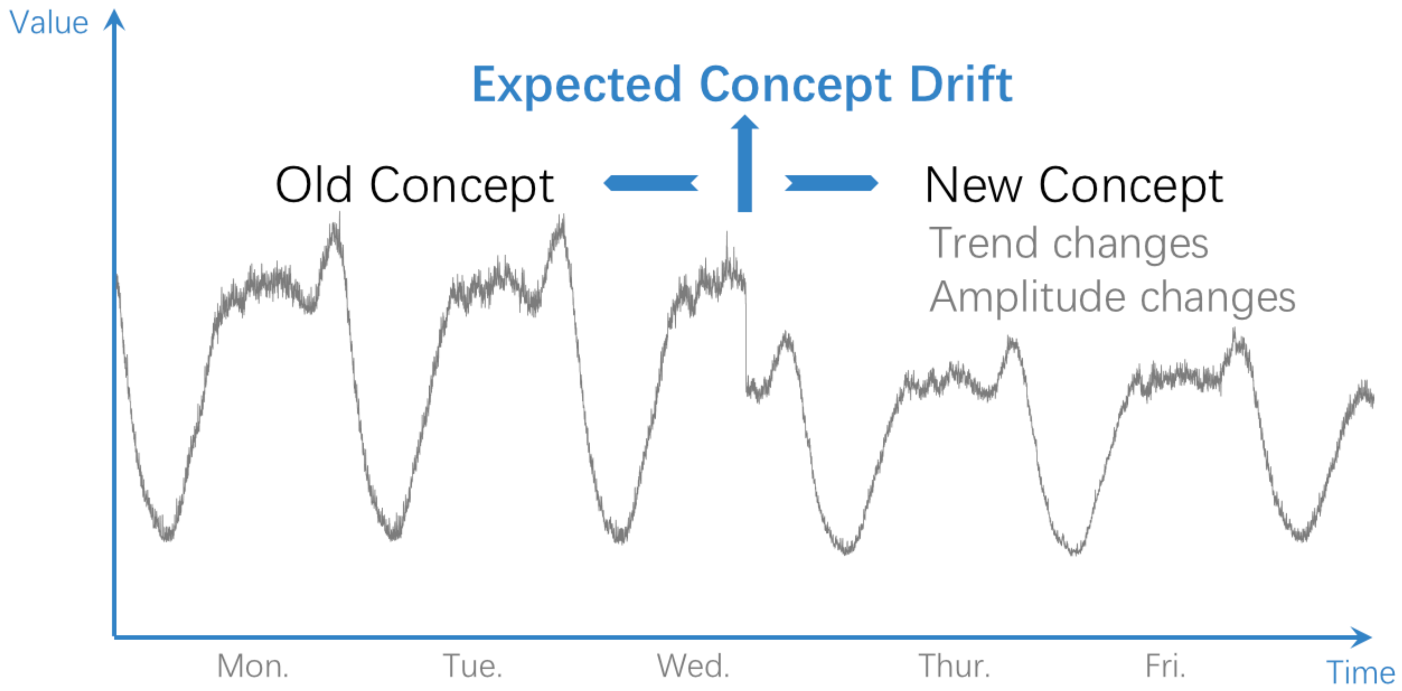
	Original DONUT [WWW2018]	ROCKA+DONUT+KPI-specific threshold
Avg. F-score	0.89	0.88
Total training time (s)	51621	5145

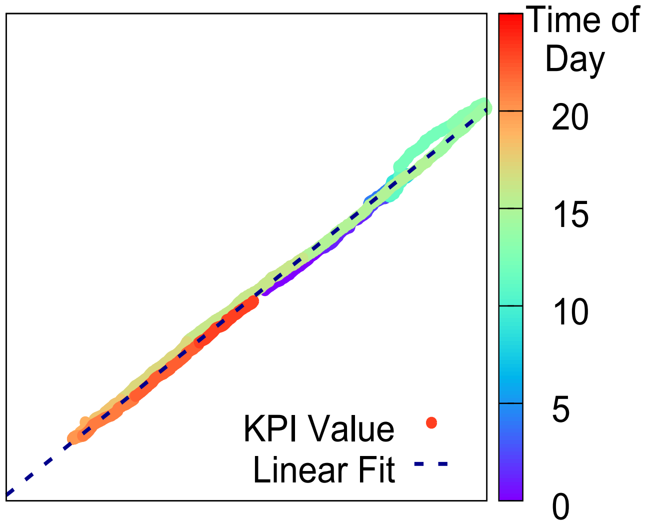
Adapt to Concept Drift

ISSRE (Class B) 2018 Best Paper

concept drift adaption improve anomaly detection F-score by 203%（**0.225 to 0.681**）

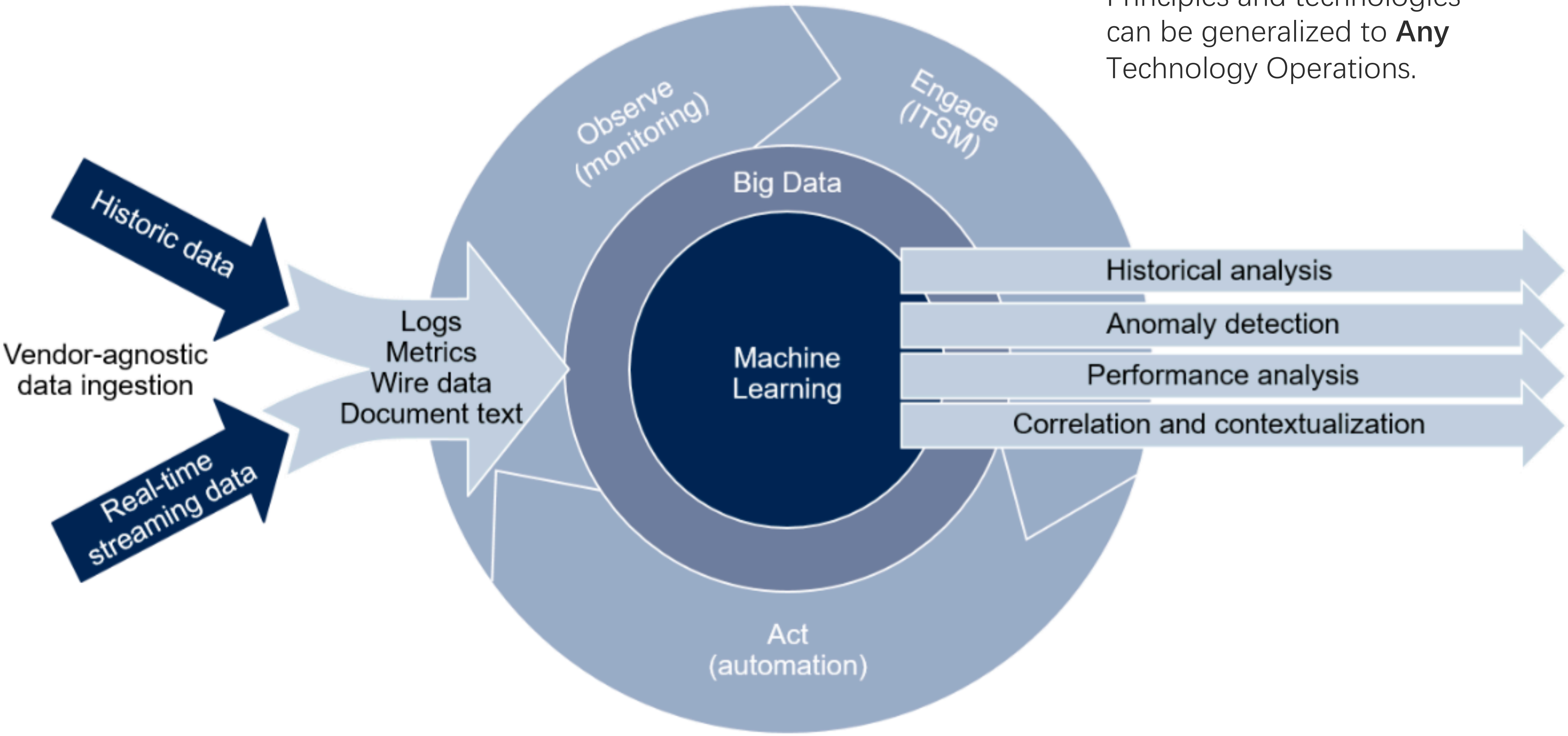
Observation: Old and New Concept Can Be Linearly Fitted





AIOps Platform Enabling Continuous ITOM

Principles and technologies can be generalized to **Any** Technology Operations.



ID: 340492

Summary

- AI is changing the world, but so far only in specific scenario of specific area in specific industry

- AI applications need be "coded" using domain (industry, area, scenario) knowledge-based "architecture"

- AIOps is a foundational technology in the increasingly digitalized world
  - What is AIOps
  - Business Value of AIOps: more revenue, less loss, more secure
  - Industry Leader's Opinion: AIOps is very promising
  - AIOps is necessary
  - AIOps is feasible
  - Defining Levels helps AIOps get accepted
  - AIOps can be very deep technologically

- AIOps is needed for all technology operations, not just IT operations.