

큐잉 모델을 활용한 리소스 활용 분포의 분석

정동영⁰, 이병준*, 정영교*, 윤희용*

⁰성균관대학교 전자전기컴퓨터공학과

e-mail: {jungdy, byungjun, jung0491, youn7147}@skku.edu^{0*}

Analysis of Resource Utilization Distribution using Queueing Model

Dong-Young Jung⁰, Byung-Jun Lee*, Young-Gyo Jung*, Hee-Yong Youn*

⁰Dept. of Electrical and Computer Engineering, Sungkyunkwan University

● 요약 ●

큐잉 이론은 스트리밍 애플리케이션들의 성능을 이해하는데 많은 이론적 도움을 제공해 줄 수 있다. 하지만 실제 애플리케이션에 큐잉 모델을 적용하기 위해서 많은 분석과 검증 또한 필요하다. 따라서 본 논문에서는 BLAST를 위한 간단한 큐잉 모델과 그 이론적 검증을 통해 수학적 분석 결과를 도출하였다.

키워드: 사용자 분포(distribution of utilization), 큐잉 시스템(queueing system)

I. 서론

큐잉 이론은 간단하게는 은행에서 고객과 상담원의 상관관계를 모델링 하는 것으로부터 웹 서버와 같은 복잡한 소프트웨어 시스템을 구성하기까지 다양한 분야에서 사용되어 오고 있다[1]. 최근에는 떠오르고 있는 고성능의 스트리밍 데이터 애플리케이션을 위해서 다양한 컴퓨팅 시스템 구조에도 적합한 큐잉 모델의 연구가 요구되고 있다. 큐잉 모델은 모델링 된 시스템의 동작에 대해 몇 가지 단순화된 가정을 하게 된다. 만약 모델링이 실행 가능하다고 보인다면 시스템의 성능 개선과 한계에 대한 논증을 시행한다.

TimeTrial 성능 모니터 [2], [3]는 성능 모델에 대한 측정과 검증을 하는데 사용된다. 측정을 위해서 도착 비율은 λ 로 나타낼 수 있고 이를 이후 단계에서 분기 확률을 측정하는데 사용한다. 서비스율은 시스템이 어떻게 동작하는지와 측정 값에 따라 결정될 수 있다. 상기 연구에서는 서버 사용률과 큐 점유율 모델링을 예측하여 수학적 분석 결과를 도출하였다.

II. 본론

BLAST [4], [5]는 생물정보학 분야에서 DNA 혹은 단백질 정보에 유사성 등을 찾는 데 사용되는 소프트웨어 툴이다. BLAST는 하나 혹은 그 이상의 쿼리 시퀀스를 또 다른 시퀀스의 데이터베이스와 비교함으로써 동작한다.

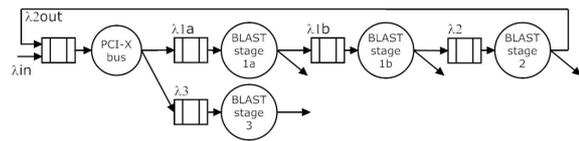


Fig. 1. BLAST Queueing Model

모델링 분석을 위해 컴퓨팅 시스템에서 널리 알려져 있는 M/M/1 시스템을 구성하였다. 이 큐잉 모델은 한 개의 큐와 한 개의 서버로 정의되고 포아송 도착 과정과 지수(Exponential) 처리 시간을 가진다. 최대 큐 용량 K는 서버 메모리에서 가질 수 있는 데이터의 총 합을 말한다.

큐잉 이론에서 각 단계에서 가질 수 있는 요소의 최대 값, 즉 큐 내에서 대기 중인 잡(Job)과 프로세싱 중인 잡을 포함하며 더 이상 대기할 수 없을 경우에 들어오는 요소는 소실된다.

$\lambda_{i,N}$ 은 입력 값에 대한 파라미터이며 1a, 1b, 2 스테이지에서 서버 활용률은 각각 ρ_{1a} , ρ_{1b} , ρ_2 로 표기한다. 이에 대한 제한 환경은 1개의 검색체를 7,964-65,400개의 세그먼트로 나누고 2.7 GBases의 데이터베이스로 구성된 시나리오를 활용한다. 표 1은 이러한 모델에 대한 입력 값이며 표 2는 그 입력에 대해 도출된 결과 값이다.

Table 1. Parameters of Queueing Model

Parameter	Value
λ_{iN}	900 MB/s
μ_{1a}	1 GB/s
μ_{1b}	2.1 Gbases/s
μ_2	130 Mseeds/s
p_{23}	0,0002

Table 2. Analytical Predictions

Parameter	Value
ρ_{1a}	0,84
ρ_{1b}	0,25
ρ_2	0,21
$N_{Q,1b}$	0,08
$N_{Q,2}$	0,06

서버의 사용률과 큐의 점유율의 예측으로 미루어볼 때, M/M/1 큐잉 모델은 적합하지 않다. 큐의 크기가 무한대이기 때문에 사용률에 대한 예측은 근사하다고 할 수 있으나 큐의 점유율과 같은 경우에는 결과 값이 크게 다를 수 있기 때문이다.

III. 결론

본 논문에서는 고성능 스트리밍 데이터 애플리케이션을 위한 간단한 큐잉 모델을 검증 해 보았다. 일반적으로 이러한 구조들은 실제 애플리케이션의 성능 예측에 큰 도움이 된다. 실제 결과와 일치하지 않는 결과가 나온다면 모델은 그 불일치 결과에 대한 이해와 그 개선사항에 대한 도움을 줄 수도 있다. 현재 명확하게 스트리밍 프로세싱 애플리케이션에 적용되는 모델링은 존재하지 않는다. 따라서 그에 근사한 모델링을 통해 큐의 점유율이 실제 큐의 용량보다 클 확률을 추정할 수 있다.

앞으로의 연구 목표는 보다 고수준의 모델링을 제안, 개선하여 직접적으로 애플리케이션에 적용할 방안과 세부적으로 개선할 부분에 대한 연구를 진행할 예정이다. 또한 하나 혹은 그 이상의 파이프라인 스테이지를 이용함으로써 얻을 수 있는 잠재적인 성능에 대한 검증을 통해 성능 향상 방안에 대해 알아 볼 것이다.

Acknowledgment

본 연구는 BK21Plus 사업, 한국연구재단 기초연구사업 (2013R1A1A2060398), 삼성전자, 미래창조과학부 및 정보통신기술 연구진흥센터의 정보통신·방송 연구개발사업 (1391105003), 미래부/정보통신방송연구개발사업의 일환으로 수행하였음.

References

- [1] G. Casale, M. Ningfang, and E. Smirmi, "Versatile models of systems using map queueing networks," in IEEE Int'l Symp. on Parallel and Distributed Processing, Apr. 2008.
- [2] R. D. Chamberlain and J. M. Lancaster, "Better languages for more effective designing," in Proc. of int'l Conf. on Engineering of Reconfigurable Systems and Algorithms, Jul. 2010.
- [3] J. M. Lancaster, J. D. Buhler, and R. D. Chamberlain, "Efficient runtime performance monitoring of FPGA-based applications," in Proc. of 22nd IEEE int'l System-on-Chip Conf., pp. 23-28, Sep. 2009.
- [4] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," Journal of Molecular Biology, vol. 215, pp. 403-10, 1990.
- [5] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, W. Miller, and D. J. Lipman, "Gapped BLAST and PSI-BLAST: A new generation of protein database search programs," Nucl. Acids Res., vol. 25, no. 17, pp. 3389-3402, Sep. 1997.