

# Dynamic Programming Knapsack 알고리즘 기반의 가상머신 통합

김민회, 박민호  
송실대학원 정보통신공학  
e-mail : [wflore@naver.com](mailto:wflore@naver.com), [mhp@ssu.ac.kr](mailto:mhp@ssu.ac.kr)

## VM Consolidation Based On Dynamic Programming Knapsack Algorithm

MinHoe Kim, SungWon Seo, MinHo Park  
Dept. of Telecommunication Engineering, SoongSil Graduate School

### 요 약

구동에 필요한 다수의 Virtual Machine을 물리적 서버 안에 Consolidation하게 구성하면, 물리적 서버의 개수를 최소화시켜 에너지 소모를 줄일 수 있다. 이 논문에서는, 하드웨어 요구량에 따른 Virtual Machine Consolidation과 시간 패턴에 따른 Virtual Machine Consolidation을 Energy Saving 관점으로 비교하고, 에너지 효율적인 Virtual Machine Consolidation 알고리즘을 제안한다.

### 1. 서론

다양한 기업들을 관리하기 위한 데이터 센터들은 기업들의 성장과 함께 커지고 있다. 그러면서 최근 각 기업마다 사용하는 데이터 센터의 에너지 소모비용도 크게 증가되어 문제로 대두되고 있다. 몇 년 전, IBM의 CFO는 에너지 비용의 급격한 상승에 대응하여 회사 내 어느 곳에서 주로 에너지를 사용하는가를 파악하기 위한 프로젝트를 진행했다. 그 결과, 데이터센터는 공간 비중이 전체의 6%에 불과하지만 에너지 사용량과 에너지 비용에서는 모두 35%를 넘게 차지하는 것으로 나타났다.

이런 에너지 사용량은 IBM 같이 기업 또는 데이터센터 아웃소싱 사업 규모가 큰 기업에만 한정된 문제가 아니다. 여러 산업 분석 기관에 따르면, 금융, 의료, 유통 및 공공기관 등 대부분의 3차 산업에서는 데이터센터가 전체 에너지 비용의 15~25%를 차지한다. Wall Street and Technology에 최근 실린 “미래의 데이터 센터(The Data Center of the Future)”라는 글에서 Citigroup의 데이터센터 계획 책임자 Jack Glass는 “Citigroup의 데이터센터는 실물 인프라 공간 비중이 1%에 불과하지만 전력 사용량은 약 25%를 차지한다.”고 밝힌 적 있다.

본 논문에서는 데이터센터 에너지 소비량을 줄이기 위한 물리적 서버 가상화 방법에 대해 제안한다. 기존 에너지 소모를 줄이기 위한 물리적 서버 가상화 방법에는 Knapsack Problem Algorithm(배낭 문제 알고리즘)을 이용한 Virtual Machine Consolidation이 있다. 이 방법은 하드웨어 요구량에 맞게 물리적 서버에 Virtual Machine을 최적으로 넣을 수 있어 Virtual Machine 숫자에 비례한 전체 물리적 서버의 개수를 최소화할 수가 있다. 그렇지만, 현

재 Virtual Machine들은 각각 하루 동안 동작하는 사용시간, 시간대 모두 다르다. 이 서비스패턴을 따라 Virtual Machine을 물리적 서버에 Consolidation하게 되면, 기존 하드웨어 자원에 따른 Virtual Machine Consolidation 방법보다 Energy Saving관점에서 더 좋은 결과가 나타난다. 그러므로, 기존 하드웨어 요구량에 따른 Virtual Machine Consolidation을 소개하고, 기존 방법과 새롭게 제안하는 Virtual Machine Consolidation방법을 비교하고 새 알고리즘을 제안한다.

### 2. Energy-Aware Consolidation

#### 2.1 하드웨어 최소 요구량에 기반한 Virtual Machine Consolidation

데이터센터의 에너지 소비량을 줄이기 위해 Virtual Machine의 Consolidation을 통해 물리적 서버 개수를 최소화시키는 방법이 있다. 데이터센터에서 사용하는 Virtual Machine들을 물리적 서버에 최대한 Consolidation 할수록 데이터 센터안의 물리적 서버의 총 개수는 줄어들 것이다. Consolidation을 통해서 줄어든 물리적 서버 수에 따라 에너지 소모도 줄어든다. 이 방법은, 물리적 서버 수가 최소면 Energy의 사용도 최소라 가정한다. Virtual Machine을 한대의 물리적 서버에 Consolidation을 하기 위해서는 몇 가지 고려할 사항이 있다. 각 Virtual Machine이 요구하는 하드웨어 자원은 다르고, 물리적 서버의 하드웨어 자원은 한정적이다. Virtual Machine이 요구하는 물리적 서버의 하드웨어 자원 총량이 물리적 서버의 한계를 넘어서면 안 된다. Virtual Machine을 물리적 서버의 한정된 하드웨어

자원에 최적의 방법으로 넣을 수 있는 방법 중 하나는 Knapsack Problem Algorithm이다. Knapsack Problem Algorithm은 n개의 물건과 각 물건 i의 무게  $w_i$ 와 가치  $v_i$ 가 주어지고, 배낭의 용량이 C일 때, 배낭에 담을 수 있는 물건의 최대 가치를 찾는 방법이다. Virtual Machine Consolidation문제를 Knapsack Problem Algorithm에 맞게 설정하면 다음 테이블 1처럼 표현된다.

테이블 1. Knapsack Problem Algorithm 요소

Notation	Description
$VM_i$	i 번째 Virtual Machine
$R^j$	물리적 서버 H/W 자원 j의 총량
$r_i^j$	$VM_i$ 의 Hardware자원 j의 요구량

물리적 서버의 하드웨어 자원은  $R^j$ 로 표시했는데, 이는 물리적 서버의 j번째 하드웨어의 총 용량을 의미한다. i번째 Virtual Machine은  $VM_i$ 로 표시하고, i번째 Virtual Machine이 원하는 물리적 서버의 j번째 하드웨어 자원의 요구량은  $r_i^j$ 이 된다. 각 Virtual Machine의 가치는 동일하므로 각  $VM_i$ 의 가치는 요구하는 하드웨어 자원에 비례한다.

알고리즘 1. Knapsack Problem Solution

```

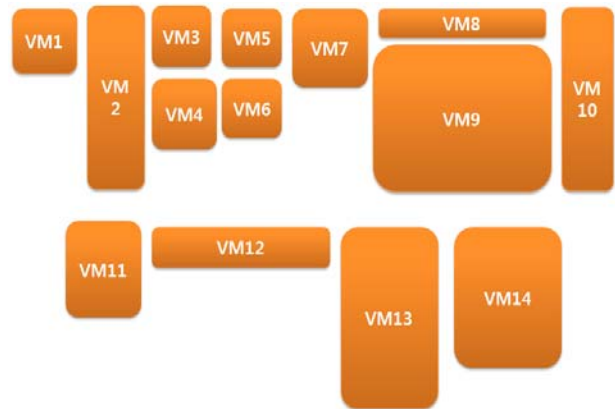
1: for i= 1 to n {
2:   for j= 1 to C{
3:     if( $r_i^j > R^j$ )
4:        $K[i, R^j] = K[i-1, R^j]$ 
5:     else
6:        $K[i, R^j] = \max\{K[i-1, R^j], K[i-1, R^j - r_i^j] + \sum r_i^j\}$ 
7:   }
8: }
9: return  $K[n, R^C]$ 
    
```

Virtual Machine을 넣을 각 물리적 서버의  $R^1$ (Hardware resource 1의 총 용량)과  $R^2$ (Hardware Resource 2의 총 용량)의 값은 600으로 가정하고 위 알고리즘 1을 진행한다. Line 1~2에서는 Virtual Machine을 1에서 n까지 하나씩 고려하여 물리적 서버의 하드웨어 j를 1에서 C까지 각각 증가시키며, 다음 라인을 수행한다. Line 3~4에서는 현재 물리적 서버에 담아보려고 고려하는 i번째 Virtual Machine의 요구하는 하드웨어 j의 자원  $r_i^j$ 가 물리적 서버의 하드웨어 j의 용량  $R^j$ 보다 크면 i번째, Virtual Machine은 물리적 서버에 담을 수 없으므로, i번째 Virtual Machine까지 고려했을 때의 최대 가치  $K[i, R^j]$ 는 (i-1)번째 Virtual Machine까지 고려했을 때의 최대 가치  $K[i-1, R^j]$ 가 된다. Line 5~6에서는 만일 현재 고려하는 i번째 Virtual Machine이 현재 물리적 서버의 하드웨어 j의 용량  $R^j$ 보다 같거나 작으면, i번째 Virtual Machine은 물리적 서버에 담을 수 있다. 그러나 현재 상태에서 i번째 Virtual Machine을 추가로 물리적 서버에 담으면 물리적 서버의 하드웨어 j의 총 용량은  $R^j - r_i^j$ 가 된다.

테이블 2.  $VM_i$ 의 자원요구량

$VM_i$	$r_i^1$	$r_i^2$
1	200	200
2	200	600
3	200	200
4	200	250
5	200	200
6	200	200
7	250	250
8	600	100
9	600	500
10	200	600
11	250	300
12	600	150
13	350	600
14	400	500

테이블 2는 알고리즘 검증을 위해 각 하드웨어 자원 요구량이 다른 14개의 Virtual Machine들을 임의로 만든 것이다.



(그림 2)  $VM_i$ 의 도식화

그림 3은  $r_i^1$ 는 가로 길이,  $r_i^2$ 는 세로 길이로 알고리즘을 통해 얻은 Consolidation 결과를 도식화 한 것이다. 하드웨어 자원 요구량에 따른 Knapsack Problem Algorithm으로 위 Virtual Machine 14개를 물리적 서버 안에 배정하면 물리적 서버 4개 안에 14개의 Virtual Machine을 모두 배치할 수 있다.



(그림 3) 하드웨어 요구량에 따른 Virtual Machine Consolidation

## 2.2 서비스 패턴을 고려한 보다 효율적인 Virtual Machine Consolidation

위 2.1의 하드웨어 최소 요구량에 기반한 Virtual Machine Consolidation에는 한 가지 문제점이 있다. 물리적 서버의 개수는 최소화 되었지만, Energy Saving 관점에서 최적이지 않은 경우가 있다. 각 Virtual Machine이 하루 동안 동작하는 시간, 즉 서비스 패턴을 고려해 설계해보면, 이전의 물리적 서버 최소화 방법보다 에너지 소모가 더 적은 상황이 있다.

테이블 3. VM<sub>i</sub>의 사용시간과 자원요구량

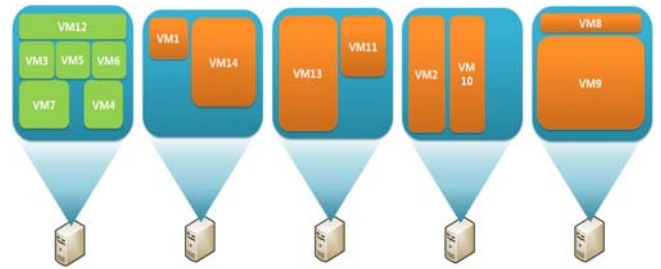
VM <sub>i</sub>	T <sub>i</sub>	R <sub>i</sub> <sup>1</sup>	R <sub>i</sub> <sup>2</sup>
1	16	200	200
2	16	200	600
3	24	200	200
4	24	200	250
5	24	200	200
6	24	200	200
7	24	250	250
8	16	600	100
9	16	600	500
10	16	200	600
11	16	250	300
12	24	600	150
13	16	350	600
14	16	400	500

테이블 3은 시간패턴에 따른 Virtual Machine Consolidation의 간단한 검증을 하기 위해, i번째 Virtual Machine의 하루 중 사용시간인 T<sub>i</sub>값을 추가했다. 서비스 패턴을 보면, 4, 5, 6, 7, 11, 12번 Virtual Machine은 하루 종일 동작하고 나머지는 0시부터 16시까지 작동한다고 가정한다.

각 물리적 서버의 시간당 전력 소비량을 P watt라고 가정한다. 기존의 Knapsack Problem Algorithm으로 물리적 서버 개수를 최소화 시킨 그림 3에서는 24시간 동작이 필요한 Virtual Machine이 각 서버마다 1개 이상 있으므로, 물리적 서버 4대 모두 24시간동안 작동되어야 한다. 그러므로, 총 사용 전력은 4 \* Pwatt \* 24hours로 96PWh가 된다.

그림 4의 서비스 패턴을 파악하여 만든 Virtual Machine Consolidation을 보면, 총 물리적 서버의 개수는 5개로 늘어났지만 시간을 고려한 Knapsack Problem Algorithm이 24시간 동작하는 가상머신 4, 5, 6, 7, 11, 12를 한 대의 물리적 서버에 모아놓았기 때문에 물리적 서버 한 대만 24시간 동작시키면 된다. 그러므로, 물리적 서버 총 사용 에너지를 계산해보면, (1 \* Pwatt \* 24hours) + (4 \* Pwatt \* 16hours)는 88PWh로 Energy Saving 관점에서 하드웨어 요구량에 따른 Virtual Machine Consolidation 방식보다 효율적인 걸 알 수 있다. 이와 같이 Energy Saving 관점에서 최대의 이익을 얻기 위해서는 각 물리적 서버의

서비스 패턴을 파악한 Virtual Machine Consolidation과정이 필요한 것을 알 수 있다.



(그림 4) 서비스 패턴에 따른 Virtual Machine Consolidation

## 2.3 제안하는 해결 방안

Virtual Machine을 시간패턴에 따라 Consolidation하기 위해서는 Knapsack Problem Algorithm을 이용한 Energy Aware Consolidation을 해야 한다. 그러므로, 다음 알고리즘을 제안한다.

### 알고리즘 2. 제안하는 Knapsack Problem Solution

- Step1** Virtual Machine를 사용시간이 긴 순서대로 Sorting 한다.
- Step2** 가장 사용시간이 긴 Virtual Machine을 물리적 서버 한대에 우선 배치한다.
- Step3** 나머지 Virtual Machine들에 대해서 Value를 정한다.
- Step4** Knapsack Problem Algorithm으로 물리적 서버 1대를 Consolidation 한다.
- Step5** go to Step1

제안하는 Knapsack Problem 알고리즘에서는 Virtual Machine을 먼저 사용시간이 긴 순서대로 Sorting한다. 가장 사용시간이 긴 Virtual Machine을 물리적 서버에 우선 배치하고, 이 Virtual Machine과 중첩되는 시간이 길수록 Value를 높게 하여, 서버에 배치한 Virtual Machine을 제외한 나머지 Virtual Machine의 Value를 정한다. Value를 정한 후, Knapsack Problem Algorithm을 통해 Value에 맞게 물리적 서버 1대를 Consolidation한다. Step1~4를 반복하여 Virtual Machine을 새로운 물리적 서버에 Consolidation하면 그림 4와 같은 결과를 얻을 수 있다.

## 3. 결론 및 향후 연구

하드웨어 요구량에 따른 Virtual Machine Consolidation은 물리적 서버의 수가 최소가 되므로, Energy Saving 관점에서 가장 좋은 예로 보인다. 그렇지만, 서비스 패턴을 고려한 Virtual Machine Consolidation은, 전의 하드웨어 요구에 따른 Virtual Machine Consolidation보다 Energy Saving 관점에서 더 뛰어난 성능을 발휘한다. 그러므로, 본문은 알고리즘2의 수정된 Knapsack Problem Algori

thm 사용을 제안한다.

향후 연구는 시간 패턴에 따른 Virtual Machine Consolidation의 알고리즘을 직접 구현해보고, 구현된 알고리즘의 성능검증을 위한 다양한 알고리즘 검증단계를 통해, 시간 패턴에 따른 Virtual Machine Consolidation의 성능을 입증해보겠다.

#### 참고문헌

- [1] James F. Kurose, Keith W. Ross "Computer Networking A Top-Down Approach" 6th Ed. Pearson
- [2] 최동욱, 김두호, 강정일, 조성우, 박종민 "실전 클라우드 가상화 구축 기술" 한빛 미디어
- [3] Melanie Rodier "Wall Street Technology. The Data Center of the Future" 2010