



To Hash, or not to Hash... That is NOT the Question

When Hashing Creates a False Sense of Security

By Jacques R. Francoeur (B.A.Sc., M.A.Sc., MBA)
ProofSpace, Inc.

Abstract

There has been extensive discussion and use precedent in the legal and security fields around hashing, its use and its value — real and perceived. Even though its robustness and usefulness are incontrovertible in the security world, hashing's perceived value in the legal field (for the purpose of establishing the authenticity of Electronically Stored Information) could benefit from some clarification. For a hash to be secure and useful for legal applications, there must be some additional mechanism to protect and preserve the unique association between the data that is hashed, a trusted time datum, and the original hash result.



Abstract

There has been extensive discussion and use precedent in the legal and security fields around hashing, its use and its value — real and perceived. Even though its robustness and usefulness are incontrovertible in the security world, hashing's perceived value in the legal field (for the purpose of establishing the authenticity of Electronically Stored Information) could benefit from some clarification. For a hash to be secure and useful for legal applications, there must be some additional mechanism to protect and preserve the unique association between the data that is hashed, a trusted time datum, and the original hash result.

Hashing is essential, but not sufficient to the demonstration of what I like to call "intrinsic authenticity", or the ability to prove at the data-level at anytime that a specific state-of-data existed at a particular time and that it has not changed since that time. This is distinct from "inferred" approaches which presume authenticity based on the presence of external factors, such as access controls and trusted insiders. Unfortunately it is clear from recent events that insiders are not always trustworthy — nor are controls always effective in keeping malicious outsiders out.

Some in the legal community do believe that hashing alone is a satisfactory method to ensure the integrity of Electronically Stored Information (ESI) over time. However, I believe that hashing alone is insufficient in demonstrating the authenticity of a record, even when stringent perimeter-based protections of the hash are in place. The key issues to clarify are: (1) what assurances does hashing alone, in fact, provide? And (2) what are the incremental benefits of cryptographically-bound hashing (i.e., Trusted Timestamping)?

One of the core legal prerequisites for the admissibility of ESI is authenticity¹ — i.e. is the record what it purports to be — an accurate representation of the record in the same state it was in at the time an assertion was made. This means that the issue is not only one of maintaining the integrity of a record (during the discovery process, for example) but of demonstrating the continued authenticity (and maintained integrity) of the record from the time the "assertion" in question was made up to the present (e.g., contact was signed, official corporate record was declared, customer advise was given).

A hash can be easily generated and placed into the metadata (or associated to the original data) but the data can just as easily be manipulated, and a fresh hash regenerated by a malicious insider (or outsider, for that matter) and reinserted. That is, if one can circumvent the controls around the record, such as would be the case in criminal or organized crime hacking, or if one is in control of the controls around the data, such as would be the case for a System Administrator, or if an executive can enter into collusion with or coerce an individual... then one can go into the records management system, falsify the record, create a new fraudulent hash of the falsified

¹ Federal Rules of Evidence: ARTICLE IX. AUTHENTICATION AND IDENTIFICATION, Rule 901. Requirement of Authentication or Identification, (a) General provision.—*"The requirement of authentication or identification as a condition precedent to admissibility is satisfied by evidence sufficient to support a finding that the matter in question is what its proponent claims."*



record and reinsert the fraudulent hash into the metadata of the falsified record. When tested, a fresh hash of the falsified record will match the fraudulent hash in the metadata of the falsified record. If this cannot be prevented or detected, then the incredible power of hashing can actually create a situation wherein falsified information that includes a hash is perceived to be “irrefutably authentic.” Mechanisms to prevent this are essential.

A hash alone is very useful in determining whether two records are identical, but insufficient in determining whether a record is unchanged. Why? Because the hash itself contains no information about the time the hash was created, or who created it. In fact, modern de-duplication applications that use hashing are premised on the first use-case, where the only question is: does a given record produce the same hash as another record? If yes, they are identical. The only notion of time is the time of comparison — not creation. However, the second use-case (is the record unchanged) requires more than just a test of integrity (i.e., hash comparison). “Is the record unchanged” begs an answer to the question — unchanged from when? It must be from the time the “assertion” in question was made, and not some other arbitrary time, such as when it has been identified relevant to litigation or regulatory proceedings.

Once a hash is generated then the issue shifts to the “chain of custody” around the record and hash and when the hash was generated. Until the hash and record content are cryptographically bound to time there is a realistic possibility that manipulation can and will occur, given that the skill necessary to perform such a manipulation is not high. If fraudulent behavior is not a concern, an alternative scenario would be as follows. Records are exchanged between parties in many business processes. During a dispute, two what are purported to be identical records show up in court as materially different, how does one establish which one is the authentic version? Both sides have purported to have the authentic record, both have a hash in the metadata that match with their record version.

The problem is that although a hash is a unique digital fingerprint of a set of data, alone it is “floating” or unanchored to any reference which is beyond the control of insiders or outsiders. Hashing is fundamental and a key step in the right direction, but it must be used in the correct way — hashing alone is “necessary but insufficient.” The best independent reference datum is time, more specifically “trusted time” that cannot be manipulated and is widely witnessed.

“Outside in” or perimeter-based approaches to preserving the chain-of-custody that depend on system or application controls (or trustworthy individuals) are complex, lower assurance, more vulnerable to challenges and therefore more costly to refute than persistent data-level intrinsic approaches derived from cryptographic hash binding — Intrinsic Authenticity. The technology for cryptographically binding hashes to trusted time stamp information is well understood and embodied in the American National Standards Institute (ANSI) ASC X9.95-2005 standard.

Some say that binding the hash is “unnecessary”, it is “far more than what is needed”, or that the “insider threat is unlikely and not a serious enough issue” to warrant the levels of assurance promised by trusted timestamping. To this the Honorable Paul W. Grimm might reply²:

“If it is critical to the success of your case to admit into evidence computer-stored records, it would be prudent to plan to authenticate the record by the most rigorous standard that may be applied.”

The most rigorous standard is an established national standard (ANSI X9.95) that practitioners need to begin to look at immediately.

² Honorable Paul W. Grimm, Chief United States Magistrate Judge, United States District Court, Maryland, *Lorraine vs. Markel Insurance Company*, May 4, 2007



ProofSpace
900 Clancy Ave NE
Grand Rapids, MI 49503
(312) 933.8823