

# BGP选路循环导致路由震荡排错案例

## 目录

- [硬件平台](#)
- [软件版本](#)
- [案例简介](#)
- [故障诊断步骤](#)
- [经验总结](#)
- [相关命令](#)
- [相关错误信息](#)
- [其他相关文档](#)

## 硬件平台

路由器

## 软件版本

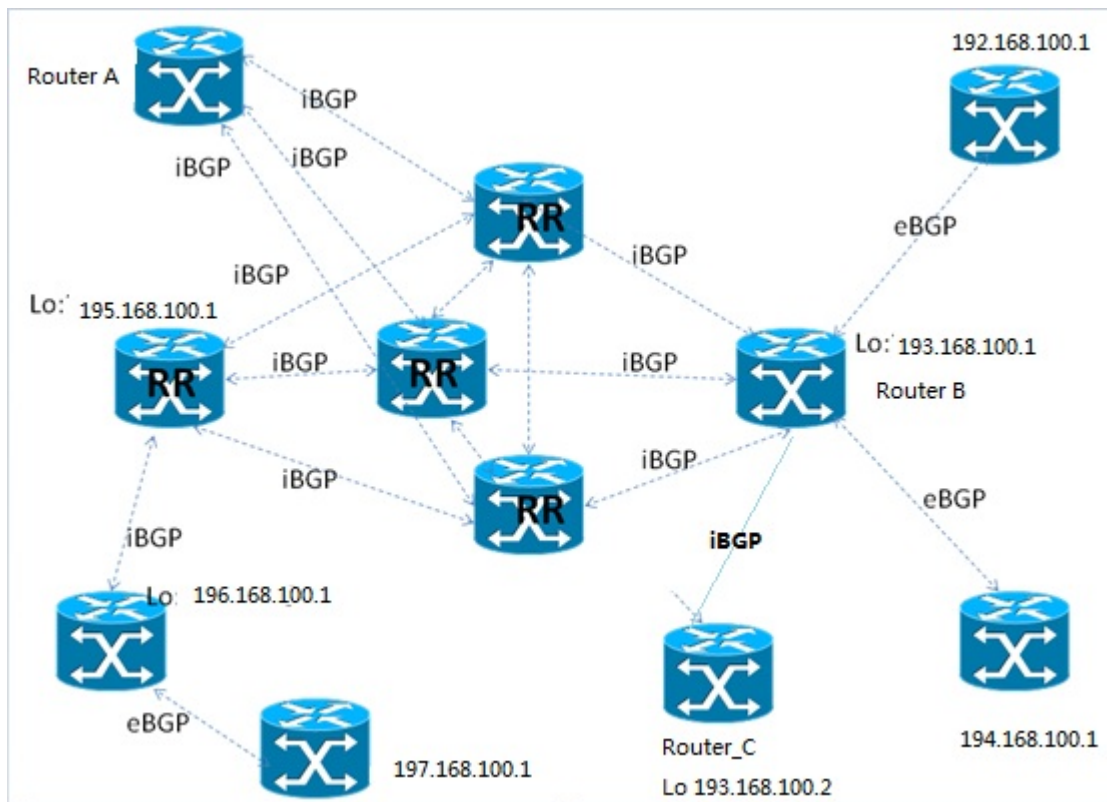
所有

## 案例简介

在大型运营商网络中，BGP的路由频繁Flapping经常会造成严重的CPU 利用率增高现象。我们有一些方法可以控制BGP 路由Flapping 对网络造成的影响，比如BGP Damping，但是有些情况BGP Damping 也不能起到很好的保护作用（BGP damping 仅对从EBGP学到的路由有效）。这时我们就需要了解flapping 产生的原因，以及如何从源头消除这种影响，以下是一个真实案例，为保护客户资料，路由器输出信息是在实验环境中抓取。

## 故障诊断步骤

这是一个传统的运营商网络架构结构，客户发现核心路由器Router A的CPU利用率一直在90%以上，并且还有多台类似位置的路由器CPU利用率同样也高达90%，通过show proc cpu | ex 0.00可以看到是由于bgp router进程引起。一般BGP Router 进程控制路由更新，由于首先报出问题的设备是Router\_A，具有直接现象，所以我们从这台路由器入手排查原因。



## 1. 确定flapping路由的来源

- a) 由于RIB表里面的条目刷新的同时会刷新该条目的计时器，如果BGP路由flapping，这些路由在RIB里面也会被频繁的刷新，因此可以通过RIB表找出flapping的路由。命令是"show ip route | in \_00:00"

```
B 101.101.96.0/23 [200/0] via 193.168.100.1, 00:00:00
B 101.101.96.0/22 [200/0] via 193.168.100.1, 00:00:00
B 101.101.98.0/23 [200/0] via 193.168.100.1, 00:00:00
B 101.101.96.0/23 [200/0] via 196.168.100.1, 00:00:00
B 101.101.96.0/22 [200/0] via 196.168.100.1, 00:00:00
B 101.101.98.0/23 [200/0] via 196.168.100.1, 00:00:00
```

反复show 这条命令，发现某些路由反复出现在路由表中，且更新时间为00:00:00,一般来说这些路由为产生Flapping的路由，由此我们可以看到通告这些flapping路由的的BGP Peer 是193.168.100.1 (Router B)，此路由器与Router A为IBGP关系。

- b) 登录193.168.100.1 ( Router\_B )，发现这些路由频繁flapping的原因是其BGP在进行选路时最优路径在不停切换。（在195.168.100.1 和192.168.100.1之间），从而导致不停此路由在不停被更新。

```
RP/0/0/CPU0:Router_B#show bgp 101.101.96.0/22
```

```
BGP routing table entry for 101.101.96.0/22
```

```
Versions:
```

```
Process bRIB/RIB SendTblVer
```

```
Speaker 188363313 188363313 Last Modified: Mar 8 09:04:45.207 for 00:00:00
```

```
Paths: (6 available, best #6) <----- 6 path 195.168.100.1 Advertised to
update-groups (with more than one peer): 0.1 0.2 0.4 0.16 Path #1: Received by
speaker 0 1001 9304 45932 55945 192.168.100.1 from 192.168.100.1 (200.200.1.1)
```

```
Origin IGP, metric 155, localpref 200, valid, external Community: 1025:32023
1025:32412 1025:32502 1025:60952 Path #2: Received by speaker 0 1002 17888 45932
55945 198.168.100.1 from 198.168.100.1 (203.192.169.249) Origin IGP, localpref
200, valid, external, multipath Community: 1025:32021 1025:32304 1025:32501
1025:60952 Path #3: Received by speaker 0 1001 9304 45932 55945 196.168.100.1 (metric
276) from 193.168.100.2 (196.168.100.1) Origin IGP, metric 0, localpref 200,
valid, internal Community: 1025:32007 1025:32412 1025:32502 1025:60952
Originator: 196.168.100.1, Cluster list: 193.168.100.2, 195.168.100.1 Path #4:
Received by speaker 0 1001 9304 45932 55945 196.168.100.1 (metric 276) from
193.168.100.3 (196.168.100.1) Origin IGP, metric 0, localpref 200, valid, internal
Community: 1025:32007 1025:32412 1025:32502 1025:60952 Originator:
196.168.100.1, Cluster list: 193.168.100.3, 195.168.100.1 Path #5: Received by
speaker 0 1001 9304 45932 55945 196.168.100.1 (metric 276) from 193.168.100.4
(196.168.100.1) Origin IGP, metric 0, localpref 200, valid, internal Community:
1025:32007 1025:32412 1025:32502 1025:60952 Originator: 196.168.100.1, Cluster
list: 193.168.100.4, 195.168.100.1 Path #6: Received by speaker 0 1002 17888 45932
55945 194.168.100.1 from 194.168.100.1 (203.192.154.9) Origin IGP, localpref 200,
valid, external, best, multipath Community: 1025:32021 1025:32304 1025:32501
1025:60952 show RP/0/0/CPU0:Router_B#show bgp 101.101.96.0/22 BGP routing table
entry for 101.101.96.0/22 Versions: Process bRIB/RIB SendTblVer Speaker
188364164 188364164 Last Modified: Mar 8 09:04:50.207 for 136y10w Paths: (3
available, best #1) 3 path 192.168.100.1 Advertised to update-groups (with more
than one peer): 0.1 0.2 0.4 0.16 Path #1: Received by speaker 0 1001 9304 45932 55945
192.168.100.1 from 192.168.100.1 (200.200.1.1) Origin IGP, metric 155, localpref
200, valid, external, best Community: 1025:32023 1025:32412 1025:32502 1025:60952
Path #2: Received by speaker 0 1002 17888 45932 55945 198.168.100.1 from
198.168.100.1 (203.192.169.249) Origin IGP, localpref 200, valid, external
Community: 1025:32021 1025:32304 1025:32501 1025:60952 Path #3: Received by speaker 0
1002 17888 45932 55945 194.168.100.1 from 194.168.100.1 (203.192.154.9) Origin
IGP, localpref 200, valid, external Community: 1025:32021 1025:32304 1025:32501
1025:60952 RP/0/0/CPU0:Router_B#show bgp 101.101.96.0/22 BGP routing table entry
for 101.101.96.0/22 Versions: Process bRIB/RIB SendTblVer Speaker
188380494 188380494 Last Modified: Mar 8 09:06:35.208 for 136y10w Paths: (6
available, best #6) Advertised to update-groups (with more than one peer): 0.1 0.2
0.4 0.16 Path #1: Received by speaker 0 1001 9304 45932 55945 192.168.100.1 from
192.168.100.1 (200.200.1.1) Origin IGP, metric 155, localpref 200, valid, external
Community: 1025:32023 1025:32412 1025:32502 1025:60952 Path #2: Received by
speaker 0 1002 17888 45932 55945 198.168.100.1 from 198.168.100.1 (203.192.169.249)
Origin IGP, localpref 200, valid, external, multipath Community: 1025:32021
1025:32304 1025:32501 1025:60952 Path #3: Received by speaker 0 1001 9304 45932 55945
196.168.100.1 (metric 276) from 193.168.100.2 (196.168.100.1) Origin IGP, metric
0, localpref 200, valid, internal Community: 1025:32007 1025:32412 1025:32502
1025:60952 Originator: 196.168.100.1, Cluster list: 193.168.100.2, 195.168.100.1
Path #4: Received by speaker 0 1001 9304 45932 55945 196.168.100.1 (metric 276) from
193.168.100.3 (196.168.100.1) Origin IGP, metric 0, localpref 200, valid, internal
Community: 1025:32007 1025:32412 1025:32502 1025:60952 Originator:
196.168.100.1, Cluster list: 193.168.100.3, 195.168.100.1 Path #5: Received by
speaker 0 1001 9304 45932 55945 196.168.100.1 (metric 276) from 193.168.100.4
(196.168.100.1) Origin IGP, metric 0, localpref 200, valid, internal Community:
1025:32007 1025:32412 1025:32502 1025:60952 Originator: 196.168.100.1, Cluster
list: 193.168.100.4, 195.168.100.1 Path #6: Received by speaker 0 1002 17888 45932
55945 194.168.100.1 from 194.168.100.1 (203.192.154.9) Origin IGP, localpref 200,
valid, external, best, multipath Community: 1025:32021 1025:32304 1025:32501
```

## 2. 路由flapping的原因

- a) 首先在Router B查看具有3条可用路径的情况，此时Router B会优选第1条路径作为最优路径（192.168.100.1），因为192.168.100.1具有最低的router ID（注意此时不会比较MED，因为它们来自不同的AS），这时Router\_B会将最优路径192.168.100.1放入路由表中，并且将此结果通告给他的IBGP邻居，包括193.168.100.2/3/4

```
RP/0/0/CPU0:Router_B#show bgp 101.101.96.0/22

BGP routing table entry for 101.101.96.0/22
Versions:
  Process    bRIB/RIB  SendTblVer
  Speaker    188379010 188379010
Last Modified: Mar  8 09:06:23.208 for 00:00:00
Paths: (3 available, best #1)
  Advertised to update-groups (with more than one peer):
    0.1 0.2 0.4 0.16
  Path #1: Received by speaker 0 1001 9304 45932 55945 192.168.100.1 from
192.168.100.1 (200.200.1.1) Origin IGP, metric 155, localpref 200, valid, external,
best Community: 1025:32023 1025:32412 1025:32502 1025:60952 Path #2: Received by
speaker 0 1002 17888 45932 55945 198.168.100.1 from 198.168.100.1 (203.192.169.249)
Origin IGP, localpref 200, valid, external Community: 1025:32021 1025:32304
1025:32501 1025:60952 Path #3: Received by speaker 0 1002 17888 45932 55945
194.168.100.1 from 194.168.100.1 (203.192.154.9) Origin IGP, localpref 200, valid,
external Community: 1025:32021 1025:32304 1025:32501 1025:60952
```

- b) Router\_C (193.168.100.2) 会学到Router\_B通告过来的此路由的更新，但是他不会将此路由作为最优路径并放入路由表，因为它的MED值比Router\_C从其他的邻居学到的另外一个路径（195.168.100.1）的MED值要高（相同的AS），因而他会选择这条路径作为最优路径（195.168.100.1）并且将此路由通告回给Router\_B，在193.168.100.3 and 193.168.100.4 上的情况也同样是如此。

```
Router_C#sh ip bgp 101.101.96.0 255.255.252.0
BGP routing table entry for 101.101.96.0/22, version 592019857
Bestpath Modifiers: deterministic-med
Paths: (6 available, best #1)
Flag: 0x820
  Advertised to update-groups:
    2 3 7 8
  1001 9304 45932 55945 196.168.100.1 (metric 274) from 195.168.100.1 (195.168.100.1)
Origin IGP, metric 0, localpref 200, valid, internal, best Community: 1025:32007
1025:32412 1025:32502 1025:60952 Originator: 196.168.100.1, Cluster list:
195.168.100.1 1001 9304 45932 55945, (Received from a RR-client) 193.168.100.1
(metric 257) from 193.168.100.1 (193.168.100.1) Origin IGP, metric 155, localpref
200, valid, internal Community: 1025:32023 1025:32412 1025:32502 1025:60952 7473
17888 45932 55945 202.84.142.11 (metric 290) from 202.84.142.11 (202.84.142.11)
Origin IGP, metric 100, localpref 200, valid, internal Community: 1025:32001
1025:32047 1025:32404 1025:32501 1025:60952 1025:65003 1002 17888 45932 55945
202.84.152.38 (metric 278) from 202.84.152.33 (202.84.152.33) Origin IGP, localpref
200, valid, internal Community: 1025:32004 1025:32301 1025:32501 1025:60952
Originator: 202.84.152.38, Cluster list: 202.84.152.33 1002 17888 45932 55945
```

```

202.84.225.8 (metric 295) from 202.84.225.6 (202.84.225.6) Origin IGP, metric 0,
localpref 200, valid, internal Community: 1025:32024 1025:32309 1025:32501 1025:60952
Originator: 202.84.225.8, Cluster list: 202.84.225.6 1002 17888 45932 55945
202.84.225.8 (metric 295) from 202.84.225.7 (202.84.225.7) Origin IGP, metric 0,
localpref 200, valid, internal Community: 1025:32024 1025:32309 1025:32501 1025:60952
Originator: 202.84.225.8, Cluster list: 202.84.225.7

```

- c) 所以Router\_B马上又会重新学到3条新的路径,分别从他的3个IBGP 邻居193.168.100.2/3/4. 因为又增加了3条新路径,此时 Router\_B需要重新对此路由进行路由选路,这时原路径1 192.168.100.1将不会再成为最优路径,因为新增加的路径3 ( 196.168.100.1 ) 与路径1 其他条件一样,但是具有更低的MED值 ( 0 ).路径3在与路径1竞争胜利后,路径3 (196.168.100.1) 也不会马上成为最优路径,因为它是一条从IBGP学到的路由 ( Internal ).而此时的路径6 (195.168.100.1) 又会在与路径3的竞争中获胜,因为相比于路径3 他是External 路由 ( 从EBGP ) 学得,会更优先。所以Router\_B 会选择路径6 (195.168.100.1) 作为最优路径,并且将这条最新的路由宣告给他的3个IBGP P邻居, 193.168.100.2/3/4.

```

RP/0/0/CPU0:Router_B#show bgp 101.101.96.0/22
BGP routing table entry for 101.101.96.0/22
Versions:
  Process    bRIB/RIB SendTblVer
  Speaker    188380494 188380494
Last Modified: Mar 8 09:06:35.208 for 136y10w
Paths: (6 available, best #6)
  Advertised to update-groups (with more than one peer):
    0.1 0.2 0.4 0.16
  Path #1: Received by speaker 0
    1001 9304 45932 55945 192.168.100.1 from 192.168.100.1 (200.200.1.1) Origin IGP,
    metric 155, localpref 200, valid, external Community: 1025:32023 1025:32412
    1025:32502 1025:60952 Path #2: Received by speaker 0 1002 17888 45932 55945
    198.168.100.1 from 198.168.100.1 (203.192.169.249) Origin IGP, localpref 200, valid,
    external, multipath Community: 1025:32021 1025:32304 1025:32501 1025:60952 Path #3:
    Received by speaker 0 1001 9304 45932 55945 196.168.100.1 (metric 276) from
    193.168.100.2 (196.168.100.1) Origin IGP, metric 0, localpref 200, valid, internal
    Community: 1025:32007 1025:32412 1025:32502 1025:60952 Originator: 196.168.100.1,
    Cluster list: 193.168.100.2, 195.168.100.1 Path #4: Received by speaker 0 1001 9304
    45932 55945 196.168.100.1 (metric 276) from 193.168.100.3 (196.168.100.1) Origin IGP,
    metric 0, localpref 200, valid, internal Community: 1025:32007 1025:32412 1025:32502
    1025:60952 Originator: 196.168.100.1, Cluster list: 193.168.100.3, 195.168.100.1 Path
    #5: Received by speaker 0 1001 9304 45932 55945 196.168.100.1 (metric 276) from
    193.168.100.4 (196.168.100.1) Origin IGP, metric 0, localpref 200, valid, internal
    Community: 1025:32007 1025:32412 1025:32502 1025:60952 Originator: 196.168.100.1,
    Cluster list: 193.168.100.4, 195.168.100.1 Path #6: Received by speaker 0 1002 17888
    45932 55945 194.168.100.1 from 194.168.100.1 (203.192.154.9) Origin IGP, localpref
    200, valid, external, best, multipath Community: 1025:32021 1025:32304 1025:32501
    1025:60952

```

- d) 在Router\_C 上 ( 193.168.100.2 ), 他会重新选择第6条路径 ( 刚从Router\_B学得 ) 最为最优路径,因为这条路径具有更低的IGP metric. 然后Router\_C会撤销之前发给Router\_B的"旧"的最优路由,并且不会再将此新最优路由发送回给Router\_B, ( 因为此时Router\_B就是最优路径,水平分割原则 ) 同样的情况也发生在 193.168.100.3 and 193.168.100.4上面

```

Router_C#sh ip bgp 101.101.96.0 255.255.252.0
BGP routing table entry for 101.101.96.0/22, version 592024168
Bestpath Modifiers: deterministic-med

```

```

Paths: (8 available, best #6)
Flag: 0x820
  Advertised to update-groups:
    2 3 4 5 7 8
  1001 9304 45932 55945
    196.168.100.1 (metric 274) from 195.168.100.1 (195.168.100.1) Origin IGP, metric 0,
localpref 200, valid, internal Community: 1025:32007 1025:32412 1025:32502 1025:60952
Originator: 196.168.100.1, Cluster list: 195.168.100.1 7473 17888 45932 55945
  202.84.142.11 (metric 290) from 202.84.142.11 (202.84.142.11) Origin IGP, metric 100,
localpref 200, valid, internal Community: 1025:32001 1025:32047 1025:32404 1025:32501
1025:60952 1025:65003 1002 17888 45932 55945
  202.84.152.38 (metric 278) from
  202.84.152.33 (202.84.152.33) Origin IGP, localpref 200, valid, internal Community:
1025:32004 1025:32301 1025:32501 1025:60952
Originator: 202.84.152.38, Cluster list:
  202.84.152.33 1002 17888 45932 55945
  202.84.225.8 (metric 295) from 202.84.225.6
(202.84.225.6) Origin IGP, metric 0, localpref 200, valid, internal Community:
1025:32024 1025:32309 1025:32501 1025:60952
Originator: 202.84.225.8, Cluster list:
  202.84.225.6 1002 17888 45932 55945
  202.84.225.8 (metric 295) from 202.84.225.7
(202.84.225.7) Origin IGP, metric 0, localpref 200, valid, internal Community:
1025:32024 1025:32309 1025:32501 1025:60952
Originator: 202.84.225.8, Cluster list:
  202.84.225.7 1002 17888 45932 55945, (Received from a RR-client)
  193.168.100.1 (metric 257) from 193.168.100.1 (193.168.100.1) Origin IGP, localpref 200, valid,
internal, best Community: 1025:32021 1025:32304 1025:32501 1025:60952
  1002 17888
45932 55945
  193.168.100.1 (metric 257) from 193.168.100.3 (193.168.100.3) Origin IGP,
localpref 200, valid, internal Community: 1025:32021 1025:32304 1025:32501 1025:60952
Originator: 193.168.100.1, Cluster list: 193.168.100.3 1002 17888 45932 55945
  193.168.100.1 (metric 257) from 193.168.100.4 (193.168.100.4) Origin IGP, localpref
200, valid, internal Community: 1025:32021 1025:32304 1025:32501 1025:60952
Originator: 193.168.100.1, Cluster list: 193.168.100.4

```

e) 所以在Router\_B上，他马上又收到193.168.100.2/3/4的撤销路由，这时他删掉从这3个邻居学到路径，又像开始一样只具有3条BGP路径，然后重新优选路径1为最优路径并进行通告，在此处产生BGP路由选路循环。从而导致他的最优路径不停在192.168.100.1和195.168.100.1之间切换，产生路由震荡。

```

RP/0/0/CPU0:Router_B#show bgp 101.101.96.0/22
Tue Mar 8 09:06:23.002 UTC
BGP routing table entry for 101.101.96.0/22
Versions:
  Process    bRIB/RIB SendTblVer
  Speaker    188379010 188379010
Last Modified: Mar 8 09:06:23.208 for 00:00:00
Paths: (3 available, best #1)
  Advertised to update-groups (with more than one peer):
    0.1 0.2 0.4 0.16
  Path #1: Received by speaker 0 1001 9304 45932 55945 192.168.100.1 from
192.168.100.1 (200.200.1.1) Origin IGP, metric 155, localpref 200, valid, external,
best Community: 1025:32023 1025:32412 1025:32502 1025:60952
  Path #2: Received by
speaker 0 1002 17888 45932 55945 198.168.100.1 from 198.168.100.1 (203.192.169.249)
Origin IGP, localpref 200, valid, external Community: 1025:32021 1025:32304
1025:32501 1025:60952
  Path #3: Received by speaker 0 1002 17888 45932 55945
194.168.100.1 from 194.168.100.1 (203.192.154.9) Origin IGP, localpref 200, valid,
external Community: 1025:32021 1025:32304 1025:32501 1025:60952

```

### 3. 解决方法：

解决方法有很多，比如调整weight，local preference，只要可以控制Router\_B不会因为其他路径加入而影响最终选路结果的方法都可以解决此类问题。

## 经验总结

BGP路由flapping 常见的原因是链路flapping或者EBGP产生flapping的路由，此类由于路由选路规则而导致的路由flapping并不常见，但无论是由于哪种原因引起，要解决问题首先需要找出flapping源头。再根据基础知识，比如在此处需要我们对BGP的选路规则有着深入的了解。然后逐条路由分析，就会得出清晰的结果。

## 相关命令

```
show ip route | in 00:00:00
```

```
show ip bgp x.x.x.x
```

```
show bgp x.x.x.x (IOS XR )
```

## 相关错误信息

无

## 其他相关文档

[BGP Best Path Selection Algorithm](#)

[BGP 排错手册](#)