

클라우드 컴퓨팅 환경의 스토리지 시스템 정보 시각화 및 Time-lapse 개념 적용에 대한 고찰*

전우진^{1O} 이봉환² 박기웅^{3*}

¹세종대학교 시스템보안연구소, ²대전대학교 정보통신공학과, ³세종대학교 정보보호학과
woojinjeon929@gmail.com, blee@dju.ac.kr, woongbak@sejong.ac.kr

Applying Time-lapse Concept into Storage System Visualization for Cloud Computing Environment

Woo-Jin Jeon^{1O} BongHwan Lee², Ki-Woong Park^{3*}

¹Sejong Univ. SysCore Lab.,

²Daejeon Univ. Dept. of Information and Communications Engineering,

³Sejong Univ. Dept. of Computer and Information Security

요 약

최근 빅데이터와 인공지능 기술이 접목되고 그 활용에 대한 수요와 신뢰성이 높아지면서 시스템 생물학, 바이오센서 등의 분야에서 장기적(Long-Term) 관점에서의 시계열 분석을 기반으로 한 다양한 연구가 수행되고 있다. 하지만 정보보안 분야에서는 스토리지 자원 관리를 위한 모니터링 기술과 디지털 포렌식을 위한 로깅 기술과 같은 단기적(Short-Term) 레벨의 시계열 정보만을 분석하여 활용하는 것에 한정되고 있다. 따라서 이와 같은 기술적 트렌드와 다양한 연구 사례들을 종합하여 보았을 때 스토리지 자원에 대한 장기적 레벨의 시계열 정보 분석은 지금까지 단기적 레벨의 시계열 정보 분석에서 얻지 못한 새로운 데이터를 추출할 수 있을 것이라 판단할 수 있다. 본 논문에서는 클라우드 환경에서의 스토리지 시스템 정보 분석에 장기적 관점에서의 시계열 정보 분석을 위한 시각화 및 타임랩스(Time-lapse) 개념을 적용하기 위하여 기존에 사용되고 있는 스토리지 정보 분석 툴의 시각화 기법을 분석하고, 시각화를 위한 용량적/연산적 오버헤드 측정을 통해 프로파일링을 수행한다. 또한 프로파일링을 통해 장기적 관점의 스토리지 시스템 정보에 타임랩스 개념을 적용하기 위하여 대용량 스토리지 정보에 적합한 데이터 시각화 표현 기법과 시스템 오버헤드에 대하여 고찰해본다.

1. 서 론

4차 산업혁명 시대로 접어든 오늘날 사물인터넷과 빅데이터, 인공지능 기술 등으로 데이터가 증가됨에 따라 데이터 분석을 통한 가치의 중요성이 높아지고 있다. 데이터량이 증가하고 I/O에 대한 부하가 높아짐에 따라 효율적인 자원 관리를 위한 시스템 모니터링 기술, 로깅 기술 등이 개발되었다. 이와 같은 기술은 시간의 흐름에 따라 주기적으로 측정되어 기록된 데이터(시스템 로그 정보 등)에 근거하여 자료의 상관관계를 분석하는 시계열 정보 분석에 기반하며, 효율적인 스토리지 자원 관리 및 속도 향상을 위해 주로 활용되고 있다. 또한 정보보안 분야에 있어서는 해킹 및 정보유출의 방지 및 사후 분석을 위해 측정 데이터 및 로깅 데이터에 대한 시계열 저장기술을 통해 디지털 포렌식 관점에서의 타임라인 분석기술을 활용한다. 기존의 스토리지 자원 관리를 위한 모니터링 기술과 디지털 포렌식을 위한 로깅 기술 등은 대부분 단기적(Short-Term) 레벨의 시계열 정보를 활용

한 사례라고 볼 수 있다. 예를 들어 정보보안 분야에서의 단기적 레벨 시계열 정보 분석은 사건(Event)을 기준으로 디스크 덤프 기술을 활용하여 이벤트들의 상관 관계를 분석하기 때문에 저장된 데이터량 대비 비교적 적은 시스템 오버헤드를 가지고 분석이 가능하다. 장기적 레벨에서의 시계열 정보는 사회 과학 계열의 재무, 경제 예측 등 사회 전반적인 추세를 분석 및 예측하는데 오래 전부터 활용되어 왔다. 특히 최근 빅데이터와 인공지능 기술이 접목되고 그 활용에 대한 수요와 신뢰성이 높아지면서 시스템 생물학, 바이오센서 등과 같은 분야에서도 장기적 레벨의 시계열 정보 분석을 통해 새로운 데이터를 생성 및 예측하기 위한 다양한 연구를 수행하고 있다[1, 2]. 이러한 기술적 트렌드와 다양한 연구 사례들을 종합하여 보았을 때 스토리지 자원에 대한 장기적 레벨의 시계열 정보 분석은 지금까지 단기적 레벨의 시계열 정보 분석에서 얻지 못한 새로운 데이터를 생성하고 더욱 전문화된 데이터를 얻을 수 있을 것이라 예상할 수 있다. 또한, 추출한 가치 있는 데이터에 대한 시각화를 통하여 심도있는 정보를 통찰(Insight)할 수 있을 것이라 예상 가능하다. 그러나 장기적 관점의 사건 또는 경향(Trend)을 분석하기 위해서는 기존에 사용되고 있는 단기적 관점의 시계열 분석을 위한 스킴(기술적 방법)이 아

* 이 논문은 한국연구재단 지원사업(2017R1C1B2003957) 및 지역혁신 창의인력 양성사업(2015H1C1A1035859) 연구수행으로 인한 결과물임
† 교신저자: 박기웅 (세종대학교 정보보호학과 교수)

닌 장기적(Long-Term) 레벨의 시계열 정보 분석에 특화된 스킴이 요구된다.

따라서 본 논문에서는 클라우드 컴퓨팅 환경에서의 스토리지 시스템 정보에 대한 장기적 관점의 시계열 정보 분석을 위한 타임랩스(Time-lapse) 개념 적용에 대하여 고찰하여 보았다. 또한 장기적인 스토리지 시스템 정보에 대한 연산 효율적인 시계열 분석을 위하여 기존의 디스크 시각화 툴을 통해 시각화에 수반되는 시스템의 용량적/연산적 측면의 오버헤드를 측정해 보았다.

본 논문의 구성은 다음과 같다. 2장에서는 클라우드 컴퓨팅 환경에서의 시계열 분석에 기반한 기술 및 선행 연구들에 대하여 알아보고, 3장에서는 기존에 사용되고 있는 디스크 분석 시각화 툴에 대하여 조사하고 디스크 시각화에 따른 용량적/연산적 오버헤드를 측정한다. 마지막 4장에서는 결론과 추후 연구에 대하여 기술한다.

2. 관련 연구

본 장에서는 클라우드 컴퓨팅 환경에서의 시계열 정보 분석에 기반한 기술에 대하여 알아보고 이와 관련된 선행 연구와 한계점을 도출한다.

2.1 클라우드 컴퓨팅 환경에서의 모니터링 수행

클라우드 컴퓨팅 환경에서의 모니터링 수행은 일반적인 로컬 컴퓨팅 환경에서 모니터링을 수행하는 것 보다 많은 정보를 분석하고 실행해야 한다. Aceto, Giuseppe 등이 저술한 논문[3]에 따르면 클라우드 모니터링 수행 시 고려해야 할 주요 측면에는 물리적인 인프라와 가상화된 리소스 성능 모니터링, 사용자에게 데이터에 대한 높은 가용성과 안정성을 보장하기 위한 모니터링, 클라우드의 인프라 및 서비스 보안 관리를 위한 모니터링 등이 있다. 그러나 이러한 주요 모니터링 측면은 클라우드 컴퓨팅을 사용하고 있는 사용자에게 빠른 확장성 및 접근성 등과 같은 고수준의 서비스를 제공하기 위한 요구사항의 단기적 레벨 모니터링이다. 따라서 이러한 연구는 장기적인 관점을 가지고 모니터링을 수행하는 환경에서는 고려되어야 하는 요구사항이 달라질 수 있기 때문에 적합하지 않다.

2.2 클라우드 컴퓨팅 환경에서의 로깅 기술

로깅 정보는 서비스 장애 모니터링, 데이터 유출 및 침해사고 발생 등과 같은 문제가 발생하였을 때 시간적 변동에 따른 로그 분석을 수행하여 문제를 해결하기 위한 증거 자료로 활용된다. 또한 다수의 인스턴스가 물리적 인프라를 공유하고 서비스가 병렬적으로 실행되는 클라우드 컴퓨팅 환경에서는 필수 구성 요소로 분류된다. 현재 많은 연구자들이 클라우드 컴퓨팅 환경의 특성을 고려한 로깅 기술에 대하여 많은 연구[4]를 진행하고 있지만, 이와 같은 연구는 단기적인 시계열 분석을 기반으로 한 로깅 기술 연구이며, 장기적 관점의 로깅 기술이 요구되는 환경에서의 사용은 적합하지 않다.

3. 스토리지 시스템의 장기적 정보 분석을 위한 시각화

및 타임랩스 개념 적용에 대한 고찰

본 장에서는 기존에 사용되고 있는 디스크 사용량 분석 시각화 툴을 조사하고 시각화에 대한 시스템 용량적/연산적 오버헤드에 대하여 측정하고, 클라우드 컴퓨팅 환경의 스토리지 시스템 정보 시각화 및 타임랩스 개념 적용에 대하여 고찰한다.

3.1 스토리지 시스템 정보 분석 툴의 시각화 방법

디스크 사용량을 분석하고 시각화하여 제공하는 스토리지 시스템 정보 분석 툴 10개의 시각화 방법에 대하여 조사해 보았다. 그 결과 대부분의 툴에서 시각화를 위한 방법으로 트리맵(Treemap) 기법을 사용하였다는 것을 알 수 있다. 트리맵[5]은 space filling 접근법을 통해 계층적 데이터 구조를 표현(시각화)하는데 적합하기 때문이다. 이러한 스토리지 정보 시각화 툴의 트리맵 기법 표현 방식을 세분화하여보면 두 가지로 나눌 수 있다.

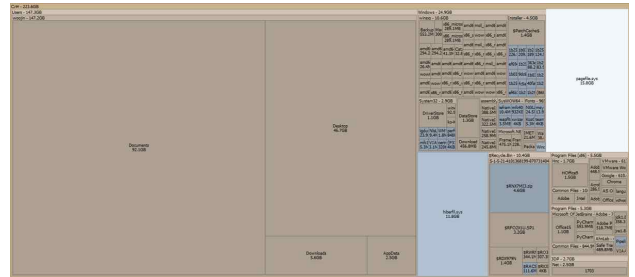


그림 1 블록형 트리맵 기법을 활용한 시각화 예시

먼저 블록 기반의 트리맵 형식이다. WinDirStat, TreeSize Free, SpaceSniffer 등과 같은 툴은 디스크의 폴더 및 파일의 트리(Tree) 구조를 블록 기반의 트리맵(Treemap) 형식으로 변환하여 디스크 구조를 시각화하였다. 블록 기반의 트리맵은 계층적 구조를 한눈에 파악할 수 있는 장점이 있다.

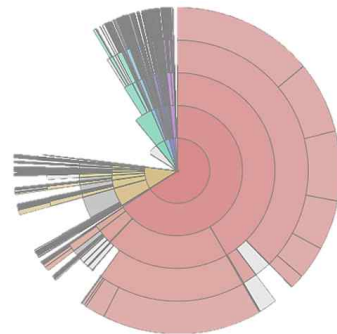


그림 2 방사형 트리맵 기법을 활용한 시각화 예시

두 번째는 방사형(Radial) 트리맵 형식이다. JDiskReport, Baobab, Duc 등과 같은 툴은 방사형 트리맵 원형 차트를 기반으로 하여 시각화 표현하였다. 방사형 트리맵은 디스크 사용량에 대한 시각화를 디스크 형태로 보여주기 때문에 전체적인 디스크 사용량에 대한 확인이 편리하다.

3.2 장기적 스토리지 시스템 정보 분석을 위한 시각화에 따른 용량적/연산적 오버헤드 측정

스토리지(하드디스크)의 물리적 단위는 1섹터(512byte)이다. 하지만 현재 사용되는 NTFS와 같은 파일 시스템에서는 데이터의 평균 크기 상승으로 인한 데이터의 효율적인 I/O 처리를 위하여 섹터 단위로 데이터를 관리하지 않고, 클러스터(4KB) 단위로 데이터를 관리한다. 예를 들어 디스크에 1MB 크기의 파일을 기록하고자 하였을 때, 섹터 단위로 데이터를 관리할 경우 2,048번의 처리가 필요하고, 클러스터 단위로 데이터를 관리할 경우 256번의 처리가 요구된다. 이와 같은 맥락으로 보았을 때 디스크 이미지의 한 픽셀 당 나타내는 데이터 량에 따라 이미지의 총 용량 역시 변화하는 것을 예상할 수 있다.



그림 3 한 픽셀 당 이미지 총 사이즈 변화

그림 3과 같이 1GB의 디스크를 대상으로 한 픽셀 당 1, 512, 1024, 2048, 4096, 65536Byte로 맵핑하여 나누어 이미지화하였을 때의 이미지 총 용량을 측정해 보았다. 1GB 파일을 한 픽셀 당 1Byte로 맵핑하였을 때 총 파일 크기는 1GB로 원본 파일 크기와 동일하게 나타났고 한 픽셀 당 65536Byte(64KB)로 맵핑하였을 때 총 파일 크기는 16384(16KB)Byte로, 한 픽셀 당 나타내는 데이터 사이즈를 어떻게 설정하느냐에 따라 총 파일 용량이 매우 상이함을 알 수 있다.

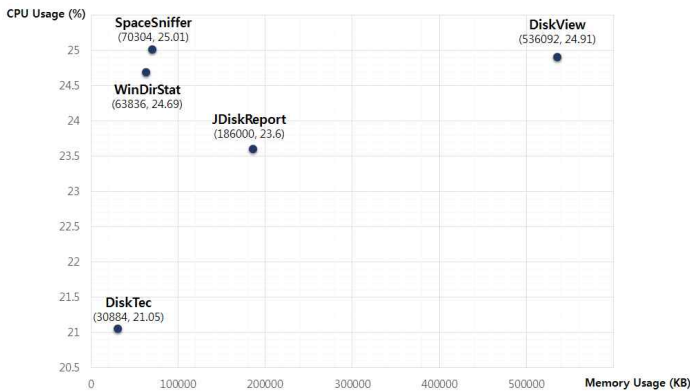


그림 4 스토리지 정보 시각화 툴의 연산적 오버헤드

그림 4에서는 기존의 스토리지 시스템 정보 분석 툴의 시각화에 따른 연산적 측면에서의 오버헤드를 측정해 보

았다. 실험은 운영체제 Windows 7 64bit 환경에서 진행되었으며, 물리 스토리지는 SSD 256GB를 사용하여 수행되었다. CPU와 Memory 사용량 측정은 Process Explorer를 사용하여 측정하였다. 위의 프로파일링 결과는 스토리지 시스템 정보 분석 툴에 따라 디스크 정보를 분석하는 단위(섹터, 클러스터)와 시각화 표현 방식에 따라 연산적 오버헤드에 차이가 있음을 나타내었다. 또한 용량적/연산적 오버헤드 측정 결과에 따라 클라우드 컴퓨팅 환경에서의 스토리지 시스템 정보를 시각화하고 장기적 레벨에서의 타임랩스 개념을 적용하기 위해서는 대용량 스토리지 시스템 정보에 대한 장기적 레벨의 데이터 추출이 필요하기 때문에 이에 대한 이해와 시각화 기법 및 연산적/용량적 오버헤드가 고려되어야 함을 알 수 있다.

4. 결 론

본 논문에서는 다양한 스토리지 시스템 정보 분석 툴의 시각화 기법을 분석하고, 데이터 시각화를 위한 용량적/연산적 오버헤드 측정을 통해 프로파일링을 수행해 보았다. 프로파일링 수행 결과 시각화를 통한 스토리지 시스템 정보 분석은 분석할 데이터의 특성과 데이터 분석 환경에 대한 오버헤드 고려가 요구되어진다는 것을 알 수 있다. 따라서 장기적 관점의 모니터링을 통하여 단기적 관점에서 얻을 수 없었던 새로운 데이터를 생성하기 위해서는, 장기적 관점의 스토리지 시스템 정보에 적합한 데이터 시각화 표현 기법과 시스템의 용량적/연산적 오버헤드에 대한 고려가 필요하다.

향후 프로파일링한 정보를 바탕으로 대용량 스토리지 시스템 정보의 시각화 및 타임랩스 개념 적용에 적합한 스킴에 대한 연구를 진행할 계획이다.

참 고 문 헌

- [1] Banaee, Hadi, Mobyen Uddin Ahmed, and Amy Loutfi. "Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges." *Sensors* 13.12 (2013): 17472-17500.
- [2] Bar-Joseph, Ziv. "Analyzing time series gene expression data." *Bioinformatics* 20.16 (2004): 2493-2503.
- [3] Aceto, Giuseppe, et al. "Cloud monitoring: A survey." *Computer Networks* 57.9 (2013): 2093-2115.
- [4] Payne, Bryan D., DP de A. Martim, and Wenke Lee. "Secure and flexible monitoring of virtual machines." *Computer Security Applications Conference, 2007. ACSAC 2007. Twenty-Third Annual. IEEE, 2007.*
- [5] Johnson, Brian, and Ben Shneiderman. "Tree-maps: A space-filling approach to the visualization of hierarchical information structures." *Proceedings of the 2nd conference on Visualization'91. IEEE Computer Society Press, 1991.*