THE LEGAL ASPECTS OF USING CRYPTOGRAPHIC HASH VALUES, IN DIGITAL FORENSICS IN LIGHT OF RECENT VULNERIBILITES AND COLLISSIONS WITHIN THESE HASH VALUES.

VERONICA SCHMITT (ACE, AMCSS)
SPECIAL INVESTIGATING UNIT

ABSTRACT

- MD5 and SHA-1 cryptographic hash algorithms are a standard practice in digital forensics that is used in the preservation of digital evidence and ensuring the integrity of the digital evidence. Recent studies have shown that both MD5 and SHA-1 have vulnerabilities and collisions. Based on this, the use of MD5 and SHA-1 hash algorithms in the practice of digital forensics to preserve and ensure the integrity of digital evidence has been questioned in certain instances.

- Using experimentation, the researcher proves the validity of using either MD5 or SHA-1 hashing algorithms to ensure the integrity of seized digital evidence, from the moment of seizure of the evidence, through to eventual presentation and use of the evidence in court; thus demonstrating that the use of hashing remains a valid forensic methodology to ensure the integrity of digital evidence.

Defining Digital Forensics

- Digital evidence is defined as information of a legal probative value that is either stored, or transmitted in a digital form.

- Another definition of digital evidence is that it is any data stored or transmitted using a computer that supports or refutes a theory of how an offence occurred, or addresses a critical element thereof such as intention or an alibi.

- Digital evidence is any digital object which contains reliable information which supports or refutes a hypothesis.

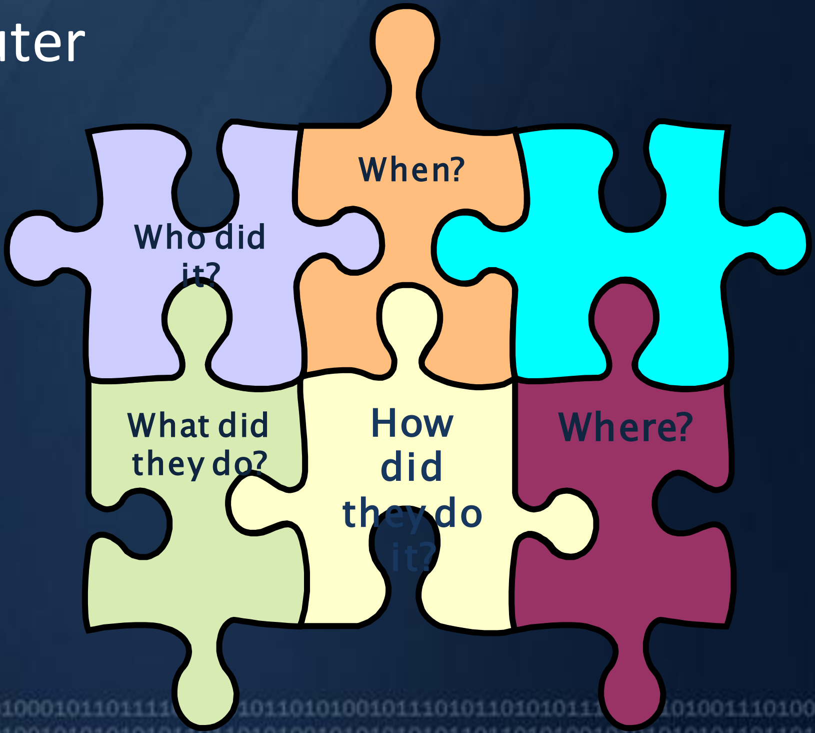
- Digital evidence includes any computer hardware (containing data), software, or data, that can be used to prove either who, what, when, where, why, and how, of an allegation being investigated.

Defining Cyber Security

- Ensuring the security of your computer or computer systems are protected from unauthorized access.

The goal of Digital Forensics

- The goal of computer forensics is to use reliable methods and procedures to perform a structured investigation while maintaining a documented chain of evidence to find out:
  - What happened on a computer
  - When it happened
  - How it happened
  - Where it happened
  - Who was responsible for it
  - Who else was involved



USES OF CRYPTOGRAPHIC MATHEMATICAL HASHES IN DIGITAL FORENSICS

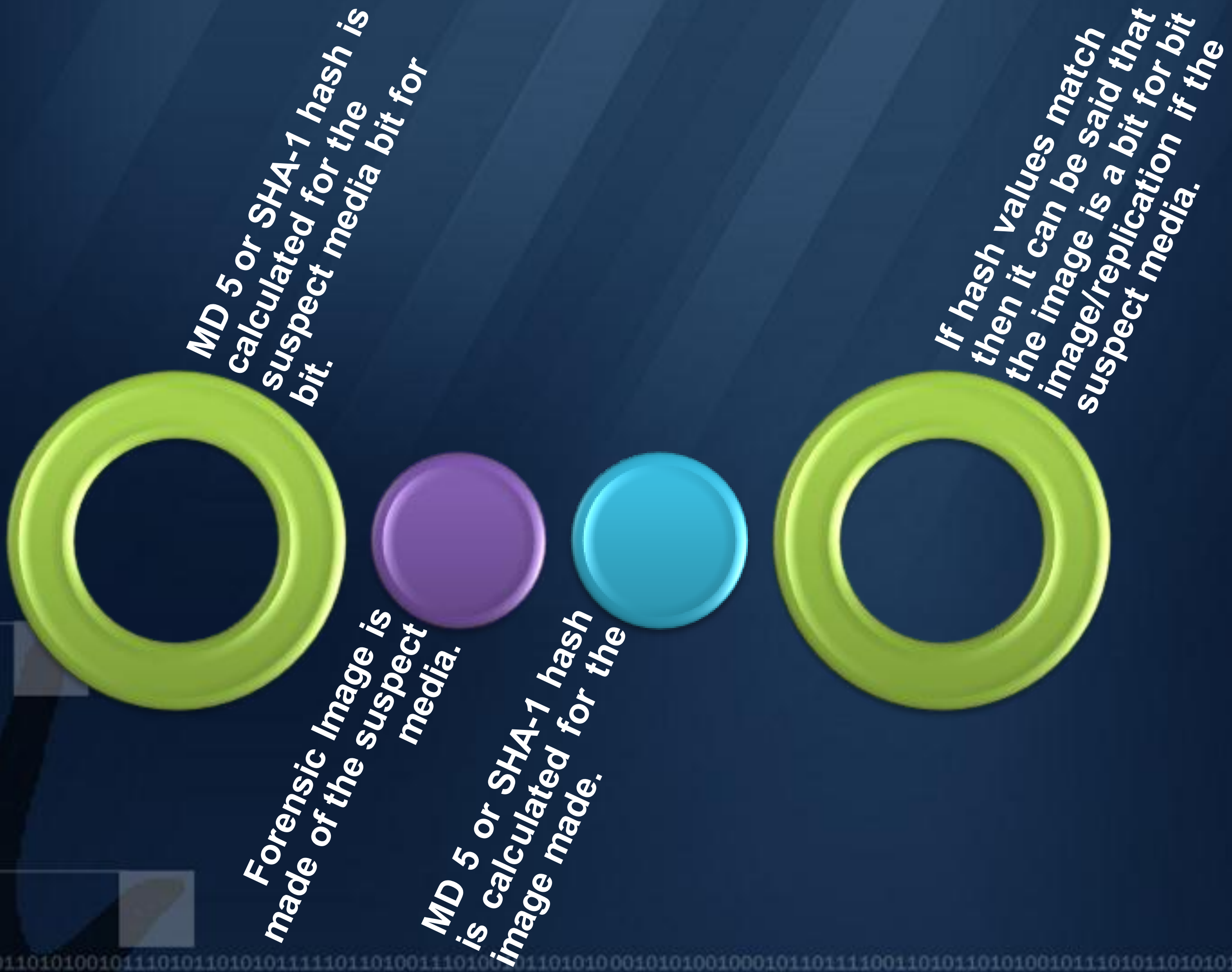
- Identification of Known Files
- Ensure Complete Forensic Images are Made
- Ensure the Integrity of Evidential Data to ensure that no Data has been Altered

Identification of Known Files

- The finding known or files from established hash sets (such as those maintained as part of the National Software Reference Library of NIST), allows the discovery of potentially incriminating data (such as child pornography images or malware) or innocent data (such as operating system files) based on the hash value of the file, rather than in the traditional sense of doing keywords and manual searches

- IT SAVES TIME ….

Ensure Complete Forensic Images are Made



MD 5 or SHA-1 hash is calculated for the suspect media bit for bit.

Forensic Image is made of the suspect media.

MD 5 or SHA-1 hash is calculated for the image made.

If hash values match then it can be said that the image is a bit for bit image/replication if the suspect media.

Ensure the Integrity of Evidential Data to ensure that no Data has been Altered

- A crucial element within digital forensics is ensuring that the digital evidence remains unaltered from the time that it has been acquired, up until it is presented in a court of law, thereby ensuring the integrity of the evidence.

- The use of cryptographic hashing such as MD5 and SHA-1 and the resulting hash values have played a critical part in ensuring that and changes or alteration of digital evidence can be identified.

- The basic premise is that even if so much is one byte in a particular file or set of data is altered after a MD5 or SHA-1 hash has been calculated for it, and it is then hashed again, it would calculate a different hash value which did not match the original.

Admissibility of Digital Evidence

- The Electronic Communications and Transactions Act 25 of 2002 address the issue of digital evidence in South African law, and have allowed the use of digital evidence as evidence in a South African court of law.
- When assigning evidential weight to digital evidence, Section 15(2) of the Electronic Communications and Transactions Act 25 of 2002 guide a court in how to evaluate the evidence.
- A key factor to be considered in this is the reliability of the digital evidence and how the integrity of it was maintained.
- In terms of current South Africa law, digital evidence, and the concept of a data message as defined in terms of the Electronic Communications and Transactions Act 25 of 2002 are synonymous. Section 1 of the Electronic Communications and Transactions Act 25 of 2002 defines data as an electronic representation of information in any form, and a data message as any data that is generated, sent, received, or stored in electronic means .

Admissibility of Digital Evidence

- The Electronic Communications and Transactions Act 25 of 2002 do not define “electronic”.
- Section 15 of the Electronic Communications and Transactions Act 25 of 2002 governs the admissibility and weight of data messages, and subsequently digital evidence.
- Section 15(1) of the Electronic Communications and Transactions Act 25 of 2002 states that a data message (and thus digital evidence) cannot be ruled inadmissible simply by virtue of the evidence being in an intangible digital format, while Section 15(2) goes on to state that information in a digital form must be given due evidential weight [12].

The role of MD 5 and SHA-1 hash values in the admissibility of Digital Evidence

- Of significant importance are Section 15(3) of the Electronic Communications and Transactions Act 25 of 2002, which sets out guidelines for a South African court to apply in assessing the evidential weight of digital evidence .
- This section requires a court to give due regard to:
- The reliability of the manner in which the data message (digital evidence), was generated, stored, or communicated.
- The reliability of the manner in which the integrity of the data message (digital evidence) was maintained.
- The manner in which the originator of the data message (digital evidence) was established.
- Any other relevant factor.

The role of MD 5 and SHA-1 hash values in the admissibility of Digital Evidence

- In essence, the use of MD5 or SHA-1 hashes of digital evidence are used as a method to demonstrate that the evidence that is presented before court is the same as that obtained initially during the investigation, and that it has not be altered or modified in any way; thus demonstrating the integrity of the evidence.

The Potential Impact of MD5 and SHA-1 Hash Collisions on the Admissibility of Digital Evidence

- The fact that MD5 and SHA-1 hashing has been potentially compromised in that files can be modified so that they produce the same hash value, raises the possibility that legal practitioners in court may argue that it does not provide adequate proof that digital evidence has not been altered from the time it has been obtained. In effect, they could argue that the digital evidence had been altered, shifting the onus onto the producing party that it had not.

THE VALIDITY OF MD5 AND SHA-1 HASHING IN ENSURING THE ADMISSIBILITY OF DIGITAL EVIDENCE

- Mathematically, the MD5 (and SHA-1) hash calculations are of such a nature that changing one bit in any item of digital evidence will cause a cascade effect during the calculation process which would produce a different hash value [2]. The researchers conducted experimentation to validate this effect

Hash Collisions

- Hash collisions describes a situation where two different data files or data sets have a hash calculation made for them, the calculated hash values are identical, even though there are clear differences in the data themselves.
- Due to the nature of hash calculations, they can only provide a number of calculated values, which can naturally result in two separate data inputs resulting in the same calculated hash value.
- The chance of two different files randomly having the same MD5 hash value is 2^128, or a 1 in 340 billion, billion, billion, billion chance.
- The chance of two different files randomly having the same SHA-1 hash value is 2^160, or a 1.46 trillion, trillion, trillion, trillion chance. Identical files and data sets when hashed should always result in the same hash values.

Hash Collisions

- Hash collisions can thus occur naturally from different data inputs; however the chance of this happening randomly is statistically infeasible. The concern from a digital forensics perspective is when hash collisions can be engineered so that two separate and different files return the same hash values.

QUATIFYING CHANCE OF HASH COLLISSIONS

- In practice this simply is a highly improbable occurrence of a birthday collision happening naturally, mathematically speaking if a file is hashed with both MD 5 (128 Bits) and SHA-1 (160 Bits) would have effectively uniqueness of 288 Bits or $1x10^{86}$.

The Experiment

- Each of these files was hashed using MD5 and SHA-1, generating two separate hash values per file. These values were then documented.

- Once the MD5 and SHA-1 hash values had been calculated for each file, each file was then modified making use of a hex editor to modify the files are a hexadecimal data level. For each file, the first byte of each file at logical offset 0x00 for the file was recorded, at then this byte was edited to read 0x23 or the ASCII symbol #. The file was then hashed using MD5 and SHA-1, generating two separate hash values per file. These values were then documented.

- The files were then each modified again, restoring the first byte at logical offset 0x00 in the file back to its original byte value. The file was then hashed using MD5 and SHA-1, generating two separate hash values per file. These values were then documented.

Conclusion

- Based on the results of the research, it can be clearly stated that the use of MD5 and SHA-1 hashing within the field of digital forensics remains a valid scientific practice.

- Legally this means that if an item of digital evidence was hashed using either MD5 or SHA-1 when it was obtained, and then hashed again using the same algorithm at a later time, and the hash values generated match, then the evidence has not been altered in the intervening time period. In other words, it the hash values match, then the integrity of the evidence from the time of acquisition to the time of presentation in court, can be relied upon.

Conclusion

- However the chance of this occurring randomly is improbable due to the significantly large numbers involved.

- While it is possible to manipulate input data in such a way that it produces two identical hash values for different inputs, the alterations have to be very specific.
- In other words to take an evidential file containing one set of information that proved or disproved an element of a matter before court, which had a specific hash value, and then manipulate it is such a way that it stated something else affecting the interpretation of the evidence, while still generating the same hash value, is computationally improbable.