람다네크워크를 통한 대규모 멀티도메인 그리드환경구현 연구

노민기*, 안성진*
*성균관대학교 컴퓨터교육학과
한국과학기술정보연구원
e-mail:mknoh@kisti.re.kr, sjahn@skku.ac.kr

The lambda network to build multi-domain intensive large-scale grid environment

Min-Ki Noh*, Sung Jin Ahn*
*Dept of Computer education Sungkyunkwan University
Korea Institute of Science Technology Information

요 약

분산된 자원의 실시간 정보교환과 그리드를 통한 효율적인 자원 재구성을 위해서는 기존의 단일 도메인에서 구성되는 네트워크와는 다르게 대규모 가상도메인(Large-Scale Multi-domain)을 위한 네트워크의 성능과 기능 향상이 필요하다. 그리드네트워크를 기반으로 활발히 진행 중인 글로벌한 연구자원을 대상으로 공유된 자원의 성능 개선과 자원 간 데이터전달의 효율 개선을 위해 TDM(Time Division Multiplexing)기반의 Multi-Point Lambda-Path Ring 구현 기술을 제안하고 이를 Multi-Domain 간 Control Plane하에서 최적의 가상도메인으로 구성 할 수 있는 기법을 제안한다.

1. 서론

첨단 어플리케이션은 슈퍼컴퓨터, 대용량 저장 장치와 같은 고성능의 자원뿐만 아니라 글로벌한 환경 하에서 고성능의 네트워크를 통해 공유되는 대규모의 연구자원을 필요로 한다. 따라서 이러한 첨단 어플리케이션의 효과적인 그리드 환경구축과 자원을 제공하기 위해서는 가능한 넓은 지역의 자원을 빠르게 데이터를 전달하여 원격의 컴퓨터와 실험 장비를 효율적으로 제어할 수 있는 고성능 네트워크 환경이 필요하다. 본 연구는 그리드기술과 구축된 환경을 효율적으로 보장하기 위해 대용량 데이터전달과 안정적 제어통신보장이 효율적으로 가능한 다지점간 광패스 구현기술을 제안하고 멀티도메인을 연계하여 대규모의 가상연구자원을 구성하고 관리하는 기법을 제안하고자 한다.

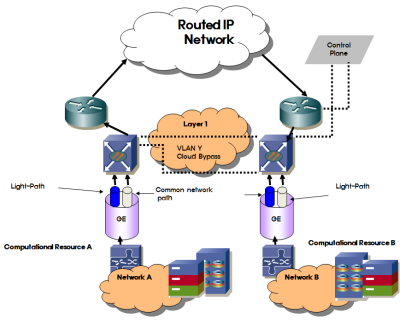
2. Grid Lambda path와 IP path의 성능 비교

그리드환경에서 분산된 자원의 경우 공유되는 자원간의 통신시간에 따라 공유자원의 성능이 변화되고 특히 특정 데이터의 전달속도는 공유자원의 전체성능에 뚜렷한 영향을 미치게 된다.

본 연구의 실험환경은 국가과학기술연구망(KREONET)의 기반을 활용해 대전과 부산에 각각 8개 노드와 스토리지를 구성하고 일반적으로 사용되는 IP라우팅 패스와 람다패스를 같은 대역폭(1Gbps)으로 구성하고 네트워크의 성능과 계산작업 완료에 소요된 시간을 대표적인 성능을 나타내는 지수로 간주하고 비교하였다. 다른 영향을 제거하기 위해 각각 접속스위치(C3750)에는 테스트노드 각 8노드만을 접속하였고, 람다패스 구성에는 MSPP(Multi Service Provisioning Platform)를 통한 신규서킷(Circuit)을 구성 후 Site A,B를 VLAN으로 구성하여 동일한 Subnet으로 동일화 하였다.

<표 1> 테스트베드 시스템 및 설정

	BW(IP)	BW(LP)	Network Interface	System Model	MTU
지역 A	1Gbps	1Gbps	1Gbps GE(NETWORK) 1000T UTP(System)	CPU(1EA), 2.0GHz Linux(2.7)	9K
지역 B	1Gbps	1Gbps	1Gbps GE(NETWORK) 1000T UTP(System)	CPU(1EA), 2.0GHz Linux(2.7)	9K

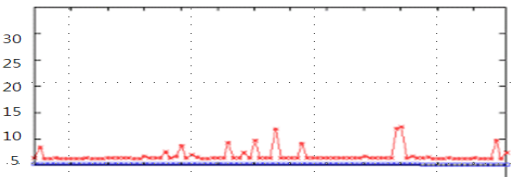


(그림 1) LP vs IP Network 성능비교 테스트베드 구성

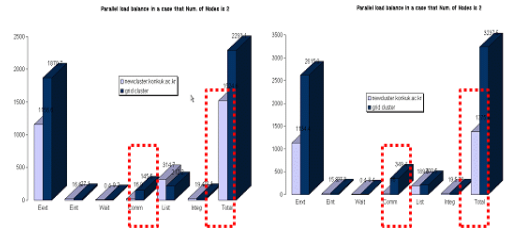
람다패스를 기반으로 연결된 네트워크와 일반 라우팅을 통한 네트워크는 iPerf를 활용한 가능대역폭 테스트에서 각각 890Mbps와 870Mbps로 실험에 소요되는 대역폭을 모두 만족하였다. 그러나 지연시간은 람다패스로 구성한

네트워크의 경우 5ms로 거의 변화가 없는 반면 일반 라우팅패스는 일정시간 11ms까지 변화를 나타냈다(그림2). 지연시간의 변화는 MPICH에서 자원공유에 필요한 통신시간(Communication total time)에서 차이를 발생하였으며, 람다패스의 공유자원간 통신 빈도가 약 20%증가함에 따라 공유자원의 활용도가 우수하게 나타났으며, 전체 계산에 소요되는 시간은 약 30%이상 절감되는 효과를 나타냈다. (그림3)

성능비교결과는 동일한 대역폭의 네트워크 환경 하에서라도 약간의 지연시간의 변화에 민감한 어플리케이션의 경우 그리드 시스템의 연계와 정보교환을 위해 필요한 통신시간에 영향을 주었으며 같은 성능의 시스템이라도 통신시간의 변화에 따라 전체 공유자원의 성능이 차이를 나타내었고, 이는 데이터전송성능뿐만 아니라 자원 간 정보공유 및 제어 데이터 전송이 그리드자원의 효율에 영향을 주며 람다네트워킹 기술은 그리드환경에 적합한 결과를 나타냈다.



(그림 2) LP vs IP Network 성능비교(RTT/ms)



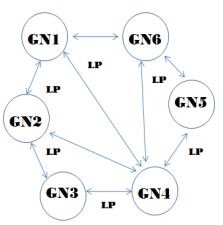
(그림 3) 그리드 시스템 성능비교 그래프

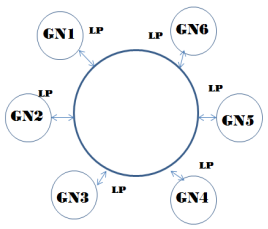
3. 대규모 Multi-Domain 그리드환경 구현을 위한 Multi-point Lambda Network 설계

(그림6)은 GN#(Grid Node number)들이 네트워크를 통해 상호 연결되어있는 일반적인 네트워크의 구조이다. 자원공유와 연계자원 정보제공을 위해 CRM(Computational Resource Manager)와 통신이 필요하며, 다 지점의 CRM간에 동시 정보를 요구하는 환경에서 각 Grid Node의 Input Port에서는 N-1 수만큼의 데이터 전달 정보 스트림을 Outport로 전달한다.(Grid Node 자신의 정보요구 스트림을 제외)

Number of Information Output Stream= Total Grid Node-1=Requirement of LPs ※On Point-to-Point Connections

이 경우(Point-to-Point) 모든 Grid Node들을 선행연구에서 검증된 람다네트워킹을 통해 성능을 보장하고자 패스를 설정할 경우 Core 네트워크자원의 용량제한에 따라 구축이 불가능하거나 혹은 불필요한 자원이 소요된다. 즉 Light-path기반의 람다네트워킹 기술이 그리드자원에 필요한 네트워킹의 요구조건은 충족시키는 반면 대규모의 Multi-Domain 환경을 구축하고자 할 때 많은 네트워킹 자원을 필요할 뿐 아니라 트래픽의 전달이 중복되어 비효율적인 네트워크 구성이 설계된다.





(그림6) Point-to-Point network (그림7) Multi-point network

(그림7)은 Core 네트워크 구간에 OXC(Optical Cross Connection)를 통해 논리적인 Ring을 구성하고 Multipoint-to-Point기반의 구조이다.

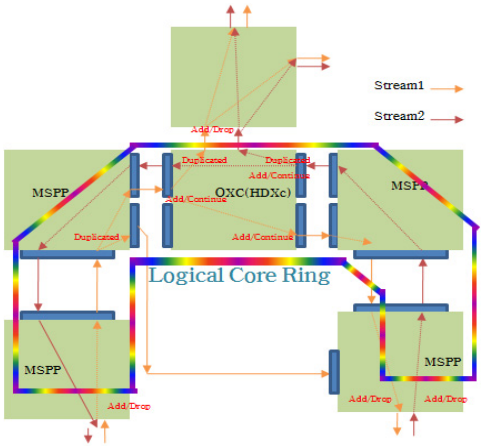
이 경우 마찬가지로 Output Port로 전달되는 자원정보스트림은 동일하게 N-1의 수만큼 전달되지만 LP를 구성하는 MSPP(Multi-Service Provision Platform)의 Stream Duplicate기능(Nortel OME 6500의 경우 TL1,Transaction Language 1, remote interface)을 통해 1:N의 스트림으로 복제되어 Core 네트워크에 전달된다. 즉 Core Ring에 접속하는 접속 LP구성만을 소요하게 되어 네트워크의 자원을 효율적으로 관리하게 되고 동시에 그리드 기반의 공유자원의 성능을 보장이 가능하다.

Number of Information Output Stream= Total Grid Node-1= Number of Link with Core ring ※On Multipoint-to-Point Connections

또한 자원의 정보교환 후 대용량 데이터전달시 각 노드는 VLAN으로 구성되므로 동일한 Subnet하에서 각 노드 경로의 정보를 가지고 있으므로, 자원연계가 필요한 노드가 연결된 패스로만 트래픽이 발생할 뿐 아니라 링 구조의 특성상 우회경로가 확보되어 효율적인 네트워크를 설계 할 수 있다.

(그림8)은 그리드 Multi-Point lambda networking의 핵심이 되는 논리적 Ring을 실제 구성하는 형태이

다. 현재 람다패스를 구성하는 OXC (또는 HDXc)은 기본적으로 'add and Drop'과 'Drop and continue' 기능을 보유하고 있다. 이 기능을 MSPP의 Stream Duplicate 기능과 조합하여 Multipoint lambda network을 구성한다.



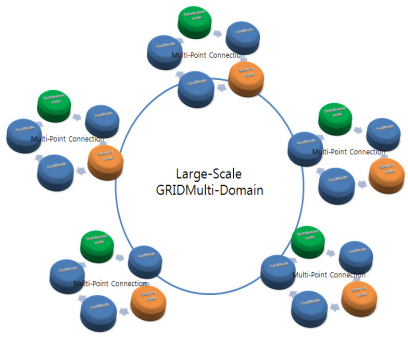
(그림8) Multipoint network on core ring

논리적 Core Ring의 구성 후 이 링에 접속한 각 Grid Node들은 각 다른 Grid Node에서 보내지는 모든 스트림(Stream1,2,...,n-1)을 받게 된다. 각 Grid Node에서 전달된 스트림에서 선택한 자원정보에 대응하여 대용량 트래픽을 전달한다. 대용량 트래픽 전송 혹은 사용자의 자원선택에 따라 각 Grid Node 간 혹은 패스간의 경로가 결정된다. 이때 Core망의 구성이 Ring형태이므로 네트워크의 상황에 따라 경로를 선택할 수 있다.

GMPLS기반 RSVP-TE(Traffic Engineering)을 통해 경로를 재구성하는데 다음과 같은 메카니즘을 전제한다.

- 터널 방식의 GMPLS 경계 노드들은 PSC LSP를 터널로 다루고 TE-link 터널을 통해 MPLS LSP를 설정한다.
- 스위치 방식에서 GMPLS 경계 노드들은 MPLS RSVP control 패킷들을 GMPLS RSVP control packet들로 변환하고 LSP를 연결한다.
- GMPLS 네트워크에서 Control Plane는 일반적으로 Data Plane과 분리되어진다.

이러한 논리적 Ring기반의 Multi-Point Lambda Networking은 그리드의 요구조건(대용량전달, 안정적 환경)을 충족함과 동시에 대규모 자원구축에 용이하여 요구 용량과 성능을 동시에 만족 시킬 수 있는 구조이다. 또한 Point-to-Point 네트워크 구성에 비해 자원의 활용도가 우수하며 복수의 전달경로선택이 가능하여 대용량 트래픽 전달이 용이 하다. 또한 대규모(Large-Scale)의 그리드환경을 구축에 가장 적합한 형태인 'Multiple network distribution node'.(그림8)의 구축이 용이하기 때문에 다른 지역(조직)의 도메인 간 상호운용성(interoperability)이 매우 우수한 구조이다



(그림9) Large-Scale Grid multi-domain

4. 결론 및 향후 연구계획

그리드 연구환경에서는 분산된 자원의 공유성능과 안정성을 위해서는 람다패스 통한 네트워킹 기반이 적합하다. 그러나 point-to-point 형태의 람다네트워킹은 소규모의 자원연동에는 적합하나 대규모 특히 도메인영역에서 벗어나 다수의 도메인이 함께 구성되는 Large-scale 도메인의 구성에는 자원소모가 심하여, Multipoint core ring을 구성하고 자원을 연동하는 방식이 적합하다.

본 논문에서는 대규모 그리드환경 구축에 필요한 Multipoint Core링의 구성방식을 제안하고 또한 도메인을 극복하여 더욱 대용량 · 고성능화 할 수 있는 Multi-domain간의 자원연계 방식을 연구하였다. 앞으로는 람다네트워킹 성능에 영향을 미치는 네트워크 파라메터(대역폭, 경로)뿐 아니라 시스템의 설정값(MTU, TCP, Window size)까지 정보교환을 통해 재조정하여 네트워크 성능을 향상 할 수 있는 방안을 연구하고자 한다.

참고문헌

[1] 황일선, "Grid with Network", 차세대 인터넷 워크샵, 2001

[2] D. Sanghi, A. K. Agrawala, B. Jain, "Experimental assessment of end-to-end

behavior on Internet", Proc. IEEE Infocom '93, San Fransisco, CA, pp. 867-874, March 1993.

[3] D. Simeonidou et al., ""Optical Network Infrastructure for Grid"", Grid Forum Draft , GFD-I.036,Oct 2004

[4] C. Qiao, M. Yoo, ""Optical Burst Switching - A new Paradigm for an Optical Internet"", Journal of High Speed Networks, Spec. Iss. On Optical Networking, vol. 8, no. 1, Jan. 2000, pp. 36-44

[6] D.K. Hunter, M.H.M Nizam, M.C. Chia, K.M. Guild, A. Tzanakaki, M.J. O'Mahony, J.D. Bainbridge, M.S.C. Stephens, R.V. Penty, I.H. White, ""WASPNET: A wavelength switched packet network"", IEEE Communications Magazine, vol.37, pp.120-129, Mar. 1999.

[7] F. Xue, Z. Pan, H. Yang, J. Yang, J. Cao, K. Okamoto, S. Kamei, V. Akella, S.J.B. Yoo,""Design and Experimental Demonstration of a Variable-Length Optical Packet Routing Systemith Unified Contention Resolution"", IEEE Journal of Lightwave Technology, vol.22, no.11,pp.2570-2581, Nov. 2004.

[8] D. Klonidis, R. Nejabati, C.(T.) Politi, M.J. O'Mahony, D. Simeonidou, ""Demonstration of a fully functional and controlled optical packet switch at 40Gb/s"", in proc. 30th European Conf. on Optical Comm., Stockholm, Sweden, PD Th4.4.5, Sep. 2004.