JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 30 No. 10, pp. 135-142, October 2025
https://doi.org/10.9708/jksci.2025.30.10.135

Design and implementation of a fuzzy hash-based insider information leakage prevention system for small and medium-sized enterprises

Dae-Won Kim*, Myung-Ho Kim**

*Ph.D Candidate, Dept. of IT Policy and Management, Soongsil University, Seoul, Korea
**Professor, Dept. of IT Policy and Management, Soongsil University, Seoul, Korea

[Abstract]

With the recent increase in document leakage incidents caused by insiders in the SME environment, there is a growing need for content-based detection technology that can complement the limitations of policy-based security systems. This paper proposes a lightweight detection system that quantitatively evaluates document modification using fuzzy hash-based similarity analysis and detects the possibility of insider leakage in real time. The system is designed to calculate similarity scores for each document by parallel application of the ssdeep, sdhash, and tlsh algorithms, and determine confidentiality based on the maximum value. Experiments were conducted using a total of 1,100 text documents and modified data with various modification rates (10–100%). The results showed that sdhash achieved the best performance across all document modification rate ranges, with an ROC-AUC of 0.99947, a precision of 0.94972, a recall of 0.92016, and an F1-score of 0.93471. These results suggest that the system is effective for real-world security responses in small and medium-sized enterprise environments, and has the potential to be applied to real-time information leakage prevention systems.

▸ **Key words**: Fuzzy hash, Insider leakage, DLP, SME, Document similarity

[요 약]

최근 중소기업 환경에서 내부자에 의한 문서 유출 사고가 증가함에 따라, 정책 기반 보안 시스템의 한계를 보완할 수 있는 내용 기반 탐지 기술의 필요성이 대두되고 있다. 본 논문은 퍼지 해시 기반 유사도 분석을 통해 문서 변형 여부를 정량적으로 평가하고, 내부자 유출 가능성을 실시간으로 탐지할 수 있는 경량 탐지 시스템을 제안한다. ssdeep, sdhash, tlsh 알고리즘을 병렬 적용하여 각 문서의 유사도 점수를 산출하고, 최댓값을 기준으로 기밀성 여부를 판별하는 방식으로 설계하였다. 실험은 총 1,100개의 텍스트 문서와 다양한 수정률(10~100%)에 따른 변형 데이터를 기반으로 수행되었으며, 그 결과 sdhash는 전체 문서 변형률 구간에서 ROC-AUC 0.99947, Precision 0.94972, Recall 0.92016, F1-score 0.93471로 가장 우수한 성능을 기록하였다. 이러한 결과는 본 시스템이 중소기업 환경에서의 실제 보안 대응에 유효하며, 실시간 정보 유출 방지 시스템에 적용 가능성을 갖는 설계임을 시사한다.

▸ **주제어**: 퍼지 해시, 내부자 유출, 데이터 유출 방지, 중소기업, 문서 유사도

• First Author: Dae-Won Kim, Corresponding Author: Myung-Ho Kim
*Dae-Won Kim (dwkim401@soongsil.ac.kr), Dept. of IT Policy and Management, Soongsil University
**Myung-Ho Kim (kmh@ssu.ac.kr), Dept. of IT Policy and Management, Soongsil University
• Received: 2025. 07. 22, Revised: 2025. 09. 24, Accepted: 2025. 09. 25.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

I. Introduction

최근 데이터 보안 환경은 외부 침입에 대한 전통적인 방어 체계를 넘어서, 내부자에 의한 기밀 정보 유출 문제로 초점이 이동하고 있다[1]. 특히 중소기업(SME) 환경에서는 보안 예산과 전담 인력이 부족한 구조적 특성상, 권한을 가진 내부자에 의해 발생하는 문서 유출에 대응하기 어려운 실정이다[2]. 중소기업의 유출 사례는, 정상 권한을 활용한 파일 반출이 탐지되지 못한 채 장기간 지속된 점에서 기술 기반 보안 대응의 한계를 드러낸다[3].

이러한 내부 위협에 대응하기 위해 다양한 기술적 접근이 시도됐다. 대표적으로 DLP(Data Loss Prevention) 시스템은 사전 정의된 정책에 따라 특정 조건을 만족하는 행위를 차단하거나 경고하는 방식으로 동작하며, UEBA(User and Entity Behavior Analytics) 기술은 비정상 행위를 기계 학습 기반으로 탐지하는 시도를 보여주고 있다[4][5]. 그러나 이들 기술은 높은 초기 구축 비용과 오탐률 문제, 무엇보다 문서 내용 기반 비교가 어려워 유사 문서 변경을 통한 유출 시나리오를 효과적으로 탐지하지 못한다는 한계를 가진다[6].

최근 퍼지 해시가 저장된 데이터뿐 아니라 이동 중인 네트워크 트래픽에도 적용할 수 있어, 다양한 네트워크 보안 장치에서 실시간 탐지를 지원할 수 있음을 보여주었다[7][8]. 이러한 점은 퍼지 해시가 단순히 사후 분석 도구에 그치지 않고, 내부 정보 유출 방지를 위한 실시간·준실시간 대응 체계에도 충분히 통합될 수 있음을 시사한다.

따라서 본 논문에서는 문서 자체의 내용 유사성을 정량적으로 분석함으로써 내부자에 의한 변형된 문서 유출을 효과적으로 탐지할 수 있는 퍼지 해시 기반 유출 탐지 시스템을 제안한다. 제안하는 시스템은 다양한 수정 비율과 위치에 대해 견고한 유사도 점수를 산출할 수 있도록 퍼지 해시 기반 알고리즘을 탐지 구조에 통합하여, 기존 보안 시스템이 갖는 탐지 불감 구간을 보완하는 것을 목표로 한다. 특히, 퍼지 해시 기법의 성능을 재검증하는 데 목적이 있는 것이 아니라, 실제 내부자 유출 시나리오에 적용했을 때 얼마나 효과적으로 탐지할 수 있는지를 실험적으로 입증하는 데 초점을 맞추었다.

본 논문의 핵심은 다음과 같다. 첫째, ssdeep, sdhash, tlsh 세 가지 퍼지 해시 알고리즘을 병렬적으로 적용하여 문서 간 유사도를 계산하고, 둘째, 최고 유사도를 기준으로 문서의 기밀성 점수를 산출한 후, 사전 정의된 임곗값에 따라 유출 의심 여부를 분류한다. 이 과정을 통해 구조적 변경이나 사소한 수정에도 민감하게 반응할 수 있는 경량 보안 시스템을 구현한다.

논문은 다음과 같이 구성된다. 제2장에서는 관련 기술과 선행 연구를 분석하고, 제3장에서는 제안하는 시스템의 구조를 설명한다. 제4장에서는 실험 설계와 성능 평가 결과를 제시하며, 제5장에서는 결론과 향후 연구 방향을 논의한다.

II. Preliminaries

최근 정보보안 환경에서는 외부 공격보다 내부자에 의한 정보 유출이 더욱 정교하고 심각한 위협으로 부각되고 있다. 내부자 위협은 조직 내부의 권한을 가진 사용자가 고의적 혹은 비고의적으로 민감 데이터를 유출하는 행위로 정의되며, 대표적인 유출 채널로는 이메일 첨부파일, 클라우드 저장소, 협업 도구를 통한 데이터 전송이 있으며, 이는 정상적인 업무 흐름을 가장해 탐지를 어렵게 만든다. 또한 USB, 외장 하드, 메모리 카드 등을 통한 휴대용 매체 반출, 프린터 출력이나 복사·붙여넣기를 통한 하드카피 유출도 주요 경로로 지적된다[4]. 이러한 위협은 기존의 외부 중심 보안 체계로는 효과적인 탐지가 어렵다는 특성을 가진다[9]. 이를 대응하기 위한 기술로는 데이터 유출 방지(DLP, Data Loss Prevention) 시스템과 사용자 및 객체 행위 분석(UEBA, User and Entity Behavior Analytics) 기반의 접근 방식이 주로 활용되고 있다. 이러한 탐지 기술은 내부자 위협의 다양성과 은폐 가능성을 고려한 탐지 로직의 필요성에 의해 등장하였다.

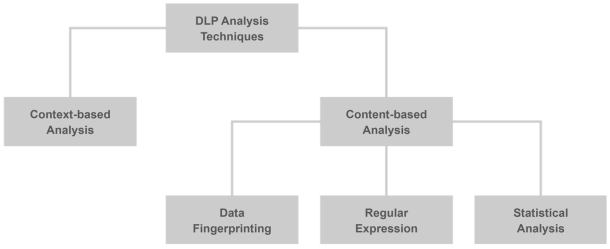

Fig. 1. Classification of DLP analysis techniques

DLP 시스템은 Fig. 1과 같이 컨텍스트 기반 분석과 콘텐츠 기반 분석으로 구분된다. 컨텍스트 기반 분석은 파일 속성이나 사용자 권한 등 메타데이터 기반 정책을 적용하고, 콘텐츠 기반 분석은 데이터 내용을 직접 분석한다. 콘텐츠 기반 분석에는 데이터 지문, 정규 표현식, 통계 분석 기법이 포함되며, 이들은 사전 정의된 정책에 따라 민감 정보 유출을 차단한다. UEBA는 통계 분석에 속하는 방식

으로, 기계 학습을 활용해 사용자 행위의 이상 패턴을 분석하고 비정상적인 접근 시도를 식별한다[4]. 반면, 시그니처 기반의 정적 탐지 방식은 정규 표현식 기법에 해당하며, 기존 공격 유형의 명세를 기반으로 신속한 탐지가 가능하나 새로운 유출 방식에는 취약하다[10]. 이러한 기존 기법들은 구축 및 운영의 복잡성, 과도한 오탐률, 실시간 탐지 어려움 등의 한계를 가진다. 특히 중소기업 환경에서는 보안 인력과 예산의 제약으로 인해 복잡한 보안 시스템의 도입 및 운영이 어렵다[11].

퍼지 해시(Fuzzy hash)는 문서 간 내용 유사도를 정량적으로 비교하는 해시 기반 기술로, 내부자에 의한 콘텐츠 변경을 감지할 가능성으로 인해 보안 탐지 분야에서 주목받고 있다. 대표적인 퍼지 해시 알고리즘인 ssdeep은 가변 길이 블록 단위의 해시를 생성하여 부분 문자열 간 유사도를 측정할 수 있으며[12], sdhash는 정보량이 많은 블록을 이용해 의미 기반의 콘텐츠 유사성을 비교한다[13]. tlsh은 Locality Sensitive Hashing 기반 알고리즘으로 문서 전체의 통계적 분포를 반영한 해시값을 생성하며, 대규모 문서 집합에 대한 경량 비교에 적합한 구조를 가진다[14]. 하지만 ssdeep은 문장 재배열과 같은 구조적 변형에 취약하며, sdhash와 tlsh는 유사도 점수의 해석 직관성이 낮다는 한계가 있다.

퍼지 해시 기반 기술은 디지털 포렌식, 악성코드 유사도 탐지, 문서 변경 추적 등 다양한 보안 응용 분야에서 활용됐다. 특히 sdhash와 tlsh는 미국 NIST의 Digital Forensics Tool Testing 프로그램에서 정밀도(Precision)와 재현율(Recall) 측면에서 우수한 평가를 받은 바 있다[15]. 최근에는 이러한 해시 알고리즘의 콘텐츠 유사도 판별 기능을 사용자 이상 행위 기반 탐지 프레임워크(UEBA)와 연계하거나, 보안 정보 및 이벤트 관리 시스템(SIEM) 내 유사도 기반 필터링에 활용하려는 시도가 일부 연구에서 이루어지고 있다[16]. 그러나 대부분의 연구는 정적 환경에서의 유사도 비교에 집중되어 있으며, 문서 구조의 변환이나 수정 비율 변화에 따른 민감도 측정은 상대적으로 부족하다.

선행 연구들은 퍼지 해시 기반 탐지 기술의 성능을 다양한 벤치마크 문서 세트를 통해 평가하였으나, 주로 단일 해시 알고리즘을 기반으로 하며, 포맷 다양성과 변형 시나리오를 충분히 반영하지 못했다는 한계가 있다[17]. 또한 실무 적용성을 고려한 실시간 탐지 구조에 대한 고려가 부족하며, 특히 중소기업의 보안 환경을 반영한 경량 설계와 운영 조건에 대한 실험적 검증이 미흡하다. 본 논문은 이러한 공백을 해소하기 위해 ssdeep, sdhash, tlsh 세 가지 퍼지 해시 알고리즘을 병렬적으로 적용하고, 변형 문서 데이터 세트 기반의 정량적 실험을 통해 각 알고리즘의 민감도와 성능 차이를 비교 분석한다. 이를 통해 기존 DLP의 콘텐츠 기반 분석을 보완하며, 실시간 유출 탐지 적용 가능성과 알고리즘별 유사도 구간의 특성 차이를 실증적으로 규명하고자 한다.

III. The Proposed Scheme

본 논문에서 제안하는 탐지 시스템은 중소기업 환경에서의 실시간 기밀문서 유출 감지를 목적으로 설계되었다. 중소기업은 전담 보안 인력이 부족하고, 대부분의 업무 문서가 텍스트 기반이며, 문서 유통 과정에 대한 통제 수단이 제한적인 특징을 갖는다. 따라서 본 시스템은 최소한의 리소스로 작동하면서도 콘텐츠 기반 유사도 탐지를 통해 내부자 유출 시나리오에 효과적으로 대응할 수 있어야 한다. 이 시스템은 문서의 콘텐츠 유사성을 정량화할 수 있는 퍼지 해시 알고리즘을 기반으로 구성되며, 사전에 등록된 기밀문서와 신규 유입 문서 간의 유사도를 비교하여 유출 가능성을 판정한다.

전체 시스템은 문서 등록 → 해시 색인화 → 문서 수집 → 유사도 산출 → 기밀성 판정 → 경고/기록의 5단계 흐름으로 구성된다. Fig. 2은 이 전체 흐름을 시각화한 시스템 흐름도이다. 먼저 기밀로 분류된 문서는 색인화 단계에서 퍼지 해시 지문이 생성되어 색인 DB에 저장된다. 이후 실시간으로 감시되는 문서는 동일한 방식으로 해시가 생성되어 색인 DB의 기존 기밀문서들과 비교되며, 각 해시 알고리즘의 유사도 점수가 산출된다. 기밀성 점수는 이 중 가장 높은 유사도를 기준으로 결정되며, 사전 정의된 임곗값을 초과하면 해당 문서는 유출 의심 문서로 분류된다. 탐지 결과는 실시간으로 기록되며, 관리자는 알림 시스템을 통해 즉시 통보받는다.

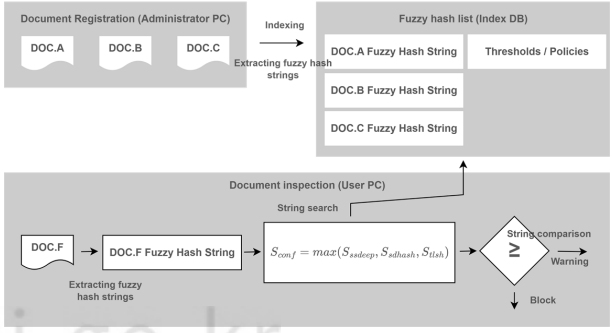

Fig. 2. Proposed Scheme Processing Flow

제안하는 시스템은 세 가지 퍼지 해시 알고리즘(ssdeep, sdhash, tlsh)을 병렬적으로 적용하여 유사도 분석의 다각화와 탐지 민감도의 향상을 도모한다. 각 알고리즘은 서로 다른 기반 기술에 의해 유사도 점수를 산출하므로, 특정 문서 유형이나 수정 방식에 따라 보완적인 효과를 제공할 수 있다. 예를 들어, ssdeep은 부분 문자열 일치를 통한 블록 기반 유사도 계산에 강점이 있으며, sdhash는 정보량(엔트로피) 기반 특징 추출을 통해 구조 변경에도 높은 일관성을 유지한다. tlsh는 로컬 민감 해시(Locality-Sensitive Hashing)를 이용하여 전체 문서의 통계적 특성을 반영한 유사도 측정을 수행한다. 이는 특정 알고리즘의 편향된 결과에 의존하지 않고 보수적인 기준으로 유출 가능성을 탐지할 수 있다.

최종 기밀성 점수 S_{conf}는 아래와 같이 정의된다:

$$S_{conf} = \max \quad (S_{ssdeep}, S_{sdhash}, S_{tlsh})$$

이러한 정의는 특정 알고리즘의 성능 편향 없이 가장 보수적인 기준으로 유출 가능성을 탐지할 수 있도록 설계되었다.

기밀성 점수를 기반으로 문서의 유출 여부를 판정하기 위해 임곗값 θ를 설정한다. 본 시스템은 사전 실험을 통해 도출된 최적의 탐지 임곗값을 전역적으로 적용하는 고정 임곗값 방식을 지원한다. 시스템 관리자는 보안 민감도에 따라 임곗값을 조정할 수 있으며, 이 값은 탐지 정책 구성 파일을 통해 시스템에 적용된다.

예를 들어, 관리자는 유출을 방지하고 싶은 기밀문서 A, B, C를 등록한다. 등록된 문서는 ssdeep, sdhash, tlsh 세 가지 퍼지 해시값을 추출하여 DB에 저장된다. 이후 사용자가 신규 문서 F를 작성하여 반출을 시도한다고 가정하자. 시스템은 이 문서에 대해 동일하게 세 가지 해시값을 생성한 뒤, 기등록된 기밀문서들과 각각 비교하여 유사도 점수를 산출한다. 만약 세 알고리즘 중 하나라도 유사도 점수가 지정된 임곗값 이상일 경우, 해당 문서는 기밀성 점수 조건을 만족하는 것으로 간주되어 유출 의심 문서로 분류된다. 이 탐지 결과는 로그로 저장됨과 동시에 관리자에게 알림 메시지로 전달되며, 사용자의 행위 로그 또한 연동 시스템에 기록된다.

기존 DLP 시스템은 규칙 기반 탐지에 의존하여 변형된 콘텐츠나 우회된 표현에 대해 탐지가 불가능한 경우가 빈번하였다. 반면, 제안 시스템은 문서의 내용 유사도를 직접 비교함으로써 오탈자, 순서 변경, 요약 등의 변형에도 일정 수준의 탐지가 가능하다. 특히 세 가지 퍼지 해시 알고리즘의 병렬 적용 구조는 개별 알고리즘의 약점을 상호 보완하여 전체적인 탐지 정확도를 향상하는데 기여한다.

IV. Evaluation

텍스트 기반 유사 문서 탐지를 위한 실험에는 Digital Corpora 프로젝트에서 구축한 공개 데이터셋인 Govdocs1을 사용하였다. Govdocs1은 주로 미국 정부 도메인(.gov)에서 수집된 약 986,278개의 파일로 구성되어 있으며, 문서, 이미지, 압축 파일 등 다양한 포맷을 포함한다. 구축 목적은 디지털 포렌식 연구와 알고리즘 평가를 위한 표준화된 데이터 제공에 있다. 파일 포맷의 분포를 살펴보면 PDF, HTML, 이미지(JPEG/GIF 등)가 다수를 차지하며, DOC, PPT, XLS, TXT 등의 오피스 문서도 일정 비율을 차지한다. 실험에서는 포맷 구조나 압축과 같은 외부 요인을 배제하고, 순수 텍스트 기반의 유사도 분석을 일관되게 수행하기 위해 .txt 형식 파일만을 실험 대상으로 제한하였다.

원본 데이터로는 총 20개의 텍스트 문서를 선정하였다. 이들은 의미론적으로 정책 지침, 기술 보고서, 산업 분석 자료, 뉴스·칼럼, 학술 논문 단락 등으로 다양하게 구성되어 있으며, 문서 길이 또한 수천 자에서 수만 자에 이르는 등 분포가 고르게 나타난다. 이러한 구성을 통해 실제 중소기업 환경에서 유통되는 메일, 보고서, 내부 문서의 다양한 양식을 포괄할 수 있도록 하였다.

실험의 목적은 퍼지 해시 기반 내부자 유출 탐지 시스템의 성능을 다양한 문서 수정 조건에서 정량적으로 평가하는 데 있다. 내부자는 민감 정보의 무단 반출을 은폐하기 위해 문서의 주요 구조를 유지하면서도 특정 구간의 내용을 치환하거나 일부 문자열을 체계적으로 변형하는 경우를 모사하기 위해 선정된 문서마다 크기를 기준으로 10%에서 100%까지 10% 단위로 수정률을 달리한 변형 문서를 생성하였다. 각 수정률에 대해 문서의 시작 위치를 0%로 하여, 수정률에 따라 가능한 모든 위치에서 일정 구간을 선택하여 수정하였고, 이때 수정 구간은 문서 크기를 벗어나지 않도록 설정하였다. 문서 수정은 기호를 제외한 숫자와 영문자를 각각 ASCII 코드상 7만큼 순환 시프트 하는 방식으로 수행하였으며(예: 0 → 7, a → h 등), 이러한 방식으로 원본을 제외한 총 1,080개의 변형 문서를 생성하였다. 실험에는 ssdeep, sdhash, tlsh 알고리즘을 적용하여 동일한 조건에서 성능을 비교하였으며, 총 20개의 원본 문서를 각각 1,100개의 문서(원본 + 변형)와 비교하여 총 22,000회의 교차 실험을 수행하였다.

비교 조건을 명확히 정의하기 위하여, 원본 문서와 동일한 문서 간 비교는 모두 동일 문서로 간주하여 "유출 의심"으로 처리하였다. 또한, 원본 대비 수정율이 30% 미만인 변형 문서 역시 원본과의 본질적 유사성이 유지된다고 판단하여 동일 문서로 분류하였다. 반면, 수정율이 30% 이상인 경우에는 구조적·내용적 차이가 뚜렷하여 새로운 문서로 간주하였다.

원본과 변형 문서 간의 유사도 판별을 위해 퍼지 해시 알고리즘(ssdeep, sdhash, tlsh)으로 산출된 점수(0-100)를 사전에 정의한 임계값(threshold, 0-100)과 비교하였다. 점수가 임계값 이상인 경우를 "유출 의심(positive)", 그 미만인 경우를 "정상(negative)"으로 레이블링하였으며, 임계값 전 구간을 변화시키며 탐지 성능을 분석하였다. 이러한 설정은 특정 단일 기준에 의존하지 않고, 임계값 변화에 따른 정밀도·재현율의 상호 균형과 알고리즘의 분류 특성을 전반적으로 평가하기 위함이다.

모든 실험은 동일한 환경에서 수행되었다. Table 1은 실험에 사용된 컴퓨터의 하드웨어 및 소프트웨어 사양을 요약한 것이다.

Table 1. System Environment

Category	Specification
Processor	Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz 1.99 GHz
Installed RAM	16.0GB (15.8GB usable)
Storage	477 GB SSD SPCC M.2 SSD
Graphics Card	Intel(R) UHD Graphics 620 (128 MB)
System Type	Windows 10 64-bit

실험에 소요된 시간은 각 퍼지 해시 알고리즘의 실제 적용 효율성을 정량적으로 평가할 수 있는 중요한 지표로 해석할 수 있다. 모든 실험은 각 원본 문서별로 프로그램을 실행하여 측정하였으며, 측정 시간에는 해시 문자열 간 유사도 점수 계산 및 프로그램 실행 시간이 포함된다. 알고리즘별로 지원하는 처리 옵션에 차이가 있었으며, 특히 tlsh의 경우 대량 처리(batch processing) 옵션을 적용하여 실험을 진행하였다. 실험 결과, ssdeep은 평균 5.74612초, sdhash는 10.41323초, tlsh는 0.33319초의 실행 시간이 소요되어 알고리즘별 처리 속도에 뚜렷한 차이가 있음을 확인할 수 있었다. 이러한 결과는 동일한 조건에서 다수의 문서 비교 작업이 요구되는 환경에서 tlsh가 상대적으로 빠른 처리 성능을 제공함을 시사한다. 특히, 문서의 해시 생성과 유사도 계산이 모두 포함된 점을 감안할 때, 전체 데이터 처리 파이프라인에서의 실제 응답 시간 및 시스템 부하에 미치는 영향을 직관적으로 파악할 수 있다. 나아가, 내부자가 이메일, 클라우드, 휴대용 매체 등 다양한 채널을 통해 원본 문서를 변형·반출하는 경우, 변형된 문서가 내부 기밀 원본과 일정 수준 이상의 유사도를 보일 때 이를 "유출 징후"로 판별할 수 있음을 실험적으로 확인하였다. 즉, 알고리즘별 탐지 성능 차이는 단순한 속도 비교를 넘어, 실제 내부 정보 유출 방지 시스템에서 탐지 신뢰성과 대응 시간에 직접적으로 연결될 수 있다.

성능 평가는 대표적인 이진 분류 기반 지표인 정밀도(Precision), 재현율(Recall), F1-score, 그리고 ROC-AUC 및 PR-AUC를 기준으로 진행되었다. 각 지표는 다음과 같이 정의된다:

- $\mathrm{Pr}ecision = \dfrac{TP}{(TP + FP)}$
- $Recall = \dfrac{TP}{(TP + FN)}$
- $F1\,Score = 2 \times \dfrac{(\mathrm{Pr}ecision \times Recall)}{(\mathrm{Pr}ecision + Recall)}$
- $ROC - AUC = \int_0^1 ROC\,Curve\,(fpr, tpr)\,dx$
- $PR - AUC = \int_0^1 PR\,Curve\,(recall, precision)\,dx$

정밀도(Precision)는 실제로 양성이라고 예측한 사례 중에서 실제 양성인 비율을 의미하고, 재현율(Recall)은 실제 양성인 사례 중에서 올바르게 양성으로 예측된 비율을 나타낸다. F1 점수(F1 Score)는 정밀도와 재현율의 조화 평균으로 계산되며, 두 지표 간의 균형을 평가하는 데 사용된다. 또한, 분류기의 임곗값 변화에 따른 종합적인 성능 평가를 위해 ROC 곡선 아래 면적(ROC-AUC)과 PR 곡선 아래 면적(PR-AUC)도 활용한다. 이들 지표는 각각의 곡선에서 얻은 면적으로, 모델의 전반적인 분류 성능을 다양한 관점에서 정량적으로 평가하는 데 중요한 역할을 한다.

본 실험에서 다양한 임곗값(0~100)을 변환시키며 각 유사도 점수 기반의 이진 분류 결과를 계산하였고, 이를 통해 ROC Curve 및 Precision-Recall Curve를 도출하였다. Table 2는 ssdeep, sdhash, tlsh 각각에 대해 F1-score가 최대가 되는 임곗값에서의 주요 분류 지표를 정리한 결과이다. sdhash는 Precision 0.94972, Recall 0.92016, F1-score 0.93471로 가장 높은 성능을 보여주었으며, 이는 변형 문서 환경에서도 오탐과 누락을 최소화할 수 있는 구조적 강점을 시사한다. ssdeep과 tlsh 역시 F1-score 0.85201, 0.88951로, 알고리즘 선택 및 임곗값 조정에 따라 다양한 현장 요건에 대응할 수 있음을 보여준

다. 또한, 각 알고리즘의 F1 최적 임곗값(ssdeep:θ=61. sdhash:θ=62, tlsh:θ=98)에서 측정한 Accuracy는 각각 0.99176, 0.99484, 0.99391로 확인되어, 세 방법 모두 0.99 내외의 높은 전체 정답률을 보였다.

Table 2. Precision, Recall, F1-score, Accuracy by Fuzzy Hash Algorithm ($\theta = threshold$)

	ssdeep ($\theta = 61$)	sdhash ($\theta = 62$)	tlsh ($\theta = 98$)
Precision	0.87129	0.94972	0.92041
Recall	0.83356	0.92016	0.86062
F1-Score	0.85201	0.93471	0.88951
Accuracy	0.99176	0.99484	0.99391

Table 3. Precision, Recall, F1-score, Accuracy by ssdeep, sdhash, and tlsh applied in parallel ($\theta = threshold$)

	Precision	Recall	F1-Score	Accuracy
S_{conf} ($\theta = 67$)	0.91885	0.94993	0.93413	0.99619

Table 3은 ssdeep, sdhash, tllsh를 병렬 적용 시 F1-Score가 최대가 되는 임곗값에서의 주요 분류 지표를 정리한 결과이다. 병렬 적용 시 ssdeep 및 tlsh 대비 F1이 각각 0.08212, 0.04462 절대 향상되었고, Recall 역시 크게 우세하다. sdhash와의 비교에서는 F1이 사실상 동등 수준(-0.00058)이지만, Recall이 병렬 적용에서 더 높아(+0.02977) 놓침(FN)을 의미 있게 감소시킨다. 반대로 Precision은 sdhash와 tlsh가 병렬 적용보다 높아, 병렬 적용은 Recall 향상을 위해 Precision를 소폭 희생하는(trad-off) 특성을 보인다. 따라서 놓침 비용이 큰 불균형 탐지 환경에서는 병렬 적용이 실무적으로 우선될 수 있다.

Fig. 3는 ssdeep, sdhash, tlsh 알고리즘에 대해 임곗값을 0부터 100까지 변화시키며 측정한 ROC 곡선을 나타낸다. 실험 결과, ssdeep의 ROC-AUC는 0.94978로 나타나, 전체 구간에서 높은 분류 성능을 보였다. sdhash와 tlsh 알고리즘은 각각 0.999947, 0.99822의 ROC-AUC를 기록하여, ssdeep 대비 더욱 뛰어난 구분력을 나타냈다. 이는 두 알고리즘이 유사 쌍과 비유사 쌍을 거의 완벽하게 구별하는 분류 능력을 보유하고 있음을 시사한다.

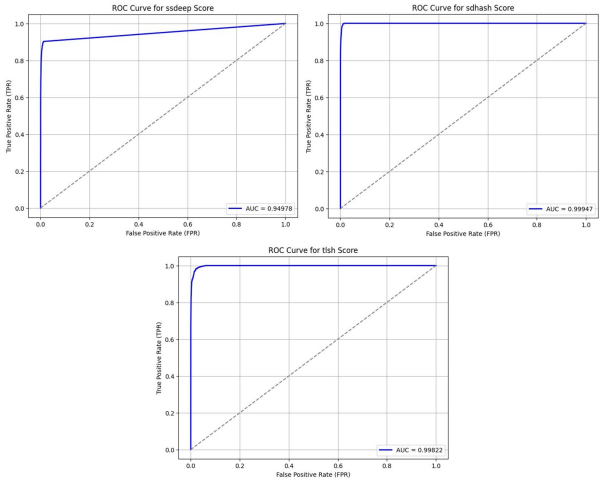

Fig. 3. ROC Curve by Algorithm

한편, Precision-Recall Curve의 AUC(Area Under Curve) 및 AP(Average Precision) 값 또한 Table 3에 요약하였다. ssdeep은 PR-AUC 0.90141, AP 0.86777로 임곗값 변화 전 구간에서 우수한 정밀도-재현율 균형을 보여주었다. sdhash의 경우 RP-AUC 0.98791, AP 0.98700을 기록하며, 실험에 사용된 문서 유사성 판별 상황에서 가장 안정적인 탐지 성능을 확인할 수 있었다. tlsh 역시 PR-AUC 0.95790, AP 0.95741을 보였으며, 임곗값 변화에 대한 성능 변동이 적고, 비교적 높은 정밀도를 유지하였다.

Table 3. Precision, Recall, and F1 Score by Fuzzy Hash Algorithm

	ssdeep	sdhash	tlsh
ROC-AUC	0.94978	0.99947	0.99822
PR-AUC	0.90141	0.98791	0.95790
AP	0.86777	0.98700	0.95741

알고리즘별 비교 분석 결과, sdhash는 정보량 기반 특징 추출 방식의 장점을 바탕으로 전체 임곗값 구간에서 가장 견고한 분류 성능을 기록하였다. 특히, PR-AUC 및 AP 지표 모두에서 0.98 이상의 값을 보임으로써, 텍스트 변형이 이루어진 다양한 환경에서도 유사 문서 탐지에 있어 높은 신뢰성을 제공함을 알 수 있다. ssdeep은 블록 기반 해시 비교 방식의 한계로 인해 일부 임계 구간에서 정밀도 및 재현율이 미세하게 저하되는 경향이 관찰되었다. tlsh는 유사도 점수 해석의 직관성 상대적으로 낮으며, 임곗값 변화에 따른 성능 민감도가 높아 운영 환경에서 임곗값 설정에 주의가 요구된다.

본 시스템은 다양한 퍼지 해시 알고리즘의 병렬 적용을 통해 문서 유사도 검출의 신뢰성과 탐지율을 극대화할 수 있음을 실험적으로 입증하였다. 특히, sdhash 기반 병렬 구조는 문서 내용 변형과 같은 실 환경 시나리오에 대해 일관성 있는 탐지 성능을 제공하였다. 다만, 본 실험은 텍스트 기반 문서에 한정되어 있으며, pdf, 이미지, 복합 문서 등 비정형 데이터에 대한 일반화 가능성에는 한계가 존재한다. 또한 임곗값의 설정에 따라 탐지 성능이 민감하게 변화하므로, 실제 적용 시에는 동적 임곗값 조정, 다단계 탐지 등의 보완 방안이 필요하다. 향후 연구에서는 문서 유형 다변화, 임곗값 자동화 등을 통해 실무 적용성을 더욱 제고할 계획이다.

V. Conclusions

본 논문은 중소기업 환경에서 발생할 수 있는 내부자에 의한 문서 유출 시나리오에 효과적으로 대응하기 위한 경량 탐지 시스템으로서, 퍼지 해시 기반 문서 유사도 판별 방식을 제안하였다. ssdeep, sdhash, tlsh 세 가지 알고리즘을 병렬 적용하여 각 문서 간 유사도를 측정하고, 최대 유사도를 기준으로 문서의 기밀성 점수를 산출함으로써, 정량적인 유사성 기반 탐지를 가능하게 하였다.

실험 결과에 따르면, sdhash는 임곗값 62에서 F1-score 0.93471 및 ROC-AUC 0.99947, tlsh는 F1-score 0.88951 및 ROC-AUC 0.99822을 기록하였다. ssdeep은 F1-score 0.85201 및 ROC-AUC 0.94978로 측정되었다. 이 결과는 퍼지 해시 기반 탐지 시스템이 실제 환경에서 높은 정확도와 안정성을 갖는다는 점을 뒷받침한다. 또한, 클래스 불균형 상황에서도 sdhash와 tlsh는 각각 PR-AUC 0.98791, 0.95790을 나타내어 데이터 분포 변화와 관계없이 탐지 신뢰도를 유지함을 확인하였다. 아울러 세 알고리즘의 병렬 적용은 임곗값 67에서 Precision 0.91885, Recall 0.94993, F1 0.93413을 기록하였고, ssdeep(θ=61) 및 tlsh(θ=98) 대비 F1이 각각 0.08212, 0.04462 향상되었다. sdhash(θ=62)와는 F1이 사실상 동등하나 Recall이 0.02977 높아 놓침(FN) 감소 관점에서 유리함을 시사한다.

이러한 결과는 본 연구의 세 가지 의의를 뒷받침한다. 첫째, 복잡한 정책 기반 DLP 시스템과 달리 단순한 퍼지 해시 기반 구조만으로도 탐지 정확도를 0.9 이상 유지할 수 있음을 보여줌으로써, 경량화된 DLP 체계의 가능성을 제시하였다. 둘째, 세 알고리즘을 비교 적용하여 환경별 최적 임곗값을 실험적으로 도출할 수 있음을 증명하였으며, 이를 통해 다양한 문서 수정 시나리오에서도 견고한 탐지가 가능함을 보였다. 셋째, 실제 유출 환경을 모의한 실험에서 sdhash와 tlsh가 높은 PR-AUC 값을 달성함으로써, 본 시스템이 실무 환경에 적용 가능한 신뢰성 있는 기밀 문서 탐지 체계임을 입증하였다.

다만, 본 연구는 텍스트 파일에 한정된 실험 설계로 인해 이미지 기반 또는 pdf 포맷과 같은 비정형 문서 유형에 대한 적용성에 제약이 존재한다. 또한 임곗값 설정에 따른 탐지 민감도는 환경에 따라 달라질 수 있으므로, 실시간 탐지를 위한 동적 임곗값 조정 알고리즘 개발이 향후 과제로 남아 있다. 앞으로는 문서 포맷 다양성 확보, 실시간 처리 능력 향상, 사용자 행위 기반 탐지 기능과의 융합을 통해 더욱 확장 가능하고 정밀한 내부자 유출 탐지 시스템으로 발전시킬 계획이다.

REFERENCES

[1] N. Saxena, R. Gajrani, M. Conti, and K. Salah, "Impact and key challenges of insider threats on organizations and critical businesses," *Electronics*, Vol. 9, No. 9, pp. 1460, Sep. 2020. DOI: 10.3390/electronics9091460

[2] A. K. Tetteh, "Cybersecurity needs for SMEs," *Issues in Information Systems*, Vol. 25, No. 1, 2024. DOI: 10.48009/1_iis_2024_120

[3] A. Moneva and R. Leukfeldt, "Insider threats among Dutch SMEs: Nature and extent of incidents, and cyber security measures," *Journal of Criminology*, Vol. 56, No. 4, pp. 416-440, 2023. DOI: 10.1177/26338076231161842

[4] S. Alneyadi, E. Sithirasenan, and V. Muthukkumarasamy, "A survey on data leakage prevention systems," *Journal of Network and Computer Applications*, Vol. 62, pp. 137-152, 2016. DOI: 10.1016/j.jnca.2016.01.008

[5] R. Yousef and M. Jazzar, "Measuring the effectiveness of user and entity behavior analytics for the prevention of insider threats," *J. Xi'an Univ. Arch. Technol*, Vol. 13, pp. 175-181, 2021. DOI: 10.37896/JXAT13.10/313918

[6] A. Ali, M. Husain, and P. Hans, "Real-Time Detection of Insider Threats Using Behavioral Analytics and Deep Evidential Clustering," *arXiv preprint arXiv:2505.15383*, 2025. DOI: 10.48550/arXiv.2505.15383 DOI: 10.1016/j.cose.2021.102221

[7] T. Göbel, F. Uhlig, and H. Baier, "Evaluation of network traffic analysis using approximate matching algorithms," *IFIP International Conference on Digital Forensics*, Cham: Springer International Publishing, pp. 89-108, 2021. DOI: 10.1007/978-3-030-88381-2_5

[8] W. Tatum, et al., "Anomaly detection in ICS networks with fuzzy hashing," *2024 Cyber Awareness and Research Symposium (CARS)*, IEEE, pp. 1-5, 2024. DOI: 10.1109/CARS61786.2024.10778919

[9] S. Yuan and X. Wu, "Deep learning for insider threat detection: Review, challenges and opportunities," *Computers & Security*, Vol. 104, 102221, 2021. DOI: 10.1016/j.cose.2021.102221

[10] N. Sarantinos, M. Theriou, K. Tzovaras, and P. Daras, "Forensic malware analysis: The value of fuzzy hashing algorithms in identifying similarities," *Proceedings of the 2016 IEEE Trustcom/BigDataSE/ISPA*, pp. 1782-1787, 2016. DOI: 10.1109/TrustCom.2016.0274

[11] M. F. Arroyabe, G. Epelde, and G. S. Ruano, "Revealing the realities of cybercrime in small and medium enterprises: Understanding fear and taxonomic perspectives," *Computers & Security*, Vol. 141, 103826, 2024. DOI: 10.1016/j.cose.2024.103826

[12] J. Kornblum, "Identifying almost identical files using context triggered piecewise hashing," *Digital Investigation*, Vol. 3, pp. 91-97, 2006. DOI: 10.1016/j.diin.2006.06.015

[13] V. Roussev, "Data fingerprinting with similarity digests," *Advances in Digital Forensics VI: Sixth IFIP WG 11.9 International Conference on Digital Forensics, Hong Kong, China, January 4-6, 2010, Revised Selected Papers*, Vol. 6, pp. 207-226, Springer Berlin Heidelberg, 2010. DOI: 10.1007/978-3-642-15506-2_15

[14] J. Oliver, C. Cheng, and Y. Chen, "TLSH--a locality sensitive hash," *Proceedings of the 2013 Fourth Cybercrime and Trustworthy Computing Workshop*, pp. 7-13, 2013. DOI: 10.1109/CTC.2013.9

[15] V. Roussev, "An evaluation of forensic similarity hashes," *Digital Investigation*, Vol. 8, pp. S34-S41, 2011. DOI: 10.1016/j.diin.2011.05.005

[16] J. Fuentes, R. G. Azevedo, J. Martínez, J. Balsa, and P. García Teodoro, "Cybersecurity threat detection based on a UEBA framework using Deep Autoencoders," *arXiv preprint arXiv:2505.11542*, 2025. DOI: 10.48550/arXiv.2505.11542

[17] M. Fleming and O. Olukoya, "A temporal analysis and evaluation of fuzzy hashing algorithms for Android malware analysis," *Forensic Science International: Digital Investigation*, Vol. 49, 301770, 2024. DOI: 10.1016/j.fsidi.2024.301770

Authors

Dae-Won Kim received the B.S. degree in Physics from Kyung Hee University, Korea, in 2003, and the M.S. degree in IT Policy and Management from Soongsil University, Korea, in 2024.

He is currently pursuing a Ph.D. in the Department of IT Policy and Management at Soongsil University, Seoul, Korea. His research interests include data leakage prevention, insider threat detection, and artificial intelligence-driven security systems in enterprise environments.

Myung-Ho Kim received the B.S. in Department of Computer Science and Engineering from Soongsil University, Korea, in 1989. M.S. and Ph.D. degrees in Department of Computer Engineering from Postech University, Korea, in 1991 and 1995, respectively. He is currently a professor in the Dept. of Software, Soongsil University. He is interested in Machine Learning, Deep Learning and Block chain.