실시간 파일 복사 유출 방지 시스템에 관한 연구

김호윤* · 김효종 · 이준연 · 신승수

동명대학교

A Study on the Real-Time File Copy Leakage Prevention System

Ho-Yoon Kim* · Hyo-Jong Kim · Jun-Yeon Lee · Seung-Soo Shin

Tongmyong University

E-mail : miask376@gmail.com / rlagywhd019@naver.com / jylee@tu.ac.kr / shinss@tu.ac.kr

요 약

ICT의 발달과 함께 데이터의 양은 증가하고, 저장하고 처리하는 기술은 중요해지고 있다. 본 연구에서는 기업, 공공장소 등에서 중요 데이터 유출 방지를 위하여 실시간 파일 복사 유출 방지 시스템에 관해 연구한다. 연구 방법으로는 데이터 유출 사례와 문제점을 분석한 뒤 데이터 유출 방지를 위해 실시간으로 이벤트를 감지하는 시스템을 제안한다. 파일 유출 방지 시스템은 기존 EDLP 시스템과 비교 분석하며 제안하는 시스템은 부하를 줄이며 이벤트를 탐지한다. 향후 연구로는 네트워크 및 다양한 경로를 통한 유출 방지에 대한 연구가 필요하다.

ABSTRACT

With the development of ICT, the amount of data increases, and the technology of storing and processing becomes important. In this study, we study real-time file copy leakage prevention system to prevent leakage of important data in enterprises, public places, etc. As a research method, we propose a system that detects events in real time to prevent data leakage after analyzing data leakage cases and problems. The file leakage prevention system compares and analyzes with the existing EDLP system, and the proposed system reduces load and detects events. Future research requires research on the prevention of leaks through networks and various channels.

키워드

Data, File, Leakage, Prevention, Hash

Ⅰ. 서 론

ICT가 발전하면서 데이터의 수집, 처리, 가공, 그리고 저장하는 등의 기술은 더욱 중요해지고 있다. 데이터의 양이 증가함에 따라 데이터 관리의 중요성과 필요성이 있으며 기업 또는 공공장소에서의 데이터 유출은 기업 또는 개인의 재산적 피해를 초래할 수 있다. 특히 기업의 경우 회사의 자산과 매출이 직결되는 중요 데이터 정보가 유출되면 피해의 규모가 크기 때문에 데이터 유출 방지를 위한 연구가 필요하다. 데이터 유출은 랜섬웨어, 악성코드 감염, 사용자 인증 우회, 메모리 영역에서의 패스워드 추출 등이 있다. 기업의 경우 내부자, 외부자, 거래 과정 등 다양한 경로의 데이터 유출이 있으나 대부분의 데이터 유출은 내부 직원들에 의해 유출된다. 데이터 유출 차단은 데이터 유출 방지(DLP: Data Loss Prevention) 기술을 이용한다. DLP는 데이터 유출 방지를 위해 이동식 디스크 또는 내·외부 저장소를 통제하는 기술이다[1].

DLP는 EDLP(Endpoint DLP), NDLP(Network DLP) 등으로 분류되며, EDLP는 PC에 통제 에이전트를 설치 후 이벤트를 수집 및 분석하여 악의적인 행위를 판별하고 차단하는 기술이다. NDLP는 네트워크상에서 발생하는 트래픽에 대한 이벤트를 수집하고 통제한다[2]. EDLP는 모든 이벤트

* speaker

및 파일에 대해 데이터를 수집 및 처리하여 병목 현상, 개인 정보 수집, 유지 관리에 대한 많은 리소스가 필요하며, NDLP는 네트워크로 송·수신되는 패킷만 탐지하여 물리적인 탐지는 불가하다[3]. 또한 네트워크의 부하를 발생시키며 외부 네트워크를 이용한 데이터 유출은 차단하지 못한다. 이를 개선하기 위해 본 논문에서는 사전에 중요 파일을 선별한 후 해시 하여 서버에 등록한다. 파일 유출 이벤트 발생 시에는 유출 파일을 먼저 암호화한 뒤 파일에 대한 해시값을 서버에서 판별하여 중요 파일일 경우 삭제 조치하고 일반 파일일 경우 복호화 작업을 진행하여 유출을 방지한다.

II. 동향 분석

2.1 DLP

개인 및 기업의 민감한 데이터의 유출은 재산적 피해는 물론 기업의 경우 이미지 추락까지 이어진다. DLP의 목적은 데이터 유출과 침해가 발생하지 않도록 한다. EDLP의 경우 네트워크 통제 기능을 포함한 DLP도 있지만, 일반적인 플랫폼과 OS에 따라 설치가 불가한 경우가 있다. 또한 Endpoint 유출패턴 업데이트, 로그 통합 관리 등의 애로사항으로 에이전트방식의 EDLP는 네트워크 유출통제가 어렵다. NDLP는 개인정보 유출 방지를 위해 이메일, 클라우드 등 네트워크를 통한 데이터 유출을 방지하기 위해 수집 및 차단의 기능을 한다. 설치와 관리가 쉬우며 사내메일, 클라우드, FTP, P2P를 통해 유출되는 데이터 유출을 방지한다.

2.2 데이터 유출 사례

대표적인 데이터 유출 사례로 2020년 12월 미국 테슬라의 전기차 관련 기밀 파일 유출과 2021년 5월 미국 연방수사국의 국가 기밀 문서 파일 유출 등이 있다[4]. 유형별 사례로는 불법적인 기술탈취, 거래 과정에서 기술유출 등이 있다. 불법적인 기술탈취에서는 내부자(전·현직 임직원 등)에 의해 대부분 데이터 유출이 일어나며, 협력업체 및 경쟁업체와 같은 외부자에 의해서도 데이터 유출이 일어난다. 거래 과정에서의 기술 유출은 불법적인 기술탈취와는 달리 합법적이고 정상적으로 이루어지는 거래 과정을 통해 기술을 유출한다. 하청업체의 거래 과정과 공동연구 과정에서 기술유출이 일어날 수 있으며, 이때의 유출은 일반적인 불법 기술 유출에 비해 기업에서 대응이 번거롭다.

국가에서는 다양한 법률을 제정하여 국내 기술 유출 방지를 위해 노력하고 있다. 국가 기술이 해외로 유출될 시 국가의 안전과 경제 발전에 큰 타격을 일으키므로 지속적인 관리·감독을 통해 유출을 사전에 방지해야 한다.

III. 파일 유출 방지 시스템

3.1 시스템 구성도

제안 시스템은 Management, Client, Server, Database로 구성된다. Management는 관리자가 중요 파일을 지정한 뒤 파일에 대한 해시값을 Database에 등록한다. Client는 PC에서 파일에 대한 이벤트 탐지, 파일 해시값 계산, 해시값 검증 및 결과에 따른 삭제와 복호화 조치를 취한다. Server는 Client로부터 수신받은 파일의 해시값을 기반으로 Database를 통한 중요 파일 여부를 검증한다. Database는 파일의 해시값과 관리자의 계정 정보 등이 저장된다. 시스템의 구성도는 Fig. 1과 같다.

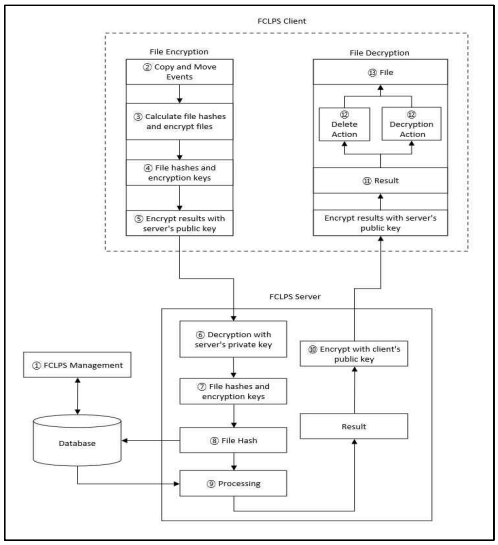

Fig. 1. System Architecture

3.2 시스템 구현 및 분석

통합개발환경은 Windows 10 Pro, CPU i9-9990, RAM 16GB, Ethernet 1000Mbps, C# .Net Framework 4.7.2, MySQL 8.0.24이다.

관리자가 Database에 파일 등록 및 해시값을 저장하기 위해 MySQL 8.0.24를 이용하고 Fig. 2와 같다.

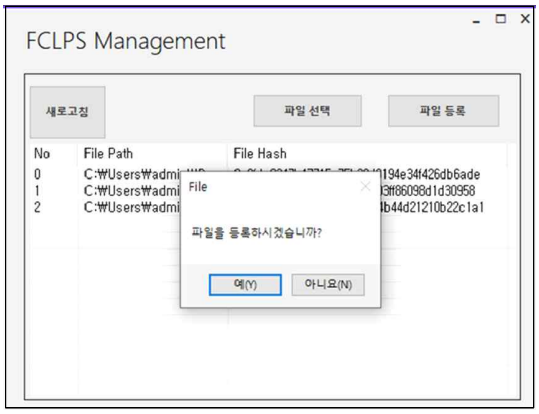

Fig. 2. Management Hash Value Registration

파일 유출 이벤트 발생 시 파일의 해시값을 계산하여 Server로부터 중요 파일의 여부를 검증하고 복호화 및 삭제 조치한다.

국내·외에서 사용 중인 EDLP는 Secure genie사의 Gradius DLP, Symantec사의 Symantec DLP, WaterWall사의 WaterWall DLP 등이 있으며 시스템을 자동 스캔하여 파일 내부 데이터에 특정 키워드나 이미지 속 단어를 분석하여 관련 문자열이 포함되어 있으면 이벤트를 차단한다. 제안하는 시스템은 키워드나 문구를 추출하지 않고 파일 자체의 해시값을 기반으로 서버를 통해 검증 후 판별한다. 이러한 방식은 선택적 이벤트를 이용한 해시값 기반으로 파일 유출을 방지한다. 과도한 리소스 사용으로 인한 시스템 부하를 방지하고 특정 이벤트에 대해서만 모니터링하고 검증한다.

Ⅳ. 결 론

4차 산업혁명 시대가 도래하면서 데이터의 양은 지속적으로 증가하고 있고 관리의 중요성은 커지고 있다. 기업의 데이터 유출은 대부분 내부자를 통해 유출되는데 중요 파일 유출을 방지하기 위해 다양한 DLP 프로그램을 도입한다. 하지만 대부분의 DLP 프로그램은 모든 파일을 일정 시간마다 스캔하고 파일 내에 키워드 또는 이미지 속 키워드가 포함되어 있으면 이벤트를 차단한다. 모든 파일을 분석하고 이벤트를 차단하는 것을 비효율적이므로 관리자에 의해 중요 파일을 사전에 선별하여 관리하는 것이 효율적이다. 이를 개선하기 위해 본 논문에서는 실시간 파일 복사 유출 방지 시스템을 설계하였다. 중요 파일 유출 이벤트에 대한 모니터링을 지속적으로 하여 시스템 부하를 최소화하고 해시값을 사전 등록 후 비교하여 판별하는 것으로 개선하였다. 본 연구에서는 EndPoint를 중점적으로 설계하여 향후 연구로는 Network망을 통한 유출 방지를 위해 패킷을 탐지하는 개선과 다양한 유출 경로를 방지하는 통합솔루션 필요하다.

Acknowledgement

* “본 논문은 부산광역시 및 부산인재평생교육진흥원의 BB21플러스 사업으로 지원된 연구임.”

* “본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학지원 사업의 연구결과로 수행되었음(2018001874004).”

References

[1] J. H. Choi, S. Y. Thew, “Monitoring System of File Outflow through Storage Devices and Printers”, *Journal of the Korea Institute of Information Security & Cryptology,* Vol. 15, No. 4, pp. 51-60, Nov. 2005.

[2] G. J. Shin, G. H. Jung, D. M. Yang, B. H. Lee, “A USB DLP Scheme for Preventing Loss of Internal Confidential Files”, *The Korea Institute of Information and Communication Engineering*, Vol. 21, No. 12, pp 2333-2340, Dec. 2017.

[3] S. J. Yoo, “A Study on DLP System for Preventing Internal Information Leakage”, *Convergence security journal,* Vol. 18, No. 5, pp. 121-126, Dec. 2018.

[4] Insight news. Tesla Confidential Leaked Case [Internet]. Available : https://www.insight.co.kr/news/322291S.