한국정보보호학회
Korea Institute of Information Security & Cryptology

2023년 한국정보보호학회 하계학술대회

CISC-S'23

Conference on Information Security and Cryptography Summer 2023

2023년 6월 22일(목)~23일(금)

강원대학교 춘천캠퍼스 60주년 기념관

주최 한국정보보호학회 Korea Institute of Information Security & Cryptology
주관 KNU 강원대학교 KANGWON NATIONAL UNIVERSITY
후원 국가정보원 NATIONAL INTELLIGENCE SERVICE, 과학기술정보통신부, 행정안전부, KISA 한국인터넷진흥원, ETRI 한국전자통신연구원 Electronics and Telecommunications Research Institute, NSR 국가보안기술연구소 National Security Research Institute, Google

이상징후 탐지를 위한 거시적 관점에서의 데이터 시각화 프레임워크

이하늘*, 김지혜*, 우승찬*, 임우협*,
신규섭*, 장서현*, 최상훈[1], 노주영[2], 박기웅†

,† 세종대학교 정보보호학과 (학부생, 교수†)
[1,2]세종대학교 정보보호학과 시스템보안연구실(연구원[1], 대학원생[2])

Data Visualization Framework from a Macro Perspective for Anomaly Detection

Haneul Lee*, Jihye Kim*, Seungchan Woo*, Woohyeop Im*, Gyuseob Shin*, Seohyun Jang*, Sang-Hoon Choi[1], Joo-Young Roh[2], Ki-Woong Park†

,† Dept. of Computer and Information Security, Sejong University (Undergraduate Student, Professor†)
[1,2]SysCore Lab. (Convergence Engineering for Intelligent Drone), Sejong University (Researcher[1], Graduate Student[2])

요 약

클라우드 네이티브 환경에서는 물리적인 호스트뿐만 아니라 컨테이너와 같은 가상의 호스트도 컴퓨팅 리소스를 소모하면서 보안 관리자의 모니터링 범위가 확대된다. 또한, 코로나19 이후 원격근무가 늘어나면서 공격 표면이 증가하고, 연결 단말의 수와 사이버 공격 수법도 고도화되면서 기존 시스템을 무력화시키고 보안 관리자의 부담을 증가시키는 문제가 대두되었다. 이에 따라, 본 논문에서는 대량의 호스트들의 성능 정보 변화를 거시적으로 확인하고, 실시간으로 시스템 성능 정보를 가시적으로 보여줌으로써 보안 관리자가 이상징후를 판단할 수 있는 프레임워크를 제안한다.

I. 서론

오늘날, 공간에 제약 없이 데이터를 필요시에 제공받을 수 있는 클라우드 서비스는 애플리케이션 구축, 배포, 및 관리를 위한 '클라우드 네이티브' 환경으로 변화하기 시작하였다. 기업들은 클라우드 네이티브 환경을 구현하기 위해 컨테이너의 도입을 확대하였으며, 대량의 컨테이너를 관리하기 위해 쿠버네티스, 도커 등의 플랫폼을 사용하기 시작하였다[1]. 이를 통해 기업들은 서비스의 확장성, 유연성 및 복원력을 제공하며 공간에 제약 없이 데이터를 필요시에 제공받을 수 있게 되었다. 실제로, 구글은 애플리케이션 관리를 위해 약 20억 개의 컨테이너를 사용하고 있다[2].

가상화 호스트의 사용 확대뿐만 아니라 개인이 소유하는 단말과 사물 인터넷 등 네트워크에 연결되는 기기의 수가 기하급수적으로 증가하면서 보안 관리자가 모니터링 해야하는 대상의 범위가 증대되었다.

또한, 코로나19 이후 원격근무 채택이 증가하며 공격 표면(Attack Surface)이 확대되어 기존의 기관에서 사용하던 폐쇄망의 이점을 감소시켰고, 증가하는 엔드포인트 단말의 수와 고도화되는 사이버 공격자들의 공격 수법은 보안 시스템을 무력화시켰다. 이로 인해, 보안 관리자의 모니터링 범위가 더욱 확대되었으며 보안 관리자의 부담을 증가시켰다.

본 논문에서는 리눅스 환경을 기준으로 시스템을 관제함으로써 이상징후를 탐지하기 위해

†교신저자: 박기웅 (세종대학교 정보보호학과 교수)

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 정보통신방송기술국제공동연구사업(Project No. RS-2022-00165794, 40%), 국방ICT융합사업(Project No. Project No.2022-0-00701, 10%), 실감콘텐츠핵심기술개발사업(Project No. RS-2023-00228996, 10%), 대학ICT연구센터 육성지원사업(Project No. 2023-2021-0-01816, 10%) 및 한국연구재단 개인기초연구과제(Project No. RS-2023-00208460, 30%)의 지원을 받아 수행된 연구임.

거시적 관점에서의 데이터 시각화 프레임워크를 제안한다. 2장에서는 시스템상에서 이상징후 탐지를 위한 시스템 모니터링 방식을 설명하고, 이를 관제하기 위한 데이터 시각화 기법에 관해 설명한다. 3장에서는 이상징후 탐지를 위한 거시적 관점에서의 데이터 시각화 프레임워크에 대해 설명하고, 4장에서 결론을 맺는다.

II. 관련 연구

2.1 시스템 모니터링

시스템 모니터링은 운영 중인 서비스에서 발생한 장애를 탐지 및 대응하기 위해 지속적인 감시와 감찰을 통해 시스템의 상황을 확인하는 과정이다. 실행 중인 시스템을 모니터링하는 방식은 시스템 성능 모니터링과 시스템 보안 모니터링 방식으로 구분된다[3].

시스템 성능 모니터링은 실행 중인 시스템의 성능이 기대하는 서비스 수준의 충족 여부를 감시하기 위해 시스템의 성능 메트릭을 수집한다. 해당 성능 메트릭은 CPU, Memory, Disk, File-System, Network, Process 정보를 사용할 수 있으며, Linux 환경의 경우 [표 1]과 같이 수집이 가능하다[4].

[표 1] Linux 성능 메트릭 수집 경로

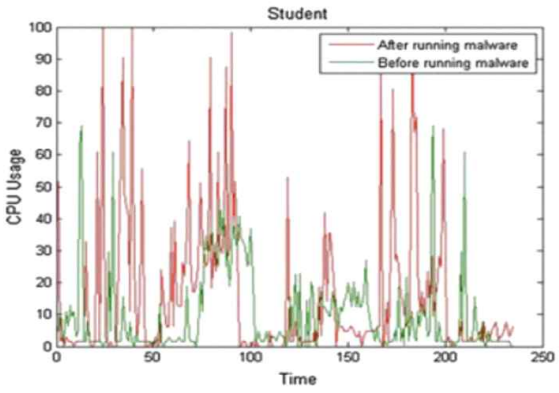
구분	수집 경로	수집 정보
CPU	/proc/cpuinfo	모델명, 코어 수
	/proc/stat	사용률
Mem	/proc/meminfo	물리 메모리 크기, 사용 가능 메모리 크기 등
Disk	/proc/partitions	파티션 정보
	/proc/diskstats	I/O에 관한 정보
FS	/proc/filesystems	현재 지원 및 지원가능한 fIle-system 종류 정보
	/proc/[pid]/mountinfo	파일의 마운트 정보
Network	/proc/net/dev	패킷의 송수신 정보
	/proc/net/fib_trie	IP에 관한 정보
Process	/proc/stat	시스템 부팅 후 실행된 process 수
	/proc/[pid]/stat	선택한 process 상태 정보

시스템 보안 모니터링은 실행 중인 시스템의 비정상 데이터를 수집하고 분석하여 이상행위를 탐지하는 방식이다. 해당 방식은 사전에 설정된 Rule Set에 따라 정해진 공격을 탐지하는 오용 탐지(Misuse Detection)와 사전에 분석한 시스템의 정상적인 특성에서 벗어나는 경우의 공격을 탐지할 수 있는 이상징후 탐지(Anomaly Detection)로 구분된다. 오용 탐지는 빠른 속도로 알려진 공격을 탐지할 수 있으나, 시그니처로 등록되지 않은 Zeroday 공격을 탐지하기 어렵다. 반면, 이상징후 탐지는 알려진 공격과 Zeroday 공격을 탐지할 수 있으며, 시그니처를 직접 등록하지 않아도 탐지할 수 있다.

2.2 데이터 시각화

데이터 시각화는 차트, 그래프와 같은 시각적 요소를 기반으로 방대한 데이터를 거시적으로 확인하기 위한 기법이다. 시스템 보안을 위해 시스템의 성능 메트릭을 시각화함으로써, 이상징후 탐지 및 악성코드 분류에 관한 연구가 제시되고 있다. 데이터를 시각화하는 기법은 시계열 그래프 기반 시각화와 이미지 기반 시각화로 구분된다[5].

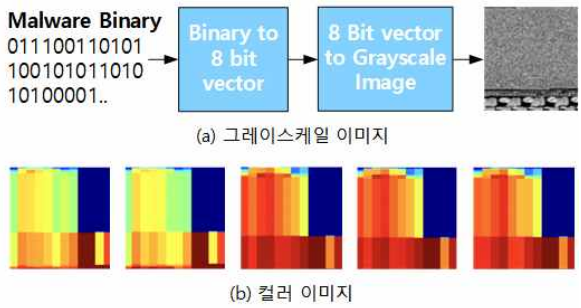
시계열 그래프 기반 시각화는 시계열로 추출한 데이터를 2차원 배열에 저장하여 정규화한 후 시계열 그래프로 표현한 기법이다. (그림 1)은 CPU 사용량을 측정하여 시계열 그래프로 표현한 연구이다[6]. 해당 연구와 같이 본 기법을 활용하여 시스템 자원의 변화 패턴을 파악함으로써 특정 시점에서의 이상징후 탐지 및 분석이 가능하다.



(그림 1) 시계열 그래프를 통한 CPU 사용량 비교

이미지 기반 시각화는 정규화 후에 그레이스케일과 컬러 이미지를 활용하는 경우로 구분된다. 그레이스케일 이미지는 회색의 이미지로 표현되며 별도의 코드 분석 없이 악성코드가 탐지

및 분류된다[7]. 컬러 이미지는 다양한 색의 이미지로 표현되며, 그레이스케일 이미지보다 위험도 구분과 같은 구체적인 악성코드 탐지 및 분류가 가능하다[8]. (그림 2)는 악성코드의 바이너리를 이미지화하여 추출한 데이터를 각 기법을 통해 이미지로 시각화한 연구이다.



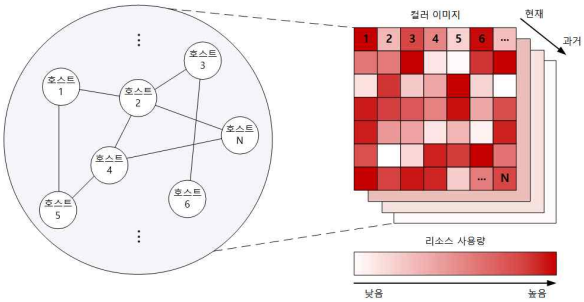
(그림 2) (a) 그레이스케일 이미지, (b) 컬러 이미지

III. 이상징후 탐지를 위한 거시적 관점에서의 데이터 시각화 프레임워크

본 논문에서는 보안 관리자가 대량의 시스템을 관제하는 경우를 고려하여, 이상징후 탐지가 용이하도록 각 시스템의 성능 정보를 시각화하는 프레임워크를 제안한다.

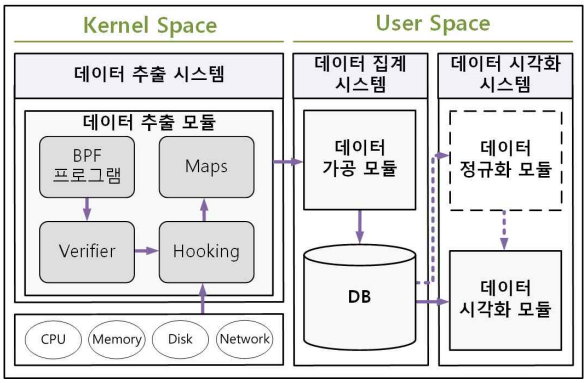
제안하는 프레임워크에서는 시스템 성능 모니터링을 기반으로 컴퓨팅 리소스 소모 패턴을 분석함으로써 이상징후를 탐지한다. 악의적인 행위를 통해 시스템에 피해 발생하는 경우, 공격의 종류에 상관없이 필연적으로 컴퓨팅 리소스를 소모한다는 특징에 따라 거시적인 탐지가 가능하다. 또한, 시계열 그래프 기반 시각화를 활용하여 성능 정보를 실시간으로 기록하고, 컬러 이미지 기반 시각화를 활용하여 이상징후를 실시간으로 판단할 수 있다.

보안 관리자가 관제해야 하는 호스트의 수가 증대되고 복잡도가 증가함에 따라, 제안하는 프레임워크를 통해 미시적 관점에서 호스트를 개별로 모니터링할 수 있을 뿐만 아니라 거시적 관점에서 시스템을 전반적으로 모니터링할 수 있다. 각 호스트의 리소스 소모량을 거시적으로 확인하기 위해 (그림 3)과 같이 호스트별 리소스 소모량을 타일 형태로 시각화한다. 악의적인 행위 혹은 시스템 결함으로 인해 리소스 소모량이 증가할수록 각 호스트의 리소스 소모량을 반영하는 이미지의 채도가 높아진다. 이는 실시간으로 반영되며 지속적으로 갱신된다. 특정 시점에 대한 호스트의 정보를 얻고자 하는 경우, 저장된 이전 기록을 통해 확인이 가능하다.



(그림 3) 거시적 관점에서의 데이터 시각화 개념도

이상징후 탐지를 위한 거시적 관점에서의 데이터 시각화 프레임워크는 데이터 추출 시스템, 데이터 집계 시스템, 데이터 시각화 시스템으로 구성된다. 해당 프레임워크는 (그림 4)와 같다.



(그림 4) 이상징후 탐지를 위한 거시적 관점에서의 데이터 시각화 프레임워크

데이터 추출 시스템에서는 데이터 추출 모듈을 통해 단말로부터 성능 정보를 추출한다. 이 경우에 실시간으로 추적 및 프로파일링이 가능한 BPF(Berkeley Packet Filter)를 사용한다. BPF 프로그램이 컴파일되는 경우, BPF Bytecode가 생성된다. 이는 Verifier에 전달되어 유효성을 검증한 후, 원하는 성능 정보를 추출한다. 추출된 데이터는 Kernel Space와 User Space 간 데이터 공유를 위해 저장 공간을 제공하는 Maps로 수집되고, 이는 데이터가 집계되기 전까지 저장된다.

데이터 집계 시스템에서는 데이터 가공 모듈을 통해 Maps로부터 추출된 성능 정보들을 집계하여 시각화를 위한 적절한 유형으로 가공한 후 데이터베이스에 저장한다.

데이터 시각화 시스템은 데이터 정규화가 필요한 경우와 아닌 경우로 구분되어 활용될 수 있다. 추출된 성능 정보 간 범위가 넓은 경우, 각 메트릭 별로 범위를 조정하기 위해 데이터 정규화 모듈을 이용할 수 있다. 최종적으로 정규화된 데이터를 시계열 그래프 및 컬러 이미지의 형태로 시각화함으로써 보안 관리자가 이상징후를 판단할 수 있다. 해당 프레임워크의 모듈에 대한 설명은 [표 2]와 같다.

[표 2] 이상징후 탐지를 위한 거시적 관점에서의 데이터 시각화 프레임워크 모듈

구분	모듈명	설명
데이터 추출 시스템	데이터 추출 모듈	단말로부터 성능 정보를 추출하는 모듈
데이터 집계 시스템	데이터 가공 모듈	성능 정보 수집 후 특정 형태로 가공하는 모듈
	데이터베이스	성능 정보를 저장해두는 저장소
데이터 시각화 시스템	데이터 정규화 모듈	시각화할 데이터를 정규화하는 모듈
	데이터 시각화 모듈	정규화된 데이터를 시각화하는 모듈

IV. 결론

본 논문에서는 이상징후 탐지를 위해 거시적 관점에서의 데이터 시각화 프레임워크를 제안하였다. 해당 프레임워크는 거시적 관점에서의 데이터 시각화를 위해 데이터 추출 시스템, 데이터 집계 시스템, 데이터 시각화 시스템으로 구분하여 나타내었다. 데이터 추출 모듈에 사용되는 BPF는 실행 중인 시스템의 성능에 미치는 영향을 최소화하며 성능 정보를 추출하기에 적합하다. 수집된 성능 정보들을 시각화하기 위해 데이터 가공 모듈을 통해 적절한 유형으로 가공하고, 데이터베이스에 저장된 정보들을 최종적으로 데이터 시각화 시스템에서 시계열 그래프 및 컬러 이미지로 나타낸다. 시계열 그래프는 대량의 호스트들의 성능 정보 변화를 거시적으로 확인하기에 적합하다. 또한, 컬러 이미지는 실시간으로 시스템 성능 정보를 가시적으로 보여주어 이상징후를 판단하기에 적합하다. 추후 연구로서 프레임워크를 기반으로 보안 모니터링 도구를 설계 및 개발하고자 한다.

[참고문헌]

[1] "[쿠버네티스①] 컨테이너 관리 복잡성 해결하며 부상", http://www.itdaily.kr/news/articleView.html?idxno=212049, 2023년 5월 2일 접속

[2] "구글 • 페이스북 • 넷플릭스 • 에어비앤비의 공통점 '오픈소스'", https://www.oss.kr/oss_guide/show/48ae2da6-7931-4256-9c6b-2c59d6077fe8, 2023년 5월 2일 접속

[3] D. N. Jha, et al, "Holistic Runtime Performance and Security-aware Monitoring in Public Cloud Environment," 2022 22nd IEEE International Symposium on Cluster, Cloud and Internet Computing (CCGrid), Taormina, Italy, pp. 1052-1059, 2022

[4] 황지후, 박기웅, "클라우드 컴퓨팅 인스펙션을 위한 시스템 모니터링 도구 메트릭 원천 분석," 2022년도 한국인터넷정보학회 춘계학술발표대회 논문집, Vol23, No.1, pp. 175-176

[5] 김준섭, "랜섬웨어의 행위-성능적 시각화를 통한 분류 프레임워크 디자인 및 구현." 국내석사학위논문 세종대학교 대학원, 2022.

[6] Poornachandran, P., et al., "Drive-by-download malware detection in hosts by analyzing system resource utilization using one class support vector machines", In Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications, pp. 129-137, 2017

[7] Nataraj, L., Karthikeyan, S., Jacob, G., Manjunath, B. S. "Malware images: visualization and automatic classification", In Proceedings of the 8th international symposium on visualization for cyber security, pp. 1-7, 2011.

[8] Shaid, S. Z. M., & Maarof, M. A. Malware behavior image for malware variant identification. In 2014 International Symposium on Biometrics and Security Technologies (ISBAST), pp. 238-243, 2014.