Automatic Traffic Scheduling for Internet Connectivity Failures

Liuqing Zhang
About me

• Liuqing Zhang

• Staff Software Engineer
• Responsible for Baidu monitoring system.
Outline

• Baidu network architecture
• Internet connectivity failures detection
• Automatic traffic scheduling
Two ways to handle the Internet connectivity failures

- Dynamic routing: BGP etc.
  - Automatically solves most of the network failures
  - Expensive

- Traffic scheduling based on DNS
  - Cheap
  - Depends on sufficiently accurate failure detection and traffic scheduling.
Baidu network architecture
Data center PoP (point of presence) network failures
ISP backbone network failures
Local ISP network failures

Before scheduling:
- Local ISP network
- ISP backbone network
- Baidu DC A

Without the local ISP network:
- ISP backbone network
- Baidu DC B
  - In another ISP

After scheduling:
- Local ISP network
- ISP backbone network
- Another ISP network

www.baidu.com
- IP address
- DNS

Baidu Cloud
The impact of Internet connectivity failures

- Internet connectivity failures occur more than 30 times per month

<table>
<thead>
<tr>
<th></th>
<th>Times per month</th>
<th>Average duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data center PoP network failures</td>
<td>12 – 20 times</td>
<td>1 hour 10 min</td>
</tr>
<tr>
<td>ISP backbone network failures</td>
<td>0 – 1 times</td>
<td>2 hour 40 min</td>
</tr>
<tr>
<td>Local ISP network failures</td>
<td>20 – 30 times</td>
<td>16 min</td>
</tr>
</tbody>
</table>

- Recovery time by manual operation
  - > 30 min
Outline

• Baidu network architecture
• Internet connectivity failures detection
• Automatic traffic scheduling
How to detect these failures

• *Falcon*: a network monitoring system

• Coverage is important. > 300 probes are required.

• Using Baidu P2P CDN host.
  • Covered all provinces in China.
The agent network situation is not stable

- The filter algorithm of abnormal probes

**Situation 1:**
20% points are abnormal.
There is only 1 detection report that the link is abnormal.

**Situation 2:**
20% points are abnormal.
There are 4 probes report that the link is abnormal.
Diagnose the fault location

- `<Province, IP>`
  - Local ISP network failures
  - ISP backbone network failures
  - Data center PoP network failures
Outline

• Baidu network architecture
• Internet connectivity failures detection
• Automatic traffic scheduling
Automatic traffic scheduling

- GTC: the Global Traffic Control system

- Falcon network monitoring System
- Traffic monitoring system
- DNS Service
- GTC Controller

Network failures
Bandwidth capacity
Traffic of DC
Current DNS Config
Commit DNS Config
The result of traffic scheduling

Request: 60G

Other Provinces

Bandwidth capacity: 100G

Baidu DC 2

Province A
Request: 30G

Province B
Request: 30G

After scheduling

DC network failures

Before scheduling

Other Provinces

Request: 60G

Bandwidth capacity: 100G

Baidu DC 3

After scheduling
The limit of service capacity

Before scheduling:
- Province A
  - PoP
  - Backend service Capacity: 2500
- DC 1

After scheduling:
- Province B
  - PoP
  - Backend service Capacity: 2500
- DC 2

1000qps

DC network failure

Overload
Capacity of backend service

- BFE:
  - Baidu Front-End, a layer-7 load balancer
- Capacity safe
Benefits

• Automatic Fault Processing
  • MTTR < 5min
• Avoid bandwidth congestion
• Avoid backend service overload
Thanks