

고성능 망에서의 고속 전송을 위한 시스템 튜닝

권윤주*, 석우진*, 변옥환*

*한국과학기술정보연구원

e-mail:{yulli, wjseok, ohbyeon}@kisti.re.kr

System Tuning for High Speed Transfer in Long Fat Network

Yoonjoo Kwon*, Woojin Seok*, Ok-Hwan Byeon*

*Korea Institute of Science Technology Information

요 약

네트워크와 연구 장비들의 발전에 따라 여러 과학 응용 분야에서 협업체계를 구축하면 원거리의 파트너들과 실시간적 연구를 진행하고 있다. 네트워크 대역폭의 크기는 현재의 상황에 충족될 수 있을 수준이지만, 데이터를 실질적으로 전송하는 프로토콜의 경우 고대역폭 요구 응용을 위해서 여러 가지 면에서 수정 및 튜닝할 필요가 있다. 시스템 단위에서 전송 프로토콜의 성능을 향상 시킬 수 있는 방법에 대해서 본 논문에서 모색해보기로 한다.

1. 서론

네트워크와 연구 장비들이 발전함에 따라 여러 과학 응용 분야에서는 협업체계를 구축하며 원거리에서 연구를 진행하고 있다. 이에 따라 기존에 메일이나, 단순한 파일 다운로드 등으로 이용되었던 네트워크는 좀 더 의미있는 연구의 한 부분으로 그 중요성을 더해가고 있다.

현재 세계 연구망의 네트워크 대역폭을 보면 10Gbps에서 100Gbps 정도의 매우 높은 대역폭으로 변모하고 있다. 고에너지 물리 분야, 천문학 분야 등에서 연구 협업 체계를 갖추기 위해 요구되는 대역폭이 이미 10Gbps를 넘어가고 있기 때문에 우리나라 뿐만이 아니라 이러한 연구의 중요성을 인지하는 나라들에서는 연구망을 급속히 증속하고 있다.

네트워크의 발전, 연구 장비의 발전 등 하드웨어적인 발판이 마련된 가운데, 여전히 병목지점으로 남아있는 시스템 단위에서 전송 프로토콜의 성능을 향상 시킬 수 있는 방법에 대해서 본 논문에서 모색해보기로 한다.

2. 관련 연구

1) 고대역폭을 요구하는 과학 응용들

<표 1> 응용별 요구 대역폭의 변화

응용연구		연도		
		'05	'06-'07	'08-'09
K-STAR 핵융합연구	데이터량(TB)	4	200	500
	링크속도 (Gbps)	0.6	5	10
고에너지물리	데이터량(TB)	100	800	3000
	링크속도 (Gbps)	5	10	40-100
가상천문대	데이터량(TB)	6	100	1000
	링크속도 (Gbps)	0.6	5	10
기후변화	데이터량(TB)	5	100	500
	링크속도 (Gbps)	1	5	10

[1]에 따르면, 고에너지 물리, 핵융합에너지, 기후변화 연구 등의 분야에서 현재에 비해 향후 5년 이내

에 데이터량과 링크속도에 있어서 각각 50배와 10배의 증가가 예상된다. 이에 따라 국내·외의 연구망들이 네트워크 인프라 증속을 추진하고 있다.

2) 성능 향상을 위한 파라미터

네트워크의 성능 향상을 위해서, 네트워크 인프라의 발전과 더불어 데이터를 전송해주는 프로토콜의 역할이 중요하다. 현재 데이터 전송을 위한 프로토콜로서, TCP(Transmission Control Protocol)을 많이 사용하고 있으나, 이는 저속에서 개발된 프로토콜로서, 현재와 같은 고대역폭에서 튜닝없이 사용하기에는 많은 한계가 있다. 프로토콜 성능을 향상 시켜줄 수 있는 파라미터에는 다음과 같은 것들이 있다.

① MTU size

2계층에서 지원하는 프레임 크기로서 Maximum Transfer Unit 이다. 이것은 전송시스템에서 한 번에 캡슐화할 수 있는 데이터 사이즈를 결정한다. <표 2>에서 보는 것처럼 MTU 사이즈는 2계층 매체마다 다르다[2]. MTU 사이즈가 클수록 더 높은 네트워크 성능을 낼 수 있으므로, MTU 사이즈가 큰 매체를 이용하는 것이 성능향상을 위해 효과적이다.

또한 성능향상을 위해서는 중간 단계에서 프래그멘테이션이 일어나지 않을 크기로 세팅하는 것이 중요하다. 프래그멘테이션하는 과정에서의 오버헤드가 성능 손실을 야기시킬 수 있기 때문이다.

<표 2> 전형적인 MTU

네트워크	MTU(bytes)
Hyperchannel	65535
16Mbps 토큰 링 (IBM)	17914
4Mbps 토큰 링 (IEEE 802.5)	4464
FDDI	4352
Ethernet	1500
IEEE 802.3/802.2	1492
x.25	576

② window size

윈도우 사이즈는 TCP 레이어에서 전송 시스템이 상대방으로부터 ACK(수신확인)를 안받고 보낼 수 있는 최대 데이터 사이즈를 의미한다. 기본 TCP 헤더의 구조에서는 "2¹⁶ -1"까지의 크기만을 지원하기 때문에, 최대 65535 즉 64KB의 데이터를 한꺼번에 전송할 수 있다. 그런데 데이터전송지역이 멀어지다

보면, 64KB크기의 데이터를 모두 전송하고 난후에 에러가 없음에도 불구하고 ACK가 도달하는 시간이 늦어져서 전송의 idle 시간을 갖게 되는 경우가 생긴다. 이러한 이유로 이것이 네트워크 성능이 한계를 가져오게 되는 하나의 원인으로 작용하게 될 수도 있다.

③ parallel stream(?)

한 스트림을 이용하여 네트워크 대역폭을 모두 소진하는 데에는 한계가 있으므로 여러 개의 스트림을 이용하여 네트워크 성능 향상을 기대할 수 있다.

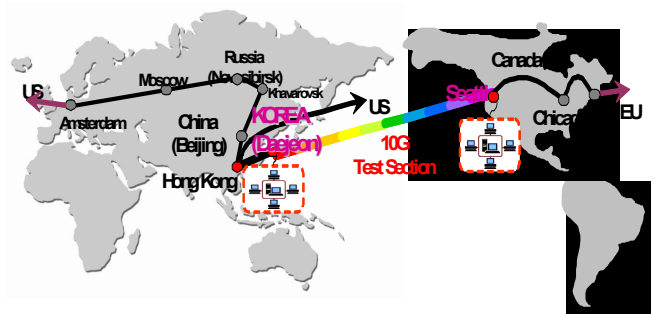
3. 성능 향상 테스트 및 결과

8월1일부터 초고속연구망이 미국지역과 10Gbps의 네트워크 인프라가 마련되었다. 이 연구망의 경우 주 목표가 응용연구자간의 협업 지원이기 때문에 10G를 충분히 사용할 수 있는 전송 프로토콜의 환경을 제공해주어야 한다. 앞서 언급한 성능 향상을 위한 시스템 파라미터에 변화를 주면서 전송량이 어떻게 변하는 지 테스트하였다.

1) 테스트 환경

이 연구망의 성능테스트를 하기 위하여 (그림 1)에서 보는 바와 같은 테스트 구간에 다음과 같은 환경으로 테스트를 수행하였다.

- 전송 프로토콜 : TCP 중 BIC-TCP
- 운영체제 : linux 2.6
- 테스트 구간 : 대전↔시애틀
- 테스트 구간의 RTT(Round Trip Time) : 184ms
- 대역폭 : 10Gbps
- 테스트 툴 : iperf 1.7 (NLANR 개발)



(그림 1) 테스트구간

2) 성능 테스트

대전 시스템에서 시애틀 시스템까지의 홉수는 3홉

이었다.

대전 시스템-대전 라우터-시애틀 라우터-시애틀 시스템

1차적으로 시스템과 라우터의 MTU 사이즈를 1500bytes로 맞추어 iperf로 TCP와 UDP에 대한 테스트를 수행하였고, 이후 MTU 사이즈를 9180bytes로 맞추어 1차와 같은 테스트를 수행하였다.

2차적으로 윈도우 사이즈에 변화를 가하면서 전송량의 차이를 비교하였다. 3차적으로는 스트림의 수를 증가시키면서 스트림 개수와 전송량의 관계를 정리하였다.

3) 성능 테스트 결과

① MTU size에 따른 변화

기본적으로 Ethernet의 경우 1500bytes의 MTU 사이즈를 갖는다. 그러나 말단 시스템, 라우터 장비 등의 발전으로 현재 9180bytes까지 지원가능하도록 되어 있다. <표 3>에서 보는 것처럼, 단위 패킷의 크기인 MTU 사이즈가 커짐에 따라 최대 전송률이 2 ~ 2.3배 정도 성능향상을 볼 수 있었다. 모든 망에 적용될 수 있는 것은 아니지만, 고속전송을 위해서는 1500bytes로는 한계를 많이 가지고 있으므로, 9KB로 MTU 사이즈로의 증가된다면 고려되어야 한다.

<표 3> MTU 변화에 따른 최대전송률

최대전송률 프로토콜	MTU: 1500	MTU : 9180
TCP	2.9Gbps	6.2Gbps
UDP	2.8Gbps	7.2Gbps

② 윈도우 사이즈에 따른 변화

(그림 2)에서 보는 것처럼 윈도우 크기의 변화에 따라 single stream으로 전송할 수 있는 throughput의 차이를 비교해볼 수 있다. 앞서 언급된 window size의 크기는 최대 64KB였으나, 대륙간 데이터전송을 하는 데 있어서 64KB는 상당히 작은 크기임으로 window size를 높여줄 수 있는 방법이 필요하다.

윈도우사이즈를 확장하기 위해서는 우선 해당 시스템이 Window Scaling 옵션이 허가되어 있어야 한다. Window Scaling 옵션이 허가된 경우 해당 시

스템은 $2^{32} - 1$, 즉 최대 4.2GB의 데이터 사이즈를 ACK 없이 전송할 수 있다[3]. 그러나 윈도우 사이즈가 맥시멈으로 하는 것이 성능향상에 있어서 최선의 방법은 아니다. 왜냐하면 단대 단 링크에서 수용할 수 있는 패킷 수 이상으로 패킷이 늘어날 경우 혼잡(congestion)을 발생시키는 하나의 이유가 되기 때문이다.

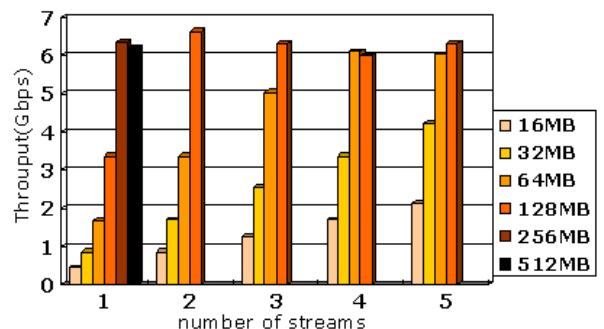
(그림 2)에서 스트림의 수가 1개 일 때의 윈도우 사이즈 별 전송량의 변화를 보면, 16MB, 32MB의 1Gbps 이하, 64MB 일 때 1.8Gbps 128MB일 때 3.3Gbps, 256MB 일 때 6.2Gbps의 전송량을 볼 수 있다. 512MB부터는 256MB의 최고점에서 조금씩 떨어지는 그래프를 보여준다.

일반적으로 최적의 윈도우 사이즈의 정의는,

$$\text{window size} = \text{RTT} * \text{bandwidth}$$

이렇게 나타낼 수 있다. 이 테스트에서 보자면 $184\text{ms} * 1/1000 * 10\text{Gbps} * 1/8 = 230\text{MB}$

일반적으로 bandwidth delay product 라고 불리는 이 값은 단대 단 간 데이터를 채울 수 있는 최대사이즈라고 볼 수 있다. 윈도우 사이즈는 ACK를 안 받고 전송 시스템에서 보낼 수 있는 최대량이기 때문에 윈도우 사이즈를 bandwidth delay product 이상 값으로 양단말 시스템을 세팅해주는 것이 (그림 2)의 결과에서 볼 수 있는 것처럼 최대의 성능을 낼 수 있다고 볼 수 있다.



(그림 2) 네트워크 성능 테스트 결과

③ stream 수에 따른 변화

(그림 2)에서 보면, 우선 윈도우 사이즈에 따라서 전송량의 차이가 보인다. 그리고 stream 수에 따라서 차이도 또한 볼 수 있다. 그러나 윈도우 사이즈가 적절한 윈도우 사이즈, 즉 single stream에서 최대

속도를 내지 못하는 경우에는 스트림 수에 따른 변화를 기대할 수 있으나, 최대 사이즈까지 도달했을 때에는 오히려 성능이 나빠지는 현상을 야기할 수도 있음을 볼 수 있다. 이 이유는 앞서 설명한 바와 같이, 단대단 간 채울 수 있는 데이터크기보다 많은 양의 패킷들이 네트워크 링크로 유입되려고 하기 때문에 그들간 충돌현상이 야기되기 때문이다.

Analysis and Tuning", IBM

4. 결론

네트워크의 발전으로 인하여 여러 가지 측면에서 협업 체계를 통한 연구의 발판이 마련되었다. 앞으로 좀 더 좋은 응용을 개발하기 위해서 응용의 발전과 독립적으로 네트워크 레이어와 전송 레이어에서의 발전이 꾸준히 진행되어야 할 것이다.

본 논문에서는 전송 프로토콜의 성능에 영향을 끼치는 파라미터들에 대해서 실험치를 바탕으로 영향 정도를 분석해보았다. 이 결과를 바탕으로 보면, MTU 사이즈와 Window 사이즈의 변화가 고속 전송을 하는 데 있어서 가장 큰 영향을 미치는 것으로 분석되었다. MTU 사이즈의 경우에는 되도록 수용할 수 있는 큰 사이즈로의 변화가 의미가 있었고, Window 사이즈의 경우에는 단대 단 간 지연(delay)과 연결되어 있는 대역폭에 따라 적절하게 조절해주는 것이 최대 성능을 내는 데 있어서 효과적이었다.

이처럼 네트워크의 성능향상을 위해서는 기본적으로 전송 프로토콜의 환경을 적절하게 구성하여 주는 것이 중요하고, 이에 더불어 각 전송 프로토콜의 메카니즘상의 문제를 풀어가는 것이 필요하다고 사료된다.

현재 여러 프로토콜이 이를 위해 출현하였고 계속 개발 중이지만, 전송 거리에 따라 좀 차이가 있지만 7Gbps를 넘는 전송 성능을 안정적으로 지원하는 프로토콜이 아직 부재하다.

향후 계획으로는 전송프로토콜의 적절한 환경 구성과 더불어 성능 향상시킬 수 있는 전송 프로토콜의 메카니즘의 변화에 대하여 연구하는 것이다.

참고문헌

- [1] 기술문서, "글로벌 과학기술 협업 연구망 사용자 분석 및 활성화 방안 연구", 한국과학기술정보연구원
- [2] W. Richard Stevens, "TCP/IP Illustrated", Vol. 1, Addison Wesley
- [3] Tung Ouyang, "Practical TCP/IP Performance