关于NE40/NE80/S8016 EACL规则增添后QoS基本失效的解决方法

周灯台 2006-08-20 发表

关于NE40/NE80/S8016 EACL规则增添后QoS基本失效的解决方法

问题现象：

xx 高校NE40每月都要进行Eacl的扩展（主要是针对国外站点进行更新），增加rule-map条目。某日进行升级，升级前版本为2226usr，升级到2321SP12usr，升级过程正常。但是在添加Eacl时，有如下提示：

[NOC-0]eacl free-ip-list fr517 permit

Send to No.2 Failed

Send to No.2 Failed

Send to No.2 Failed

Send to No.2 Failed

QoS: IPC_Send Error!

QoS: IPC_Send Error!

QoS: IPC_Send Error!

slot .2 QoS=IPC_Send Error!

QOS: Download build tree to slot 2 flag error!

按照升级后的配置加载设备后发现所有对QoS的操作返回IPC错误。进一步测试后确定当Eacl规则应用到两个端口上时可以正常工作，应用到三个端口上计算的时间大约为40s左右，当应用到4个端口上返回IPC错误，此时已经不能做QoS的操作，不能自动恢复。

NE40版本:VRP3.10-232112

原因分析：

根据现场的描述核对代码，QoS返回错误的原因时IPC调用超时，日志的断言也从另一个方面证明了当时接口板侧CPU占用率高，没有及时应答IPC消息。接口板CPU是被拿一个任务占用还需要继续分析。

通过实验室搭建测试环境复现现场问题，升级失败的原因已经彻底定位，总结起来有两个主要原因：

1、 现场配置的EACL规则中包含了较多的any节点，使得根据配置生成的SMT树的总节点数超出了接口板的处理上限。

2、 一般情况接口板都设置为小路由模式，虽然会造成SMT树节点丢失，但不会造成接口板的异常，EACL的删除、查询都可以正常进行。但是现场的接口板设置为大路由模式，导致接口板内存耗尽，nps任务被挂起，主控板和np的通讯中断。

原因1：对于接口板ACL节点规格问题，是由NP芯片的固有缺陷造成。由于ACL至少需要IP源地址、IP目的地址、报文协议类型、端口号（协议端口号）、TCP信息、UDP信息等。更加广义的流分类甚至需要基于二层信息，4层以上的信息。并且在实际应用中，ACL节点又存在大量的包含关系。这样就决定的ACL的算法非常复杂。在NP芯片的实现中，采用了SMT树的结构来计算ACL节点，该算法的关键点采用了NBT（Next Bit Test）算法，由于该算法为IBM提供的专利算法（NE40采用了IBM的转发引擎raineir），不能对外公布。此算法涉及到大量中间处理和数学算法操作，和单独的规则关系不大，主要决定于规则之间的相互关系。通过该算法无法对用户实际配置的ACL起到指导作用。根据实测结果，所配置的rule-map中的any节点越多，树的深度会越大，所产生的中间节点就越多。在测试中产生过配置121条规则，产生了超过10240个nodes的现象。

原因2：大小路由模式是转发引擎raineir的两种不同工作方式。小路由模式为普通工作方式，可以达到raineir标称的转发速度，接口板单板的路由数不超过20万条。大路由模式是为了支持低速接口要求更大路由规模的需求，使用了raineir的扩展内存，可以支持50万路由，但是转发速度降低。由于raineir的内存扩大，因此支持的SMT树的nodes数目也相应增加，当nodes数目过大时会造成接口板的CP内存耗尽，NPS任务（负责CP和raineir之间通讯）被挂起，导致所有对raineir操作的命令都不能执行，接口板处于异常状态。

解决方法：

针对目前情况，建议可以通过如下方式进行操作，尽可能的规避用户在配置中带来的不便：

a. 优化EACL规则的配置，尽量减少实际产生的中间节点。

b. 减少每个单板的应用端口数目。按照升级后的配置，推荐应用两个端口。

c. 使用LPUF单板，设置大路由模式。LPUF单板的CP内存较大，可以支持4端口规

则的下发。
d. 针对SMT树的节点数目的问题，后续版本中将在nodes节点满的情况下返回报错信息，以提醒用户修改配置。