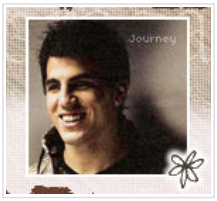
个人资料



slvher

访问： 162149次
积分： 2625
等级： BLOG 5
排名： 第9644名

原创： 99篇 转载： 1篇
译文： 0篇 评论： 30条

文章搜索

文章分类

Linux (32)
C (9)
Python (22)
GoLang (6)
Database (9)
NetworkProgramming (4)
CompilerPrinciples (5)
ComputerArchitecture (3)
NoSQL (6)
WebServer (6)
RecSys (12)

文章存档

2016年01月 (1)
2015年12月 (5)
2015年08月 (2)
2015年07月 (2)
2015年06月 (8)

展开

阅读排行

【Git笔记】"error:14090 (9444)
【Redis笔记】 第5篇: re (8121)

【Linux网络编程笔记】TCP短连接产生大量TIME_WAIT导致无法对外建立新TCP连接的原因及解决方法—实践篇

标签： Linux TCPIP Network Programming

2013-05-17 21:59 4485人阅读 评论(1) 收藏 举报

本文章已收录于： 计算机网络知识库

分类： NetworkProgramming（3）

上篇笔记主要介绍了与TIME_WAIT相关的基础知识，本文则从实践出发，说明如何解决文章标题提出的问题。

1. 查看系统网络配置和当前**TCP**状态

在定位并处理应用程序出现的网络问题时，了解系统默认网络配置是非常必要的。以x86_64平台Linux kernelversion 2.6.9的机器为例，ipv4网络协议的默认配置可以在/proc/sys/net/ipv4/下查看，其中与TCP协议栈相关的配置项均以tcp_xxx命名，关于这些配置项的含义，请参考这里的文档，此外，还可以查看linux源码树中提供的官方文档（src/linux/Documentation/ip-sysctl.txt）。下面列出我机器上几个需重点关注的配置项及其默认值：

```
[plain]
cat /proc/sys/net/ipv4/ip_local_port_range      32768   61000
cat /proc/sys/net/ipv4/tcp_max_syn_backlog      1024
cat /proc/sys/net/ipv4/tcp_syn_retries          5
cat /proc/sys/net/ipv4/tcp_max_tw_buckets       180000
cat /proc/sys/net/ipv4/tcp_tw_recycle           0
cat /proc/sys/net/ipv4/tcp_tw_reuse             0
```

其中，前3项分别说明了local port的分配范围（默认的可用端口数不到3w）、incomplete connection queue的最大长度以及3次握手时SYN的最大重试次数，这3项配置的含义，有个概念即可。后3项配置的含义则需要理解，因为它们在定位、解决问题过程中要用到，下面进行重点说明。

1) tcp_max_tw_buckets

这篇文档是这样描述的：Maximal number of time wait sockets held by system simultaneously. If this number is exceeded TIME_WAIT socket is immediately destroyed and warning is printed. This limit exists only to prevent simple DoS attacks, you must not lower the limit artificially, but rather increase it (probably, after increasing installed memory), if network conditions require more than default value (180000).

可见，该配置项用来防范简单的DoS攻击，在某些情况下，可以适当调大，但绝对不应调小，否则，后果自负。。。

2) tcp_tw_recycle

Enable fast recycling of sockets in TIME-WAIT status. The defaultvalue is 0 (disabled). It should not be changed without advice/request of technical experts.

该配置项可用于快速回收处于TIME_WAIT状态的socket以便重新分配。默认是关闭的，必要时可以开启该配置。但是开启该配置项后，有一些需要注意的地方，本文后面会提到。

3) tcp_tw_reuse

Allow to reuse TIME-WAIT sockets for new connections when it is safe from protocol viewpoint. The default value is 0. It should not

【Redis笔记】 第3篇: re (5564)
【Linux学习笔记】Linux (5380)
【MySQL实践经验】LO/ (5024)
【Linux网络编程笔记】T (4970)
【Linux网络编程笔记】T (4484)
【Linux学习笔记】kill及 (3872)
【Linux学习笔记】用nc (3803)
OpenSSL中RC4加解密 (2983)

评论排行

vc+mapx开发的程序"建 (8)
【Linux学习笔记】Linux (3)
【Python笔记】Python (2)
vim7.3编译报错error: ca (2)
【Python笔记】python第 (2)
【Redis笔记】 第5篇: re (2)
【Python笔记】如何用C (2)
内存写越界导致破坏堆结 (2)
【Linux网络编程笔记】T (1)
关于Windows编程中进程 (1)

推荐文章

*Android RocooFix 热修复框架
* android6.0源码分析之Camera API2.0下的初始化流程分析
*Android_GestureDetector手势滑动使用
*Android MaterialList源码解析
*Android官方开发文档Training系列课程中文版：创建自定义View之View的创建

最新评论

【Linux学习笔记】kill及kill -9的
liyang910910: 请教大神一个问题，如果服务没有对TERM信号进行处理，使用kill给服务进程发送TERM信号时，进程...
【网络编程笔记】Linux系统常见
Trust_FreeDom: 图都展示不出来
【Linux网络编程笔记】TCP短连
hust_dxxxd: 谢谢楼主，关于为什么要TIME_WAIT的两个原因将的很清楚！
【Python笔记】python第三方库
zlp1992: 知道解决方法了见：http://taoo.iteye.com/blog/1826912
【Python笔记】python第三方库
zlp1992: 楼主，源码安装libxml2时提示：/usr/bin/ld: /usr/local/lib/pyth...
【Linux学习笔记】Linux C中内
luop911123: 您好，请问第1大节第4小节中"constraints可以是gcc支持的各种约束方式"这句话是什么意思...
【Linux学习笔记】Linux C中内
纯黑老白: 多谢分析, 比很多网上的资料好理解.
vc+mapx开发的程序"建立空文档
feng840401917: @magicrb1:嗯嗯，谢谢我的没有解决，但是我下载了一个别人的就没有这个问题，我就直接在他的上面...
基于Java的开源日志库log4j调研
牛迁迁: 感谢楼主的分享，望多多指点：http://blog.csdn.net/u010028869

be changed without advice/request of technical experts.

开启该选项后，kernel会复用处于TIME_WAIT状态的socket，当然复用的前提是"从协议角度来看，复用是安全的"。关于"在什么情况下，协议认为复用是安全的"这个问题，这篇文章从linux kernel源码中挖出了答案，感兴趣的同学可以查看。

2. 网络问题定位思路

参考前篇笔记开始处描述的线上实际问题，收到某台机器无法对外建立新连接的报警时，排查定位问题过程如下：

用netstat –at | grep "TIME_WAIT"统计发现，当时出问题的那台机器上共有10w+处于TIME_WAIT状态的TCP连接，进一步分析发现，由报警模块引起的TIME_WAIT连接有2w+。将netstat输出的统计结果重定位到文件中继续分析，一般会看到本机的port被大量占用。

由本文前面介绍的系统配置项可知，tcp_max_tw_buckets默认值为18w，而ip_local_port_range范围不到3w，大量的TIME_WAIT状态使得local port在TIME_WAIT持续期间不能被再次分配，即没有可用的**local port**，这将是导致新建连接失败的最大原因。

在这里提醒大家：上面的结论只是我们的初步判断，具体原因还需要根据代码的异常返回值（如socket api的返回值及errno等）和模块日志做进一步确认。无法建立新连接的原因还可能是被其它模块列入黑名单了，本人就有过这方面的教训：程序中用libcurl api请求下游模块失败，初步定位发现机器TIME_WAIT状态很多，于是没仔细分析curl输出日志就认为是TIME_WAIT引起的问题，导致浪费了很多时间，折腾了半天发现不对劲后才想起，下游模块有防攻击机制，而发起请求的机器ip被不在下游模块的访问白名单内，高峰期上游模块通过curl请求下游的次数太过频繁被列入黑名单，新建连接时被下游模块的TCP层直接以RST包断开连接，导致curl api返回"Recv failure: Connection reset by peer"的错误。惨痛的教训呀 =_=

另外，关于何时发送RST包，《Unix Network Programming Volume 1》第4.3节做了说明，作为笔记，摘出如下：

An RST is a type of TCP segment that is sent by TCP when somethingis wrong.**Three conditions that generatean RST are:**

1) when a SYN arrives for a port that has no listening server;

2) when TCP wants to abort an existing connection;

3) when TCP receives a segment for a connection that does not exist. (TCPv1 [pp.246–250] contains additional information.)

3. 解决方法

可以用两种思路来解决机器TIME_WAIT过多导致无法对外建立新TCP连接的问题。

3.1 修改系统配置

具体来说，需要修改本文前面介绍的tcp_max_tw_buckets、tcp_tw_recycle、tcp_tw_reuse这三个配置项。

1）将tcp_max_tw_buckets调大，从本文第一部分可知，其默认值为18w（不同内核可能有所不同，需以机器实际配置为准），根据文档，我们可以适当调大，至于上限是多少，文档没有给出说明，我也不清楚。个人认为这种方法只能对TIME_WAIT过多的问题起到缓解作用，随着访问压力的持续，该出现的问题迟早还是会出现，治标不治本。

2）开启tcp_tw_recycle选项：在shell终端输入命令"echo 1 > /proc/sys/net/ipv4/tcp_tw_recycle"可以开启该配置。

需要明确的是：其实TIME_WAIT状态的socket是否被快速回收是由tcp_tw_recycle和tcp_timestamps两个配置项共同决定的，只不过由于tcp_timestamps默认就是开启的，故大多数文章只提到设置tcp_tw_recycle为1。更详细的说明（分析kernel源码）可参见这篇文章。

还需要特别注意的是：当client与server之间有如NAT这类网络转换设备时，开启tcp_tw_recycle选项可能会导致server端drop（直接发送RST）来自client的SYN包。具体的案例及原因分析，可以参考这里、这里或这里以及这里的分析，本文不再赘述。

3）开启tcp_tw_reuse选项：echo1 > /proc/sys/net/ipv4/tcp_tw_reuse。该选项也是与tcp_timestamps共同起作用的，另外socket reuse也是有条件的，具体说明请参见这篇文章。查了很多资料，与在用到NAT或FireWall的网络环境下开启tcp_tw_recycle后可能带来副作用相比，貌似没有发现tcp_tw_reuse引起的网络问题。

3.2 修改应用程序

具体来说，可以细分为两种方式：

1）将TCP短连接改造为长连接。通常情况下，如果发起连接的目标也是自己可控制的服务器时，它们自己的TCP通信最好采用长连接，避免大量TCP短连接每次建立/释放产生的各种开销；如果建立连接的目标是不受自己控制的机器时，能否使用长连接就需要考虑对方机器是否支持长连接方式了。

2）通过getsockopt/setsockoptapi设置socket的SO_LINGER选项，关于SO_LINGER选项的设置方法，

！

vc+mapx开发的程序"建立空文档
magicrb1: @feng840401917:解决了，我的问题就是mapx控件安装的不对才导致这个错误的，你可以试试...

《UNP Volume1》一书7.5节给出了详细说明，想深入理解的同学可以去查阅该教材，也可以参考这篇文章，讲的还算清楚。

4. 需要补充说明的问题

我们说TIME_WAIT过多可能引起无法对外建立新连接，其实有一个例外但比较常见的情况：S模块作为WebServer部署在服务器上，绑定本地某个端口；客户端与S间为短连接，每次交互完成后由S主动断开连接。这样，当客户端并发访问次数很高时，S模块所在的机器可能会有大量处于TIME_WAIT状态的TCP连接。但由于服务器模块绑定了端口，故在这种情况下，并不会引起"由于TIME_WAIT过多导致无法建立新连接"的问题。也就是说，本文讨论的情况，通常只会在每次由操作系统分配随机端口的程序运行的机器上出现（每次分配随机端口，导致后面无端口可用）。

【参考资料】

1. ipsysctl-tutorial之tcpvariables
2. proc_sys_net_ipv4
3. linux+v3.2.8/Documentation/networking/ip-sysctl.txt
4. 系列文章—tcp短连接TIME_WAIT问题解决方法大全1-5
5. 打开tcp_tw_recycle引起的一个问题
6. dropping of connections with tcp_tw_recycle = 1
7. tcp_tw_recycle和nat造成syn_ack问题

========== EOF ==========

上一篇 【Linux网络编程笔记】TCP短连接产生大量TIME_WAIT导致无法对外建立新TCP连接的原因及解决方法—基础知识篇
下一篇 【Linux学习笔记】kill及kill -9的用法及如何实现进程的优雅退出

我的同类文章

NetworkProgramming（3）

- 【网络编程笔记】Linux系统… 2015-06-29 阅读 689
- 【Linux网络编程笔记】TCP… 2013-05-17 阅读 4963
- 【网络编程基础笔记】struct… 2013-04-26 阅读 1153

猜你在找

查看评论

1楼 wooogooo 2015-02-27 20:46发表

Thanks!

发表评论

用户名： liu1989feng

评论内容：

提交

* 以上用户言论只代表其个人观点，不代表CSDN网站的观点或立场

核心技术类目

全部主题 Hadoop AWS 移动游戏 Java Android iOS Swift 智能硬件 Docker OpenStack
VPN Spark ERP IE10 Eclipse CRM JavaScript 数据库 Ubuntu NFC WAP jQuery
BI HTML5 Spring Apache .NET API HTML SDK IIS Fedora XML LBS Unity
Splashtop UML components Windows Mobile Rails QEMU KDE Cassandra CloudStack
FTC coremail OPhone CouchBase 云计算 iOS6 Rackspace Web App SpringSide Maemo
Compuware 大数据 aptech Perl Tornado Ruby Hibernate ThinkPHP HBase Pure Solr
Angular Cloud Foundry Redis Scala Django Bootstrap