

知 某客户bmp监控中发现某重复路由上报

BGP 赵阳 2021-11-23 发表

组网及说明

(拓扑由下往上依次为TOR—LEAF—SPINE---DCI) ,1台TOR会上联到8台S9820-8C上, 全EBGP互联, 同一级S9820-8C as号是一致, 每台TOR一个as号。

问题描述

设备侧配置bmp监控路由功能，且有配置route-mode loc-rib，根据官网手册，该命令配置之后，可以达到如下效果：本功能配置后会先向监控服务器发送一次BGP所有的路由，之后路由更新时，仅向监控服务器发送BGP优选后的最优更新路由信息。

但并没有起到相应效果，发现有许多非最优路由由还在重复上报：

主机:10.32.21.0	邻居:10.32.40.110	可达路由: [10.32.4.15/32, 10.32.4.14/32]	as: [4203204020, 4203201000, 4203204014]	time:2021-11-15 13:03:57
主机:10.32.21.0	邻居:10.32.40.38	可达路由: [10.32.4.14/32]	as: [4203204000, 4203201000, 4203204014]	time:2021-11-15 13:03:57
主机:10.32.21.0	邻居:10.32.40.38	可达路由: [10.32.4.14/32]	as: [4203204000, 4203201000, 4203204014]	time:2021-11-15 13:04:07
主机:10.32.21.0	邻居:10.32.40.74	可达路由: [10.32.4.15/32, 10.32.4.14/32]	as: [4203204010, 4203201000, 4203204014]	time:2021-11-15 13:04:07
主机:10.32.21.0	邻居:10.32.40.86	可达路由: [10.32.4.14/32]	as: [4203204020, 4203201000, 4203204014]	time:2021-11-15 13:04:07
主机:10.32.21.0	邻居:10.32.40.90	可达路由: [10.32.4.15/32, 10.32.4.14/32]	as: [4203204020, 4203201000, 4203204014]	time:2021-11-15 13:04:07
主机:10.32.21.0	邻居:10.32.40.110	可达路由: [10.32.4.15/32, 10.32.4.14/32]	as: [4203204020, 4203201000, 4203204014]	time:2021-11-15 13:04:07
主机:10.32.21.0	邻居:10.32.40.142	可达路由: [10.32.4.15/32, 10.32.4.14/32]	as: [4203204036, 4203201000, 4203204014]	time:2021-11-15 13:04:07
主机:10.32.21.0	邻居:10.32.40.150	可达路由: [10.32.4.15/32]	as: [4203204000, 4203201000, 4203204014]	time:2021-11-15 13:04:07
主机:10.32.21.0	邻居:10.32.40.38	可达路由: [10.32.4.14/32]	as: [4203204000, 4203201000, 4203204014]	time:2021-11-15 13:04:27
主机:10.32.21.0	邻居:10.32.40.38	可达路由: [10.32.4.14/32]	as: [4203204000, 4203201000, 4203204014]	time:2021-11-15 13:04:37
主机:10.32.21.0	邻居:10.32.40.42	可达路由: [10.32.4.14/32]	as: [4203204010, 4203201000, 4203204014]	time:2021-11-15 13:04:37
主机:10.32.21.0	邻居:10.32.40.38	可达路由: [10.32.4.14/32]	as: [4203204000, 4203201000, 4203204014]	time:2021-11-15 13:04:57
主机:10.32.21.0	邻居:10.32.40.74	可达路由: [10.32.4.15/32, 10.32.4.14/32]	as: [4203204010, 4203201000, 4203204014]	time:2021-11-15 13:04:57
主机:10.32.21.0	邻居:10.32.40.6	可达路由: [10.32.4.14/32]	as: [4203204000, 4203201000, 4203204014]	time:2021-11-15 13:04:57
主机:10.32.21.0	邻居:10.32.40.86	可达路由: [10.32.4.14/32]	as: [4203204020, 4203201000, 4203204014]	time:2021-11-15 13:04:57
主机:10.32.21.0	邻居:10.32.40.98	可达路由: [10.32.4.15/32, 10.32.4.14/32]	as: [4203204010, 4203201000, 4203204014]	time:2021-11-15 13:04:57
主机:10.32.21.0	邻居:10.32.40.42	可达路由: [10.32.4.14/32]	as: [4203204010, 4203201000, 4203204014]	time:2021-11-15 13:04:57
主机:10.32.21.0	邻居:10.32.40.38	可达路由: [10.32.4.14/32]	as: [4203204000, 4203201000, 4203204014]	time:2021-11-15 13:05:07
主机:10.32.21.0	邻居:10.32.40.86	可达路由: [10.32.4.14/32]	as: [4203204020, 4203201000, 4203204014]	time:2021-11-15 13:05:07
主机:10.32.21.0	邻居:10.32.40.34	可达路由: [10.32.4.14/32]	as: [4203204000, 4203201000, 4203204014]	time:2021-11-15 13:05:27
主机:10.32.21.0	邻居:10.32.40.38	可达路由: [10.32.4.14/32]	as: [4203204000, 4203201000, 4203204014]	time:2021-11-15 13:05:27
主机:10.32.21.0	邻居:10.32.40.10	可达路由: [10.32.4.14/32]	as: [4203204002, 4203201000, 4203204014]	time:2021-11-15 13:05:27
主机:10.32.21.0	邻居:10.32.40.42	可达路由: [10.32.4.14/32]	as: [4203204010, 4203201000, 4203204014]	time:2021-11-15 13:05:27
主机:10.32.21.0	邻居:10.32.40.38	可达路由: [10.32.4.14/32]	as: [4203204000, 4203201000, 4203204014]	time:2021-11-15 13:05:37
主机:10.32.21.0	邻居:10.32.40.86	可达路由: [10.32.4.14/32]	as: [4203204020, 4203201000, 4203204014]	time:2021-11-15 13:05:37
主机:10.32.21.0	邻居:10.32.40.110	可达路由: [10.32.4.15/32, 10.32.4.14/32]	as: [4203204010, 4203201000, 4203204014]	time:2021-11-15 13:05:57
主机:10.32.21.0	邻居:10.32.40.38	可达路由: [10.32.4.14/32]	as: [4203204000, 4203201000, 4203204014]	time:2021-11-15 13:05:57
主机:10.32.21.0	邻居:10.32.40.42	可达路由: [10.32.4.14/32]	as: [4203204010, 4203201000, 4203204014]	time:2021-11-15 13:05:57
主机:10.32.21.0	邻居:10.32.40.34	可达路由: [10.32.4.14/32]	as: [4203204000, 4203201000, 4203204014]	time:2021-11-15 13:06:07
主机:10.32.21.0	邻居:10.32.40.6	可达路由: [10.32.4.14/32]	as: [4203204000, 4203201000, 4203204014]	time:2021-11-15 13:06:07
主机:10.32.21.0	邻居:10.32.40.38	可达路由: [10.32.4.14/32]	as: [4203204000, 4203201000, 4203204014]	time:2021-11-15 13:06:07
主机:10.32.21.0	邻居:10.32.40.86	可达路由: [10.32.4.14/32]	as: [4203204020, 4203201000, 4203204014]	time:2021-11-15 13:06:27

过程分析

- 1、由于现网设备（TOR）已经存在路由，该设备（TOR）bmp配置监控了大量的邻居（所有SPINE和TOR）信息；
- 2、当bmp连接刚刚建立时，由于route-mode默认是adj-rib-in pre-policy模式，该设备（TOR）会向所有监控的邻居发送route-refresh报文，所有监控的邻居收到route-refresh报文后，同时向该设备（TOR）重新发送所有路由信息（无论是否配置loc-rib命令，bmp刚刚建立连接时都会先进行本过程）
- 3、该设备（TOR）收到所有的路由信息都会发送给bmp server，由于同时数据量比较大，TCP发送较慢，bgp发送速度比tcp发送速度快，造成bmp报文累积，当累积数达到缓冲区限制1万条时，设备会主动断开bmp连接；
- 4、Bmp连接断开前，bmp报文信息包含的路由信息正好是10.32.4.14和10.32.4.15；
- 5、Bmp断连接后，会主动重连server，然后再重复执行步骤2、3、4；最终导致该现象的发生

解决方法

在解决问题版本正式推出之前，可以通过把配置peer bmp server去掉，只配置route-mode loc-rib的方式，只监控最优路由信息，不监控邻居路由信息来规避本问题，后面也会推出新版本，在bmp刚进行连接时，只上报过滤后的最优的路由，无需重新将所有路由率先报备一次。

