为什么您的应用程序仅使用10Mbps，即使链路为1Gbps?

目录

[简介](#)
[背景信息](#)
[问题概述](#)
[带宽延迟产品](#)
[验证](#)
[解决方案](#)
[如何区分两个地点之间的往返时间(RTT)?](#)

简介

本文档介绍与高速、高延迟网络相关的问题。它从BDP中推导出计算给定条件下实际带宽使用的公式。

背景信息

随着越来越多的企业已经或正在构建地理上分散的数据中心，并通过高速链路将数据中心互联。对更好地利用带宽的需求正在增加。

带宽延迟产品(BDP)已在Internet上发布多年。然而，在这个问题看起来如何方面，没有现实的例子。BDP公式侧重于TCP窗口大小。它不提供根据距离计算可能的带宽使用量的方法。本文档简要介绍BDP并演示问题和解决方法。本文还推导了在给定条件下计算带宽使用的公式。

问题概述

您的公司有两个数据中心。您的公司将业务关键型数据从一个数据中心备份到另一个数据中心。备份管理员报告，由于网络速度慢，他们无法在备份窗口内完成备份。作为网络管理员，您被指派调查网络速度缓慢问题。您知道以下因素：

- 这两个数据中心相距1000千米。
- 这些数据中心通过1Gbps链路互连。

经调查，您注意到：

- 有足够的可用带宽。
- 不存在网络硬件或软件问题。

- 备份应用程序仅使用约10Mbps的带宽，即使剩余的990Mbps带宽是免费的。
- 备份应用程序使用TCP传输数据。

带宽延迟产品

为了回答备份应用程序仅使用10Mbps的问题，它引入了带宽延迟产品(BDP)。

BDP只是声明：

BDP（位）= total_available_bandwidth（位/秒）x round_trip_time（秒）

或者，因为RWIN/BDP通常以字节为单位，而延迟以毫秒为单位：

BDP（字节）= total_available_bandwidth(KBytes/sec)x round_trip_time(ms)

这意味着TCP窗口是一个缓冲区，它确定在服务器停止之前可以传输多少数据，并等待接收数据包的确认。吞吐量实质上受BDP的约束。如果BDP（或RWIN）低于延迟和可用带宽的乘积，则无法填写行，因为客户端无法以足够快的速度发送确认。传输不能超过（RWIN /延迟）值，因此TCP窗口(RWIN)需要足够大以适合maximum_available_bandwidth x maximum_endesided_delay。

使用上述公式。导出的带宽计算公式为：

带宽使用(Kbps)= BDP（字节）/RTT(ms)* 8

注意：此公式计算最大理论带宽使用量。它不考虑操作系统的数据包传输时间，因为它涉及许多因素，例如可用内存、网卡驱动程序、本地网卡速度、缓存，有时甚至磁盘速度。因此，当TCP窗口大小较大时，计算的带宽将大于实际带宽。当TCP窗口大小非常大时，偏差也会很大。

使用派生的公式，您可以回答为什么备份应用程序只能使用10Mbps的问题，方法是执行以下计算。

- 1000KM的RTT一般为~15，因此RTT=15ms
- 默认情况下，Windows 2003操作系统Windows大小为17,520字节。BDP=17,520字节
- 将这些数字放入公式中：

带宽使用(Kbps)=17520/15*8。

结果是9344Kbps或9.344Mbps。9.344Mbps加上TCP和IP报头。最终结果是约10Mbps。

验证

作为网络管理员，您理论上已经回答了这个问题。现在，你需要在现实世界中证实这个理论。

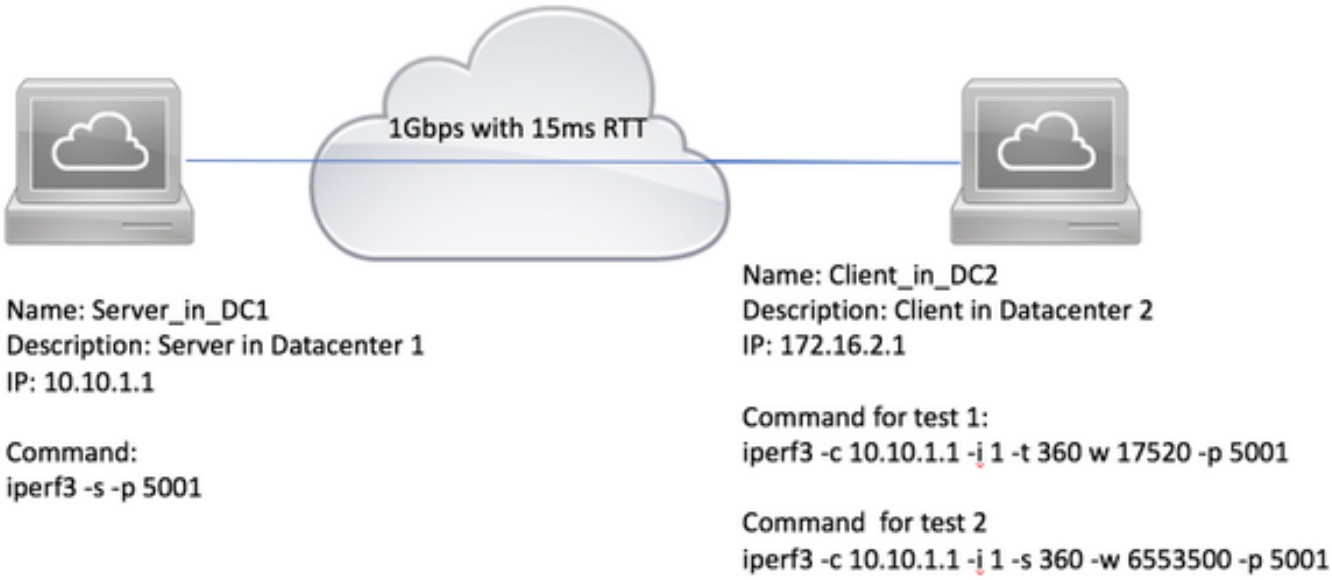
您可以使用任何网络性能测试工具来确认该理论。您决定运行**iperf**来演示问题和解决方法。

实验设置如下：

1. 数据中心1中IP地址为10.10.1.1的服务器。

2. 数据中心2中IP地址为172.16.2.1的客户端。

拓扑如图所示：



请按照以下步骤验证：

1. 在10.10.1.1上运行**iperf3 -s -p 5001**，使其成为服务器并侦听TCP端口5001。

2. 使用默认TCP窗口大小17,520字节进行测试。在**172.16.2.1上运行iperf3 -c 10.10.1.1 -** i 1 **-t 360 -w 17520 -p 5001**，使其成为客户端。此命令告知iperf在端口5001上连接到服务器，运行360秒，并在TCP windows大小为17,520字节时每1秒报告一次带宽使用情况。

3. 要使用自定义TCP窗口大小（如6,553,500字节）进行测试，请运行**iperf3 -c 10.10.1.1 -i 1 -t 360 -w 6553500 -p 5001**

这是默认TCP窗口大小为17,520字节的实验测试结果。您可以看到带宽使用率约为10Mbps。

```
C:\Tools>iperf3.exe  -c 10.10.1.1  -t 360  -p 5001 -i 1 -w 17520

Connecting to host 10.10.1.1, port 5001

[  4] local 172.16.2.1 port 49650 connected to 10.10.1.1 port 5001

[ ID] Interval           Transfer     Bandwidth

[ ID] Interval           Transfer     Bandwidth

[  4]   0.00-1.00   sec  1.30 MBytes  10.9 Mbits/sec

[  4]   1.00-2.02   sec   919 KBytes  7.41 Mbits/sec

[  4]   2.02-3.02   sec  1.28 MBytes  10.7 Mbits/sec

[  4]   3.02-4.02   sec  1.14 MBytes  9.59 Mbits/sec

[  4]   4.02-5.01   sec  1.24 MBytes  10.4 Mbits/sec

[  4]   5.01-6.01   sec  1.33 MBytes  11.3 Mbits/sec
```

```
[  4]   6.01-7.01   sec  1.15 MBytes  9.65 Mbits/sec

[  4]   7.01-8.01   sec  1.12 MBytes  9.36 Mbits/sec

[  4]   8.01-9.01   sec  1.22 MBytes  10.3 Mbits/sec

[  4]   9.01-10.01  sec  1.13 MBytes  9.49 Mbits/sec

[  4]  10.01-11.01  sec  1.30 MBytes  10.8 Mbits/sec

[  4]  11.01-12.01  sec  1.17 MBytes  9.84 Mbits/sec

[  4]  12.01-13.01  sec  1.13 MBytes  9.48 Mbits/sec

[  4]  13.01-14.01  sec  1.28 MBytes  10.7 Mbits/sec

[  4]  14.01-15.01  sec  1.40 MBytes  11.8 Mbits/sec

[  4]  15.01-16.01  sec  1.24 MBytes  10.4 Mbits/sec

[  4]  16.01-17.01  sec  1.30 MBytes  10.9 Mbits/sec

[  4]  17.01-18.01  sec  1.17 MBytes  9.78 Mbits/sec
```

这是TCP窗口大小为6,553,500字节的实验测试结果。您可以看到带宽使用率约为200Mbps。

```
C:\Tools>iperf3.exe  -c 10.10.1.1  -t 360  -p 5001 -i 1 -w 6553500

Connecting to host 10.10.1.1, port 5001

[  4] local 172.16.2.1 port 61492 connected to 10.10.1.1 port 5001

[ ID] Interval           Transfer     Bandwidth

[  4]   0.00-1.00   sec  29.1 MBytes   244 Mbits/sec

[  4]   1.00-2.00   sec  25.4 MBytes   213 Mbits/sec

[  4]   2.00-3.00   sec  26.9 MBytes   226 Mbits/sec

[  4]   3.00-4.00   sec  18.2 MBytes   152 Mbits/sec

[  4]   4.00-5.00   sec  25.8 MBytes   217 Mbits/sec

[  4]   5.00-6.00   sec  28.8 MBytes   241 Mbits/sec

[  4]   6.00-7.00   sec  26.1 MBytes   219 Mbits/sec

[  4]   7.00-8.00   sec  21.1 MBytes   177 Mbits/sec

[  4]   8.00-9.00   sec  22.5 MBytes   189 Mbits/sec

[  4]   9.00-9.42   sec  9.54 MBytes   190 Mbits/sec
```

解决方案

从软件开发的角度看，多线程运行多个并发TCP会话可以提高带宽使用率。但是，网络管理员或系统管理员修改源代码并不可行。您可以微调操作系统。

RFC1323为高性能TCP定义了多个TCP扩展，包括Window Scale选项和选择性ACK。它们由主操作系统实施。但是，默认情况下，某些操作系统会禁用它们，甚至会写入TCP/IP堆栈来支持它们。

- 默认情况下，这些操作系统禁用RFC1323:Windows 2000、Windows 2003、Windows XP和内核低于2.6.8的Linux。

  如果您在Microsoft Windows系统上遇到问题，请通过此链接微调TCP。
https://support.microsoft.com/en-au/kb/224829。

  有关其他操作系统，请参阅供应商关于如何配置它们的文档。

- 默认情况下，这些操作系统启用RFC1323:Windows 2008及更高版本、Windows Vista及更高版本、带内核2.6.8及更高版本的Linux。您可能需要应用补丁来改进这些功能。在某些情况下，需要禁用它们。请参阅供应商有关如何禁用它们的文档。

- 某些设备构建在Microsoft Windows 2000、Windows 2003或嵌入式操作系统之上。例如NAS、医疗保健硬件。请检查供应商的文档以验证RFC1323是否已启用。

如何区分两个地点之间的往返时间(RTT)?

通常，RTT与距离关联。下表列出了距离及其相关RTT。在正常网络条件下，您也可以使用ping测试来获取有关RTT的一些想法。

距离（千米）	RTT(ms)
1,000	15
4,000	50
8,000	120

> **注意：**以上仅是指南，实际RTT时间可能不同。此外，延迟还受所使用技术的影响。例如，3G延迟可以频繁地为100ms，而不管距离如何。卫星也是如此。