플래시 메모리 기반 저장장치에서 디지털 포렌식을 위한 데이터 무결성에 영향을 주는 특성 및 기술 연구

이현섭*
백석대학교 컴퓨터공학부 교수

A Study on Characteristics and Techniques that Affect Data Integrity for Digital Forensic on Flash Memory-Based Storage Devices

Hyun-Seob Lee*
Professor, Division of Computer Engineering, Baekseok University

요 약 디지털 포렌식에서 가장 중요하게 여기는 특징 중 하나는 무결성이다. 무결성은 데이터가 변조되지 않았음을 의미한다. 디지털 포렌식 과정에서 증거를 수집하는데 이 증거가 나중에 변조되었다면 증거로 사용될 수 없다. 아날로그 증거물은 사진을 찍어놓는 방식 등을 통해 변조된 사실을 쉽게 파악할 수 있다. 그러나 저장매체 속의 데이터 즉, 디지털 증거는 눈에 보이지 않기 때문에 변조되었는지 알기가 어렵다. 그래서 이 증거 데이터가 증거 수집 단계에서 법정 제출까지의 과정 중 변조가 되지 않았음을 증명하기 위해 해시값을 사용한다. 해시값은 증거 수집 단계에서 저장 데이터로부터 수집한다. 그러나 NAND 플래시 메모리는 내부적인 동작의 특성 때문에 시간이 지나면 물리적 데이터 형상이 수집 단계와 달라질 수 있다. 본 논문에서는 고의적인 데이터 훼손을 시도하지 않더라도 플래시 메모리의 물리적 형상이 변경될 수 있는 플래시 메모리의 특성 및 기술들을 연구한다.

주제어 : 메모리, 해시, 디지털 포렌식, 저장 시스템, 데이터 무결성

Abstract One of the most important characteristics of digital forensics is integrity. Integrity means that the data has not been tampered with. If evidence is collected during digital forensic and later tampered with, it cannot be used as evidence. With analog evidence, it's easy to see if it's been tampered with, for example, by taking a picture of it. However, the data on the storage media, or digital evidence, is invisible, so it is difficult to tell if it has been tampered with. Therefore, hash values are used to prove that the evidence data has not been tampered with during the process of collecting evidence and submitting it to the court. The hash value is collected from the stored data during the evidence collection phase. However, due to the internal behavior of NAND flash memory, the physical data shape may change over time from the acquisition phase. In this paper, we study the characteristics and techniques of flash memory that can cause the physical shape of flash memory to change even if no intentional data corruption is attempted.

Key Words : memory, hash, digital forensic, storage system, data integrity

*This paper was supported by 2023 Baekseok University Research Fund
*교신저자 : 이현섭(hyunseob@bu.ac.kr)
접수일 2023년 3월 23일 수정일 2023년 5월 4일 심사완료일 2023년 5월 7일

1. 서론

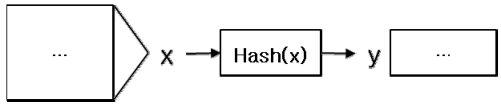
디지털 포렌식은 범죄 수사에서 적용되는 과학적 증거 수집 및 분석기법으로, 디지털 데이터를 수집하여 범행과 관련된 증거를 확보하는 수사 기법이다. 디지털 포렌식에서 가장 중요한 요소 중 한가지는 무결성이다. 무결성은 데이터가 변조되지 않았음을 의미하는 것으로, 초기 수집된 데이터가 법정에 제출될 때까지 데이터 무결성은 지켜져야 한다. 이를 위해 디지털 데이터 기반 증거자료는 디지털 포렌식 절차에 따라 데이터 수집, 보관, 분석, 제출되어야 한다. 데이터 무결성을 위한 대표적인 방법은 해시를 이용하는 것이다. 이 방법은 초기 단계에서 수집된 데이터를 위변조 없이 법정까지 유지와 관리했음을 증명하기 위해 물리적 저장장치로부터 해시값을 추출하고 이후 분석 단계에서 추출한 해시값을 비교하여 데이터의 형상 변경을 비교하는 방법이다.[1-4]

그러나 디지털 포렌식 절차에 따라 디지털 증거를 관리하더라도 데이터를 유지하고 있는 물리적 환경이 변경되거나 오염될 경우 데이터의 물리적 형상이 변경될 수 있다. 특히 NAND 플래시 메모리 기반 저장장치는 플래시 메모리의 물리적 특징 때문에 저장장치 내부에서 물리적인 데이터의 위치와 물리적 형상이 변경될 수 있다. 이러한 현상은 초기 데이터 수집 단계에서 플래시 메모리의 물리적 위치로부터 데이터 해시값을 추출한 이후 저장장치 내부에서 데이터의 위치 변경으로 인해 데이터의 최종 형상이 변경된 것으로 판정될 수 있는 것을 의미한다. 이 경우 수집된 데이터는 법정에서의 증거의 가치를 상실할 수 있다. 따라서 의도하지 않은 데이터 형상 변경에 영향을 줄 수 있는 플래시 메모리의 특성을 연구하는 것은 디지털 포렌식스의 무결성 및 관련 연구를 위한 중요한 기초연구이다. 따라서 본 논문에서는 플래시 메모리의 특성을 살펴보고 고의적인 시도가 없더라도 저장장치에서 물리적인 데이터의 형상이 변경될 수 있는 기술들을 연구한다.

2. 배경 및 문제점

2.1 해시 알고리즘의 역할

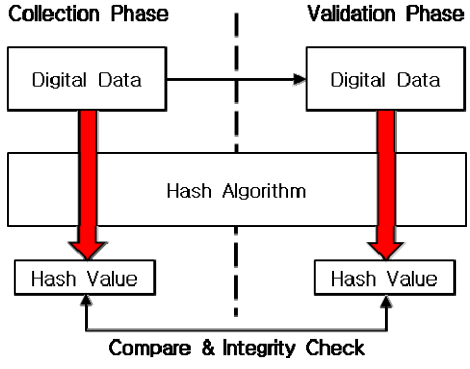
Fig. 1은 해시 알고리즘을 위한 해시 함수의 기본적인 역할을 보여주고 있다. 해시 알고리즘은 임의의 데이터 x를 고정된 길이의 데이터 y로 멥핑한다. 또한, x를 y로 멥핑하는 것은 가능하지만, y에서 x로 변환하는 것은 불가능한 방향성을 가진 알고리즘이다. 대표적인 해시 알고리즘은 MD5[5, 6], SHA1[7, 8], SHA256[9, 10], SHA512[11, 12]등이 있다. 해시 알고리즘은 데이터의 크기와 관계없이 고정된 크기의 데이터로 멥핑하기 때문에 디지털 데이터의 전자지문 역할을 할 수 있다.



[Fig. 1] Hash Algorithm

2.2 해시를 이용한 데이터 무결성 검증

Fig. 2는 데이터 수집 단계와 증거 검증 단계에서 사용된 증거 데이터가 위변조가 없음을 증명하기 위한 무결성 확인과정을 보여주고 있다. 그림과 같이 증거 수집 단계에서 디지털 데이터로부터 해시 알고리즘을 이용하여 추출한 해시값과 검증 단계에서 증거 데이터로부터 같은 해시 알고리즘으로 추출된 해시값을 비교하여 수사 과정에서의 데이터의 변조가 없었는지 무결성 검증을 수행한다. 그러나 수집된 증거 데이터가 저장되어있는 저장장치의 물리적 형상이 변경될 경우 증거 데이터의 무결성에 영향을 줄 수 있다. 특히 플래시 메모리 기반 저장장치는 메모리의 특성과 내부 동작 때문에 의도하지 않은 물리적 형상 변경이 발생할 수 있다.[13-16] 따라서 이러한 특성을 고려한 데이터 무결성 검증 방법을 연구해야 한다. 본 논문에서는 플래시 메모리 기반 저장장치에서 의도하지 않은 형상변경이 발생할 수 있는 특징을 연구하여 정리한다.

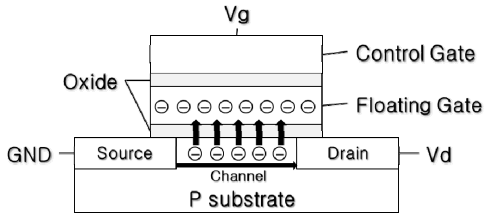


[Fig. 2] Integrity Check

3. 데이터 무결성에 영향을 주는 특징

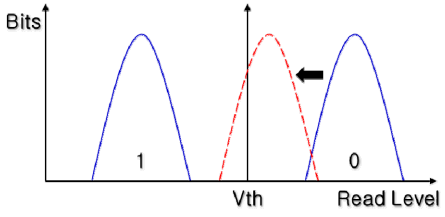
3.1 읽기 동작으로 인한 산포 변화

플래시 메모리는 NOR 플래시 메모리와 NAND 플래시 메모리로 구분된다. 각 셀이 병렬로 연결된 NOR 플래시 메모리와 비교하여 NAND 플래시 메모리는 직렬로 연결되어있다. 따라서 NOR 플래시 메모리는 랜덤한 접근이 가능하지만, NAND 플래시 메모리는 순차적인 접근을 해야 하는 특징이 있다.



[Fig. 3] Structure of Cell

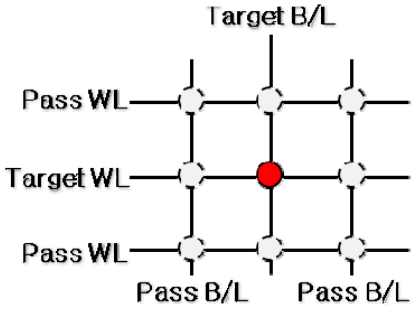
Fig. 3는 플래시 메모리 셀의 구조를 보여주고 있다. 각 셀은 회로 기판 위에 소스, 드레인으로 구성되어 있다. 그리고 그 위에 옥사이드와 실리콘 게이트의 중첩된 구조로 되어 있다. SLC(Single Level Cell)의 경우 하나의 셀은 하나의 비트를 저장할 수 있다. 플래시 메모리 셀에서 데이터 쓰기 동작의 과정은 Vg에 높은 전압을 가하여 전류 기판에 흐르는 전자를 옥사이드를 넘어서 플로팅 게이트로 넘어오게 하는 것이다. 그리고 읽기 동작의 과정은 Vg와 Vd에 일정 전압을 가하여 소스와 드레인 사이에 전자들이 모여 전류가 흐르는 수준으로 채널이 형성되는지를 관찰하는 것이다. 만약 쓰기 동작이 수행되었을 경우 플로팅 게이트에 음극 전자가 모여있기 때문에 적은 전력으로도 소스와 드레인 사이 전류가 흐를 수 있는 채널을 형성할 수 있다. 그러나 쓰기 동작이 선행되지 않으면 채널이 형성되지 않아서 소스와 드레인 사이 전류가 흐르지 못한다.



[Fig. 4] Movement of Elections

Fig. 4는 플래시 메모리 동작으로 인해 셀에서 이동된 산포의 형상을 보여주고 있다. 각 셀은 게이트에 유지하고 있는 전자의 양에 따라 전압이 변경된다. 그리고 셀에서 유지하고 있는 데이터 정보는 그림에서 보여주는 것과 같이 문턱 전압(Vth)을 기준으로 좌측과 우측 중 산포를 통해 판별할 수 있다. 기본적으로, 플래시 메모리는 쓰기 동작을 수행한 경우 바닥 기판의 전자들이 플로팅 게이트로 넘어가 유지되기 때문에 산포들이 우측에 위치하고, 지우기 동작을 수행할 경우 플로팅 게이트의 전자들이 바닥 그라운드로 넘어가기 때문에 좌측으로 이동한다. 그러나 우측으로 이동한 산포는 지우기 동작을 수행하지 않아도 읽기 동작 장시간 반복적으로 수행한 경우 하단에 흐르는 전류로부터 영향을 받아서 누적된 데미지를 받는다. 즉 반복된 읽기 동작으로 셀의 산포 일부가 우측에서 좌측으로 이동할 수 있다. 이렇게 누적된 손상으로 인한 데이터 손실을 막기 위해 일정 레벨 이상 좌측으로 이동된 셀을 포함한 페이지나 일정 횟수 이상 읽기 동작이 수행된 페이지는 다른 위치의 비어있는 페이지에 복사하여 산포를 정상화 하는 레텐션(Retention) 동작을 수행한다. 이러한 리텐션 동작은 플래시 메모리의 외부로부터 데이터 쓰기 명령을 받지 않더라도 데이터 신뢰성을 유지하기 위해 내부적으로 발생한다. 결과적으로 사용자가 의도하지 않은 데이터 복사 및 이동 때문에 물리적 변경이 발생한다. 그리고 이 동작이 디지털 포렌식의 증거 수집 단계와 증거 검증 단계 사이에서 수행될 경우 물리적 형상이 변경될 수 있기 때문에 증거의 무결성에 영향을 미칠 수 있다.

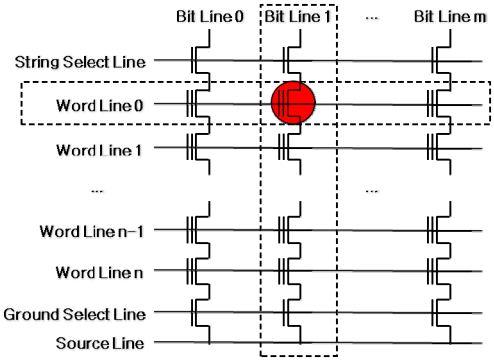
3.2 쓰기 동작으로 인한 인접 셀의 산포 변화



[Fig. 5] Word Line and Bit Line

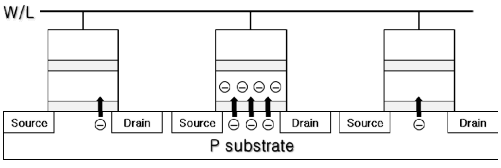
Fig. 5는 NAND 플래시 메모리 기반 저장장치를 구성하기 위해 비트라인(B/L)과 워드라인(W/L)으로 연결

된 셀들의 구조를 보여주고 있다. 그림과 같이 모든 셀들은 기판 위에 비트라인과 워드라인으로 직열 연결되어 있다. 따라서 각각의 셀 동작을 제어하기 위해서는 셀과 연결된 라인들의 조합을 이용한다. 즉 특정 셀을 조작하기 위해서는 관련이 없는 라인을 제외하고 타겟 비트라인과 타겟 워드라인에만 전압을 인가하여 교차하는 위치의 셀을 제어한다.



[Fig. 6] Connection of Cell

Fig. 6은 셀들이 직렬로 연결된 플래시 메모리 기반 저장장치의 구조를 보여주고 있다. 그림에서는 동그라미로 표시된 셀을 제어하기 위해 비트라인 1번과 워드라인 0번에 전압을 가해야 한다. 그런데 같은 라인에 직렬로 연결된 셀은 타겟 셀과 동일한 전압의 영향을 받기 때문에 산포의 변화가 발생할 수 있다.

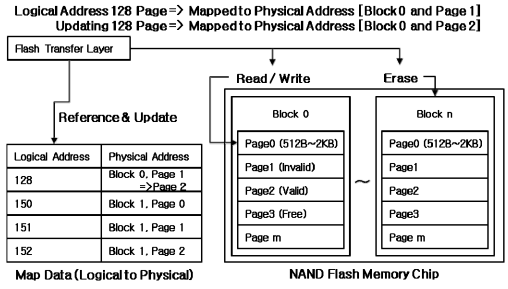


[Fig. 7] Impact of Neighboring Cells

Fig. 7은 쓰기 동작을 수행하는 동안 동일 워드라인에 연결된 셀들의 영향을 보여주고 있다. 쓰기 동작은 비트라인으로 선택된 셀의 상단 워드라인에 강한 전압을 인가하고, 동시에 비트라인을 통해 전류가 흐르도록 하여 음의 속성을 가진 전자들이 실리콘 소재의 옥사이드를 넘어서 플로팅 게이트에 적재되도록 유도하는 동작이다. 이때 그림과 같이 동일한 워드라인에 연결된 이웃 셀들은 동일하게 강한 전압을 인가받게 된다. 따라서 비트 라인에 전류가 흐르지 않더라도 소량의 전자가 플로팅 게이트로 전이되어 셀의 산포에 변화를 줄 수 있다. 이러한 산포 변화는 예측하지 못한 시점에 의도하지 않은 데이터 변화와 불량을 일으킬 수 있기 때문에 디지털 포렌식을 위한 증거 무결성에 영향을 미칠 수 있다.

3.3 가비지 컬렉션으로 인한 물리적 형상 변화

플래시 메모리 기반 저장장치에서 데이터를 저장하기 위한 블록들은 여러 개의 페이지로 구성되어 있다. 플래시 메모리의 특징 중 한 가지는 데이터를 쓰기 전 지우기 연산을 수행해야 한다는 것이다. 그런데 읽고 쓰는 단위는 페이지 단위이지만 지우기 단위는 블록 단위로 동작하기 때문에 데이터의 업데이트가 발생하면 블록 내에 데이터들을 백업하고 복구하기 위해 많은 비용을 소비한다. 따라서 이러한 고비용의 동작을 방지하기 위해 FTL(Flash Transfer Layer)를 사용한다.

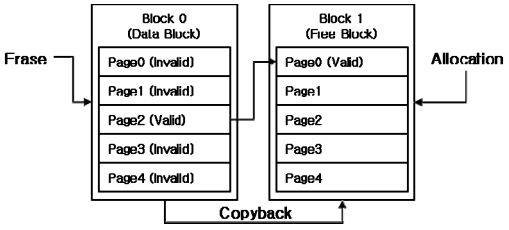


[Fig. 8] Flash Transfer Layer

Fig. 8은 FTL의 동작을 보여주고 있다. FTL은 논리적인 주소를 물리적인 주소로 멥핑하는 방법을 사용하여 업데이트가 발생해도 대용량의 데이터 이동을 지연시킨다. 그림의 예에서 논리주소 128번 페이지는 물리주소 0번 블록의 1번 페이지에 멥핑되어 있었다. 그런데 128번 페이지에 업데이트가 발생했을 때 FTL은 이전 페이지를 무효화(Invalid) 하고 비여있던 2번 페이지에 업데이트 된 데이터를 쓴다. 그다음 이 정보를 멥핑 테이블에 반영한다. FTL은 이러한 방법을 통해 고비용의 데이터 업데이트 동작을 한 번의 쓰기 동작으로 처리할 수 있다. 그러나 빈번한 업데이트 때문에 누적된 무효화 된 페이지를 정리하기 위한 동작이 필요하다.

Fig. 9는 무효화 된 페이지를 정리하기 위한 가비지 컬렉션 동작을 보여주고 있다. FTL은 내부 정책에 따라 다수의 무효화 된 페이지를 보유하고 있는 블록으로부터 유효한 데이터 페이지를 새로운 블록으로 이동시키고 이전 블록을 삭제하는 동작을 통해 무효화 된 페이지를 정

리한다. 그림의 예제에서는 0번 블록 내에 유효한 페이지 2번을 새롭게 할당된 블록 1번의 비어있는 0번 페이지에 복사하고 이전 블록인 0번 블록을 삭제하였다. 이러한 동작을 통해 FTL은 4개의 무효화 된 페이지를 사용 가능한 페이지로 확보할 수 있다.



[Fig. 9] Garbage Collection

가비지 컬렉션을 위한 FTL의 동작은 플래시 메모리 저장장치의 내부적으로 동작하기 저장장치 외부에서 동작을 예측할 수 없다. 따라서 디지털 포렌식의 증거 수집 단계에서 확인할 수 있는 저장장치의 물리적 형상과 검증 단계의 물리적 형상은 서로 다를 수 있다.

3.4 플래시 메모리 블록의 수명

FTL의 또 다른 역할 중 하나는 불량이 발생한 블록을 정상 블록과 교체하는 것이다. 플래시 메모리 블록은 실리콘 옥사이드에 전자가 이동하는 도중 충돌 때문에 균열이 발생하면 플로팅 게이트에 정상적인 전자유지가 불가한 불량이 발생한다. 그리고 이러한 전자의 이동은 쓰기 동작과 지우기 동작을 수행할 때 발생하기 때문에 실리콘의 내구력을 강화해도 일정 횟수 이상의 쓰기 동작과 지우기 동작을 반복하면 필연적으로 불량이 발생한다. 이렇게 블록에 불량이 발생하면 FTL은 불량 블록을 생산 과정에서 추가로 준비해 놓은 예비 블록과 교체하여 고장을 회복한다. 그리고 이러한 블록 교체 과정에서 블록 내에 균열이 발생하지 않은 페이지의 데이터는 새로운 예비 블록으로 복사되어 복구된다. 그러나 균열이 발생한 페이지의 데이터 손실은 피할 수 없다. 결과적으로 이러한 고장회복 동작은 플래시 메모리의 물리적 형상에 영향을 줄 수 있다.

3.5 백그라운드 미디어 검사

플래시 메모리 기반 저장장치는 정해진 규칙에 따라 내부적으로 데이터가 저장된 블록들에 대해 읽기 동작을 수행한다. 이 읽기 동작은 외부로 데이터를 전송하기 위한 동작이 아니라 미디어의 정상 동작 여부를 확인하기 위한 검사 동작이다. 그러나 검사 동작 수행 중 앞에서 설명한 리텐션 동작이나 가비지 컬렉션 동작이 동시에 수행될 수 있다. 예를 들어 한 페이지당 1000번의 읽기 동작이 수행될 때마다 리텐션 동작을 수행하도록 정책이 정해져 있다면 외부의 개입 없이 내부적으로 데이터가 저장된 물리적 페이지의 위치가 변경될 수 있다. 그 밖에도 백그라운드 미디어 검사를 위한 읽기 동작이 수행되는 사이에 가비지 컬렉션 또한 동시에 동작할 수 있다. 이러한 동작은 앞에서 언급한 물리적 위치를 변경할 수 있는 또 다른 동작 특징이다. 마지막으로 미디어 검사 중 불량이 발견될 경우 내부적으로 불량을 복구하기 위한 알고리즘이 수행된다. 이러한 동작들은 결과적으로 백그라운드 미디어 검사가 동작하기 때문에 발생할 수 있는 형상 변화이다. 따라서 디지털 포렌식을 위한 증거 수집 이후에도 외부 명령 없이 물리적인 형상의 변경이 발생할 수 있다.

4. 결론

본 논문에서는 디지털 포렌식을 위해 낸드 플래시 메모리 기반 저장장치의 데이터를 증거로 수집한 경우 외부의 명령이 없어도 내부적 동작 때문에 증거의 무결성에 영향을 줄 수 있는 플래시 메모리의 특징을 연구했다. 조사된 연구 결과는 플래시 메모리의 신뢰성을 위한 기술이지만 저장장치의 물리적 형상에 영향을 줄 수 있다. 따라서 증거 수집 단계와 검증 단계에서 데이터 형상이 동일해야 하는 디지털 포렌식에서는 플래시 메모리의 특징이 반드시 고려되어야 한다. 향후에는 물리적 형상 변경이 발생하더라도 증거의 무결성을 지킬 수 있는 증거 수집 방법을 연구할 예정이다.

REFERENCES

[1] H.J.Seong, J.H.Jung, K.R.Park and S.J.Cho, "Research Trends in Vehicle Digital Forensics Focused on Infotainment Systems and Mobile Devices," *Journal of KIISE*, Vol.41, No.1, pp.38-45, 2023.

[2] J.O.Lee and T.S.Shin, "Forensics for Android and Linux-based file system on IoT platform," *Journal of Digital Contents Society*, Vol.23, No.2, pp.335-342, 2023.

[3] W.K.Jung and S.J.Lee, "Measures to maintain the admissibility of evidence for taking over digital evidence in accordance with the adjustment of the police · prosecution investigation authority," *Journal of Digital Forensics*, Vol.16, No.2, pp.126-141, 2022.

[4] H.J.Jung and S.J.Lee, "Digital forensic technology trends in the Internet of Things era," *Journal of KIISE*, Vol.38, No.9, pp.33-39, 2020.

[5] S.B.Suhaili, C.C.A.Niam, Z.M.Zainn and N.Julai, "Design and Implementation of MD5 Hash Function Algorithm Using Verilog HDL," *Proceedings of the 12th National Technical Seminar on Unmanned System Technology 2020*, Vol.770, pp.499-510, 2021.

[6] U.Kumar and V.C.Venkaiah, "A New Modified MD5-224 Bits Hash Function and an Efficient Message Authentication Code Based on Quasigroups," *Cyber Security, Privacy and Networking*, Vol.370, 2022.

[7] F.Zhai, P.Tao, B.Xu, X.Liang and Y.Cao, "Research and System Design of Remote Comparison Method for Embedded Device Files," *2022 International Symposium on Advances in Informatics, Electronics and Education (ISAIEE)*. pp.137-141, 2022.

[8] M.Ali, A.Ismail, H.Elgohary, S.Darwish and S.Mesbah, "A Procedure for Tracing Chain of Custody in Digital Image Forensics: A Paradigm Based on Grey Hash and Blockchain," *Symmetry*, Vol.14 No.2, pp.344, 2022.

[9] Z.Wang, X.Dong, Y.Kang and H.Chen, "Parallel SHA-256 on SW26010 many-core processor for hashing of multiple messages," *The Journal of Supercomputing*, Vol.79, pp.2332-2355, 2022.

[10] E.A.Adeniy, P.B.Falola, M.S.Maashi, M.Aljebreen and S.Bharany," *Information*, Vol.13, No.10, pp.442, 2022.

[11] M.Yang, Y.Zhang, B.Yang, H.Wang, S.Yin, S.Wei, L.Liu, "A SHA-512 Hardware Implementation Based on Block RAM Storage Structure," *2022 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, pp.132-135, 2022.

[12] D.P.Purba, "Analisa Dan Perbandingan Algoritma Whirpool Dan Sha-512 Dalam Penyandian Data Gambar," *Bulletin of Artificial Intelligence*, Vol.1, No.1, pp.8-12, 2022.

[13] H.S.Lee, "A Prediction-Based Data Read Ahead Policy using Decision Tree for improving the performance of NAND flash memory based storage devices," *The Korea Internet of Things Society*, Vol.8, No.4, pp.9-15, 2022.

[14] H.S.Lee, "A Safety IO Throttling Method Inducting Differential End of Life to Improving the Reliability of Big Data Maintenance in the SSD based RAID," *The Society of Digital Policy & Management,* Vol.20, No.5, pp.593-598, 2022.

[15] H.S.Lee, "Performance analysis and prediction through various over-provision on NAND flash memory based storage," *The Society of Digital Policy & Management*, Vol.20, No.3, pp.343-348, 2022.

[16] H.S.Lee, "A method for optimizing lifetime prediction of a storage device using the frequency of occurrence of defects in NAND flash memory," *The Korea Internet of Things Society*, Vol.7, No.4, pp.9-14, 2021.

이 현 섭(Hyun-Seob Lee) **[종신회원]**

- 2013년 2월 : 한양대학교 컴퓨터공학과 (공학 박사)
- 2012년 3월 ~ 2021년 2월 : 삼성전자 책임연구원
- 2021년 3월 ~ 현재 : 백석대학교 컴퓨터공학부 조교수

〈관심분야〉

인공지능, 저장시스템, 임베디드 시스템