為什麼即使鏈路為1Gbps，您的應用程式也只使用10Mbps?

目錄

簡介

本文描述與高速、高延遲網路相關的問題。它從BDP中匯出一個公式，用以計算給定條件下的實際頻寬使用情況。

背景資訊

隨著越來越多的企業已經或正在建立地理上分散的資料中心，並通過高速鏈路互連這些資料中心。對更好地利用頻寬的需求日益增加。

寬頻延遲產品(BDP)已經在網際網路上發佈多年。然而，關於這個問題看起來如何，沒有真實案例。BDP公式重點關注TCP視窗大小。它不能為我們提供根據距離計算可能的頻寬使用量的方法。本檔案簡要說明BDP，並演示問題和解決方案。本文還推導了給定條件下頻寬使用率的計算公式。

問題概述

您的公司有兩個資料中心。您的公司將業務關鍵資料從一個資料中心備份到另一個資料中心。備份管理員報告由於網路速度緩慢，無法在備份視窗內完成備份。作為網路管理員，您被指派調查網路速度慢的問題。您知道以下因素：

- 這兩個資料中心相距1000公里。
- 這些資料中心通過1Gbps鏈路互連。

經調查，你已注意到：

- 可用頻寬充足。
- 不存在網路硬體或軟體問題。

- 備份應用程式僅使用約10Mbps的頻寬，即使其餘的990Mbps頻寬是免費的。
- 備份應用程式使用TCP傳輸資料。

頻寬延遲產品

為了回答備份應用程式僅使用10 Mbps的問題，引入了頻寬延遲產品(BDP)。

發展局只是說：

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

或者，因為RWIN/BDP通常以位元組為單位，而延遲以毫秒為單位：

BDP（位元組）= total_available_bandwidth(KBytes/sec)x round_trip_time（毫秒）

這表示TCP視窗是一個緩衝區，用於確定在伺服器停止並等待接收封包確認之前可傳輸多少資料。吞吐量基本上受BDP的限制。如果BDP（或RWIN）低於延遲和可用頻寬的乘積，則無法填充線路，因為客戶端無法以足夠快的速度傳送回確認。傳輸不能超過(RWIN / latency)值，因此TCP視窗(RWIN)需要足夠大，以滿足maximum_available_bandwidth x maximum_expected_delay。

上面有公式。推導的頻寬計算公式為：

頻寬使用情況(Kbps)=BDP（位元組）/RTT（毫秒）* 8

附註：此公式計算最大理論頻寬使用量。它不考慮作業系統的資料包傳輸時間，因為它涉及許多因素，例如可用記憶體、NIC驅動程式、本地NIC速度、快取，有時甚至是磁碟速度。因此，當TCP視窗大小較大時，計算的頻寬將大於實際頻寬。當TCP視窗非常大時，偏差也會很大。

通過推導的公式，您可以回答為什麼備份應用程式只能使用10Mbps的問題，方法如下：

- 1000KM的RTT一般為~15，所以RTT=15ms
- 預設情況下，Windows 2003作業系統Windows大小為17,520位元組。因此BDP = 17,520位元組
- 將這些數字輸入公式：

頻寬使用率(Kbps)=17520/15*8。

結果是9344Kbps或9.344Mbps。9.344Mbps，加上TCP和IP報頭。最終結果為~10Mbps。

驗證

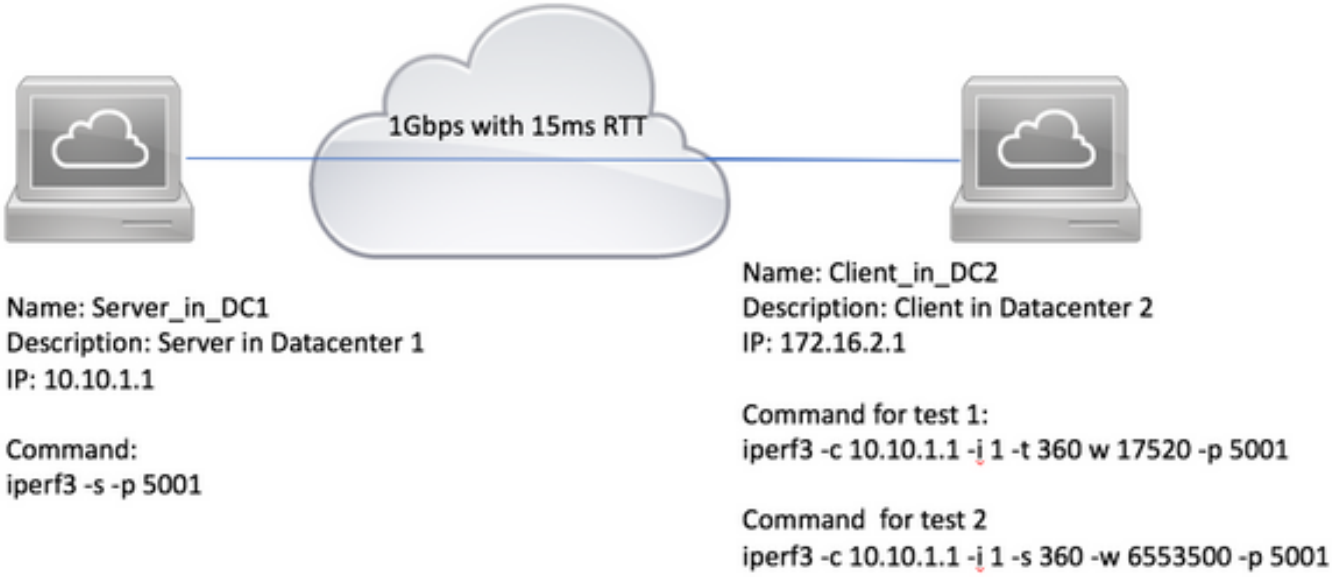
作為網路管理員，您理論上已經回答了這個問題。現在你要在現實世界中確認這個理論。

您可以使用任何網路效能測試工具來驗證此理論。您已決定運行**iperf**來演示問題和解決方案。

以下是實驗室設定：

1. 資料中心1中IP地址為10.10.1.1的伺服器。
2. 資料中心2中的IP地址為172.16.2.1的客戶端。

拓撲如圖所示：



請按照以下步驟進行驗證：

1. 在10.10.1.1上運行**iperf3 -s -p 5001**，使其成為伺服器並在TCP埠5001上偵聽。

2. 使用預設TCP視窗大小17,520位元組進行測試。在172.16.2.1上運行**iperf3 -c 10.1.1 -i 1 -t 360 -w 17520 -p 5001**以使其成為客戶端。此命令通知iperf連線到埠5001上的伺服器，運行時間為360秒，並且每1秒報告頻寬使用量，TCP windows大小為17,520位元組。

3. 要測試自定義TCP視窗大小（如6,553,500位元組），請運行**iperf3 -c 10.10.1.1 -i 1 -t 360 -w 6553500 -p 5001**

這是預設TCP視窗大小為17,520位元組的實驗室測試結果。您可以看到頻寬使用率約為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
```

這是TCP視窗大小為6,553,500位元組的實驗室測試結果。您可以看到頻寬使用率約為200 Mbps。

```
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
```

解決方案

從軟體開發角度看，多執行緒運行多個併發TCP分段可以改善頻寬使用情況。但是，網路或系統管

理員修改原始碼並不現實。您可以微調作業系統。

RFC1323為高效能TCP定義了多個TCP擴展。這些擴展包括視窗縮放選項和選擇性ACK。它們由主作業系統實施。但是，預設情況下，某些作業系統會禁用它們，即使TCP/IP協定棧也編寫以支援它們。

- 這些作業系統預設禁用RFC1323:Windows 2000、Windows 2003、Windows XP和Linux的核心版本低於2.6.8。

如果您在Microsoft Windows系統上遇到問題，請點選此連結微調TCP。 https://support.microsoft.com/en-au/kb/224829。

對於其他作業系統，請參閱供應商有關如何配置它們的文檔。

- 預設情況下，這些作業系統啟用RFC1323:Windows 2008及更高版本、Windows Vista及更高版本、核心2.6.8及更高版本的Linux。您可能需要應用修補程式來改善這些功能。在某些情況下，需要禁用它們。有關如何禁用它們，請參閱供應商文檔。

- 某些裝置構建在Microsoft Windows 2000、Windows 2003或嵌入式作業系統之上。例如NAS、醫療保健硬體。請檢查供應商的文檔以驗證RFC1323是否已啟用。

如何得知兩個地點之間的往返時間(RTT)?

一般來說，RTT與距離有關。下表列出距離及其相關RTT。在正常網路情況下，您還可以使用ping測試來瞭解RTT。

距離(KM)	RTT(ms)
1,000	15
4,000	50
8,000	120

> **附註**：以上僅作為指南，實際RTT時間可以變化。此外，所用技術也會影響延遲。例如，無論距離如何，3G延遲通常可以達到100ms。衛星也是如此。