형사정책연구 Korean Criminological Review
제29권 제2호(통권 제114호, 2018 · 여름)

해시함수의 형사법적 고찰

김기범*

국 | 문 | 요 | 약

해시함수는 1990년대 중후반부터 사이버수사와 디지털포렌식 분야에서 사용된 이래 지금까지 디지털증거의 동일성을 입증하는 수단으로 활용되고 있다. 2007년 대법원이 일심회 사건에서 증거능력의 인정요건으로 동일성을 제시하였고, 이어 왕재산 사건과 이석기 사건 등에서 해시값을 동일성 입증 수단으로 인정함에 따라 압수 · 수색 과정에서 많이 활용되고 있다. 특히, 2011년 형사소송법 개정(제106조 제3항)으로 디지털증거에 대한 출력 · 복제가 압수원칙이 되면서 동일성 입증을 위한 중요한 절차로 자리매김 하였다.

그런데 대법원은 해시함수를 동일성 입증의 중요한 수단으로 인정하면서도 해시함수의 안전성, 허용되는 기술규격, 기술적 한계에 대해서 충분한 논의하지 않았다. 대법원 판례에 따르면 증거능력이 부정될 가능성이 높은 경우 즉, 해시값이 기술적 이유로 변경되거나 원본이 실시간으로 변하여 해시값이 변경될 수밖에 없는 경우 그리고 해시값을 생성할 수 없는 경우 어떻게 평가해야하는지 모호하다. 과학적 증거의 논의가 주로 지문, 거짓말탐지기, DNA, 혈흔 등을 중심으로 이루어져 해시함수에 대한 학계의 관심도 부족했다. 나아가 해시함수가 동일성 입증에 관한 원칙으로 규정되는 것이 적절한지에 대한 의문도 있다.

따라서 본 논문에서는 동일성 증명수단으로서 해시함수의 안전성에 대해 평가하고, 형사절차에서 해시값 생성절차와 동일성 입증수단으로서의 한계에 대하여 살펴보았다. 그 결과 현재 사용 중인 해시함수인 MD5, SHA-1이 안전성에 대한 논란에도 증거법적으로 인정하는데 문제가 없다고 보았다. 이어 우리나라와 미국의 판례를 분석하여 해시값이 법정에서 널리 인용되고 있다는 사실을 확인하고, 해시함수를 둘러싼 3가지의 법적 · 기술적 쟁점을 도출하였다. 즉, 해시함수가 과학적 증거에 해당하는지, 동일성 입증의 원칙적 방법으로 적절한지, 오류율의 개념은 어떻게 정의되는지를 살펴보았다. 마지막으로 형사절차에서 해시함수의 활용성과 신뢰성 제고를 위해 기술적 문제로 해시값이 변경되는 경우 증거능력을 인정받기 위한 다양한 방법을 개발할 것을 제안하였고, 나아가 해시도구의 기술표준 마련과 메타데이터의 동일성까지 보장할 수 있는 고성능 해시도구 개발을 촉구하였다.

❖ 주제어 : 해시함수, 디지털증거, 증거능력, 과학적 증거, 범죄수사, 형사절차, 해시도구, 동일성, 무결성

* 경찰대학 경찰학과 교수

Ⅰ. 서론

해시함수(Hash function)는 주로 데이터 위·변조 검사, 데이터의 빠른 검색, 통신의 안전성 증명, 패스워드의 안전한 보관 등에 사용되었는데 최근에는 전자서명, 불법 저작물 차단, 홈페이지 해킹여부 판단, 암호화폐의 신뢰성 확보, 전자투표 등으로 활용범위가 확대되고 있다. 수사기관은 1990년대 중후반부터 사이버수사와 디지털포렌식 분야에서 해시함수를 사용하기 시작하였고, 2007년 대법원에서 처음으로 디지털증거의 증거능력 인정요건으로 동일성을 제시하면서 증거법 측면에서 주목받기 시작하였다. 이어 2011년 형사소송법에서 디지털증거에 대한 출력·복제를 압수원칙으로 규정(제106조 제3항)하면서 복제 사례가 증가하여 해시함수는 동일성 입증의 주요한 기술로 자리를 잡았다.

그런데 대법원은 해시함수를 동일성 입증의 중요한 수단으로 인정하면서도 해시함수의 안전성, 허용되는 기술규격, 기술적 한계 등에 대해서 충분히 논의하지 않았다. 대법원 판례에 따르면 증거능력이 부정될 가능성이 높은 경우 즉, 해시값이 기술적 이유로 변경되거나 원본이 실시간으로 변하여 해시값이 변경될 수밖에 없는 경우 그리고 해시값을 생성할 수 없는 경우 어떻게 처리해야 하는지 애매하다.

과학적 증거의 논의가 주로 지문, 거짓말탐지기, DNA, 혈흔 등을 중심으로 이루어져 해시함수에 대한 학계의 관심도 부족했다. 뿐만 아니라 해시함수를 동일성 입증에 관한 원칙으로 규정하는 것이 적절한지에 대한 의문도 제기되고 있다.

따라서 본 논문에서는 동일성 입증 수단으로서 해시함수의 안전성을 평가해 보고, 형사절차에서 수행되고 있는 해시값 생성절차와 동일성 증명에서의 한계에 대해 살펴보고자 한다. 이어 우리나라와 미국의 해시함수에 대한 판례를 분석하여 법적·기술적 쟁점을 도출하고 이를 바탕으로 형사절차에서 해시함수의 활용성과 신뢰성을 제고할 수 있는 법적·기술적 개선방안을 제안하고자 한다. 본 연구는 그간의 학문적 영역의 문제를 극복하고 해시함수에 대한 기술분야와 형사법 분야를 학제적으로 접근하여 문제해결을 시도하였다는데 의미가 있다.

Ⅱ. 동일성 입증 수단으로서 해시함수의 안전성

1. 개념

해시함수(Hash Function)는 임의의 비트열을 고정된 길이의 비트열로 변환시켜 주는 함수[1]로 해시함수에 의한 결과값을 해시값(Hash value)이라고 한다. 해시값에서 원래의 데이터를 찾아낼 수 없다고 하여 일방향 함수(One-way Function)라고 불리기도 한다. 해시값은 입력하는 데이터의 개수와 용량에 관계없이 그 해시함수가 정해 놓은 일정한 크기의 비트열로 표현된다. 예를 들어, 임의의 파일 1개를 해시함수인 MD5에 입력하면 해시값은 "2e9a3018f506f60e8d5faaaa6824ed7a"으로 32자리로 나타나고, SHA-1에 입력하면 "7f50ea679932a58d840083aac2ae2035e439ed6f"으로 40자리로 표현된다.

해시함수는 Ronald Rivest가 1989년에 MD2를 개발하고 1992년에 MD5를 개발하면서 활용되기 시작했다.[2] 이어 1993년에 미국 국가안보국(NSA)이 SHA계열의 함수를 개발한 이래 수많은 해시함수가 소개되었고, 우리나라도 2014년에 국가보안기술연구소에서 LSH(Lightweight Secure Hash)를 개발하여 표준으로 채택하였다.[3] 해시함수는 크게 암호학적 해시와 비암호학적 해시로 나눌 수 있는데 전자는 안전성을 검증하기 위해 일정한 특징을 요구하는 함수로 MD4, MD5[4], SHA-1[5], SHA-3, HAS-160 등이 대표적이고, 후자는 안전성 검증이 필요 없는 함수로 데이터베이스의 색인속도를 높이거나 전송오류 검사(Cyclic Redundancy Check, CRC)

1) 이상진, 「디지털 포렌식 개론」(개정판), 이룬출판사, 2015, 283면

2) Fr´ed´eric Muller, "The MD2 Hash Function Is Not One-Way", Advances in Cryptology - ASIACRYPT 2004, 10th International Conference on the Theory and Application of Cryptology and Information Security, 2004, pp.215~216.

3) 한국인터넷진흥원 홈페이지(https://seed.kisa.or.kr/iwt/ko/sup/EgovLshInfo.do, 2018.5.10. 최종 확인).

4) Rivest, Ronald, "The MD5 message-digest algorithm", 1992; RFC 1321, "The MD5 Message-Digest Algorithm", Internet Request for Comments 1321, 1992.

5) National Institute of Standards and Technology, "Secure hash standard", U.S. Department of Commerce, FIPS 180-1, 1995.

를 위해 사용되는데 CRC32가 대표적이다. 일반적으로 디지털포렌식에서 사용하는 해시함수는 암호학적 해시함수(Cryptographic Hash Function)를 말한다.

2. 특징

해시함수는 데이터를 압축하는 성질(Compression, 압축성)과 계산을 쉽게 할 수 있다는 성질(Ease of computation, 용이성)을 가진다. 여기에 역상 저항성(Preimage Resistance), 제2역상 저항성(2nd-Preimage Resistance), 충돌 저항성(Collision Resistance)을 만족해야 한다.[6)]

역상 저항성은 함수 h에 대한 임의의 출력 y가 주어지고 출력 y를 생성하게 된 입력 x는 알려지지 않은 상황에서 h(x)=y를 만족시키는 임의의 역상 x를 찾는 것이 불가능하다는 성질이다. 제2역상 저항성은 입력 x가 주어졌을 때 h(x)=h(x′)을 만족하는 x와는 다른 입력 x′(제2 역상)를 찾는 것이 불가능하다는 성질이다. 마지막으로 충돌저항성은 h(x)=h(x′)를 만족시키는 임의의 서로 다른 두 입력 x, x′를 찾아내는 것이 불가능하다는 성질을 말한다. “불가능하다”는 의미는 수학적으로 “계산할 수 없다”가 아니라 유한한 자원을 동원해서 계산하는 것이 “사실상 불가능하다”는 것을 말한다.[7)]

디지털증거의 동일성 입증은 위에서 설명한 3가지 특징 중에서 ‘제2역상 저항성’의 성질을 이용하는 것이다. 즉, 디지털증거 x를 해시함수 h에 입력하여 해시값 y=h(x)를 생성했을 때 h(x)와 동일한 해시값을 갖는 다른 디지털증거 x′를 찾는 것이 어렵기 때문에 원본과 사본의 해시값이 같으면 동일성을 인정하는 것이다.

3. 안전성

가. 불안전하다는 논거

해시함수는 초기에 MD4, HAVAL-128, RIPEMD를 사용하였는데, 제2역상과

6) 이상진, 「디지털 포렌식 개론」(개정판), 이룬출판사, 2015, 285면.
7) 이상진, 「디지털 포렌식 개론」(개정판), 이룬출판사, 2015, 285면.

충돌쌍을 찾을 수 있는 취약점이 발견되어 MD5, SHA계열의 함수를 개발하였다.[8] 이후 해시함수에 대해 생일자 공격[9], 전수조사 공격, 역상을 미리 계산하여 DB화 하는 공격(Rainbow Table Attack) 등 다양한 방법이 제시되었다. 그럼에도 MD5, SHA1의 수학적 안전성은 유지되었으나 중국의 Wang 교수 등이 2005년에 MD5에서 충돌쌍을 찾았고,[10] 미국의 구글도 2017년에 내용이 다른 2개의 PDF 파일을 이용하여 동일한 해시값을 생성해 SHA-1에서 충돌쌍을 찾아 안전성에 대한 우려가 제기되었다.[11] 그 결과 미국 국가안보국(NSA), 국립표준기술연구소(NIST)는 암호 사용 권고안에서 MD5를 제외하였고,[12] 안전성이 높은 SHA-3을 표준 해시함수로 채택하였다.[13] 구글도 SHA-3, SHA-256과 같이 보다 안전한 해시함수를 쓸 것을

8) 박재유, "디지털증거의 무결성 입증시 MD5 해시함수 사용에 관한 재고찰", (사)한국디지털포렌식학회 2016 디지털 포렌식 기술 워크샵, 2016, 41면.

9) 생일 역설(Birthday Paradox)은 사람이 임의로 모였을 때 그 중에 생일이 같은 두 명이 존재할 확률을 구하는 문제를 말한다. 생일이 가능한 경우는 365개이므로 366명 이상의 사람이 모인다면 비둘기집 원리에 따라 생일이 같은 두 명이 반드시 존재하며, 23명 이상만 모여도 그 중 두 명의 생일이 같은 확률은 1/2를 넘는다. 얼핏 생각하기에는 생일이 365가지이므로 임의의 두 사람의 생일이 같을 확률은 1/365이고, 따라서 365명쯤 모여야 생일이 같은 경우가 있을 것이라고 생각하지만 실제로 23명만 모여도 생일이 같은 두 사람이 나올 확률이 50%를 넘고, 57명이 모이면 99%를 넘게 된다. 이와 비슷하게 해시값이 같은 두 입력값을 찾는 것 역시 모든 입력값을 계산하지 않아도 충분히 높은 확률로 찾을 수 있는데 이러한 해시함수 공격을 생일공격(birthday attack)이라고 한다.(위키백과 참고) 이를 정리하면 생일공격(Birthday Attack)에서 해시함수의 출력길이가 n비트이면 안전성은 $2^{n/2}$을 넘을 수 없기 때문에 MD4, MD5는 128비트로 안정성은 $2^{128/2}$을 넘을 수 없다. 입력데이터가 $2^{128/2}$개 있을 경우 해시값이 동일한 쌍이 나올 확률이 50%가 된다는 것으로 컴퓨팅 파워가 발달한 지금보더라도 그 경우의 수는 천문학적이다.(E. H. Mckinney, "Generalized Birthday Problem", The American Mathematical Monthly Vol. 73 No.4, Apr 1966, pp.385~387)

10) X. Wang, H. Yu, "How to Break MD5 and Other Hash Functions", Advances in Cryptology – Proceedings of EUROCRYPT 2005, LNCS 3494, 2005, pp.19~35.

11) Marc Stevens, Elie Bursztein, Pierre Karpman, Ange Albertini, Yarik Markov, Alex Petit Bianco, Clement Baisse, "Announcing the first Sha1 collision", Google Security Blog, 2017. 2. 23. (https://security.googleblog.com/2017/02/announcing-first-sha1-collision.html; https//shattered.io/, 2018.5.10. 최종확인).

12) Elaine Barker, Allen Roginsky, "Transitions: Recommendation for Transitioning the Use of Cryptographic Algorithms and Key Lengths", NIST Special Publication 800-131A, 2011.

13) Morris J. Dworkin, "SHA-3 Standard: Permutation-Based Hash and Extendable-Output Functions", U.S. Department of Commerce, National Institute of Standards and Technology, FIPS PUB 202, 2014.

권장하였다.[14]

암호학계의 이러한 논의는 디지털포렌식에도 영향을 미쳐 MD5를 디지털증거의 동일성을 입증하는 수단으로 활용하는 것은 적절하지 않다는 주장이 제기되었다.[15] 비록 충돌 저항성에 대한 문제이기는 하나 이 역시 제2역상 저항성과 관련이 있고,[16] 전수조사를 하는 것보다 빠른 탐색방법이 제시되어 안전하지 않다는 것이다.[17] 여기에 해시함수의 안전성 기준이 80비트에서 128비트로 바뀌자 디지털포렌식에서도 SHA-3를 사용하는 것이 바람직하다는 견해도 제기되고 있다.[18]

나. 안전하다는 논거

앞서 소개한 사례는 모두 충돌 저항성이 공격받은 사례로 디지털증거의 동일성 입증에 사용되는 제2역상 저항성에 대한 공격[19]이 아니기 때문에 안전성에 영향이 없다. 게다가 ① 여전히 역상 저항성에 관한 효율적인 공격방법이 제시되지 않았고, ② 특정한 입력 Block에 한정된 결과로 수사현장에서 이러한 전제조건이 실현되리라고 가정하기 어려우며 ③ 짧은 문자열이나 메시지 블록이 아니라 파일이나 이미징 형태를 취급하기 때문에 충돌 가능성은 여전히 희박하다.[20] 수학적으로 볼 때

14) Marc Stevens, Elie Bursztein, Pierre Karpman, Ange Albertini, Yarik Markov, Alex Petit Bianco, Clement Baisse, “Announcing the first Sha1 collision”, Google Security Blog, 2017. 2. 23. (https://security.googleblog.com/2017/02/announcing-first-sha1-collision.html; https//shattered.io/, 2018.5.10. 최종확인).

15) Stevens, M., Lenstra, A.K. and de Weger, B. “Chosen-prefix collisions for MD5 and applications”, Int. J. Applied Cryptography,Vol.2 No.4, 2012, pp.322～359.

16) Rogaway, Phillip, and Thomas Shrimpton, “Cryptographic hash-function basics: Definitions, implications, and separations for preimage resistance, second-preimage resistance, and collision resistance”, International Workshop on Fast Software Encryption, 2004, pp.1～20.

17) Kelsey, John, and Bruce Schneier, “Second preimages on n-bit hash functions for much less than 2n work”, Annual International Conference on the Theory and Applications of Cryptographic Techniques. Springer Berlin Heidelberg, 2005, pp.1～15.

18) 이상진, 「디지털 포렌식 개론」(개정판), 이룬출판사, 287면.

19) 이상진, 「디지털 포렌식 개론」(개정판), 이룬출판사, 286면; Eric Thompson, “MD5 collisions and the impact on computer forensics”, Digital Investigation, Volume2, Issue1, 2005, pp.36～40.

20) 박재유, “디지털증거의 무결성 입증시 MD5 해시함수 사용에 관한 재고찰”, 한국디지털포렌식학회 2016 디지털 포렌식 기술 워크샵, 2016, 5~6면; Eric Thompson. "MD5 collisions and the impact on computer forensics." Digital Investigation Volume2, Issue1, 2005, pp.36～40.

충돌저항성이 보다 공격당하기 쉬운 조건이기 때문에 충돌저항성이 공격을 받았다고 하여 역상 저항성과 제2역상 저항성이 취약하다고 할 수 없다.[21]

따라서 디지털포렌식에서 MD5, SHA-1의 사용을 폐기하고 SHA-256과 같이 안전성이 높은 해시함수를 사용해야 하거나 MD5나 SHA-1을 단독으로 사용하지 않고 복수 이상을 사용하여야 한다는 주장은 성급하다. 이에 우리나라를 비롯하여 미국 등 다수의 국가에서 MD5, SHA-1을 여전히 사용하고 있다.

다. 소결

안전성은 법원에서 디지털증거의 동일성을 입증하는 수단으로 해시함수를 인정할 것인지 여부를 결정하는 가장 중요한 요소이다. 앞서 살펴본 바와 같이 해시함수에 대한 충돌쌍 공격이 성공하여 수학적 안전성에 대한 우려가 제기되고 있는 것은 분명하다. 하지만 MD5, SHA-1에 대한 제2역상 저항성 공격은 보고된 바가 없기 때문에 안전하지 않다고 볼 수 없다. 법정에서 제2역상 저항성 공격으로 인해 해시값이 똑같은 다른 증거가 제시되거나 그로 인하여 증거능력이 부정된 사례도 없다.

나아가 법정에서 인정하는 안전성은 수학적 안전성과 일치하지 않는다. 해시값이 같은 2개의 증거가 있더라도 그 파일이 당해 사건과 관련성이 높을 수 없고, 관련성이 있어도 법정에서 그 내용을 확인하여 증거법적 평가를 할 수 있기 때문에 무조건 배척해서는 안 된다. 이는 해시값이 다른 경우에도 마찬가지이다. 다만, 현재의 해시함수에 대한 암호학적 공격은 계속될 것이므로 향후 컴퓨팅 파워를 고려하면 현재의 MD5, SHA-1이 영원히 안전하다고 볼 수는 없다.

---

21) 이상진, 「디지털 포렌식 개론」(개정판), 이룬출판사, 2015, 285면; 충돌저항성은 제2역상 저항성과 달리 서로 다른 두 개의 입력값 x, x′를 임의적으로 선택하기 때문에 특정한 값이 주어진 상태인 제2역상 저항성에 비해 공격당할 가능성이 더 높은 것이다.

Ⅲ. 형사소송에서 해시값 생성과 동일성 입증의 한계

1. 통상적인 해시값 생성

디지털포렌식 전용도구인 Encase, FTK, X-ways와 경찰청에서 개발한 자체도구인 폴해시(Pol-Hash)는 수많은 해시함수를 내장하고 있는데 수사기관은 이 중에서 주로 MD5, SHA-1, SHA-256을 사용하고 있다.[22] 과거에 수사기관은 전용도구에 내장된 해시함수를 사용하였으나 이제는 자체 도구를 개발하여 사용할 해시함수를 직접 선택하고 있어 어떠한 기준으로 무엇을 선택했는지가 관심의 대상이 된다.

수사실무에서 동일성 입증은 복제본을 만든 다음 원본과 복제본의 해시값을 생성하여 비교한 다음 이를 기재한 확인서 2부를 작성하여 피처분자에게 서명하도록 한 후에 피처분자와 수사기관이 각 1부씩 보관하는 방법으로 이루어진다. 해시도구에 데이터를 입력하면 해시값이 자동적으로 생성되고 이때 동일한 데이터와 동일한 해시함수를 사용하면 그 결과로 도출되는 해시값은 항상 같기 때문에 법률적으로 다툴 쟁점이 없다. 다만 해시값이 변경되어 동일성 입증이 곤란한 경우나 해시값으로 동일성을 입증할 수 없는 경우가 문제가 된다.

2. 해시값이 변경되어 동일성 입증이 곤란한 경우

원본이나 사본에 베드섹터(Bad sector)가 있으면 그 상태에 따라 해시값이 매번 다르게 나타날 수 있다. 해시도구가 베드섹터(Bad sector)를 어떻게 읽었느냐에 따라서 입력값이 달라지기 때문이다. 한글파일을 암호화하는 경우 원본 파일, 암호화된 파일, 암호를 해제한 파일, 암호를 변경한 파일의 해시값이 모두 다르다.[23] 복제

22) 래리 다니엘,라즈 다니엘(번역 백제현), 「포렌식 전문가와 법률가를 위한 디지털 포렌식」, SYNGRESS·BJ퍼블릭, 2012, 220면; 경찰청, “2007 현장 수사관을 위한 디지털증거 압수수색 길라잡이”, 2017, 29면.

23) 이렇게 볼 때 법원의 압수·수색영장 별지인 “압수 대상 및 방법의 제한”에서 전자정보 봉인 및 개봉 방법 중의 하나로 “수사기관과 피압수자 등 쌍방이 암호를 설정하는 방법”을 적시하고 있는데 이는 피압수자가 암호 진술을 거부하면 증거를 확보할 수 없다는 점 외에도 해시값이 변경되는 문제를 내포하고 있다. 수사실무에서 원본에 암호를 설정한

본에 대한 해시값을 생성하였는데, 운반 과정에서 복제본 안에 저장되어 있는 개별 파일에 기술적 문제가 발생한 경우 전체의 해시값이 변경되기도 한다. 유사한 사례로 법원의 영장 별지에 따라 복제본에서 혐의사실과 관계없는 정보를 삭제하면 복제본 전체의 해시값이 달라지게 된다.[24] 이때 복제본에서 압수파일에 대한 해시값을 별도로 생성하지 않았다면 복제본을 가지고 있지 않는 이상 그 파일이 최초에 압수되었던 상태 그대로임을 입증할 수 없게 된다.

서버가 운영 중이면 해당 파일이 실시간으로 변경될 수 있어 압수 직후에 해시값을 생성하더라도 다를 수 있다. 스마트폰을 압수하여 물리적인 통제 상태에 놓였다 하더라도 피의자 또는 공범이 킬 스위치(Kill switch)를 이용하거나 원격에서 조작하는 방법으로 데이터를 변경하면 해시값이 달라진다. 물리 메모리 영역에서 수집하는 경우 정보가 계속 변경되고, 휘발성이 있기 때문에 마찬가지의 경우가 발생한다.[25]

MS-Office 2003 이전 버전에서는 엑셀 파일을 열람만 해도 원본의 해시값이 변경된다.[26] CD 등 보조기억장치는 아무런 접촉이 없어도 장시간 경과할 경우 데이터에 손상이 발생하여 해시값이 변경될 수 있다. IoT 등 임베디드(Embedded) 장비는 제조사 마다 하드웨어의 인터페이스가 다르기 때문에 디지털증거를 추출하기 어렵고, 그 과정에서 데이터가 변경되기도 한다. SSD(Solid State Drive) 메모리는 가비지 컬렉션(Garbage Collection) 기능이 있어 불필요한 데이터를 수시로 삭제하기 때문에 원본이 계속 변경된다.

다음에 해시값을 산출할 수 있으나 이는 이미 원본을 변경한 것이기 때문에 문제가 있다. 그렇다면 원본을 복사한 다음 위의 여러 가지 경우에 해당하는 파일을 만들어 각각의 해시값을 산출한 다음에 모두를 압수해야 하는 번거로움이 있어 수사현장에서 활용하기 어렵다.

24) 이완규, “디지털 증거 압수 절차상 피압수자 참여 방식과 관련성 범위 밖의 별건 증거 압수 방법”, 형사법의 신동향 통권 제48호, 2015.9, 117면; 법원의 영장 별지인 “압수 대상 및 방법의 제한”에서 복제한 저장매체에 대해 혐의사실과 관련된 전자정보만을 출력 또는 복사하여야 하고, 이와 같은 방법으로 증거 수집이 완료되고 복제한 저장매체를 보전할 필요성이 소멸된 후에는 혐의사실과 관련 없는 전자정보를 지체없이 삭제, 폐기하도로 규정하고 있다.

25) 서기민, 장기식, 김기범, “디지털 증거의 무결성 : 모든 디지털 증거의 무결성을 입증할 수 있는가?”, 디지털포렌식연구 제7호, 2010, 9면.

26) 경찰청, “2007 현장 수사관을 위한 디지털증거 압수수색 길라잡이”, 2017, 5면.

수사관·분석관이 증거를 수집하고 분석하여 법정에 제출하는 과정에서 실수나 조작미숙으로 해시값이 변경되는 경우도 종종 발생한다. 디지털증거를 보관, 운반하는 과정에서 전류나 물리적 충격으로 변경될 수 있고, 파일의 특성을 충분히 숙지하지 못해 열람만으로 해시값이 변경되기도 한다.

3. 해시값으로 동일성을 입증할 수 없는 경우

수사기관이 압수·수색영장으로 포털회사로부터 전자우편 내용을 회신 받거나 전기통신사업자로부터 특정기지국에 대한 역발신 자료를 회신 받을 때 해시값을 생성하지 않기 때문에 동일성을 입증할 수 없다. 제3자의 시스템에 저장되어 있는 원본과 수사기관에게 회신해준 사본 간의 동일성을 증명할 방법이 없다. 제3자가 전자우편이나 역발신 자료를 편집하여 제출하는 경우가 대부분이라면 해시값 비교도 의미가 없다. 형사소송법상 당연히 증거능력이 있는 통상문서(제315조)라고 해석한다 하더라도 디지털증거이고 복제되어 전달된다면 동일성 입증은 필요하다.

전기통신감청은 시간이 경과함에 따라 원본이 소멸되고 별도의 저장장치에 남아 있는 증거밖에 없기 때문에 원본과 사본에서 각각의 해시값을 생성해 비교하는 것이 불가능하다.

수사기관이 공개된 게시판이나 P2P 사이트 등에서 범죄혐의와 관련된 내용을 갈무리하거나 다운받을 때에도 해시값을 산출하지 않아 동일성 입증이 안 된다. 수사기관이 인터넷 공간에서 채증을 하거나 피의자로부터 디지털증거를 임의제출을 받을 때에도 해시값을 산출하고 있지 않다. 이때 해시값 산출이 필요한지 논란이 될 수 있는데 최근에 법원에서 임의 제출하는 경우에도 영장과 같은 엄격한 절차를 요구하기는 어렵지만 제출자로부터 원본이 조작되지 않았다는 취지의 확인을 받은 후 원본을 압수하거나 복사하는 경우에 사본을 봉인하여 무결성에 대한 합리적인 의심이 들지 않을 정도의 증명이 필요하다고 판결하였다.[27]

나아가 현재 사용하는 해시도구는 메타데이터의 동일성을 증명하지 못한다. 파일을 복제하면 생성시간 등 메타데이터가 변경되고, 복제가 언제 생성되었는지를 확

27) 서울중앙지방법원 2014. 4. 25. 선고 2013고합805 판결.

인할 수 있는 시점확인 서비스(Time Stamping Service) 기능도 없다. 해시값을 생성할 때 데이터 영역에 있는 정보만을 입력값으로 사용하고, 메타데이터 영역에 있는 데이터는 활용하지 않기 때문이다. 그래서 파일을 압수한 후에 파일의 이름, 확장자, 시간을 변경해도 해시값은 바뀌지 않기 때문에 동일성을 담보할 수 없다. 해시값은 디스크, 파티션, 파일 단위로 생성할 수 있는데, 디스크와 파티션은 메타데이터까지 포함하여 해시값을 산출하는 반면 파일은 메타데이터를 포함하지 않기 때문이다.

4. 소결

일반적으로 해시함수는 디지털증거의 동일성을 입증하는데 훌륭한 기술이라고 평가할 수 있다. 하지만 앞서 살펴본 바와 같이 해시값이 변경되어 입증이 곤란한 경우나 해시값만으로는 동일성을 입증할 수 없는 경우가 존재한다. 전자는 원본이나 사본에 베드섹터가 있는 경우나 수사관의 조작실수가 해당되고, 후자는 제3자 대상 전자우편 압수나 메타데이터가 포함될 것이다.

그런데 해시값으로 동일성을 입증할 때에 증거법적으로 보면 원본과 사본의 해시값이 동일하다는 것은 전자정보 확인서[28]에 의해 증명되기 때문에 결국 서면에 의해 입증이 이루어진다. 이때 해시값의 개념을 숙지하고, 해시값이 같으면 원본과 사본의 동일성이 인정된다는 사실을 이해하고 서명하는 경우가 얼마나 될지 의문이다.

따라서 이러한 경우에 디지털증거의 동일성 입증을 어떻게 해야 하는지, 동일성 입증을 위한 다른 방법은 없는지, 그렇게 했을 경우 증거능력은 인정되는지가 중요한 쟁점이 될 수밖에 없다.

28) 경찰청 훈령「디지털 증거 수집 및 처리 등에 관한 규칙」에 근거하고 있는 [서식1] 전자정보확인서, [서식2] 전자정보 확인서(간이)를 이용하고 있다.

Ⅳ. 국내·외 판례 및 법적·기술적 쟁점

1. 국내·외 판례 비교

가. 우리나라

2005년 녹음테이프의 동일성 입증에 관하여 대법원은 "대화내용을 녹음한 원본이거나 혹은 원본으로부터 복사한 사본일 경우에는 복사과정에서 편집되는 등의 인위적 개작 없이 원본의 내용 그대로 복사된 사본임이 입증되어야만 하고, 그러한 입증이 없는 경우에는 쉽게 그 증거능력을 인정할 수 없다."고 판결하였다.[29] 동일성 입증이 없는 경우에 증거능력을 부정하는 것이 아니라 쉽게 인정할 수 없다고 판시한 것으로 보아 그 정황에 대하여 다양한 방법으로 입증할 수 있으면 증거능력을 인정할 여지가 있다는 취지로 해석된다. 이때는 녹음테이프가 아날로그 방식이기 때문에 동일성 입증 수단으로 해시함수를 사용하지 않았다.

그 후 해시함수는 2007년 일심회 사건에서 디지털증거의 증거능력 인정요건을 제시할 때 등장한다. 대법원은 저장매체를 봉인하여 수사관서로 반출한 다음 세계적으로 인정받는 포렌식 도구(EnCase)[30]를 이용하여 이미징을 하고, 원본의 해시값과 복제본의 해시값이 동일하다는 근거를 들어 출력된 문건에 대한 동일성과 증거능력을 인정하였다.[31] 이어 왕재산 사건에서 대법원은 더 나아가 디지털증거의 동일성 입증 방법으로 해시함수를 원칙으로 규정하였다. 즉 "출력 문건과 정보저장매체에 저장된 자료가 동일하고 정보저장매체 원본이 문건 출력 시까지 변경되지 않았다는 점은, 피압수·수색 당사자가 정보저장매체 원본과 '하드카피' 또는 '이미징'한 매체의 해시(Hash) 값이 동일하다는 취지로 서명한 확인서면을 교부받아 법원에 제출하는 방법에 의하여 증명하는 것이 원칙"이라고 보았다.[32] 예외적인 사유로 해

29) 대법원 2005. 12. 23. 선고 2005도2945 판결; 대법원 2007. 3. 15. 선고 2006도8869 판결.

30) EnCase는 대법원 판결의 1심판결을 확인하고 저자가 추가한 것이다; 서울중앙지방법원 2007. 4. 16. 선고 2006고합1365, 2006고합1363(병합), 2006고합1364(병합), 2006고합1366(병합), 2006고합1367(병합) 판결.

31) 대법원 2007.12.13. 선고 2007도7257 판결.

32) 대법원 2013.7.26. 선고 2013도2511 판결.

시값을 이용한 방법에 의한 증명이 불가능하거나 현저히 곤란한 경우에 "정보저장매체 원본에 대한 압수, 봉인, 봉인해제, '하드카피' 또는 '이미징' 등 일련의 절차에 참여한 수사관이나 전문가 등의 증언에 의해 정보저장매체 원본과 '하드카피' 또는 '이미징'한 매체 사이의 해시 값이 동일하다거나 정보저장매체 원본이 최초 압수 시부터 밀봉되어 증거 제출 시까지 전혀 변경되지 않았다는 등의 사정을 증명하는 방법 또는 법원이 그 원본에 저장된 자료와 증거로 제출된 출력 문건을 대조하는 방법 등으로도 그와 같은 무결성·동일성을 인정할 수 있으며, 반드시 압수·수색 과정을 촬영한 영상녹화물 재생 등의 방법으로만 증명"하여야 하는 것은 아니라고 판시하였다.[33] 그런데 당시 고등법원 판결을 보면 "동일성과 표리의 관계에 있는 무결성(보관의 연속성)이 담보되는 한, 반드시 동일성 입증을 위하여 해쉬값을 비교 확인하는 절차가 요구된다고 할 수 없고, 법정에 제출된 원본과 출력문건을 직접 비교한 검증결과 등 제반사정을 종합하여 그 동일성 여부에 관하여 자유심증주의에 따라 객관적이고 합리적으로 판단하면 된다."고 판결[34]하여 다양한 방법을 종합적으로 고려하도록 하였다. 이렇게 보면 동일성 입증의 원칙과 예외에 관한 개념은 대법원에서 새롭게 정립한 것으로 보인다. 이후 이석기 사건에서 대법원은 "녹음파일의 생성과 전달 및 보관 등의 절차에 관여한 사람의 증언이나 진술, 원본이나 사본 파일 생성 직후의 해시(Hash)값과의 비교, 녹음파일에 대한 검증·감정 결과 등 제반 사정을 종합"하도록 판시하였고,[35] 이어 조세범처벌법위반 사건에서도 동일한 취지로 판결[36]하여 다양한 가능성을 열어 두었다.

나. 미국

미국 연방증거규칙 제901(b)(4)는 증거의 진정을 인정할 수 있는 사유 중의 하나로 "독특한 특성과 이에 유사한 것(Distinctive Characteristics) - 모든 정황과 함께 채택된 외관, 내용, 실질, 내부적 패턴이나 다른 독특한 특성"을 규정하고, 여기에

33) 대법원 2013.7.26. 선고 2013도2511 판결.
34) 서울고등법원 2013. 2. 8. 선고 2012노805 판결.
35) 대법원 2015. 1. 22. 선고 2014도10978 판결.
36) 대법원 2018. 2. 8. 선고 2017도13263 판결.

해시값이 포함된다고 해석하였다.[37] 폴 그림(Paul Grimm) 판사는 2007년 Lorraine v. Markel 사건[38]에서 디지털증거의 증거능력 인정요건으로 ① 관련성(Relevant) ② 진정성(Authentic) ③ 전문증거가 아니거나 예외사유 ④ 원본이거나 복사본일 경우 추가적인 조치 필요(Original or Duplicate) ⑤ 부당한 편견의 부재(Unfair Prejudice)를 제시하였는데, 여기에서 ② 진정성(Authentic) 증명 방법으로 해시값과 메타데이터[39]를 언급하였다.[40] 미국 법무부의 컴퓨터 압수수색 매뉴얼에서도 해시값이나 이와 유사한 포렌식 식별자를 사용하여 원본을 포렌식적으로 의미 있게 복사(copy)한 경우 진정성을 인정할 수 있다고 설명하고 있다.[41]

미국 법원은 디지털증거의 동일성 입증 수단으로 개별적 판결을 통해 MD5[42]와 SHA-1[43] 등을 인정하여 왔다. 법원에서 개별적인 해시함수에 대해서 심사하였다는 측면에서 우리나라보다 논의가 더 진행되었다고 볼 수 있다. 다수의 판결문에서 전문가의 증언을 인용하여 해시함수가 일치할 확률을 99.99××라고 표현하면서 중복될 가능성이 없다고 서술하고 있으나 이는 수학적 근거가 아닌 기술적으로 불가능하다는 사실을 은유적으로 표현한 것이다.[44] 한편, 해시함수가 아동음란물 탐지에도 활용되었는데, 미국 제8항소법원은 스페인으로부터 미국 소재 피고인의 컴퓨터에 저장되어 있는 아동음란물의 해시값을 전달받아 이를 근거로 영장을 발부받아 영장을 집행한 사안에서 피고인 측이 해시함수가 수색영장 발부를 위한 상당한 근거가 될 수 없다고 주장하면서 다른 두 개의 파일이 충돌(collide)하거나 중복(override)되는 해

37) Lorraine v. Markel American Ins. CO., 241 F.R.D. 534, 546-47 & n.23 (D. Md. 2007).

38) Lorraine v. Markel American Ins. CO., 241 F.R.D. 534, 546-47 & n.23 (D. Md. 2007).

39) 전현욱, 김기범, 조성용, Emilio C. VIANO, "사이버범죄의 수사 효율성 강화를 위한 법제 개선 방안 연구", 경제·인문사회연구회 미래사회 협동연구총서 15-17-01, 2015, 355~356면.

40) Lorraine v. Markel American Ins. CO., 241 F.R.D. 534, 546-47 & n.23 (D. Md. 2007).

41) H. Marshall Jarrett Director(EOUSA), Michael W.Bailie Director(OLE), Ed Hagan Assistant Director(OLE), "SEARCHING AND SEIZED COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS", Office of Legal Education Executive Office for United States Attorneys, 2009, p.199.

42) United States v. Crist, 627 F. Supp. 2d 575, 578, 585 (MDPA 2008).

43) United States v. Glassgow, 682 F.3d 1107, 1110 n.2 (8th Cir. 2012); United States v. Miknevich, 638 F.3d 178, 181 n.1 (3rd Cir. 2011).

44) United States v. Glassgow, 682 F.3d 1107, 1110 n.2 (8th Cir. 2012); State v. Mahan, 2011 Ohio 5154, n.2 (Court of Appeals, 8th Appellate Dist. Ohio 2011).

시값을 가질 수 있다고 주장하였으나 법원은 이를 받아들이지 않았다.[45]

다. 소결

우리나라와 미국의 판례를 살펴보면 해시함수가 디지털증거의 동일성을 입증하는 기술로 널리 활용되고, 법원에서도 신뢰하고 있는 것은 분명하다. 현재가지 MD5, SHA-1에 대한 과학적 신뢰성도 논란의 여지는 없어 보인다. 그런데 우리나라 대법원은 동일성 입증에 엄격한 태도를 취하고 있어 해시값이 변경된 경우 증거능력을 인정하지 않을 가능성이 높다. 대법원 판례에서 동일성을 입증하기 위한 증거능력 인정요건으로 제시하고 있기 때문이다. 반면 해시값으로 동일성을 증명할 수 없는 경우에는 대법원에서 다양한 입증방법을 허용하고 있어 증거능력을 인정받을 수 있는 여지가 생겼다고 보여진다. 다만, 동일성 입증의 원칙으로 해시값을 규정하였다가 최근에 다양한 방법으로 입증할 수 있다고 판결하여 대법원의 입장이 변경된 것인지가 모호한 실정이다.

2. 법적·기술적 쟁점

가. 과학적 증거 여부

증거능력을 평가하기 위해서는 해시함수가 먼저 과학적 증거에 해당하는지 살펴보아야 한다. 일반적으로 과학적 증거는 특정한 자연과학의 영역에서 활용되는 원리와 기술을 응용하여 수집된 증거로 특정 영역의 전문가가 과학적 방법으로 증거를 분석하여 현출한 내용물을 말한다.[46] 사실적 측면에서 일정한 사상 또는 작용에 관하여 통상의 오감에 의한 인식을 뛰어넘는 수단과 방법을 사용하여 인지하고 분석한 것을 말하고, 규범적 측면에서 그와 같은 인지와 분석의 기초가 되는 원리로서 과학적 근거가 있고 그 수단과 방법이 타당성과 신뢰성을 가지고 있어서 인지·분석에 의한 판단결과가 형사재판에서 증거로 허용되는 것을 말한다.[47] 증거능력 및 증

45) United States v. Cartier, 543 F.3d 442, 444(8th Cir. 2008).

46) 김준성, "형사소송에서 과학적 증거의 허용범위", 강원법학 제40권, 강원대학교 비교법학연구소, 2013, 298면.

47) 권영법, "과학적 증거의 허용성 - 전문가증인의 허용성 문제와 관련 쟁점의 검토를 중심으

명력의 인정여부에 따라 크게 ① 가장 강한 정도의 과학적 증거[48], ② 반증의 여지가 있는 과학적 증거[49], ③ 과학과 전문가의 주관적 분석이 결합한 증거[50], ④ 실험의 근간을 이루는 이론의 적정성을 전제 하는 과학적 증거[51]로 구분할 수 있는데,[52] 이는 과학적 증거의 스펙트럼이 다양하고 그에 따라서 증거능력과 증명력을 판단할 때는 결국 개별적으로 심사할 수밖에 없다는 것을 보여준다.

그런데 해시함수는 지문·DNA·거짓말탐지기와 달리 유죄를 입증하는 증거가 아니라 디지털증거의 동일성을 입증하는 기술에 해당한다. 거짓말탐지기결과, 필적감정, 위드마크(Widmark), 지문과 달리[53] 수학적 알고리즘으로 전문가 수준의 지식·기술·경험을 요구하지 않는다. 수사현장에서도 사이버수사관, 디지털분석관 뿐만 아니라 일반 수사관도 사용하고 있다. 동일한 증거에 동일한 해시도구를 사용하면 누가 사용해도 동일한 해시값이 산출되기 때문에 사용자의 지식·경험에 따른 오류가 발생하지 않는다. 따라서 앞서 정의한 개념을 전제로 하면 해시함수는 과학적 증거에 해당한다고 볼 수 없다. 이는 과학적 증거에 포함되지 않는 것이 아니라 과학적 증거로 논의하는 것 자체가 타당하지 않다는 것을 의미한다.

하지만 기술적 문제로 해시함수가 변경되는 경우 디지털증거를 과학적 증거의 영역에서 논의할 수 있을 것이다. 디지털증거의 추출, 복제, 운반, 보관 등의 과정에서 해시값이 변경될 수 있는데 이는 해시함수의 알고리즘의 문제라기보다는 하드웨어, 운영체제, 파일시스템, 네트워크를 비롯한 이미징 기술과 관련되어 발생하는 문제로 봐야 한다. 이 경우에는 과학적·기술적 원리와 방법으로 원인을 규명해야 하고, 전문가의 지식·경험·기술이 필요할 뿐만 아니라 전문가 별로 해석에서 차이가 발생할

로-", 법조 통권 제667호, 법조협회, 2012, 85면; 신이철, "과학적인 감정자료를 기초로 한 감정서의 증거능력", 외법논집 제33권 제1호, 한국외국어대학교 법학연구소, 2009, 551면.

48) 대법원 2010.3.25. 선고 2009도14772 판결.

49) 대법원 2009.3.12. 선고 2008도8486 판결.

50) 대법원 2000.12.22. 선고 99도4036 판결.

51) 대법원 2005.5.26. 선고 2005도130 판결.

52) 조병구, "과학적 증거에 대한 증거채부결정 – 합리적 증거결정 기준의 모색-", 형사법실무연구, 재판자료 제123집(2011년하), 법원도서관, 2011, 612면.

53) 이웅혁, 이성기, "형사재판상 과학적 증거의 기준과 국내 발전방향 - 최근 미국에서의 과학적 증거의 개혁 논의를 중심으로 -", 형사정책 제23권 제1호, 한국형사정책학회, 2011, 310~313면, 320면.

수 있다. 따라서 과학적 증거에서 논의처럼 오류 발생빈도가 최소화될 수 있는 영역에서 그 활용범위나 수준이 결정되고, 그 범위 안에서 증거로서 가치가 있고, 구체적인 신뢰성을 통해 과학적 확실성이 합리적인 정도로 인정되면 증거능력을 인정할 수 있을 것이다.[54)]

나. 동일성 입증의 원칙으로서의 적절성

대법원은 왕재산 사건에서 동일성 입증의 원칙적 방법으로 해시함수를 규정하고 조사자 증언·진술, 보관연속성 증명, 법정에서 원본과 사본의 대조, 영상촬영물 등을 예외로 규정하였지만, 이석기 사건과 조세범처벌법 사건에서 다양한 입증방법을 종합적으로 검토하도록 규정하였다. 해석에서 모호함이 있지만 수사기관은 여전히 '원칙과 예외'의 기준을 활용하고 있다. 해시함수가 낮은 비용으로 높은 신뢰성을 가져오기 때문에 원칙으로 규정하여 활용하는 것도 분명 의미가 있다.

하지만 '원칙과 예외'는 예외적인 방법이 자칫 부차적인 수단으로 전락할 우려가 있고, 다양한 입증 방법을 만들어야할 동인을 제거하게 된다. 그래서 '원칙과 예외'는 수사현장에서 여건과 상황에 따라서 판단할 문제이지 대법원에서 규정할 사안은 아니다. 물리적 증거에서 증거능력이 인정되는 관련자 증언·진술, 검증·감정결과, 영상촬영물 등이 디지털증거에서 예외적 요건으로 인용되는 것도 납득하기 어렵다. 현장을 고스란히 촬영한 영상녹화물이나 복제본을 밀봉하여 보관연속성(Chain of Custody)이 증명된 디지털증거의 동일성 증명방법이 해시값보다 못하다고 할 수 없다. 즉, 영상촬영물은 누구의 컴퓨터에서 어떤 폴더의 파일을 복제하였는지, 수사기관의 억압은 있었는지, 참여인은 누구인지, 어떠한 수사관이 조작하고 무슨 도구를 사용했는지까지 증명할 수 있기 때문에 그 무엇보다도 증거가치가 높다. 나아가 해시함수라는 특정기술을 원칙으로 규정하는 것은 동일성 입증을 위한 새로운 기술의 개발을 가로막는 문제도 있다.

다. 오류율(Error rate) 측정

54) 황만성, "과학적 증거의 증거능력과 증명력-유전자감식결과를 중심으로", 형사정책연구 제18권 제3호(통권 제71호), 2007·가을호, 798면.

해시함수가 과학적 증거라고 볼 수 없어도 그 기술의 신뢰성에 대한 평가는 필요하고, 이를 위해서 오류율을 이해하여야 한다. 오류율은 특정 기술이나 방법이 잘못된 결과를 제시할 수 있는 확률로 적용되는 원리나 기술, 방법이 과학적으로 타당한지를 판단하는데 활용된다.[55] Daubert Test에서 제시된 개념(known error rate)으로 과학적 증거의 증거능력을 인정하는데 중요한 요소로 작용한다.[56] 일반적 오류율은 일정 증거에 사용되는 표준화된 절차와 특정 장비에서 발생하는 것을 말하고, 구체적 오류율은 문제가 되는 증거를 분석하고 실험한 방법이나 기술의 일반적 오류율뿐만 아니라 이를 행한 실험실의 조건, 분석가의 오류율도 함께 고려한 것을 말한다.[57]

이렇게 볼 때 동일성을 입증하는데 사용되는 해시함수의 오류율은 제2역상을 찾을 가능성을 의미한다고 볼 수 있다. 특정한 증거의 해시값과 일치하는 해시값을 갖는 또 다른 증거를 찾아낼 가능성을 말한다. 그런데 해시함수의 오류율은 DNA와 달리 확률(%) 형태로 표현할 수 없다. 해시값을 산출할 수 있는 모든 경우의 수는 비트 수(n)에 따라 2^n으로 한정되지만 입력에 사용되는 디지털증거의 수는 무한하기 때문에 확률적으로 표현이 안 된다. 다만, 이론상 MD5는 해시값의 총 개수가 2^{128}개이고, SHA-1은 2^{160}개, SHA-256은 2^{256}개이기 때문에 SHA-256의 오류율이 보다 낮다고 할 수 있을 것이다. 나아가 해시함수는 데이터가 1bit만 바뀌어도 해시값이 완전히 변경되는 쇄도효과(Avalanche Effect)를 가지고 있어 두 개의 디지털증거에 대한 해시값인 h(x)와 h(x')의 유사도를 확률적으로 비교해봐야 의미가 없다. 따라서 해시값이 100% 일치하는 것이 아니라면 나머지는 수학적·확률적으로 의미가 없어 사실상 0%라고 보는 것이 타당할 것이다.

55) 이웅혁, 이성기, "형사재판상 과학적 증거의 기준과 국내 발전방향 – 최근 미국에서의 과학적 증거의 개혁 논의를 중심으로 -", 형사정책 제23권 제1호, 2011.6, 328면.

56) United States v. Crisp, 324 F. 3d 261(4th Cir. 2003).

57) 이웅혁, 이성기, "형사재판상 과학적 증거의 기준과 국내 발전방향 – 최근 미국에서의 과학적 증거의 개혁 논의를 중심으로 -", 형사정책 제23권 제1호, 2011.6, 328면; 지문, 혈흔형태분석, 족적 등은 전문가 숙련도에 따라 오류율이 달라지기 때문에 일반적 오류율뿐만 아니라 구체적 오류율까지 심사해야 한다.

Ⅴ. 해시함수 활용성과 신뢰성 개선방안

1. 해시값이 변경된 경우 증거능력 인정 방안

해시값이 정상적으로 생성된 경우 동일성 증명에 아무런 문제가 없지만, 해시값이 쉽게 변경되거나 해시값으로 동일성을 증명할 수 없는 경우에는 다양한 입증 방법을 개발하고 활용해야 한다. 대법원에서 동일성 입증 방법으로 해시값 뿐만 아니라 사람의 증언이나 진술, 보관연속성, 영상녹화물 등 다양한 방법을 인정하고 있는 만큼 해시값 변경이 가능한 다양한 기술적 상황에 대한 실험과 검증을 통해 인과관계를 규명하는 노력을 해야 한다. 하드웨어의 문제로 해시값이 다른 경우 동일한 조건의 하드웨어에서 반복적으로 해시값을 산출하여 변경된 해시값이 반복적으로 산출된다는 것을 증명하면 될 것이다. IoT장비와 같이 원본의 데이터 추출을 위하여 부득이 증거의 손상이 발생하는 경우나 원본 자체와 동일성을 비교할 수 없는 전기통신감청의 경우에는 전문가를 참관시키는 것도 방법일 것이다. 나아가 이러한 상황을 체계화하여 기술표준으로 정립하여 집행방법을 객관적으로 인정받는 노력도 필요하다.

이와 같은 노력을 한다 하더라도 증거능력이 인정되지 않는다면 의미가 없을 것이다. 하지만 해시값이 변경된 경우 그 과정에 대한 합리적 의심이 없을 정도의 증명을 한다면 증거능력이 인정할 수 있을 것이다. 해시값이 변경되었다고 무조건 증거능력을 부정하는 것이 아니라 다른 방법으로도 증명이 되지 않을 경우에 증거능력을 부정해야 한다. 제주도지사 사건의 대법원 판결[58]에서도 위법수집증거 배제법칙의 예외적인 경우에 증거능력을 인정할 수 있는 여지를 두고 있다. 즉, “수사기관의 절차위반 행위가 적법절차의 실질적인 내용을 침해하는 경우에 해당하지 아니하고, 오히려 그 증거의 증거능력을 배제하는 것이 헌법과 형사소송법이 형사소송에 관한 절차조항을 마련하여 적법절차의 원칙과 실체적 진실규명의 조화를 도모하고 이를 통하여 형사사법정의를 실현하려 한 취지에 반하는 결과를 초래하는 것으로 평가되는 예외적인 경우라면 법원은 그 증거를 유죄인정의 증거”로 사용할 수 있다

58) 대법원 2007.11.15. 선고 2007도3061 판결.

는 것이다.[59] 동일성 입증은 증거능력 인정 요건으로 검사가 그 존재에 대하여 구체적으로 주장·입증하여야 하는 것이지만 소송상의 사실에 관한 것으로 엄격한 증명을 요하지 않고, 자유로운 증명으로 족하기 때문에 반드시 법률상 증거능력 있고 적법한 증거조사를 거친 증거에 의하여야 하는 것이 아니다.[60] 나아가 최근 마약사건에 대한 대법원 판결에서도 경찰관이 피고인의 머리카락을 뽑고, 소변을 증거물병에 담은 다음에 그 자리에서 별도의 봉인 조치를 하지 않고 밖으로 반출하였고, 이후 조작훼손·첨가를 막기 위하여 어떠한 조처가 행해졌고 누구의 손을 거쳐 국립과학수사연구원에 전달되었는지 확인할 수 없는 사안에서 합리적 의심의 여지가 없을 정도로 진실임을 확신하게 하는 증명력을 가진 증거라고 보기 어렵다며 증명력 문제로 판단한 사례도 있다.[61]

따라서 수사기관이 의도적으로 해시값을 조작한 경우에는 증거능력을 당연히 부정해야겠지만 증거의 저장·처리 과정에서 기술적 문제로 변경된 경우에는 기술적 착오 내지 결함으로 보아 증거능력을 인정하고, 증명력의 문제로 판단해서 법관의 자유판단에 맡기는 것이 필요하다.[62] 다만, 기술적 문제와 수사관의 의도적 변경을 구분하기 어렵기 때문에 무조건적으로 또는 쉽게 증거능력을 인정할 수는 없을 것이다.

2. 해시도구의 기술표준 마련 및 성능 개발

디지털증거의 동일성을 증명하는데 사용할 수 있는 해시도구의 기술 표준이 필요하다. 가장 먼저 해시함수는 종류가 많고, 안전성도 다양하기 때문에 법정에서 허용될 수 있는 기술 규격을 만들어야 한다. 이때 경찰·검찰 등 수사기관, 국립과학수사연구원 등 법과학 기관, 디지털포렌식 관련 학회, 기술표준 기관 등 다양한 이해관계자들이 참여하여야 할 것이다. 나아가 해시값이 변경되는 경우에 대비하여 다양

59) 대법원 2007.11.15. 선고 2007도3061 판결.

60) 서울고등법원 2013. 2. 8. 선고 2012노805 판결; 대법원 2001.9.4. 선고 2000도1743 판결.

61) 대법원 2018. 2.8. 선고 2017도14222 판결.

62) 전현욱, 김기범, 조성용, Emilio C. VIANO, "사이버범죄의 수사 효율성 강화를 위한 법제 개선 방안 연구", 경제·인문사회연구회 미래사회 협동연구총서 15-17-01, 2015, 414면.

한 기능을 도입하여야 한다. 해시값을 산출하는 속도를 높이기 위해 복제본을 분할하여 해시값을 산출하고 그 해시값을 다시 입력값으로 사용하여 파일 전체의 해시값을 생성하는 병렬처리 방법도 있다.[63] 복제본 중 일부가 손상되었거나 복제본 중에서 혐의사실과 관련성이 없는 증거를 삭제한 경우 손상·삭제된 증거의 해시값을 알고 있다면 정상적인 증거의 해시값과 이중해시를 하여 복제본의 해시값과 비교해 동일성을 입증할 수도 있다.[64] 이러한 기능을 이용하여 해시값끼리 해싱하여 해시값을 계산할 수 있는 도구를 개발해야 한다.[65]

또한 메타데이터 영역까지 동일성을 입증할 수 있는 해시도구가 시급하다. 파일을 복제할 때 누락되는 메타데이터를 확보하고, 이를 통해 메타데이터가 인위적으로 조작되었다는 공격을 차단할 수 있을 것이다. 유닉스 계열의 dd나 파일을 복사할 때 $MFT/entry에 접근하는 기능을 포함한 해시도구가 필요하고, 이에 대한 신뢰성을 평가할 수 있는 기준과 절차도 함께 마련하여야 한다.

Ⅵ. 결론

이상에서 해시함수의 형사법적 함의에 대하여 알아보았다. 해시함수는 동일성 입증에서 중요한 기술인 점은 분명하지만 해시값이 변경되는 경우나 해시값을 산출할 수 없는 경우가 발생하기 때문에 이에 대한 법·기술적 대책마련이 필요하다. 이를

63) 신용학, 김도원, 이창훈, 김종성, "암호학적 관점에서의 EWF 파일 이미징 효율성 개선 방안 연구", 정보보호학회논문지 Vol.26 No.4, 2016.8, 917면.

64) 이상미, "관련성 없는 디지털증거 삭제시 이중해쉬를 이용한 무결성 입증 방안", 서울대 융학과학기술대학원 석사학위논문, 2016.2, 23~29면; 신용학, 김도원, 이창훈, 김종성, "암호학적 관점에서의 EWF 파일 이미징 효율성 개선 방안 연구", 정보보호학회논문지 Vol.26 No.4, 2016.8, 917면.

65) 이상미, "관련성 없는 디지털증거 삭제시 이중해쉬를 이용한 무결성 입증 방안", 서울대 융학과학기술대학원 석사학위논문, 2016.2, 35면; 나아가 해당 논문에서는 효율적인 해시값 계산이 가능한 해시 트리 구성의 문제는 디지털증거의 구조에 따라 달라질 것인데, 어떤 기준으로 트리화할 것인지, 효율적인 트리구조는 어떻게 구성할 것이인지, FAT·NTFS 등 파일 시스템에 따라 트리 구조 형성은 어떤 방법을 해야 하는지 등에 관한 추가적인 논의가 필요하다고 제언하고 있다.

위해 대법원은 동일성 입증에 관하여 '원칙과 예외' 프레임을 폐기하고, 다양한 입증방법을 강구하도록 하여야 한다. 해시함수는 그 자체로는 수학적 알고리즘이기 때문에 과학적 증거라고 볼 수 없지만 해시값이 기술적 문제로 변경되는 경우에는 과학적 증거로 볼 수 있어 증거능력을 인정받을 여지가 있다.

이에 대한 실천방안으로 해시값이 변경되는 경우에도 다양한 실험과 검증을 통하여 인과관계를 입증하는 노력이 필요하다. 이를 바탕으로 기술적 문제나 조작실수 등에 의한 경우라면 그 과정이 합리적 의심 없이 증명될 경우에는 증거능력을 인정할 수 있을 것이다. 나아가 해시도구에 대한 기술표준을 마련하고, 메타데이터의 동일성을 입증할 수 있는 해시도구를 만들고 그에 대한 검증 절차와 방법까지 제도화할 것을 제안하였다.

참고문헌

1. 국내문헌

경찰청, “2007 현장 수사관을 위한 디지털증거 압수수색 길라잡이”, 2017

권영법, “과학적 증거의 허용성 - 전문가증인의 허용성 문제와 관련 쟁점의 검토를 중심으로-”, 법조 통권 제667호, 법조협회, 2012

김준성, “형사소송에서 과학적 증거의 허용범위”, 강원법학 제40권, 강원대학교 비교법학연구소, 2013

래리 다니엘,라즈 다니엘(번역 백제현), 「포렌식 전문가와 법률가를 위한 디지털 포렌식」, SYNGRESS·BJ퍼블릭, 2012

박재유, “디지털증거의 무결성 입증시 MD5 해시함수 사용에 관한 재고찰”, (사)한국디지털포렌식학회 2016 디지털 포렌식 기술 워크샵, 2016

서기민, 장기식, 김기범, “디지털 증거의 무결성 : 모든 디지털 증거의 무결성을 입증할 수 있는가?”, 디지털포렌식연구 제7호, 2010

신용학, 김도원, 이창훈, 김종성, “암호학적 관점에서의 EWF 파일 이미징 효율성 개선 방안 연구”, 정보보호학회논문지 Vol.26 No.4, 2016

이상미, “관련성 없는 디지털증거 삭제시 이중해쉬를 이용한 무결성 입증 방안”, 서울대 융학과학기술대학원 석사학위논문, 2016

이상진, 「디지털 포렌식 개론」(개정판), 이룬출판사, 2015

이완규, “디지털 증거 압수 절차상 피압수자 참여 방식과 관련성 범위 밖의 별건 증거 압수 방법”, 형사법의 신동향 통권 제48호, 2015

이웅혁, 이성기, “형사재판상 과학적 증거의 기준과 국내 발전방향 - 최근 미국에서의 과학적 증거의 개혁 논의를 중심으로 -”, 형사정책 제23권 제1호, 한국형사정책학회, 2011

신이철, “과학적인 감정자료를 기초로 한 감정서의 증거능력”, 외법논집 제33권 제1호, 한국외국어대학교 법학연구소, 2009

전현욱, 김기범, 조성용, Emilio C. VIANO, “사이버범죄의 수사 효율성 강화를 위

한 법제 개선 방안 연구", 경제·인문사회연구회 미래사회 협동연구총서 15-17-01, 2015

조병구, "과학적 증거에 대한 증거채부결정 – 합리적 증거결정 기준의 모색-", 형사법실무연구, 재판자료 제123집(2011년하), 법원도서관, 2011

황만성, "과학적 증거의 증거능력과 증명력-유전자감식결과를 중심으로", 형사정책연구 제18권 제3호(통권 제71호), 2007·가을호

2. 국외 문헌

E. H. Mckinney, "Generalized Birthday Problem", The American Mathematical Monthly Vol. 73 No.4, Apr 1966

Elaine Barker, Allen Roginsky, "Transitions: Recommendation for Transitioning the Use of Cryptographic Algorithms and Key Lengths", NIST Special Publication 800-131A, 2011

Eric Thompson, "MD5 collisions and the impact on computer forensics", Digital Investigation, Volume2 Issue1, 2005

Fr´ed´eric Muller, "The MD2 Hash Function Is Not One-Way", Advances in Cryptology - ASIACRYPT 2004, 10th International Conference on the Theory and Application of Cryptology and Information Security, 2004

H. Marshall Jarrett Director(EOUSA), Michael W.Bailie Director(OLE), Ed Hagan Assistant Director(OLE), "SEARCHING AND SEIZED COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS", Office of Legal Education Executive Office for United States Attorneys, 2009

Kelsey, John, and Bruce Schneier, "Second preimages on n-bit hash functions for muchless than 2n work." Annual International Conference on the Theory and Applications of Cryptographic Techniques. Springer Berlin Heidelberg, 2005

Morris J. Dworkin, “SHA-3 Standard: Permutation-Based Hash and Extendable-Output Functions”, U.S. Department of Commerce, National Institute of Standards and Technology, FIPS PUB 202, 2014

National Institute of Standards and Technology, “Secure hash standard”, U.S. Department of Commerce, FIPS 180-1, 1995

RFC 1321, “The MD5 Message-D igest Algorithm”, Internet Request for Comments 1321, 1992

Rivest, Ronald, “The MD5 message-digest algorithm”, 1992

Rogaway, Phillip, and Thomas Shrimpton, “Cryptographic hash-function basics: Definitions, implications, and separations for preimage resistance, second-preimage resistance, and collision resistance”, International Workshop on Fast Software Encryption, 2004

Stevens, M., Lenstra, A.K. and de Weger, B. “Chosen-prefix collisions for MD5 and applications”, Int. J. Applied Cryptography, Vol.2 No.4, 2012

X. Wang, H. Yu, “How to Break MD5 and Other Hash Functions”, Advances in Cryptology – Proceedings of EUROCRYPT 2005, LNCS 3494, 2005

A Study on Hash Function in Criminal Procedure

Kim, Gi-bum*

The hash function is used in verifying the integrity of the digital evidence in the field of digital forensics and cyber investigation since mid-1990s. The identification requirement was introduced by the Supreme Court ruling on 일심회 case in 2007. Subsequently the hash value was accepted as the method of providing the identification in 왕재산 case, as well as in 이석기 case. Now the hash function is commonly used in the scenes of search and seizure to secure the identification. It importance of using hash function elevated further, with the recent amendment of the Criminal Procedure Law in 2011 which introduced printing and copying as the main method of seizure of digital evidence.

However, the Supreme Court did not provide guidance or discussed enough in details on matters of security of hash function, acceptable technical specifications, technical limitations. The ruling does not give guidances in situations where the original data is in constant change or inability to produce stable hash value. We have seen lack of discussions on hash function as the academic interest in the field of scientific evidence was largely limited to fingerprints, polygraphs, DNA and other biological objects. The question remains unanswered whether to accept the hash function as the standard method of providing identification.

In this paper we evaluated the security of the hash function as the method to provide identification, and discussed the processes and limitations when used in the criminal procedure. We concluded that hash functions which are being questioned of their security, such as MD5 and SHA-1, are safe to be accepted in the context of criminal procedure.

* First author: Professor, Korean National Police University

We have confirmed that the hash function is widely used in the courts of law in the Republic of Korea and the United States, and we identified three issues surrounding hash functions: whether the hash value falls in the category of scientific evidence, whether it is suitable to adopt it as standard method to provide identification, and how to define the error rates for hash function.

In conclusion, we suggest that various alternative methods should be considered in order to secure admissibility of the evidence to the court in case of absence of identification provided by hash function. Furthermore we suggest to develop standard technical specifications of the hash functions and increase functionality of hash tools to include metadata from the file systems.

❖ Keyword: Hash Function, Digital Evidence, Admissibility of Evidence, Scientific Evidence, Crime Investigation, Hash Tool, Identification, Integrity, Chain of Custody

투고일 : 5월 31일 / 심사일 : 6월 21일 / 게재확정일: 6월 21일