Draining the flood

a combat against alert fatigue

Yu Chen

The Alert Flood in Baidu

- the amount of alerts is high
  - More than 100 alerts per person per day
    - Day time: ~75% in 17 hours
    - Night time: ~25% in 7 hours
- Highly Redundant
  - # effective alerts / # alert SMS < 0.15

Bai du 百度

Observations & Solutions

Observation	Reason	Solution
Duplicate ratio: 58%	• Persistent alerts • Correlated alerts	• Alert grouping
Attention ratio: 25% (at night time)	• Over-aggressive alert importance	• Alert importance level – Delivery behavior • Level calibration
Receivers per alert: 3	• In-effective oncall procedure	• Oncall schedule and escalation
Single instance alerts: 88%	• > 40% only requires simple operations to recover	• Automatic self-healing

Baidu 百度

Alert Grouping

- Simple grouping
  - Remove simple duplicates
- Cross-module patterns
  - Reveal underlying issues
- Network connectivity detection
  - Suppress alert surge

Bai du 百度

Simple Grouping

- Grouping based on natural dimensions
  - Alert rule name
  - Deployment structure
    - Product, Module, Cluster, Instance
    - Datacenter, machine

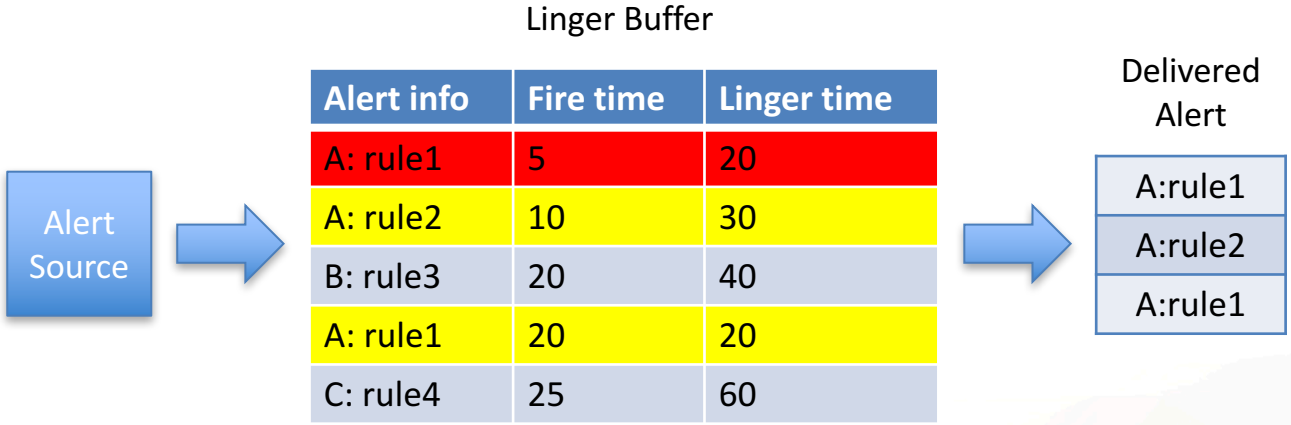
Grouping Result

{group.ab-zxcvq.AB.all:instance:B_zxcvq_FATAL}{总体异常实例比例:1.36054%}{异常(2):0.opr-zty5-zxcvq-000-cc.AB.bjdc,1.opr-zty5-zxcvq-000-cc.AB.bjdc}{05-02 16:49:36 - 16:54:09} {http://dwz.cn/… }

- Rule name
  - group.ab-zxcvq.AB.all:instance:B_zxcvq_FATAL
  - Instance level alert
- Ratio
  - 1.36054%
- Instance list
  - 0.opr-zty5-zxcvq-000-cc.AB.bjdc
  - 1.opr-zty5-zxcvq-000-cc.AB.bjdc
- Time
  - 05-02 16:49:36 - 16:54:09
- Link to detail page
  - http://dwz.cn/…

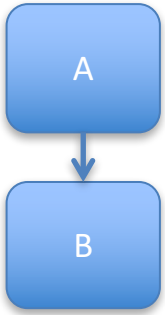
Baidu 百度

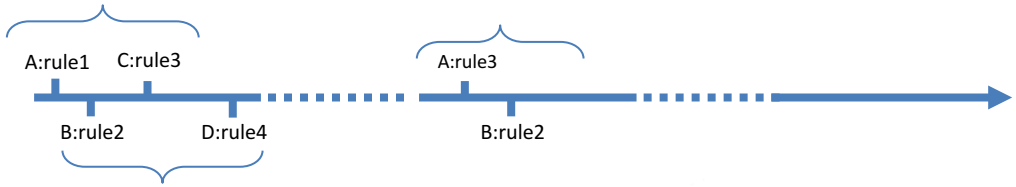
Delivery with Grouping



Cross-Module Patterns

- Caller / Callee
  - Both alerts when callee is in trouble
- Association rule mining
  - Transaction window starting from every alert





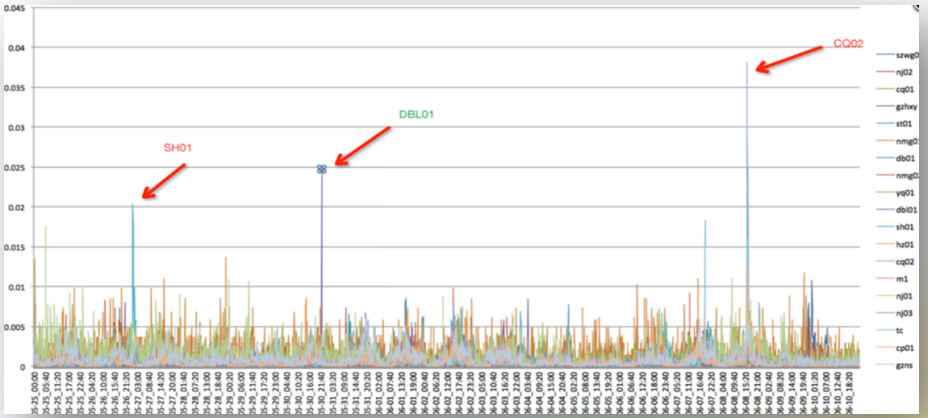
A:rule1, B:rule2, C:rule3
B:rule2, C:rule3, D:rule4
……
C:rule3, B:rule2



M:ruleX→ N:ruleY

Network Connectivity

- Network device failure can caused a lot of alerts
- Should trigger alerts for
  - Most rules
  - Most products
- Heuristic rule $\text{score} = \sqrt{\frac{\#\text{ alerting rules}}{\text{total rules}} \cdot \frac{\#\text{ alerting products}}{\text{total products}}}$



Linger Time

- Configurable
  - Different among alert rules
- Extra delay to receive alerts
  - Less punctual
- Need better ways to balance

Bai du 百度

Attention Ratio

- Check existence in interval $[t_{\text{alert}}, t_{\text{alert}} + t_{\Delta}]$
  - Access log of the monitoring system
    - View alert detail
    - View relevant curves
  - Login log of the production machine
- Exist: alert is attended
- Absent: alert is ignored
- Only applied to night time

Bai du 百度

Alert Calibration

- Importance levels
  - Critical: SMS + Phone to all receivers
  - Major: SMS + Escalation
  - Warning: SMS without Escalation
  - Notice: Mail
- Attention ratio should be compatible to levels
  - Push from mangers

Alert Receivers

- Typical receivers of an alert
  - Primary oncall engineer
  - Secondary oncall engineer
  - Oncall engineer lead
  - Senior engineer
  - Manager
- Primary oncall engineer handles alerts usually
  - But alerts always sent to all

Baidu 百度

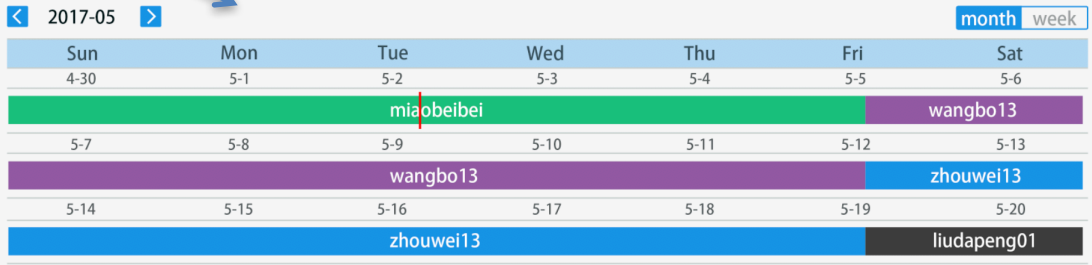
Oncall Escalation

- Alerting stages
  - One fixed stage
    - primary, secondary
  - Zero or more escalation stages



Oncall Escalation

Oncall schedule



Importance level: Major

Call phones on alert: No

	Receiver	Unclaimed after(min)	Unsolved after(min)	Call phones
Fixed stage	#z_ODA_lihui	5	10	yes
Escalation stage	oda	10	20	yes

Automatic Self-healing

- Lazy log purge
  - Set an alert on disk free space
  - Delete some log when alert triggers
- Granularity
  - Instance level
    - “bin_control restart”
  - Module/Cluster level
    - “curl master.a.com”
- Alert
  - will not deliver
  - view alert log

```
{
  "rule_name": "proc_net_fd_num_lt_1_or_ge_1200",

  "formula": "proc_net_fd_num < 1 ||
    proc_net_fd_num >= 1200",
  "linger_time": 60,
  "level": "warning",
  "delivery": "#z_ODA",

  "cb_action": "exec_script('/home/work
    /experiment-plat/bin/bin_control restart',
    'work')"
}

{
  "callback": [
    {
    "object": "service",
    "type": "post",
    "action": "http://productx.baidu.com/alert
      .php",
    "timeout": "10",
    "token": "12afecd710adc"
    }
  ]
}
```

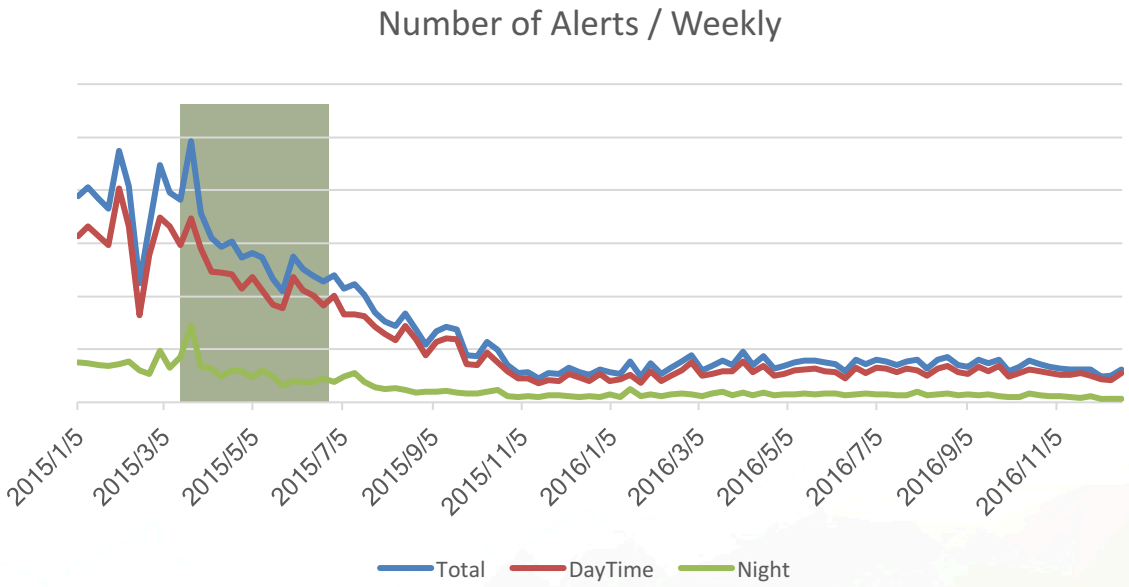
Baidu 百度

Management Support

- Alert importance calibration
  - Lower importance level
- Oncall escalation
  - Include attention ratio into work evaluation

Bai du 百度

Decrease by 85%



Remarks

- Reducing redundant alerts
  - Mining alert correlation for grouping
  - Estimate attention ratio for importance calibration
  - Receiver escalation mechanism
  - Alert self-healing mechanism
- Helpful on understanding root causes of issues

Bai du 百度

THANKS

chenyu07@baidu.com

