

결함 발생을 고려한 대규모 주문형 멀티미디어 서비스 시스템의 큐잉 네트워크 모델 설계

박 기 진[○] 김 성 수

아주대학교 정보통신대학 정보및컴퓨터공학부

A Design of Queuing Network Model for Large-scale Multimedia-On-Demand Service Systems with Unreliable Components

Kiejin Park and Sungsoo Kim

Division of Information and Computer Engineering, Ajou University

요 약

멀티미디어 서버를 구성하고 있는 자원(프로세서, 디스크, 네트워크 등)의 결함 발생을 고려한 큐잉 네트워크 모델을 개발하였다. 이를 적용하여 고객 수용 능력 및 멀티미디어 서비스 시스템의 성능과 관련된 파라미터를 분석하였으며, 시뮬레이션 패키지를 활용한 모의 실험을 통하여 개발된 큐잉 네트워크 모델의 정확성을 검증하였다. 고객 도착률과 자원 고장률 변동에 대한 서버 자원의 이용률과 큐의 길이를 계산함으로써 효율적인 시스템 구성 자원의 부분별 용량 산정이 가능하게 되었으며, 또한 데이터 패킷 응답 시간 분포 및 고장률 변화에 따른 시스템의 신뢰도 수준을 계산하였고, 이러한 성능 관련 파라미터의 확보를 통해서 멀티미디어 서비스 질(Quality of Service)에 중요한 영향을 미치는 서버 설계 요소를 파악할 수 있었다.

1. 서론

수만 명이 연결되어 있는 멀티미디어 서비스 시스템에서 대용량 서버의 고장으로 인한 파급 효과는 분산 시스템 구조와는 비교가 되지 않을 정도로 막대한 경제적 손실을 초래할 수도 있다. 하드웨어 및 병렬처리 기술의 발전으로 인해서 대규모 고성능 멀티미디어 서버가 등장했으나 서버의 신뢰도(Reliability), 가용도(Availability), 안전도(Safety), 보수도(Maintainability) 등의 의존도(Dependability)에 대한 고려는 이제 시작 단계에 불과하다. 현재까지 멀티미디어 서버에 관한 연구는 주로 클라이언트의 요청에 대한 멀티미디어 데이터의 실시간 처리가 주를 이루었으며, 향후에는 멀티미디어 서비스 시스템 구성 요소의 결함 발생을 고려하는 연구의 필요성 [1]이 검증할 것으로 판단된다.

본 연구에서는 멀티미디어 서비스 시스템의 운영을 피드백을 갖는 열린 큐잉 네트워크로 모델링한 연구 결과[2]에 기초하여, 서버 자원의 결함 발생이 서비스 시스템에 미치는 영향을 분석할 수 있는 큐잉 모델을 개발하였으며, 시뮬레이션 패키지를 활용한 모의 실험을 통하여 개발된 큐잉 네트워크 모델의 정확성을 검증하였다.

서버 자원의 평균 고장 시간 간격(Mean Time To Failure)을 고려한 자원 이용률과 큐 길이에 관한 계산을 통해서, "서버 구성 부품들의 신뢰도를 높이는 것이 중요하지? 혹은 결함이 발생했을 때, 이를 신속히 해결하는 고장 수리 능력을 보유하는 것이 중요하지?"에 대한 평가 척도를 제시하였다. 또한 서비스 시스템의 의존도 계산을 위한 신뢰도 및 결함 발생을 고

려한 데이터 패킷 지연 확률을 구하였다.

본 논문의 2장에서는 결함 발생을 고려한 멀티미디어 서비스 시스템 구조를 설명하였으며, 3장에서는 제안된 멀티미디어 시스템 구조의 큐잉 네트워크 모델을 제시한 후, 개발된 모델에서 얻은 해석적 결과와 모의실험에서 얻은 결과를 상호 비교 분석하였다. 마지막으로 결론에서는 큐잉 네트워크 모델의 활용 방안에 관하여 논하였다.

2. 결함 발생을 고려한 멀티미디어 서비스 시스템 구조

본 연구에서 가정하고 있는 결함 발생을 고려한 멀티미디어 서비스 시스템은 그림 1과 같이 다수의 클라이언트가 고속 네트워크를 통해 대용량 서버에 접속하며, 다수의 프로세서와 디

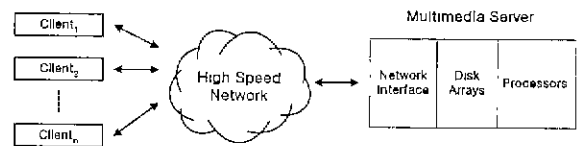


그림 1 대용량 서버로 구성된 멀티미디어 서비스 시스템 구조

스크 및 네트워크 인터페이스로 이루어진 서버는 이들의 요청에 실시간으로 답하는 구조를 가지고 있다. 이와 같은 시스템 구조를 채택한 이유는 향후 컴퓨터 기술의 발전으로 인한 상용 대용량 서버의 등장이 확실하며, 시스템의 부하에 따라 서버의 용량을 변경하는 확장성이 제공되리라 보기 때문이다[3].

한편 연결이 이루어지면 실시간 데이터 전송이 계속해서 보

본 연구는 1998년도 정보통신연구관리단 대학기초연구지원사업에 의한 결과임.

장 되어야 하는 멀티미디어 서비스 특성을 고려하여, 클라이언트의 파일 요청에 대한 데이터 패킷의 전달 과정을 그림 2와 같은 피드백을 갖는 운영 구조로 모델링 하였다.

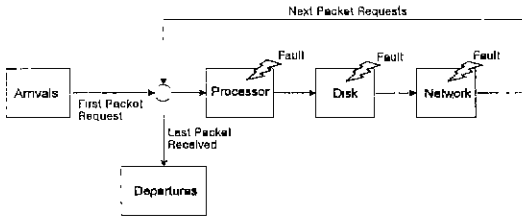


그림 2 결합 발생은 고려한 멀티미디어 데이터 전송 서비스

3. 큐잉 네트워크 모델

초고속 통신망에 연결된 멀티미디어 서버 디스크에 저장되어 있는 파일의 전송 요청(Arrival)과 이 요청에 대한 데이터 패킷 전송 서비스가 포이송 프로세스를 따른다고 가정할 경우, 서비스 시스템 안으로의 도착과 서버의 디스크 자원에 대한 접근 및 네트워크로의 출력은 그림 3과 같이 복수개의 M/M/1 큐로 구성된 피드백을 갖는 열린 큐잉 네트워크(Open Queueing Network)로 표현될 수 있다[4].

클라이언트는 λ_c 의 도착률로 멀티미디어 서비스를 요청한 후, 원하는 데이터 패킷을 모두 전송 받을 때까지 서버내의 자원(프로세서, 디스크, 네트워크)을 할당받게 되며, P_0 의 확률로 서비스를 종료하게 된다. 프로세서, 디스크, 네트워크 대역폭에 따라 각 자원의 서비스율이 결정되며, 클라이언트가 요청한 데이터를 보관하고 있는 디스크로의 분기 확률(P_1, P_2, \dots, P_m)은 특정 멀티미디어 데이터의 저장 중복도에 따라 결정된다.

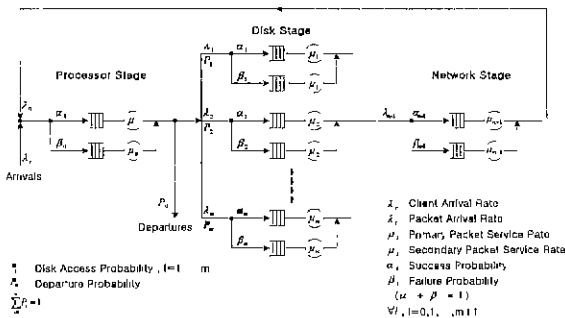


그림 3 멀티미디어 서비스 시스템의 큐잉 네트워크 모델

멀티미디어 서버 자원에 결합이 발생할 경우, 서비스 Stage의 서비스율이 저하되는 결과를 초래하게 된다. 이와 같은 결합 발생 상태를 고려하기 위해서, 데이터 패킷 요청에 대한 서비스 시간의 변화를 그림 3의 각 서비스 Stage와 같은 병렬-선택 구조로 표현하였다. 예를 들면, $\alpha_i=0.9999999$ 일 경우 1,000,000개중 1개의 패킷은 디스크 서비스 Stage의 일시적 결합으로 인해 서비스율이 μ_i' 로 변화되는 것을 의미한다. 즉 정상 동작 상태일 경우, 각 서비스 Stage에 입력되는 데이터 패킷 요청은 α_i 의 확률로 μ_i 의 서비스율을 선택하게 되지만, β_i 의 확률로 결합이 발생하였을 경우에는 μ_i' 로 데이터 패킷의 서비스율이 변경된다.

3.1 큐잉 파라미터의 계산

(1) 도착률

클라이언트의 데이터 패킷 요청을 서비스할 수 있는 각 큐로의 평균 도착 횟수(V)와 도착률(λ)은 아래와 같이 계산된다.

$$V_i = \begin{cases} \alpha_0 * \frac{1}{P_0} & , i=0 \\ \alpha_i * \frac{P_i}{P_i} & , i=1, 2, \dots, m \\ \alpha_i * \sum_{j=1}^m \frac{P_j}{P_j} & , i=m+1 \end{cases} \quad V_i' = \begin{cases} \beta_0 * \frac{1}{P_0} & , i=0 \\ \beta_i * \frac{P_i}{P_i} & , i=1, 2, \dots, m \\ \beta_i * \sum_{j=1}^m \frac{P_j}{P_j} & , i=m+1 \end{cases}$$

$$\lambda_i = \lambda_c V_i \quad , i=0, 1, \dots, m+1 \quad \lambda_i' = \lambda_c V_i' \quad , i=0, 1, \dots, m+1$$

(2) 서비스율

프로세서, 디스크 및 네트워크의 서비스율(μ)은 각 자원의 대역폭을 전송되는 데이터 패킷 크기로 나눈 값으로부터 구하며, 결합이 발생한 경우의 서비스율(μ')은 정상 동작 경우의 서비스율에서 일정 비율(Service_Delay)만큼 감소하게 된다.

$$\mu = \frac{Bandwidth}{Packet_Size} \quad \mu_i' = Service_Delay * \mu_i \quad , i=0, 1, \dots, m+1$$

(3) 자원 이용률 및 큐 길이

구해진 도착률과 서비스율을 적용하여 평형상태(Steady State)에 있는 M/M/1 큐의 서버내 각 자원 이용률(ρ)을 계산할 수 있다

$$\rho_i = \frac{\lambda_i V_i}{\mu_i} = \frac{\lambda_i}{\mu_i} \quad \rho_i' = \frac{\lambda_i V_i'}{\mu_i'} = \frac{\lambda_i}{\mu_i'} \quad , i=0, 1, \dots, m+1$$

서버내 각 자원에 대한 평균 큐 길이 E(Q_i) 및 전체 시스템의 큐 길이 E(Q)는 아래와 같이 구해질 수 있다.

$$E(Q_i) = \frac{\rho_i^2}{1-\rho_i} \quad E(Q_i') = \frac{\rho_i'^2}{1-\rho_i'} \quad , i=0, 1, \dots, m+1 \quad E(Q) = \sum_{i=0}^{m+1} \left(\frac{\rho_i^2}{1-\rho_i} + \frac{\rho_i'^2}{1-\rho_i'} \right)$$

(4) 패킷 지연 확률(Packet Delay Probability)

실시간 연속성이 보장되어야 하는 데이터 패킷 전송에 지연이 발생할 가능성을 계산하기 위해서 패킷 응답 시간 확률밀도 함수(Response Time PDF)를 구하였다. 데이터 패킷 한 개가 멀티미디어 서비스 시스템 내에 머무르는 시간은 프로세서, 디스크 및 네트워크 각 Stage에서의 응답 시간을 합하여 구하며, M/M/1 큐로 구성된 각 서버 자원의 응답시간 확률 밀도 함수는 다음과 같다.

- Processor $r_p(t) = (\mu_0 - \lambda_0) e^{-(\mu_0 - \lambda_0)t}$
 $r_p'(t) = (\mu_0' - \lambda_0') e^{-(\mu_0' - \lambda_0')t}$
- Disk $r_{D_i}(t) = (\mu_i - \lambda_i) e^{-(\mu_i - \lambda_i)t} \quad , i=1, 2, \dots, m$
 $r_{D_i}'(t) = (\mu_i' - \lambda_i') e^{-(\mu_i' - \lambda_i')t} \quad , i=1, 2, \dots, m$
- Network $r_v(t) = (\mu_{m+1} - \lambda_{m+1}) e^{-(\mu_{m+1} - \lambda_{m+1})t}$
 $r_v'(t) = (\mu_{m+1}' - \lambda_{m+1}') e^{-(\mu_{m+1}' - \lambda_{m+1}')t}$

전체 시스템 응답 시간의 확률밀도함수 r(t)는 다음과 같다.

$$r(t) = r_p(t) + r_p'(t) + \sum_{i=1}^m (r_{D_i}(t) + r_{D_i}'(t)) + r_v(t) + r_v'(t) \quad (1)$$

식(1)의 확률 밀도 함수 합의 분포는 Convolution과 Laplace Transform 연산을 통해서 구할 수 있으며, 데이터 패킷의 지연 확률은 다음과 같이 계산된다.

$$P(Delay) = \int_0^{\infty} r(t) dt \quad , t \geq 0 \quad r = \text{패킷재생시간}$$

(5) 신뢰도 모델링

그림 3의 각 서비스 Stage 정상동작 확률(α)에 기초한 멀티미디어 서비스 시스템의 신뢰도 계산 과정은 다음과 같다. 먼저 각 서비스 Stage에서의 고장률(Failure Rate: λ_i)은 다음과 같이 계산될 수 있다.

$$Transfer_Time = \frac{Packet_Size}{MPEG_Stream_Rate} * (1-\alpha)^{-1}, \quad \lambda_i = \frac{1}{Transfer_Time}$$

또한 그림 3의 각각의 서비스 Stage는 서로 독립적으로 작동하므로 전체 서비스 시스템의 신뢰도는 다음과 같이 계산된다.

$$R_{system}(t) = e^{-\sum_{i=1}^n \lambda_i t}, \quad t \geq 0, \quad \lambda_i, Failure_Rate$$

3.2 성능 분석

멀티미디어 서비스 시스템의 성능을 평가하기 위해서 표 1의 파라미터에 기초하여, 큐잉 및 시뮬레이션 분석을 수행하였다. 고장률이 멀티미디어 서비스 시스템의 자원 이용률 및 패킷 지연 확률에 미치는 영향을 그림 4, 그림 5에 나타내었으며, 그림 6에는 큐 길이에 대한 두 모델간의 결과를 서로 비교하였다. 멀티미디어 서비스 시스템의 시뮬레이션 수행 결과(점선)와 큐잉 네트워크 모델의 해석적 방법으로 계산한 값(실선)이 거의 일치함을 알 수 있다. 그림 7에서는 결합 확률에 대한 상영시간 길이별 신뢰도 변화를 표시했다

표 1 성능 분석에 사용된 파라미터

항목	값	비고
Arrival_Rate(λ_i)	0.1, 0.2, 0.25, 0.30	초당 도착하는 클라이언트 수
Movie_Play_Time	7300 sec	영화 상영 길이 (2시간)
Processor_Bandwidth	15 Gbps	프로세서 처리 속도
MPEG_Stream_Rate	4.5 Mbps	초당 요청 데이터량
Packet_Size	250 KB	전송되는 데이터 패킷의 크기
Disk_Group	4	디스크 그룹의 수
Disk_Bandwidth	1 Gbps	디스크 개당 10Mbps의 대역폭
Network_Bandwidth	9 Gbps	네트워크 처리 속도
Failure_Probability	$10^{-4}, 10^{-5}$	패킷 당 고장 확률
Service_Delay	0.9	서비스권의 감소 비율

4. 결론

결합 발생을 고려하는 멀티미디어 서비스 시스템의 성능 분석을 위하여, 큐잉 이론을 적용한 시스템 분석 모델을 개발하였다. 고객 도착률과 자원 고장률 변동에 대한 프로세서, 디스크, 네트워크 등과 같은 시스템 자원의 이용률과 큐의 길이 및 데이터 패킷 응답 시간 분포 및 고장률 변화에 따른 시스템의 신뢰도를 계산함으로써 비용 효율적인 시스템 구성 자원의 부분별 용량 산정에 필요한 수치를 구하였다. 또한 시뮬레이션을 활용한 모의실험을 통해, 제안된 성능 분석 결과의 정확성을 검증함으로써, 효율적인 용량 계획에 관한 의사 결정에 적용 가능하게 되었다.

참고 문헌

[1] J. Laprie, "Dependable Computing. Concepts, Limits, Challenges," in Proc. IEEE FTCS-25(Special Issue), LA, CA, pp. 42-57, June 1995.
 [2] 박기진, 정지영, 김용규, 박주용, 김성수, "대규모 주문형

멀티미디어 서비스 시스템의 큐잉네트워크 모델 설계," 1998년 한국정보과학회 춘계학술발표논문집, vol. 25, no. 1, pp. 237-239, 1998.

[3] R. Tewari, D. Diasand and H. Vin, "High Availability in Clustered Multimedia Servers," ICDE, pp. 645-654, 1996.
 [4] L. Kleinrock, *Queueing Systems Volume1: Theory*. John Wiley & Sons Inc., 1975.

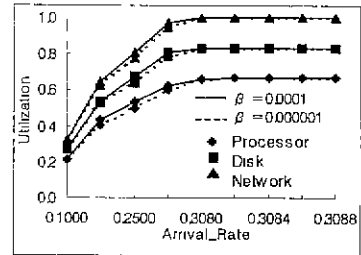


그림 4 고장률에 대한 이용률의 변화

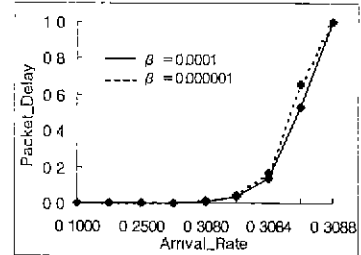


그림 5 고장률에 대한 패킷지연확률의 변화

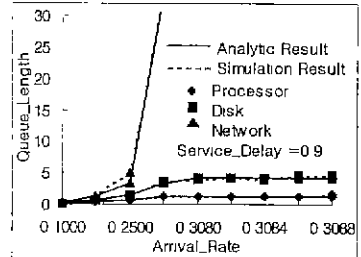


그림 6 서비스 지연률에 대한 큐길이의 변화

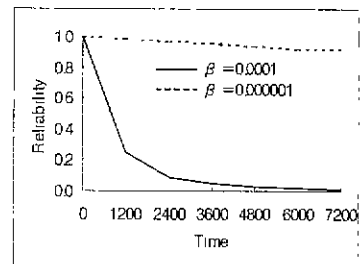


그림 7 시간에 대한 신뢰도 변화