

통신 네트워크에 대한 시뮬레이션 수행결과의 신뢰도에 관한 연구

이중숙⁰, 박형우, 정해덕^{*}
한국과학기술정보연구원 슈퍼컴퓨팅센터 그리드연구실
{jsruthlee⁰, hwpark}@kisti.re.kr

^{*}Department of Computer Science, University of Canterbury, NZ

A Study on the Credible Simulation Results of Telecommunication Networks

JongSuk Ruth Lee⁰, Hyoungwoo Park, Haeduck Jeong^{*}
Grid Technology Research Department, Supercomputing Center
Korea Institute of Science and Technology Information
^{*}Department of Computer Science, University of Canterbury, NZ

요 약

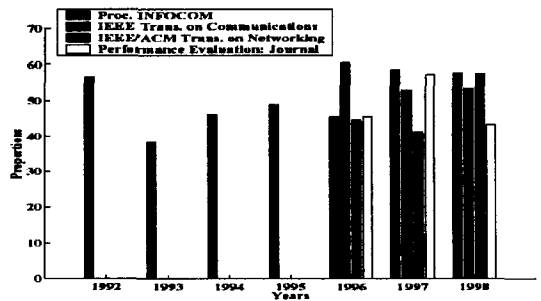
본 논문에서는 신뢰도가 높은 시뮬레이션 최종결과를 얻기 위한 2가지 중요 요건인 (1) 독립적이고 균일하게 분포된(independent and uniformly distributed) 난수를 발생시키는 적절한 난수 발생기(pseudo-random number generator)의 사용과 (2) 적절한 시뮬레이션 데이터의 분석 기법에 대해서 고려하였다. 통신 네트워크 분야의 저명한 논문지들에 발표된 논문들을 조사하였다. 조사한 결과에 의하면 컴퓨터 시뮬레이션 기법이 아주 일반적인 연구 기법으로 사용되어지고 있었으나, 시뮬레이션이 통계적 실험임을 고려하지 않은 결과물도 아주 많았다. 따라서, 본 논문에서는 시뮬레이션 결과의 신뢰도를 높이기 위한 지침을 언급하였다.

1. 서 론

초고속 네트워크(high performance network) 기반의 컴퓨팅 환경이 발전함에 따라 지리적으로 분산되어 있는 슈퍼컴퓨터, 첨단연구장비, 차세대 응용, 첨단 네트워크 그리고 전문 연구 인력을 밀접하게 연계하여 마치 "하나의 로컬 시스템(a single virtual system)"처럼 활용할 수 있게 하는 그리드 컴퓨팅(Grid computing)이 21세기의 시작과 함께 전 세계적으로 활발하게 연구되어지고 있다. 그리드 컴퓨팅을 가능하게 하는 핵심요소 중에 하나가 통신 네트워크의 기술이다. 따라서, 그리드 컴퓨팅 연구는 통신 네트워크의 성능측정(performance evaluation)을 수행하기 위해 설계된 다양하고 사용하기 편리한 소프트웨어 패키지들을 이용한 과학적 계산 분야에서의 연구에 의해서 확실하게 발전 될 수 있다. 기초 과학 및 공학 분야에서 뿐만 아니라, 통신 네트워크 연구분야에서 컴퓨터를 이용하여 연구를 수행할 경우, 과학적인 탐구를 수행하기 위한 새로운 패러다임으로서 실험적인 연구 방법인 이론적 연구(theoretical studies)와 실험(experimentation)을 통한 연구 방법과 함께 컴퓨터 시뮬레이션 기법이 널리 사용 되고 있다. 하지만 이는 컴퓨터 시뮬레이션이 단순히 컴퓨터 프로그래밍의 실습이라는 인식을 심어주게 되었다. 이러한 시뮬레이션 패키지들은 일례로, 통신 네트워크의 시뮬레이션 모델을 설계하는 것을 단순화된 아이콘(네트워크의 가장 기본적인 기능을 나타내는 블록을 가리킴)을 적절한 위치에 놓는 것으로 축소시킨 다음, 시뮬레이션을 시작하는 것으로 시뮬레이션 프로그래밍을 심각할 정도로 단순화 시켰

다.

컴퓨터 시뮬레이션은 이미 과학자나 엔지니어들에게 아주 광범위하게 활용되어지고 있는 연구기법이다. (그림 1)은 통신 네트워크 분야에서 잘 알려진 논문지들(Proceedings of IEEE INFOCOM, IEEE Transactions on Communications, IEEE/ACM Transactions on Networking, Performance Evaluation Journal)에 발표된 2,246개의 논문들을 조사하여 얻은 것이다. 이 그림에서 보여주는 것처럼 조사된 논문들의 51% 이상이 시뮬레이션 기법을 활용하여 그들의 연구를 수행하였다.



[그림 1] 컴퓨터 시뮬레이션 기법을 이용하여 연구를 수행한 논문들에 대한 비율

이러한 성능측정을 위한 방법으로써 시뮬레이션 기법에 대한 인기는 모든 시뮬레이션 툴킷 개발자나 이용자들이에 의해 공유되지 않고 있다. 때문에 통신 네트워크를

포함한 다양한 동적 시스템들의 성능측정의 도구로서 컴퓨터 시뮬레이션 기법이 오용되고 있다 할 수 있다. 또한, 이러한 현상들이 광범위하게 퍼지는 것은 시뮬레이션으로 얻은 수행 결과에 대한 아주 심각한 신뢰도의 문제이다. 이는 다시 말하면, 컴퓨터 시뮬레이션 기법으로 성능측정을 수행한 동적 시스템의 대부분 결과들을 믿기가 힘들다는 것이다. 이러한 문제는 1994년과 1996년에 개최된 국제 Winter Simulation Conference와 1996년에 개최된 IEEE INFOCOM 그리고 시뮬레이션 사설[1]에서 논의되었으나 별다른 변화가 없는 실정이다. 본 논문에서는 통신 네트워크의 성능측정을 위한 확률적 이산 시뮬레이션(stochastic discrete-event simulation) 기법에 대해서 연구한다. 특히, 신뢰도가 높은 시뮬레이션을 수행하기 위한 적절한 난수 발생기(pseudo-random number generator)의 사용과 시뮬레이션 데이터에 대한 분석 방법에 대해 논의한다. 그리고 시뮬레이션 기법을 이용하여 얻은 결과물들의 신뢰도 문제와 이를 해결하기 위한 몇몇 지침을 정리하고자 한다.

2. 컴퓨터 시뮬레이션 결과의 신뢰성(Credibility)

컴퓨터 시뮬레이션 전문가들은 시뮬레이션의 모델링 단계의 비용은 시뮬레이션 프로젝트를 성공적으로 이끌기 위한 전체 비용의 30-40% 정도라고 평가한다. 컴퓨터 시뮬레이션 기법에 기인한 성능평가 연구의 첫 번째 단계는 타당한 시뮬레이션 모델을 사용하는 것이다. 통신 네트워크의 경우, 네트워크의 타당한 개념적인 모델이란 네트워크의 내부 메카니즘(internal mechanism), 한계들(limitations), 프로세스들의 확률적 특성(stochastic characteristics of processes)등을 적절히 고려한 것을 의미한다. 타당한 시뮬레이션 모델을 설계하는 방법에 대한 지침은 [2]에서 찾을 수 있다. 그러나, 이는 시뮬레이션 기법을 이용하여 신뢰도가 높은 최종결과를 얻기 위한 첫 단계라는 점이다. 다음 단계로는 타당하게 잘 만들어진 시뮬레이션 모델이 타당한 시뮬레이션 실험에 사용되어 지도록 하는 것이다. 이 단계에서 어떠한 시뮬레이션 실험의 타당성을 확보하기 위해서는 적절한 난수 발생기(pseudo-random number generator)의 사용과 시뮬레이션 데이터에 대한 적절한 분석 방법을 잘 적용해야 한다.

2.1 적절한 난수(pseudo-random number)의 사용

컴퓨터 시뮬레이션에서 가장 일반적으로 이용되는 난수는 연산적인 난수 생성기(algorithmic pseudo-random number generator)에서 생성된 것이다. 연산적인 난수 생성기는 일정 주기(cycle)를 가지고 난수를 생성시키고, 그 주기가 끝나면 처음부터 다시 같은 난수를 생성하게 된다. 난수 생성에 관한 이론적인 배경은 [3]에 잘 정리되어 있다. 그리고 지난 50여년간 많은 난수 생성기가 제안되었다. 그 중에서 가장 널리 쓰이는 난수 생성기는 multiplicative linear congruential 난수 생성기이다. 이 난수 생성기에 대한 자세한 내용은 [2]에서

찾을 수 있으며, 이것의 주기는 $2^{31} - 1$ 로 GPSS, SIMSCRIPTS II.5, SIMAN, SLAM II[2] 등에서 사용되어 지고 있다.

지금까지 이 난수 생성기를 이용한 시뮬레이션 수행은 별다른 문제가 없었다. 하지만, 이것이 모든 경우에 적용되는 것은 아니다. 난수 생성기에 대한 매우 사려깊은 사용자라면 현실 세계의 응용연구에서 난수를 사용할 경우 잠정적으로 매우 심각한 문제가 발생 할 수 있다는 것을 인식하여야 한다. 1990년 수 백 MHz의 워크스테이션에서 $2^{31} - 1$ 개의 난수를 발생시키는데 12분 정도가 소요되었다. 하지만, 현재 GHz급의 PC에서는 수십 초 이내에 난수가 바닥나게 된다. 따라서 그리드 환경에서 각종 응용 프로그램을 수행할 경우 지금까지 아무문제 없이 사용되었던 난수 발생기는 아주 심각한 문제를 야기시킬 수 있다. 그러나, 다행스럽게도 $2^{185} - 2^{377}$ 그리고 $2^{19937} - 1$ 의 아주 긴 주기를 갖는 여러 난수 발생기들이 제안되었다[4, 5]. 하지만 이것이 난수 발생기와 관련된 모든 문제를 해결했다고는 할 수 없다. 예를 들자면, 분산 병렬 시뮬레이션을 그리드 환경에서 수행 할 경우 한 개의 난수 발생기에서 생성된 연속적인 난수들을 여러 개의 스트림으로 나누어서 이용할 때 생기는 상관관계(correlation) 등의 문제를 조심스럽게 고려해야 한다. 이럴 경우, 시뮬레이션으로 얻은 최종결과가 잘못 해석될 수 있기 때문이다. 어떠한 난수 발생기를 선택해야 하는지에 대한 지침은 [6]에서 찾을 수 있다.

2.2 시뮬레이션 데이터의 분석

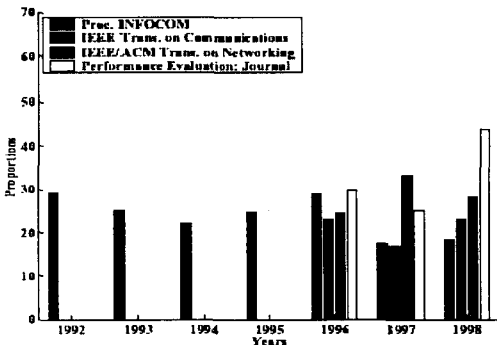
랜덤 프로세스들(random processes)을 가지고 시뮬레이션을 수행한다는 것은 통계적인 실험(statistical experiments)을 수행한다는 것과 같은 의미이다. 따라서, 시뮬레이션 실험으로 얻은 시뮬레이션 데이터는 통계학적 기법들을 이용하여 분석하는 것이 당연한 것이다. 그렇지 않을 경우 네덜란드 Tilburg 대학의 J. Kleijnen 교수가 [7]에서 경고한 것처럼 막대한 양의 데이터(mass of data)가 순식간에 쓰레기(mess)로 바뀔 수 있다. 어떠한 통계적 실험의 최종 결과에 대한 통계적 오차(statistical error) 또는 주어진 최종 추정치(estimate)의 정확도(accuracy)는 일반적으로 주어진 신뢰수준(confidence level)에서의 신뢰구간(confidence interval)으로 측정된다. 정확하게 잘 구현된 시뮬레이션의 신뢰구간 폭(width)은 시뮬레이션이 진행되면서 즉 수집된 데이터의 수가 증가하면서 점차적으로 감소된다.

하지만, 어느 정도의 최종 신뢰도를 원하느냐에 따라서 시뮬레이션의 전체 수행 길이가 정해 질 수 있다. 이러한 시나리오에 가장 적합한 시뮬레이션 기법은 순차적 시뮬레이션(sequential simulation)이다. 순차적 시뮬레이션이란 연속된 검사점(checkpoint)에서 시뮬레이션 실행자가 원하는 정도의 신뢰도에 도달했는지를 체크해서 만족하면 시뮬레이션을 중단하고, 아니면 다음 검사점에 도달할 때까지 시뮬레이션을 계속하는 것을 의미한다. 이러한 절차로 시뮬레이션을 수행한 후 수집된 데이터를 분석하는 방법도 매우 다양하게 제안되어서 사용

되어 지고 있다[2, 8]. 예를 들면, batch means 이용한 분석 방법, spectral 기법을 이용한 분석 방법, regenerative cycle을 이용한 분석 방법 등이다[9]. 이때 어떠한 방법으로 시뮬레이션 데이터를 분석할 것인가를 결정하는 것도 시뮬레이션 전문가가 아니라면 매우 어려운 일이다. 또한 상업화된 많은 시뮬레이션 툴킷들도 데이터 분석을 위한 방법들을 다양하게 제공하지 않고 있어서, 이들 시뮬레이션 툴킷을 이용하여 얻은 시뮬레이션 결과도 신뢰도에 있어서 다소 의심스럽다.

3. 시뮬레이션 기법을 이용하여 얻은 결과물들의 신뢰도 문제와 이를 해결하기 위한 제안

오늘날 시뮬레이션 결과의 오차를 측정하고 최소화할 수 있는 통계적 분석 기법을 훈련받은 연구자는 쉽게 찾을 수 있다. 하지만, (그림 2)에서 보여주는 것처럼 많은 연구자들이 시뮬레이션 기법으로 얻어진 결과가 고유의 랜덤(random)한 성질을 가지고 있다는 것을 고려하지 않고 있음을 알 수 있다. 시뮬레이션 기법을 이용한 연구자 중에 약 23.5% 만이 시뮬레이션이 확률적 실험임을 인식하고 있다. 이는 통신 네트워크에 대한 시뮬레이션 연구가 신뢰도에 매우 심각한 문제가 있음을 나타내는 것이다. 이러한 문제를 극복하기 위해서 필요한 간단한 지침을 정리하면 다음과 같다. 첫 번째로 시뮬레이션은 과학적 실험이므로 반복해서 실험 가능해야 하므로 시뮬레이션에서 이용된 난수 발생기(pseudo-random number generator)와 어떠한 종류의 시뮬레이션 시나리오를 적용했는지를 꼭 밝혀야 한다. 또한, 두 번째로 시뮬레이션 데이터들의 분석에 적용된 분석 방법 및 최종 결과에 허용된 통계적 오차(statistical error)를 독자에게 알려야 한다.



[그림 2] 시뮬레이션 기법을 이용한 논문들 중 통계적 방법으로 분석된 논문들의 비율

4. 결론

본 논문에서는 신뢰도가 높은 시뮬레이션 최종결과를 얻기 위한 2가지 중요 요건인 독립적이고 균일하게 분포된(independent and uniformly distributed) 난수를 받

생시키는 적절한 난수 발생기(pseudo-random number generator)의 사용과 적절한 시뮬레이션 데이터의 분석 기법에 대해서 살펴보았다. 통신 네트워크 분야의 저명한 논문지들에 발표된 논문들을 조사하였다. 조사한 결과에 의하면 컴퓨터 시뮬레이션 기법이 아주 일반적인 연구 기법으로 사용되어지고 있음을 알 수 있었다. 그러나, 많은 연구자들이 시뮬레이션에 대한 이해가 제대로 되지 않았음도 알 수 있었다. 하지만, 이는 앞에서 언급한 지침들을 따른다면 쉽게 개선될 것이다. 물론, 통신 네트워크의 시뮬레이션은 종종 엄청난 양의 CPU 시간을 요구한다. 이러한 지나치게 긴 시뮬레이션 수행 시간은 시뮬레이션 모델의 개발 및 검증하는 과정을 어렵게 만드는 장애 요인이 된다. 따라서, 지금까지 통신 네트워크에 대한 시뮬레이션 수행을 speedup하고 시뮬레이션을 아주 손쉽게 할 수 있는 완전 자동화된 시뮬레이션 툴킷 연구가 많은 연구자들에 의해 진행되고 있다. 하지만, 전세계적으로 활발하게 연구되어지고 있는 그리드 컴퓨팅 환경에서 수행될 수 있는 시뮬레이션에 대한 연구는 아직 미미한 실정이다.

5. 참고문헌

- [1] B. Gaither, "Empty Empiricism," ACM Performance Evaluation Review, Vol. 18(2), pp. 2-3, 1990.
- [2] A. M. Law and W. D. Kelton, "Simulation Modeling and Analysis," 2nd ed., New York: McGraw-Hill, 1991.
- [3] D. E. Knuth, "Art of Programming," Vol 2: Seminumerical Algorithms," 3rd ed., Reading, MA: Addison-Wesley, 1998.
- [4] P. L'Ecuyer, "Good Parameters and Implementations for Combined Multiple Recursive Random Number Generators," Operations Research, Vol. 47, pp. 159-164, 1999.
- [5] M. Matsumoto and T. Nishimura, "Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generators," ACM Transactions on Modeling and Computer Simulation, Vol. 8(1), pp. 3-30, 1998.
- [6] R. Jain, "The Art of Computer Systems Performance Analysis," New York: Wiley, 1991.
- [7] J. Kleijnen, "The Role of Statistical Methodology in Simulation," Methodology in Systems Modeling and Simulation, B. P. Zeigler et al., eds. North-Holland, Amsterdam, 1979.
- [8] K. Pawlikowski, "Steady-State Simulation of Queueing Processes: A Survey of Problems and Solutions," ACM Computing Surveys, Vol. 2, pp. 123-170, 1990.
- [9] J. R. Lee, "On Automated Sequential Steady-State Simulation," PhD Thesis, University of Canterbury, New Zealand, 2000.