

캠퍼스내 레거시 웹서버 통합 운영을 위한 클라우드 컴퓨팅의 최적용량 및 전환이익 분석

이구연*, 최창열*, 최황규*, 장민**, 윤재구**

요약

홈페이지 서비스를 주로 하는 웹 서버들의 평균 이용률이 대체로 낮고 업무 부하의 변화도 크지 않은 편이다. 또한 대부분의 사용자가 구성원 등으로 국한되고 새로운 서버를 구매할 때 최적용량 보다는 최대용량을 선택하는 경향이 있어 과도한 서버를 독점 운용하게 된다. 웹 서버의 운용에는 초기 비용뿐 아니라 평상시 운용비용도 포함된다. 평균 이용률이 낮은 웹 서버들을 클라우드화하여 통합 운영하면 사용자에게 동일한 품질을 제공하면서도 총 비용을 줄일 수 있을 것이다. 본 논문에서는 캠퍼스내의 여러 웹 서버들을 클라우드화하여 통합 운영할 때 기존 웹 서버와 동일한 서비스품질을 제공하기 위한 최적용량을 산출하고 그때 절약되는 비용을 분석한다. 전통적인 웹 서버들과 통합 클라우드를 위한 서비스 모델을 정립하고 사용자 요구 행태에 따른 최적용량을 대기이론으로 분석하고 비용 절감 정도를 제시한다. 제시된 서비스 모델과 분석 결과는 중대규모의 대학 캠퍼스는 물론 비슷한 규모의 기업, 기관들에 쉽게 응용할 수 있다.

키워드 : 클라우드, 캠퍼스 웹서버, 통합 웹서버, 성능 분석, 전환 이익

Conversion Profit and Optimal Capacity of Cloud Computer for Integrating Legacy Campus Web Servers

Goo Yeon Lee*, Chang Yeol Choi*, Hwang Kyu Choi*, Min Jang**, Jae Ku Yoon**

Abstract

Cloud computing helps users to save a significant amount of cost that is related to infrastructure investment, management, and maintenance. In this paper, we study the conversion planning from campus legacy web servers into an integrated cloud computing web server system. We also analyze the conversion profit when campus web servers are integrated into a cloud computer. We first investigate the cost of legacy system model of campus web servers operated by individual laboratories, departments, institutes and so on. Next, we set up a cloud computer model for the integrated web services meeting the same performance requirements. Then, we derive the conversion profit. From the result of the derivation, we see that the conversion can be effectively applied to and adopted by mid or large sized campuses and similar institutions that provide web services

Keywords : cloud, campus web server, integrated web server, performance analysis, conversion profit

※ 교신저자(Corresponding Author): Goo Yeon Lee
접수일:2013년 11월 11일, 수정일:2013년 12월 15일
완료일:2014년 04월 30일

* 강원대학교 컴퓨터정보통신공학전공 교수
Tel: +82-33-250-6394 , Fax: +82-33-252-6390
email: leegyeon@kangwon.ac.kr

** (주) 더존비즈온

▣ 본 연구는 (주) 더존비즈온의 지원을 받아 수행되었음.

1. 서론

최근 IT 산업에서 큰 관심을 끌고 있는 분야 중의 하나가 클라우드 컴퓨팅 서비스이다. 클라우드 컴퓨팅은 서버, 네트워크, 스토리지, 소프트웨어 그리고 인터넷으로 제공되는 다양한 어플리케이션 등의 정보 자원을 필요할 때 임대하여 활용한다. 조직의 정보시스템을 직접 구축하여

운용해 오던 전통적인 접근에서 간편한 임대로 변경하는 것으로써 사용자뿐 아니라 IT 생태계 전반에 큰 영향을 미치고 있다. 기존 정보시스템을 클라우드로 전환하면 비용을 절감할 수 있을 뿐 아니라 전문기업이 이미 확보하고 있는 서비스나 서비스컴포넌트를 이용하여 조직에 필요한 정보서비스시스템을 신속히 만들 수 있어 조직의 민첩성을 증대시킬 수 있는 장점이 기대된다.

그러나 기존 응용(legacy applications)을 클라우드 환경으로 전환하거나 신규로 진입할 의사결정은 그렇게 간단하지가 않다. 대부분의 클라우드 서비스 제공자는 비슷한 서비스를 서로 다른 프라이싱 모델로 제공하고 있으나, 대부분의 경우 성능 자료는 간과하고 있으며, 제시하지 못하는 상황이며, 이에 따라 사용자는 기존 응용을 클라우드 환경에서 운용할 때의 비용과 클라우드로 전환할 대상에 대해서도 확신을 하지 못한다. 현실적으로 모든 데이터센터를 클라우드로 옮겨가는 것은 거의 불가능할 뿐 아니라 클라우드 공급자도 이것을 권장하지 않는다. 궁극적으로 클라우드 컴퓨팅 환경으로 전환되더라도, 기관이 보유한 애플리케이션 중에서 클라우드로 전환함으로써 비용을 절감할 수 있는 애플리케이션을 선별하여 점진적으로 클라우드화 하는 것이 효과적이다. 잠재적인 사용량이 불확실하거나, 매월 짧은 기간 동안 집중적으로 자원이 사용된 후 나머지 시간에는 유휴 상태로 유지되거나, 지속적으로 일정한 수준의 자원이 사용되지만 특정 기간에만 작업량이 매우 많은 애플리케이션 등이 그 대상이 될 수 있다.

전통적인 웹 서비스 환경에서 서버의 평균 이용율은 4~18%로 낮은 편이다[1]. 하나의 웹 서버를 운영할 때의 비용은 서버 도입 시의 하드웨어, 소프트웨어 비용뿐 아니라 평상시의 운용 비용도 포함된다. 이 비용은 대체로 관리기관이 부담하지만 낭비 요소가 있어 보인다. 특히 대부분의 사용자가 업무 담당자로 한정되어 있고 업무부하의 분포가 특정 시점에만 집중되는 환경에서는 업무 부하량에 비해 과도한 서버를 독점적으로 사용하게 된다. 또한 업무가 추가되어 새로운 서버를 구매해야 할 때 최적용량을 산정하기보다는 최대용량의 서버를 선택하는 경향이 있다. 따라서 평균 이용률이 낮은 웹 서버들을 클라우드링하여 통합 운영하면 사용자에게 동일한

품질을 제공하면서도 총 비용을 줄일 수 있다.

본 논문에서는 대학이 보유한 수많은 웹 서버들을 클라우드링하여 통합 운영할 때 기존 웹서버와 동일한 서비스 품질을 제공하기 위한 최적용량을 산출하고 그때 절약되는 비용을 분석한다. 먼저 전통적인 웹 서버들과 이들을 통합한 클라우드 컴퓨팅을 위한 서비스 모델을 정립하고 사용자 요구 행태에 따른 최적용량을 대기이론으로 분석하고 비용 절감 정도를 제시한다. 대학 캠퍼스의 수많은 서버들은 대체로 각 부서나 학과의 홈페이지 서비스에 활용되고 있다. 이 서버들을 클라우드링하여 통합 운영하면 사용자들에게 동일한 서비스 품질을 제공하면서도 시스템의 유지, 운영에 필요한 총 비용을 줄일 수 있게 된다. 제시된 서비스 모델과 분석 결과는 중대규모의 대학교는 물론 웹서비스를 많이 하는 비슷한 규모의 기업, 기관들에 쉽게 응용할 수 있다. 본 논문의 구성은 다음과 같다. 2장에서 관련연구를 살펴보고, 3장에서는 웹서버와 통합 클라우드 시스템의 모델을 정립한다. 대기이론으로 사용자 요구 행태에 따른 최적용량을 분석, 검토하고 비용 절감 정도를 4장에서 제시한다. 그리고 5장에서 결론을 기술한다.

2. 관련 연구

클라우드 컴퓨팅의 성능과 비용은 모델링, 시뮬레이션, 테스트베드를 이용한 측정을 통해 분석되고 있다[2]. 대규모 멀티티어 시스템을 효율적으로 설계하는데 필요한 가이드라인을 제시하는 MDCSim[3]은 시뮬레이션 인프라로서 클러스터 기반 데이터 센터의 성능과 소비전력을 다각도로 분석할 수 있게 한다. 전통적인 데이터센터 환경을 IaaS로 전환할 때의 성능과 확장성에 대한 예측과 평가도 중요하다[4][5]. [4]은 엔터프라이즈 수준의 가상화 환경에서 N 티어의 응용 벤치마크 프로그램인 RUBBoS를 이용하여 자원이용률, 사용자요구특성, 성능 등을 분석하고 문맥교환 오버헤드, 네트워크 드라이버 처리 부담 등을 시스템 차원의 병목요소로 언급하였다. CloudGuide[6]은 전통적인 어플리케이션을 서로 다른 클라우드 공급자에서 실행할 때 각각의 비용을 추정하고 타겟 어플리케이션이 요구

하는 컴퓨팅 자원을 대기이론[7]으로 예측하는 도구이다. 즉 사용자가 요구하는 특정한 성능 수준을 만족하면서 기존응용을 다양한 클라우드 공급자에서 실행할 때의 비용을 추정하고, 새로운 작업부하에 적합한 클라우드 구성방안도 찾을 수 있다. 그러나 요구되는 자원의 타입과 수를 입력하여 추정비용을 산출하고, 응용이 요구하는 성능조건을 만족시키는데 필요한 클라우드 자원량을 사용자들이 추정해야한다. 그리고 성능과 비용 벤치마크를 이용하여 많은 정보를 얻을 수는 있지만 벤치마크 결과를 응용별 요구사항과 연관지우는 작업도 필요하다. [8]에서는 레거시 응용을 클라우드 환경으로 전환해야 할 것인가에 대한 답을 구하고자 하였으며 British Telecom에서 운영하는 세 가지 응용들을 대상으로 전환 방법과 실체를 다루고 있다. Koch 등은 클라우드 자원을 교육기관에 제공할 때 더욱 정제된 수요예측 방법을 제안하고 그 효과를 평가하였으며[9], 응용프로그램의 특성을 바탕으로 서비스품질을 손상하지 않고도 비용에 최적화하는 동적 자원할당 기법들을 수치해석으로 분석하였다.

3. 웹 서비스 시스템의 모델링

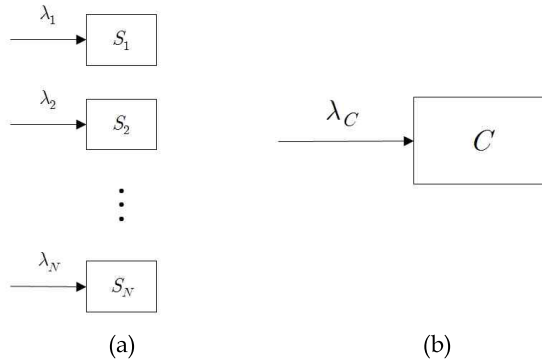
웹 서비스는 인터페이스, 응용서버, 데이터서버의 3단계를 거치는데, 사용자가 요청하는 트랜잭션은 인터페이스 단계에서 해석되어 해당하는 응용서버로 전달, 처리된다. 전형적인 인터넷 응용인 N 티어 작업부하를 클라우드로 전환할 때의 성능이슈와 비용을 분석하기 위한 모델을 설정한다.

3.1 레거시 웹서버 시스템 모델

대학의 각 기관들이 개별적으로 보유하고 웹 서비스를 제공하는 레거시 웹 서버 시스템을 (그림 1)의 (a)와 같이 나타낼 수 있다. (그림 1)의 (a)에서의 서버 $S_i (i=1 \sim N)$ 는 캠퍼스내에서의 개별 서버들을 나타낸다. 각 웹 서버의 사용자 트래픽과 성능에 관련된 파라미터는 다음과 같다. i 번째 웹서버에 도착하는 사용자 웹 트랜잭션은 도착률 λ_i 인 Poisson 분포로 가정한다. 포아슨 분포는 네트워크의 트래픽을 분석할 때

가장 널리 사용되는 분포로서 많은 사용자들이 서로 간에 독립적으로 트래픽을 생성할 때 적합한 모델이다.

(그림 1) (a) 레거시 웹 서버 모델 (b) 클라우드 컴퓨터 (C)를 이용한 통합 웹 서비스 모델



(Figure 1) (a) Legacy web server system (b) Clouding web server system

또한 웹서버의 트랜잭션 처리시간은 랜덤변수 a_i 로 가정한다. 웹서버의 트랜잭션 처리률은 μ_i 로 정의하며 이는 $1/\bar{a}_i$ 로 나타낼 수 있다 (즉 $\mu_i = 1/\bar{a}_i$). 여기서 \bar{a}_i 는 a_i 의 평균값을 표시한다. 주어진 사용자 웹 트랜잭션 트래픽에 대하여 μ_i 는 웹서버 컴퓨터의 처리용량으로 해석할 수 있다. 또한 랜덤변수 a_i 에 대한 분산값으로 σ_i^2 를 사용한다. 웹서버의 성능척도로는 서비스에 대한 성능으로서 가장 널리 사용되는 파라미터인 평균 응답시간을 사용한다.

많은 웹서버들이 비슷한 홈페이지를 운영하고 사용자들은 여러 웹서버를 돌아다니면서 웹 서핑을 하므로 본 연구에서는 캠퍼스내의 모든 웹서버가 동일한 평균 응답시간을 제공한다고 가정한다. 현실적으로는 웹서버들이 서로 다른 응답시간을 보이겠지만, 클라우드로의 통합 웹서버와의 성능을 비교하기 위해서는 기존의 웹서버들이 동일한 평균 응답시간을 제공하도록 설계되었다고 단순화 할 필요가 있기 때문이다. 레거시 시스템의 평균 응답시간을 T_L 이라고 표시하자. 유사한 홈페이지를 운영하는 웹서버들의 처리용량은 서로 다르겠지만, 각 웹 트랜잭션이 요

구하는 컴퓨터시스템의 처리용량의 분포 패턴은 스케일링만 차이가 있지 동일한 패턴을 가진다고 볼 수 있다. 그러므로 웹 서버 S_i 는 독자적인 처리용량 $\mu_i(1/\bar{a}_i)$ 및 σ_i^2 를 갖지만, 본 가정하에서는 모든 웹 서버는 동일한 coefficient of variation $C_L = \frac{\sigma_i}{a_i}$ ($i=1 \sim N$)를 갖게 된다. 또한 웹서버 시스템 S_i 가 갖는 TCO는 다음과 같다.

$$TCO_i = CCF_i + OE_i, \quad (i=1 \sim N) \quad (1)$$

여기서 TCO_i , CCF_i , OE_i 는 각각 i 번째 웹서버에서의 Total Cost of Ownership, Cost of Computer Facility 및 Operation Expenses를 나타낸다.

3.2 클라우드 컴퓨팅을 이용한 통합 웹 서비스

(그림 1)의 (b)는 학내 레거시 웹서버 시스템을 클라우드링으로 통합한 시스템 C 를 나타낸다. 레거시 시스템에서 웹서버 별로 분산되던 사용자 트래픽은 통합되어 도착률 λ_C 인 포아송 분포로 나타낼 수 있다. 이는 여러 개의 포아송 분포의 합은 역시 포아송으로 나타낼 수 있다는 포아송 분포의 메모리리스 특성에 기인한다. 캠퍼스내의 전체 트랜잭션 수는 변함이 없으므로 다음과 같이 나타낼 수 있다.

$$\lambda_C = \sum_{i=1}^N \lambda_i \quad (2)$$

또한 클라우드 컴퓨터의 웹 트랜잭션 처리시간은 랜덤변수 a_C 로 가정하고, 클라우드 컴퓨터의 웹 트랜잭션 처리율을 μ_C 로 정의하면 이는 $1/\bar{a}_C$ 로 나타낼 수 있다 (즉 $\mu_C = 1/\bar{a}_C$). 여기서 \bar{a}_C 는 a_C 의 평균값을 표시한다. 주어진 전체 사용자의 총 웹 트랜잭션 트래픽에 대하여 μ_C 는 클라우드 컴퓨터의 처리용량으로 해석할 수 있다. 또한 랜덤변수 a_C 에 대한 분산값으로 σ_C^2 를 사용한다. 그러면 coefficient of variation는

$C_C = \frac{\sigma_C}{a_C}$ 와 같이 나타낼 수 있다. 클라우드 컴퓨터의 웹서비스에 대한 성능척도로서 역시 평균 응답시간을 가정하며 T_C 라고 표시하기로 하자.

클라우드링 컴퓨터 C 는 다음과 같은 TCO를 갖는다.

$$TCO_C = CCF_C + OE_C \quad (3)$$

여기서 TCO_C , CCF_C , OE_C 는 클라우드 컴퓨터에서의 Total Cost of Ownership, Cost of Computer Facility 및 Operation Expenses를 나타낸다.

3.3 웹 서버의 운영비용 (CCF)

CCF 에는 여러 요소들, 즉 기본적으로 컴퓨터 facility에 대한 구입비용을 내구연한으로 나눈 값뿐만 아니라, 사용 중의 업그레이드 비용, 고장 대체 비용 등이 포함된다. 컴퓨터 facility 구입 비용에는 CPU 성능, 메모리, 하드디스크 그리고 성능향상을 위해 부가하는 보조 모듈에 대한 비용 등이 포함된다. 하지만 본 논문에서는 이 CCF 비용이 최종적으로 나타나는 웹 트랜잭션 처리 용량에 대한 함수로서 산정될 수 있다고 가정한다. 즉 웹 트랜잭션 처리 용량에는 여러 하드웨어, 소프트웨어 등 컴퓨터 성능에 관련된 요소들이 복합적으로 작용한 결과가 반영되기 때문이다. 그러므로 웹 트랜잭션 처리용량을 μ 로 나타낼 때, CCF 는 다음과 같은 관계로 표현할 수 있다.

$$CCF = g(\mu) \quad (4)$$

여기서 $g(\cdot)$ 는 평가 당시의 컴퓨터 facility의 시세를 반영하는 함수로, 기술의 발전이나, 시장의 수요 공급에 따라 조사, 규정되는 관계 함수이다. 임의의 단조증가함수도 $g(\cdot)$ 가 될 수 있으나, 본 논문에서는 4장의 성능 수치 추출을 위해 상식적인 두 개의 함수를 활용한다. 물론 실제 시세를 반영하는 함수가 조사되면 그 함수를 본 논문의 분석과정에서의 수식 전개의 변화없이 그대로 적용시킬 수 있음에 유의하기 바란다.

첫 번째 적용할 함수는 다음의 1차 비례 함수이다.

$$g(\mu) = A \cdot \mu \quad (5)$$

즉 CCF의 비용이 컴퓨터의 처리용량에 따라 비례상수 A와 같이 선형 증가하는 경우로서, 처리용량이 3배가 되면 CCF도 3배가 되는 단순 관계를 나타낸다. 그러나 현실적으로 처리용량이 3배가 되더라도 비용은 3배까지는 되지 않는다. 이런 경우는 처리용량이 증가하더라도 비용은 위로 볼록한 함수의 형태를 띠게 되며, 이러한 특성을 나타내는 함수의 예로서 다음 함수를 두번째로 적용한다.

$$g(\mu) = B \cdot \log(1 + \mu) \quad (6)$$

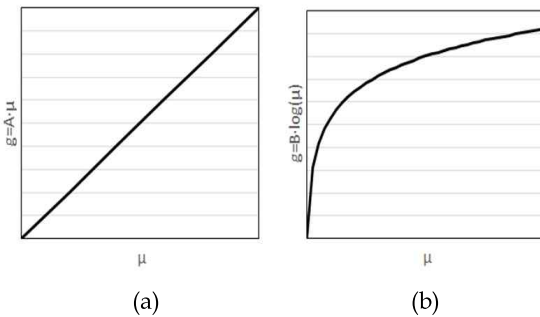
단 여기서 B는 비례상수이다.

지금까지 논의한 두 가지 시세 함수는 (그림 2)와 같다.

(그림 2) 컴퓨터 facility의 시세 함수 $g(\cdot)$ 의

두가지 적용 예 :

(a) $g(\mu) = A \cdot \mu$ (b) $g(\mu) = B \cdot \log(1 + \mu)$



(Figure 2) Two examples of $g(\cdot)$:

(a) $g(\mu) = A \cdot \mu$ (b) $g(\mu) = B \cdot \log(1 + \mu)$

4. 성능 분석

4.1 레거시 시스템의 용량 분석

성능최도인 평균 응답시간 T_L 및 coefficient of variation C_L 이 주어졌을 때 이를 제공하기

위한 웹서버 시스템 S_i 의 용량은 M/G/1 대기이론 모델을 적용하여 다음과 같이 구할 수 있다 [7].

$$T_L = \bar{a}_i + \frac{\rho_i \bar{a}_i (1 + C_L^2)}{2(1 - \rho_i)}, \quad (7)$$

단 여기서 $\rho_i = \lambda_i \bar{a}_i$ 의 관계를 갖는다.

위의 식을 풀어 정리하면 다음과 같다.

$$2T_L(1 - \lambda_i \bar{a}_i) = 2\bar{a}_i(1 - \lambda_i \bar{a}_i) + \lambda_i (\bar{a}_i)^2 (1 + C_L^2) \quad (8)$$

정리하면,

$$\lambda_i (C_L^2 - 1) (\bar{a}_i)^2 + 2(1 + \lambda_i T_L) \bar{a}_i - 2T_L = 0. \quad (9)$$

윗 식에서 \bar{a}_i 를 구하면 다음과 같다.

$C_L = 1$ 일 때 (coefficient of variation이 1인 경우로 지수분포의 처리시간이 한 예가 됨)는 1차식이 되므로 간단히 구할 수 있다.

$$\bar{a}_i = \frac{T_L}{1 + \lambda_i T_L} \quad (10)$$

$C_L \neq 1$ 일 때는 다음과 같이 구하여 진다.

$$\bar{a}_i = \frac{-(1 + \lambda_i T_L) \pm \sqrt{\lambda_i^2 T_L^2 + 2C_L^2 \lambda_i T_L + 1}}{\lambda_i (C_L^2 - 1)} \quad (11)$$

$C_L > 1$ 일 때 (coefficient of variation이 1보다 큰 경우로 hyperexponential 분포의 처리시간이 한 예가 될 수 있음), \bar{a}_i 는 양수이어야 하므로 위의 두 근중에서 다음의 하나의 근만 가지게 된다.

$$\bar{a}_i = \frac{-(1 + \lambda_i T_L) + \sqrt{\lambda_i^2 T_L^2 + 2C_L^2 \lambda_i T_L + 1}}{\lambda_i (C_L^2 - 1)} \quad (12)$$

또한 $C_L < 1$ 일 때 (coefficient of variation이 1보다 작은 경우로 Erlangian 분포 또는 상수 분포 처리시간이 예가 될 수 있음)는,

$0 < 1 - C_L^2 < 1$, $1 + \lambda_i T_L > 1$,
 $1 < \sqrt{\lambda_i^2 T_L^2 + 2C_L^2 \lambda_i T_L + 1} < 1 + \lambda_i T_L$ 인 상태에
 서, $\rho_i = \lambda_i \bar{a}_i$ 는 1보다 작아야 하므로 (1보다 크
 면 시스템 발산함), 물리적으로 의미 있는 답은
 다음과 같다.

$$\bar{a}_i = \frac{-(1 + \lambda_i T_L) + \sqrt{\lambda_i^2 T_L^2 + 2C_L^2 \lambda_i T_L + 1}}{\lambda_i (C_L^2 - 1)} \quad (13)$$

그러므로 위의 식들로부터 \bar{a}_i 는 다음과 같이
 정리될 수 있다.

$$\bar{a}_i = \begin{cases} \frac{T_L}{1 + \lambda_i T_L} & \text{for } C_L = 1 \\ \frac{-(1 + \lambda_i T_L) + \sqrt{\lambda_i^2 T_L^2 + 2C_L^2 \lambda_i T_L + 1}}{\lambda_i (C_L^2 - 1)} & \text{for } C_L \neq 1 \end{cases} \quad (14)$$

즉 웹서버 시스템 S_i 의 처리용량 μ_i 는 다음과
 같다.

$$\mu_i = \frac{1}{\bar{a}_i} = \begin{cases} \frac{1 + \lambda_i T_L}{T_L} & \text{for } C_L = 1 \\ \frac{\lambda_i (C_L^2 - 1)}{-(1 + \lambda_i T_L) + \sqrt{\lambda_i^2 T_L^2 + 2C_L^2 \lambda_i T_L + 1}} & \text{for } C_L \neq 1 \end{cases} \quad (15)$$

이에 따른 웹서버 시스템 S_i 의 CCF는 다음과
 같다.

$$CCF_i = g(\mu_i) \quad (16)$$

4.2 통합 클라우드 시스템의 용량 분석

본 절에서는 클라우드링으로 통합하여 웹 서비
 스를 수행할 때의 용량을 분석한다. 성능적도인
 평균 응답시간 T_C 및 웹 트랜잭션의 처리용량
 요구특성 C_C 가 주어졌을 때 또한 입력되는 웹
 트랜잭션 입력이 포아슨 분포로서 도착률이
 $\lambda_C = \sum_{i=1}^N \lambda_i$ 일 때 M/G/1 대기이론 모델을 적용
 시킨 4.1절의 유도과정을 그대로 따라가면, 다음
 식을 얻을 수 있다.

$$\bar{a}_C = \begin{cases} \frac{T_C}{1 + \lambda_C T_C} & \text{for } C_C = 1 \\ \frac{-(1 + \lambda_C T_C) + \sqrt{\lambda_C^2 T_C^2 + 2C_C^2 \lambda_C T_C + 1}}{\lambda_C (C_C^2 - 1)} & \text{for } C_C \neq 1 \end{cases} \quad (17)$$

그리고 웹서버를 통합한 클라우드 시스템 C
 의 처리용량 μ_C 는 다음과 같이 구해진다.

$$\mu_C = \frac{1}{\bar{a}_C} = \begin{cases} \frac{1 + \lambda_C T_C}{T_C} & \text{for } C_C = 1 \\ \frac{\lambda_C (C_C^2 - 1)}{-(1 + \lambda_C T_C) + \sqrt{\lambda_C^2 T_C^2 + 2C_C^2 \lambda_C T_C + 1}} & \text{for } C_C \neq 1 \end{cases} \quad (18)$$

이에 따른 웹서버 통합 클라우드 시스템 C 의
 CCF는 다음과 같다.

$$CCF_C = g(\mu_C) \quad (19)$$

4.3 TCO 전환 이익

학내 레거시 웹 서버 시스템 모델의 총 TCO
 는 다음과 같다.

$$\begin{aligned} TCO_L &= \sum_{i=1}^N TCO_i = \sum_{i=1}^N [CCF_i + OE_i] \\ &= \sum_{i=1}^N [g(\mu_i) + OE_i] \end{aligned} \quad (20)$$

반면 클라우드링 컴퓨터를 이용한 통합 웹 서비
 스 모델의 TCO는 다음과 같다.

$$TCO_C = CCF_C + OE_C = g(\mu_C) + OE_C \quad (21)$$

그러므로 $T_L = T_C$ 로 하여 동일한 웹 서비스
 품질을 제공하면서 학내 레거시 웹서버 시스템
 에서 통합 웹서비스 클라우드 모델로 전환할 경
 우의 전환이익을 구할 수 있다.

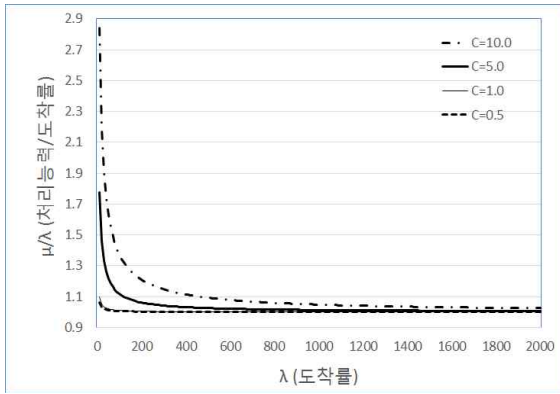
$$\text{전환 이익} = TCO_L - TCO_C \quad (22)$$

4.4 분석 및 토의

4.4.1 웹 트랜잭션 도착률과 컴퓨터 처리용
 량과의 관계

사용자가 요청하는 웹 트랜잭션들의 도착률과 컴퓨터 처리용량과의 관계를 살펴 볼 필요가 있다(그림 3). 이는 레거시 시스템과 통합 클라우드 컴퓨터에 모두 똑같이 적용된다.

(그림 3) 웹 트랜잭션 도착률과 컴퓨터 처리용량과의 관계



(Figure 3) Relationship between computer processing capacity and web transaction arrival rate

(그림 3)에서 x축은 웹 트랜잭션의 초당 도착률 λ 를 나타낸다. y축은 주어진 도착률에서 평균응답시간 T 를 1초로 가정할 때 이를 충족시키는데 필요한 컴퓨터 처리용량을 도착률에 대한 비율로 나타내었다. 또한 coefficient of variation이 각각 0.5, 1.0, 5.0, 10.0 일 때의 관계를 표시하였다. 그래프로부터 도착률이 낮을수록, 또한 coefficient of variation이 클수록 주어진 웹 트랜잭션 트래픽에 대비하여 준비해야할 컴퓨터 처리용량이 커짐을 알 수 있다. 예를 들어, $C=10.0$ 일 때 도착률이 초당 10건이면 원칙적으로 처리용량이 10정도인 컴퓨터가 필요하지만, $T=1$ 초의 평균응답시간을 맞추려면 (그림 3)의 경우 이보다 약 3배가 많은 28.4의 처리용량을 가진 컴퓨터가 필요하다. 다만 도착률이 커지면서 이 비율이 낮아져 거의 1의 값으로 수렴한다. 이는 컴퓨터 이용률이 1에 수렴하게 되면 대기시간이 급격히 커진다는 대기이론의 결과에 어긋나는 것처럼 보이나, 본 경우에는 도착률이 커짐에 따라 처리용량이 같이 증가하여, 하나의 웹 트랜잭션의 처리시간도 처리용량에 반비례하여 급격히 작아지므로, $T=1$ 초라는 시간은 하나

의 웹 트랜잭션의 평균 처리시간을 기준으로 보면 아주 큰 숫자가 되어 컴퓨팅 이용률이 1에 수렴하게 되면 대기 시간이 급격히 커진다는 대기이론의 결과와 부합하게 된다.

또한 (그림 3)의 그래프에서 coefficient of variation가 작게 되면 도착률에 대한 처리능력의 비율이 초기부터 1의 값으로 수렴함을 볼 수 있다. 이는 작은 값의 coefficient of variation는 웹 트랜잭션 처리시간의 변이가 작다는 것을 의미하며, 처리시간의 불규칙성이 작아지면서 대기시간의 변동폭이 줄어들어 전체적으로 평균 대기시간이 짧아지게 되기 때문이다. 하지만 일반적으로 웹 트랜잭션 처리시간의 변동폭은 크며, 이는 아주 간단한 웹 화면 처리에서부터 멀티미디어 데이터 등 대량 데이터의 처리 및 하드디스크 접속시간 차이, 데이터베이스 접속에 따른 검색시간 차이, 생성되는 쓰레드 개수에 따른 처리시간의 변동 등에 기인한다. 그러므로 웹 트랜잭션 처리시간의 coefficient of variation는 큰 값을 가지게 된다.

나아가 (그림 3)의 그래프로부터 여러 웹 서버들을 웹 서버 통합 클라우드링으로 전환할 때, 준비해야하는 컴퓨터 처리용량 면에서도 유리함을 알 수 있다. $C=10.0$ 일 때 웹 트랜잭션의 도착률이 10인 웹 서버의 처리용량은 28.4이 되는데, 이러한 서버 컴퓨터가 10대인 경우 준비해야하는 총 처리 용량은 $28.4 \times 10 = 284$ 가 된다. 그러나 통합 클라우드링으로 전환하였을 경우에 도착률은 $10 \times 10 = 100$ 이 되므로 도착률이 100인 클라우드 컴퓨터의 처리용량을 구하면 되는데, 이 경우 처리용량은 137이 된다. 즉 처리용량 면에서 약 절반 정도만 준비하면 됨을 알 수 있다.

4.4.2 레거시 웹 서버 시스템과 웹 서비스 통합 클라우드링 컴퓨터의 비용

본 절에서는 레거시 웹 서버 시스템과 웹 서비스 통합 클라우드 컴퓨터의 비용을 구체적으로 비교한다. 레거시 웹 서버의 개수 N 은 실제로 학내에서 구동되는 웹 서버들의 개수이며, 본 절에서는 100으로 가정한다. 또한 모든 서버가 입력되는 웹 트랜잭션의 트래픽도 동일하고, 트랜잭션의 처리시간 분포도 동일하다고 가정하여 결과값 추출의 편의를 꾀한다. 그리고 coefficient of variation은 $C_L=10$ 이고, 모든 웹서버 컴퓨터

의 처리용량은 평균 응답시간 T_L 로서 결정되었다고 가정한다. 그러면 다음과 같은 관계식을 얻게 된다.

$$\begin{aligned} \lambda_L &= \lambda_1 = \lambda_2 = \dots = \lambda_N, \quad \bar{a}_L = \bar{a}_1 = \bar{a}_2 = \dots = \bar{a}_N, \\ \mu_L &= \mu_1 = \mu_2 = \dots = \mu_N \end{aligned} \quad (23)$$

4.1절의 결과로부터 하나의 웹 서버가 필요로 하는 처리용량은 다음과 같다.

$$\mu_L = \frac{\lambda_L(C_L^2 - 1)}{-(1 + \lambda_L T_L) + \sqrt{\lambda_L^2 T_L^2 + 2C_L^2 \lambda_L T_L + 1}} \quad (24)$$

또한 모든 웹서버 컴퓨터의 OE 도 동일하다고 가정한다. 즉

$$OE_L = OE_1 = OE_2 = \dots = OE_N \quad (25)$$

그러면 레거시 웹 서버 시스템의 총 TCO는 다음과 같이 구할 수 있다.

$$\begin{aligned} TCO_L &= \sum_{i=1}^N TCO_i = \sum_{i=1}^N [CCF_i + OE_i] \\ &= N \cdot [g(\mu_L) + OE_L] \end{aligned} \quad (26)$$

레거시 웹서버 시스템을 통합 클라우드 시스템으로 전환할 때 $\lambda_C = \sum_{i=1}^N \lambda_i = N \cdot \lambda_L$ 및 $C_C = 10$ 이 성립하며, 클라우드 컴퓨터의 처리용량은 다음과 같다.

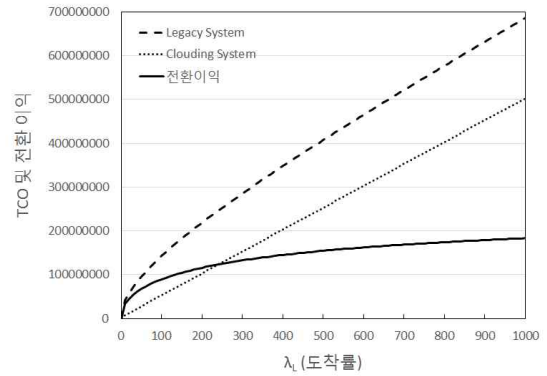
$$\mu_C = \frac{\lambda_C(C_C^2 - 1)}{-(1 + \lambda_C T_C) + \sqrt{\lambda_C^2 T_C^2 + 2C_C^2 \lambda_C T_C + 1}} \quad (27)$$

또한 TCO는 다음과 같이 구해진다.

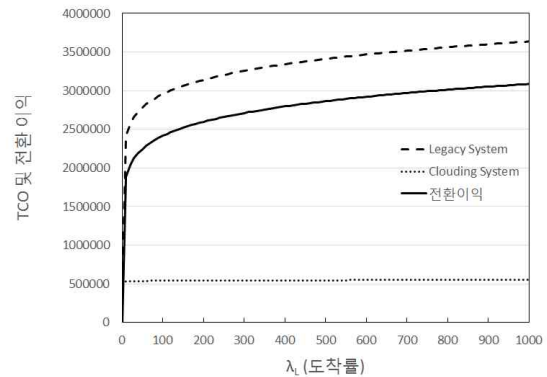
$$TCO_C = CCF_C + OE_C = g(\mu_C) + OE_C \quad (28)$$

(그림 4)의 그래프에 TCO_L 과 TCO_C 를 비교하고 그 때의 전환이익을 보였다. 이때 동일한 서비스 환경, 즉 같은 평균 응답시간($T_L = T_C$)에서 비교를 하였다.

(그림 4) $N=100, A=5000$ 원/월, $B=10000$ 원/월, $T_L = T_C=0.1$ 초, $C_L = C_C=10, OE_L=5000$ 원/월, $OE_C=500,000$ 원/월 일때의 레거시 시스템 및 클라우드 시스템에서의 TCO 및 클라우드 컴퓨터 시스템으로의 전환이익



(a) $g(\mu) = A \cdot \mu$



(b) $g(\mu) = B \cdot \log(1 + \mu)$

(Figure 4) TCOs and conversion profit when $N=100, A=5000$ KRW/month, $B=10000$ KRW/month, $T_L = T_C=0.1$ sec, $C_L = C_C=10, OE_L=5000$ KRW/month, $OE_C=500,000$ KRW/month

레거시 웹 서비스를 통합 클라우드 시스템으로 전환할 때 얻는 이익은 크게 세가지 요소로 분석된다. 첫째는 전체 레거시 웹 서버의 운영 경비(OE)보다 통합 클라우드 시스템의 OE가 더 작은 경우로서 웹 서버의 개수 N 에 따라 다르게 나타날 수 있다. 즉 N 이 클 때는 전환이익이 충분히 나타날 수 있으나, N 이 작으면 오히려 전

환 손실이 발생할 수도 있다. 일반적으로 개별 레거시 웹 서버의 OE 보다 클라우드 컴퓨팅의 OE가 많이 크기 때문에 어느 정도 이상의 N 개 웹 서버가 클라우드링으로 통합되면 의미가 있게 된다. 두 번째는 클라우드링으로 통합했을 때 처리용량의 다중화에 따른 전환 이익이다. 이는 레거시 시스템의 개별 웹 서버에서는 웹 트랜잭션을 서비스 할 때와 서비스하지 않을 때의 유휴 상태로 있을 경우로 구성되나, 통합 시에는 다중화 효과로 인하여 유휴 상태가 줄어들게 됨으로서 처리용량의 효율이 높아진다. 이 전환이익은 요구되는 평균응답시간이 작을수록 더 커지게 되는 데, 요구되는 평균 응답시간이 작을수록 처리시간의 불확실성에 대해 준비해야할 처리용량이 더 커지므로, 이를 클라우드 컴퓨터로 통합하면 다중화 효과가 극대화 된다. 하지만 요구되는 평균 응답시간이 크면, 처리시간의 불확실성이 작아져서 다중화 효과는 줄어든다.

세 번째는 처리용량과 CCF와의 관계에서 나오는 처리용량 규모에 따른 전환 이익이다. 특별한 경우가 아니면 컴퓨터의 처리용량이 n 배가 되더라도 CCF가 n 배로 되지는 않는다. 본 논문에서는 시세함수로서 $g(\mu) = A \cdot \mu$ 및 $g(\mu) = B \cdot \log(1 + \mu)$ 를 가정하였는데, 전자는 처리용량 규모에 의한 전환이익이 없게 되며, 후자는 처리용량 규모에 따른 전환 이익이 존재한다.

(그림 4)의 그래프는 $T_L = T_C = 0.1$ 초인 경우로서 웹 서비스의 평균응답시간이 아주 짧은 경우이다. 실제로 웹 서비스의 응답시간은 웹 서버의 응답시간에 네트워크 전송 시간, 클라이언트 컴퓨터의 처리시간 등을 포함하므로, 주어진 T_L 및 T_C 보다 더 길리는 것이 일반적이므로, 이 경우도 체감 응답시간은 $T_L = T_C = 0.1$ 초 보다 커지게 된다. $T_L = T_C = 0.1$ 은 웹 트랜잭션의 처리시간에 비해서 크지 않은 시간이므로, 클라우드 컴퓨터로의 통합 웹서비스 전환에 대한 다중화의 효과가 웹 트랜잭션의 도착률이 증가함에 따라 명확하게 나타나는 경우이다. (그림 4)(a)는 $g(\mu) = A \cdot \mu$ 의 경우로서 컴퓨터의 처리용량과 비용이 선형적으로 증가하는 관계로서 처리용량 규모에 따른 비용 절감 효과는 없으며, 다중화에 따른 효과가 주가 된다. 반면 (그림 4)(b)는 $g(\mu) = B \cdot \log(1 + \mu)$ 의 경우이며 다중화 효과와

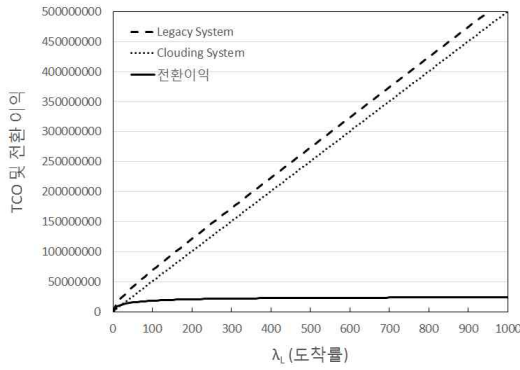
처리용량 규모에 따른 효과가 더해짐으로써 전환이익이 크게 나타남을 알 수 있다. (그림 4) 및 이어 논의될 (그림 5)와 (6)의 그래프에서는 다중화 효과와 처리용량 규모에 따른 비용 절감 요소를 중심으로 분석하기 위하여 OE(운영비용)에 의한 전환 이익은 0이 되도록 하였다. 본 분석에 사용된 수치에 따르면 웹 서버 N 이 100개를 넘어서면 운영비용에 의한 전환이익이 발생한다.

(그림 5)와 (6)의 그래프는 $T_L = T_C = 1.0$ 초 및 2.0초일 때, 즉 웹 서비스의 평균응답시간을 적절하게 그리고 다소 크게 설정한 경우의 전환이익을 보인다. 하나의 웹 트랜잭션을 처리하는 시간으로 보면 $T_L = T_C = 1.0$ 초 및 2.0초는 아주 큰 값으로서 다중화에 따른 전환이익 효과(레거시 시스템의 처리용량의 총합과 통합 클라우드 컴퓨터의 처리용량의 차이)는 도착률(λ_L)이 200 정도에서 포화상태에 접근하여 이후에는 거의 일정한 값을 갖는다. 이러한 결과는 (그림 5) 및 (6)의 그래프에서 전반적으로 나타난다. 특히 (그림 5) 및 (6)의 (a) $g(\mu) = A \cdot \mu$ 그래프들에서는 전환이익이 바로 포화상태에 들어가며, 이에 따라 전환이익도 매월 2천4천백만원(그림 5(a)) 및 1천2백만원(그림 6(a))에 수렴하고 있음을 알 수 있다. 다만 (b) $g(\mu) = B \cdot \log(1 + \mu)$ 그래프에서는 처리용량 규모에 따른 효과가 반영되어 도착률(λ_L)이 증가함에 따라 전환이익이 완만하지만 계속 증가함을 알 수 있다.

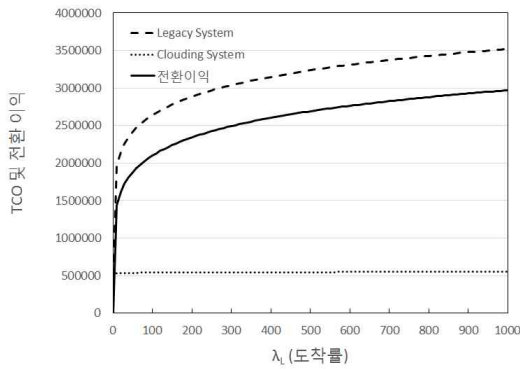
4. 결론

클라우드 컴퓨팅 서비스에 대한 관심이 많아지면서 기존의 전통적인 응용을 효과적으로 클라우드 환경으로 전환할 방안들을 모색하고 있다. 전환할 대상 응용을 선정하는 것 뿐 아니라 전면적으로 또는 부분적으로 전환할 것인가를 결정할 때도 성능과 비용을 중요하게 고려해야 한다.

(그림 5) $N=100$, $A=5000$ 원/월,
 $B=10000$ 원/월, $T_L = T_C=1.0$ 초, $C_L = C_C=10$,
 $OE_L=5000$ 원/월, $OE_C=500,000$ 원/월 일때의
 레거시 시스템 및 클라우드 시스템에서의 TCO
 및 클라우드 컴퓨터 시스템으로의 전환이익



(a) $g(\mu) = A \cdot \mu$

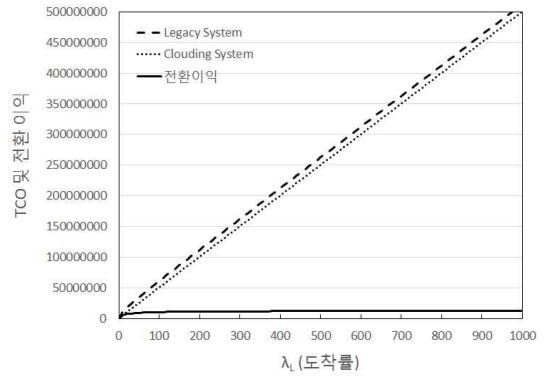


(b) $g(\mu) = B \cdot \log(1 + \mu)$

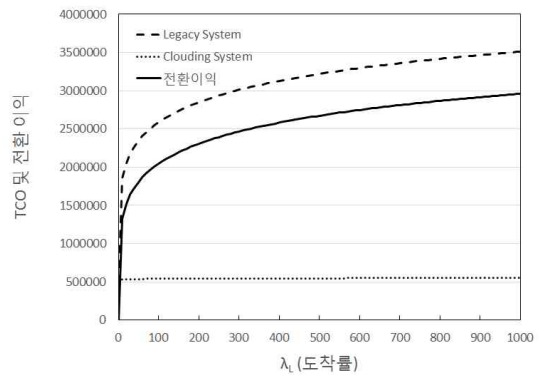
(Figure 5) TCOs and conversion profit when
 $N=100$, $A=5000$ KRW/month,
 $B=10000$ KRW/month, $T_L = T_C=1.0$ sec,
 $C_L = C_C=10$, $OE_L=5000$ KRW/month,
 $OE_C=500,000$ KRW/month

클라우드 컴퓨팅의 성능과 비용은 모델링, 시뮬레이션, 측정의 방법으로 인프라, 플랫폼, 서비스 수준에서 광범하게 분석돼 왔으나 대체로 그 방법이 복잡하고 많이 사용하는 실제응용을 다룬 연구는 많지가 않다.

(그림 6) $N=100$, $A=5000$ 원/월,
 $B=10000$ 원/월, $T_L = T_C=2.0$ 초, $C_L = C_C=10$,
 $OE_L=5000$ 원/월, $OE_C=500,000$ 원/월 일때의
 레거시 시스템 및 클라우드 시스템에서의 TCO
 및 클라우드 컴퓨터 시스템으로의 전환이익



(a) $g(\mu) = A \cdot \mu$



(b) $g(\mu) = B \cdot \log(1 + \mu)$

(Figure 6) TCOs and conversion profit when
 $N=100$, $A=5000$ KRW/month,
 $B=10000$ KRW/month, $T_L = T_C=2.0$ sec,
 $C_L = C_C=10$, $OE_L=5000$ KRW/month,
 $OE_C=500,000$ KRW/month

본 논문에서는 홈페이지 서비스를 주로 하는 웹 서버들을 통합하여 클라우드 환경으로 전환할 때 요구되는 최적용량과 전환이익을 분석하였다. 홈페이지 서버들의 서비스 모델을 정립하고, 기존 웹 서버와 동일한 서비스품질을 지원할 수 있는 클라우드 컴퓨터의 최적용량 관계식을 찾고 그 때의 전환이익을 분석하였다. 서버의 평

균 이용률이 대체로 낮고 평상시 업무부하의 변화도 크지 않은 특성을 바탕으로 사용자가 요청하는 웹 트랜잭션의 도착률과 컴퓨터 처리용량과의 관계를 분석한 결과, 요구되는 클라우드 컴퓨터의 처리용량은 평균 응답시간에 따라 현격히 줄어들 수 있음을 보였고, 다중화 효과와 처리용량 규모가 비용 절감에 크게 기여하는 요소임을 확인하였다. 제시된 웹 서비스 모델과 분석 결과는 중대규모의 대학 캠퍼스는 물론 서비스 환경과 업무 특성이 비슷한 기업, 기관들에 쉽게 적용할 수 있고 클라우드 서비스 제공자들 또한 계획단계에서 간편하게 활용할 수 있다. 상위 계층 응용 서비스들의 자원요구 특성과 클라우드 플랫폼의 다양한 자원관리 정책 등을 면밀하게 고려하는 모델링과 분석은 또 다른 과제로 남긴다.

References

[1] Michael Armbrust, et al., "Above the Clouds: A Berkeley View of Cloud Computing," Technical Report No. UCB/EECS-2009-28, Feb. 2009.

[2] Georgia Sakellari, George Loukas, "A survey of mathematical models, simulation approaches and testbeds used for research in cloud computing," Simulation Modelling Practice and Theory, 2013.

[3] Seung-Hwan Lim, Bikash Sharma, Gunwoo Nam, Eun Kyoung Kim, and Chita R. Das, "MDCSim: A Multi-tier Data Center Simulation Platform," Department of Computer Science and Engineering, The Technical Report CSE 09-007, Pennsylvania State University.

[4] Simon Malkowski, et al., "Challenges and Opportunities in Consolidation at High Resource Utilization: Non-monotonic Response Time Variations in n-Tier Applications," 2012 IEEE Fifth International Conference on Cloud Computing(CLOUD), June 2012.

[5] C. Curino, E. P. Jones, et al., "Workload-aware database monitoring and consolidation," SIGMOD '11.

[6] Siew Huei Liew and Ya-Yunn Su, "CloudGuide: Help

ing Users Estimate Cloud Deployment Cost and Performance for Legacy Web Applications," 2012 IEEE 4th International Conference on Cloud Computing Technology and Science, 2012.

[7] L. Kleinrock, Queueing Systems, Vol. 1: Theory (Wiley, 1975).

[8] Quang Hieu Vu and Rasool Asal, "Legacy Application Migration to the Cloud: Practicability and Methodology," 2012 IEEE Eight' World Congress on Services.

[9] Fernando Koch, Marcos D. Assuncao, Marco A. S. Netto, "A cost analysis of cloud computing for education," GECON'12 Proceedings of the 9th international conference on Economics of Grids, Clouds, Systems, and Services, pp.182-196, ACM, 2012.

[10] Massimiliano Rak, Antonio Cuomo, and Umberto Villano, "Cost/Performance Evaluation for Cloud Applications Using Simulation," 2013 Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, IEEE, 2013.



이 구 연

1986년 : 서울대학교 전자공학과 (학사)
 1988년 : KAIST 전기및전자공학과 (석사)
 1993년 : KAIST 전기및전자공학과 (박사)
 1993년~1996년: 디지털정보통신 연구소
 1996년: 삼성전자
 1997년~현재: 강원대학교 컴퓨터학부 교수
 관심분야 : 이동통신, 네트워크보안, 인터넷, 초고속 통신망, ad-hoc 네트워크



최창열

1979년 : 경북대학교 전자공학과 (학사)
1981년 : 경북대학교 전자공학과 (석사)
1995년 : 서울대학교 컴퓨터공학과 (박사)

1984년~1996년 : ETRI 컴퓨터연구단 책임연구원/연구실장

1996년~현재 : 강원대학교 컴퓨터학부 교수
관심분야 : 컴퓨터시스템아키텍처, 임베디드시스템, 클라우드컴퓨팅



윤재구

2009년 : 한국방송통신대학교 컴퓨터공학과(학사)

1993년~2000년 : 인텍소프트
2000년~2000년 : CJ시스템즈
2000년~2002년 : 성은정보
2002년~현재 : 더존비즈온 이사
관심분야 : 빅데이터 컴퓨팅, 데이터마이닝, 데이터베이스



최황규

1984년 : 경북대학교 전자공학과 (학사)
1986년 : KAIST 전기및전자공학과 (공학석사)
1989년 : KAIST 전기및전자공학과 (공학박사)

1990년~ 현재 : 강원대학교 IT대학 컴퓨터정보통신공학전공 교수

관심분야 : 데이터베이스시스템, 멀티미디어시스템, 클라우드컴퓨팅



장민

1993년 : POSTECH 컴퓨터공학과 (학사)
1995년 : POSTECH 컴퓨공학과 (석사)
2000년 : POSTECH 컴퓨터공학과 (박사)

2000 - 2003 : LG전자 IMT2000 선임연구원
2003 - 2006 : 신지소프트 연구소장
2006 - 2010 : 비원플러스 연구 본부장
2010 - 현재 : 더존비즈온 기획실 이사
관심분야 : Data mining, Big data, Cloud Computing, Mobile Computing