Published in "Journal of advances in information technology", 2019, vol. 10, no. 2, pp. 41-47, which should be cited to refer to this work.
DOI: 10.12720/jait.10.2.41-47

Blockchain-based Digital Evidence Inventory

David Billard
HES-SO - Geneva School of Business Administration, Geneva, Switzerland
Email: David.Billard@hesge.ch

***Abstract*— This paper proposes the use of a blockchain-based structure in order to store evidences in a digital forensics investigation. The traditional chain of evidence is augmented with properties of immutability and traceability, thanks to a cryptographic process. The blockchain is constructed by forensics experts by adding evidences through the process.**

Since the blockchain is immutable, it can be shared among the different parties involved in a prosecution in order to review the chain of evidence and build their case. Furthermore, the blockchain structure can be applied to other forensics fields, like drugs, firearms, NDA.

This blockchain is called a Digital Evidence Inventory (DEI) and is part of a wider framework encompassing a Forensics Confidence Rating (FCR) structure, in order to give experts the ability to rate the level of confidence for each evidence and a Global Digital Timeline (GDT) to order evidence through time. The whole framework is called '*Aldiana*'.

***Index Terms*— Digital forensics, digital evidence, e-evidence, blockchain technology, legal evidence admissibility, data provenance**

I. Introduction

This work aims at concurring to the work of justice by comforting court rulers and parties about the confidence and traceability they should expect from digital evidence. The work takes advantage from new advances in blockchain technology and cryptography in order to provide digital forensics investigators with tools to collect, order and produce e-evidence with associated metrics and in a secured manner.

During the process, the forensic practitioner builds three data structures:

(1) The *Digital Evidence Inventory* (DEI), based on a blockchain technology, in order to capture evidence. This DEI is immutable and can be used by every party in a case. Each party has access to the same knowledge about the digital evidences.

(2) The *Forensics Confidence Rating* (FCR) structure. With the FCR, the practitioner grades the e-evidence, based on a categorization of data and data provenance. This rating is subject to modification, depending on the unfolding of the case.

(3) The *Global Digital Timeline* (GDT) to order evidence through time. It is the experience of the author that magistrates and lawyers are particularly sensible to the order of events. It is of utmost importance for the forensics practitioner to provide them with a timeline composed of e-evidence.

The whole framework, called '*Aldiana'*, has been outlined in general terms in [1]. This paper focuses on the DEI, which is the central part of the system.

The paper is structured as follows: after some related works on digital forensics processing, provenance data and blockchain, we detail the notions used in the *Digital Evidence Inventory* (DEI), namely the transactions, blocks and mining algorithm. Then we present in a more concise manner the *Global Digital Timeline* (GDT) and the *Forensics Confidence Rating* (FCR) structures. We then conclude this paper with the works in progress.

II. Related works

The work presented in this paper is at the crossroad of digital forensics, data provenance and blockchain.

A. *Digital forensics*

Digital forensics appears as a research field in the mid-1990. In 1999, a first structuration of the forensics handling of digital evidence is proposed in [2]. The process is composed of four phases (subsequent researches have refined these steps):

(1) *Identification phase*: the digital traces that may contain evidence are identified in a unique manner;

(2) *Preservation phase*: the digital objects are protected in order to be eventually analyzed by others. Typically, acquisition, or imaging, of data supports is done during the preservation phase;

(3) *Analysis phase*: the investigators try to make sense of digital traces;

(4) *Presentation phase*: the methodology and the findings are written in a report intended to non-specialists (lawyers, laymen).

Our work is focused on the Preservation and Analysis phases, although mechanisms for verification can be triggered in the Presentation phase. The raw evidence (acquired image) is added to the DEI in the Preservation phase and additional evidence is added to the DEI in the Analysis phase.

B. *Qualification of digital evidence*

To the best of our knowledge, few literature has exposed a framework that, at the same time, is usable by the expert to characterize the e-evidences and by the courts to base their judgements on facts with a measurable degree of certainty.

One of the most accomplished work in this area can be found in [3]. It follows the lessons learned from the Daubert case [4] concerning the generally accepted guidelines for evaluating scientific evidence that include quantifying the technique's potential rate of error, and the work from Judge Pollack [5] calling for more rigorous requirements. In [3], the author voices the opinion that forensic examiners have a duty to estimate how closely the measured values represented in their data approximate reality.

In a prospective essay on the future of digital forensics, [6] emphasis the fact that the research community should work to develop digital forensic techniques that produce reportable rates for error or certainty when they are run.

If we leave for a moment the digital world to the physical world, most authors and in particular [7] state that evidence admissibility should be determined on the basis of the reliability and accuracy of the process involved.

Most work rely on the validity of the process. Although important, most of the processes at the origin of e-evidence are unavailable for analysis. Either because they are unknown, or simply because the source code of the software governing the data creation is closed, or too complex to analyze.

But two main aspects can be detailed:

(1) characterizing how the e-evidence was collected and

(2) what measure of its relevance and confidence one can tag to the e-evidence.

As a matter of fact, battles in courts seldom question the existence of the e-evidence, but rather the reason of its existence. For instance, if e-evidence includes pedo-pornographic images, the debate will focus on why the images were there, with defendants usually incriminating viruses, advertisements on a web page that the suspect did not volunteer, etc.

One can note that e-evidence can also be the absence of data. For instance, when the system log files have been voluntarily deleted from a computer.

C. *Data provenance and blockchain*

Data provenance dates back in early 2000 and matured around 2006 with the Open Provenance Model [8]. Data provenance is the representation of the origin of data, and its subsequent alterations.

III. Digital Evidence Inventory

A. *Overview*

The Digital Evidence Inventory is used to capture e-evidence inside an immutable blockchain and forms a traceable e-evidence bag. It takes its structure from multiple advances in digital forensics and cryptography:

- The digital evidence lifecycle [2];
- Data provenance [9] and Scrybe framework [10];
- CASE [11] and DFAX [12] modeling;
- Bitcoin protocol [13].

The DEI is part of the preservation phase of the digital evidence lifecycle since it will record the digital acquisitions. It is also part of the analysis part, since new digital evidence, embedded in the acquired evidence, will be added to the blockchain. Finally, the presentation part will ultimately refer to the DEI in order to explain the findings.

The DEI is based on the Scrybe Provenance Framework [9]. In Scrybe, the authors present the components necessary to preserve "provenance data", which means preserving how data was derived. In their work, the authors define a model based on blockchain technology, with a lightweight mining and distributed consensus.

Data provenance, or data lineage, is a research field getting more attention thanks to the wide acceptation of the blockchain technology. Although data provenance is mainly used in debugging cases, it can be applied very successfully to digital forensics, since it provides a historical record of the data, its origins and, in our case, the several possible alterations.

Blockchain, by its immutable nature, is the ideal candidate for supporting data provenance in a forensics environment. When a digital evidence is added to the blockchain, as a transaction, it is validated by the users of the blockchain by the commit of its block. Once committed, the digital evidence cannot be further altered or removed.

A digital evidence containing other digital evidences, will be materialized by generating new transactions, linked to the original transaction. The mechanism is similar to the Unspent Transaction Outputs (UTXO) from the bitcoin, or other crypto-currency protocol. Excepted that no crypto-currency is used.

B. *Transactions in DEI*

A transaction concerns a digital evidence item that needs to be joined to a case.

When a disk is imaged, the provenance of the digital evidence - for instance the log copy describing the acquisition process, the case identification, time of acquisition and technical details - is added to a transaction. At this level, the content of the digital evidence itself is represented by a hash, or a set of hashes.

When a digital evidence is found inside the disk image, it is added to a transaction, represented as a CASE object [11] or an XML token with the Digital Forensic Analysis eXpression (DFAX) format [12].

At creation time, transactions are an output of the previous transaction referring to the embedding evidence. Excepted the original transaction in the origin block (that contains the description of the forensic lab, or police force department), there exist none "alone" transaction, they are all linked together.

For the sake of clarity, we are using the same examples that are also used in [11]. In TABLE I we represent the identification of an example case in CASE format.

TABLE I. Example of a CASE record

```
{
"@id": "investigation-4586742a-710a-454f-bcb8-b60e230eclb2",
"@type": "Investigation",
"name": "Crime A",
"focus": "Murder",
"description": "In Mycenae, Atreus killed two sons of Thyestes,
cooked them (except for their hands and heads}, fed them to
Thyestes, and then taunted Thyestes with his murdered sons'
hands and heads.",
"object": ["thyestes-uuid", "victiml-uuid", "role-
relationship1-uuid"]
},
```

The “@id” is the unique reference of the object. From an implementation point of view, this reference can be the same as the transaction identifier. It can be generated at creation time.

The next section presents the process of validating a block of transactions.

C. *Transaction validation (mining) algorithm*

The Scrybe [10] framework achieves a lightweight mining with a rapid and small footprint algorithm that can be summarized as follows:

(1) Each miner, from a total of *N* miners, generates a random number.

(2) Each miner broadcasts the hash of their random number.

(3) Once all hashes have been broadcasted, each miner broadcasts its own random number.

(4) Each miner verifies the hashes and calculates

$$\mathit{Elected\_miner} = \mathit{sum} \,\%\, N$$

where *sum* is the sum of all the random numbers

(5) The miner with an id equal to *Elected_miner* creates the new block and broadcasts it to all other miners.

The proposed model fits our need for the DEI, since we can map almost directly all our components:

- The *transaction* represents a digital evidence that is linked to a case. The provenance of the digital evidence includes at least a case identification, time of acquisition and technical details. The digital evidence itself is represented by a hash (or multiple hashes) of its content, and its location. In some cases, when the evidence is small in size, it can be directly stored into the blockchain in CASE format. By the model, the user’s cryptographic signature is added to the transaction and therefore indicates the investigator identity.
- The *block* is a collection of transactions. Transactions are validated through the commit of their block.
- The *miners* are the digital forensics investigators working in the same laboratory or office. Contrary to the bitcoin miners, they don’t need to present a proof of work, since there is nothing to gain out of mining. The “reward” for mining is to obtain an immutable blockchain.

The version number that was part of the bitcoin transaction is removed in [10], although we think it might be valuable, should the protocol adapts or specializes.

In order to clarify the difference between Bitcoin and Scrybe protocols, TABLE II, resp. TABLE III, characterizes the elements of the blockchain used for the Bitcoin resp. Scrybe protocol.

TABLE II. COMPONENTS OF A BITCOIN PROTOCOL

Component	Bitcoin protocol
Transaction	Proof of *present* ownership *Value* oriented
Block	Difficulty and nonce
Mining	Resource intensive
Security	Based on *computational difficulty*

TABLE III. COMPONENTS OF A SCRYBE PROTOCOL

Component	Scrybe protocol
Transaction	Proof of *past* ownership *Data* oriented
Block	Cryptographically signed
Mining	Simple selection algorithm
Security	Based on *cryptographic signature*

The case of TABLE I is a block with only one transaction containing the identification of the case. Each subsequent transaction concerning the same case will use this original transaction as input. Every subsequent e-evidence is recorded into the blockchain on the form of CASE objects. For instance, TABLE IV presents a simplified view of the CASE object associated to the imaging of the partition 6 of the hard drive ‘Cassandra’. Thanks to the CASE format, we know the partition length, the file system type and the data content hash.

D. *Linking transactions*

We want to add this image of a partition to the DEI. In order to comply with the Scrybe protocol, we need to add the components in red (see TABLE V). They all relate to the transaction processing, and thus begin with “tx”.

- *@TxId* is the transaction ID, in the form of a public address;
- *@TxUserId* is the public key of the user submitting the transaction;
- *@TxUserSig* is the signature of the user submitting the transaction;
- *@TxPreviousEntry* is the ID of the previous transaction. This previous transaction represents the provenance of the digital evidence;
- *@TxTimeStamp* is the timestamp when the transaction was submitted.

Suppose now that during the Analysis phase, another digital evidence, a Windows Word file worth of interest, is found inside the partition 06 of Cassandra image, then the corresponding transaction to add in the DEI could be similar as the transaction depicted in TABLE VI.

The *@TxPreviousEntry* refers to the *@TxId* of the first transaction, concerning the partition 06 of Cassandra image (*@TxId* = 2F40C7442654).

TABLE IV. EXAMPLE OF A CASE OBJECT FOR A PARTITION

```
{
"@id": "cassandra-image-partition6-uuid"
"@Type": "Trace",
"propertyBundle": [
{
  "@Type": "DiskPartition",
  "DiskPartitionType": "MSDOS",
  "PartitionID": "06",
  "PartitionOffset": "63",
  "PartitionLength": "24523563",
}
{
  "@Type": "FileSystem",
  "FileSystemType": "EXT3",
}
{
  "@Type": "ContentData",
  "sizeinBytes": 245235000,
  "hash": [
  {
  "@Type": "Hash",
  "hashMethod": "SHA256",
  "hashValue":
"7ea081166336119da78ee4bbdbd06840b94efe28988a2bdb0bcf2387a481e283"
  }],
}]}
```

TABLE V. EXAMPLE OF A TRANSACTION FOR A PARTITION

```
{
"@TxId": "2F40C7442654"
"@TxUserId": "3048024100C9"
"@TxUserSig": "18FACF8DEB2D"
"@TxPreviousEntry": "18FACF8DEB2D"
"@TxTimeStamp": "2017-06-22T08:12:19.32Z"
"@id": "cassandra-image-partition6-uuid"
"@Type": "Trace",
"propertyBundle": [
{
  "@Type": "DiskPartition",
  "DiskPartitionType": "MSDOS",
  "PartitionID": "06",
  "PartitionOffset": "63",
  "PartitionLength": "24523563",
}
{
  "@Type": "FileSystem",
  "FileSystemType": "EXT3",
}
{
  "@Type": "ContentData",
  "sizeinBytes": 245235000,
  "hash": [
  {
  "@Type": "Hash",
  "hashMethod": "SHA256",
  "hashValue":
"7ea081166336119da78ee4bbdbd06840b94efe28988a2bdb0bcf2387a481e283"
  }],
}]}
```

TABLE VI. EXAMPLE OF A TRANSACTION FOR A FILE

```
{
"@TxId": "eb9c562e907f"
"@TxUserId": "3048024100C9"
"@TxUserSig": "18FACF8DEB2D"
"@TxPreviousEntry": "2F40C7442654"
"@TxTimeStamp": "2017-06-23T10:12:19.32Z"
"@Type": "File",
"createdTime": "2017-06-22T08:12:19.32Z",
"extensionn: "docx",
"fileName": "AthensDemocraty.docx",
"filePath": "C:/evidence/AthensDemocraty.docx",
"isDirectory": false,
"sizeinBytes": 980500
},
{
"@Type": "ContentData",
"hash": [
  {
  "@Type": "Hash",
  "hashMethod": "SHA256",
  "hashValue":
"e9e8f47b2070704ddf53f7fded0c69d6637c5ca9573ce6b4235f4309464a0bc4"
  }],
"sizeinBytes": 980500
}
```

E. Blocks

Blocks are very similar to the Bitcoin blocks, excepted that they do not have any field for the proof of work protocol, since this protocol is not used. Hence, the difficulty target and the nonce are no more in the block. However, the signature of the successful miner is added to the block.

For our small example, if we suppose that the two transactions are validated inside the same block, then the block structure could be as depicted in TABLE VII.

When the block is processed by the “miners”, the winning miner adds his signature to the block in the field *@BlkMinerSig*, before broadcasting it to the network.

TABLE VII. EXAMPLE OF A MINED BLOCK

```
{
"@BlkVersion": "1.0",
"@BlkPreviousHash": "397a5481e7",
"@BlkTimeStamp": "2017-06-23T12:15:23.32Z"
"@BlkTransactionCount": "2",
"@BlkTransactionList": [
  {"@TxId": "2F40C7442654"},
  {"@TxId": "eb9c562e907f"}]
"@BlkMinerSig": "7a759a01a6c0"
}
```

F. Using the DEI

The DEI can be primarily used to verify the provenance of a digital evidence. Each transaction concerns a digital evidence and transactions are validated inside blocks. Blocks are back-chained one to the other and transactions are back-chained to their originating evidence.

Furthermore, since every transaction is signed by the user submitting the evidence, a chain of evidence is obtained. By implementing adapted procedures inside a forensic laboratory, for instance users signing blocks are not users signing transactions, a double-verification is achieved.

If the DEI is restricted to a single case, then it is possible to give access to the parties. Therefore, the whole case can be processed and verified by all involved parties, limiting the risk of biais.

The different attacks on the mining protocol are not part of the scope of this paper.

IV. GLOBAL DIGITAL TIMELINE (GDT)

The Global Digital Timeline (GDT) is a data structure associated to the DEI. It is a simple key-value database, where the key is a date, or more precisely a timestamp, and the value is a pair consisting of:

- a *reference* to an evidence (a transaction, in the blockchain terminology) in the DEI;
- a *label* tagging the evidence.

A key is not unique, since multiple evidence may share the same timestamp.

With this data structure, it is possible to extract rapidly meaningful information for a given period of time and to present the result even to a non-specialist. Table 4 presents the content of the GDT applied to an imaginary example.

In order to explain the data structures devised in this section, we take a small example of e-evidence, in TABLE VIII.

This e-evidence is taken (and modified) from a real case. It is a list of USB devices connected to a computer. This list comes from the USBSTOR Windows registry hive. TABLE IX gives the representation of this evidence as DEI transactions, and TABLE X presents the content of the GDT.

TABLE VIII. EXAMPLE OF E-EVIDENCE

Serial #	Name	User	Last connection	First connection
42014287	S3300		04.11.2016 08:52:50	
7299803F	Kingston Data-Traveler 2.0 USB Device	BadGuy	08.11.2016 12:30:11	2016.05.17 12:45:57
182127000	USB Flash Memory USB Device	BadGuy	18.07.2016 12:15:16	2016.07.18 08:39:50

TABLE IX. TRANSACTIONS FOR USBSTOR EVIDENCE

```
{
"@TxId": "b8fc4661d646"
…
"@Type": "RegistryKey",
"RegKey: "USBSTOR"
},
{
"@Type": "ContentData",
"LastCon": "04.11.2016 08:52:50"
}
{
"@TxId": "9f423577e25a"
…
"@Type": "RegistryKey",
"RegKey: "USBSTOR"
},
{
"@Type": "ContentData",
"LastCon": "18.07.2016 12:15:16",
"FirstCon": "18.07.2016 08:39:50"
}
{
"@TxId": "645bad667141"
…
"@Type": "RegistryKey",
"RegKey: "USBSTOR"
},
{
"@Type": "ContentData",
"LastCon": "08.11.2016 12:30:11",
"FirstCon": "17.05.2016 12:45:57"
}
```

TABLE X. EXAMPLE OF GLOBAL DIGITAL TIMELINE RECORDS

Key	TransactionID	Label
04.11.2016 08:52:50	b8fc4661d646	LastCon
18.07.2016 12:15:16	9f423577e25a	LastCon
18.07.2016 08:39:50	9f423577e25a	FirstCon
08.11.2016 12:30:11	645bad667141	LastCon
17.05.2016 12:45:57	645bad667141	FirstCon

This table can be ordered by the key, to find contemporary elements, or by the label if only one kind of element is sought. This structure is not immutable and is flexible enough for easy processing. Each item can be checked against the e-evidence referenced in the DEI.

V. FORENSICS CONFIDENCE RATING (FCR)

The Forensics Confidence Rating (FCR) is also a key-value database, where the key is a pair consisting of:

- a *reference* to an evidence (a transaction, in the blockchain terminology) in the DEI;
- a *timestamp* of the time the rating was issued.

The value is the rating associated to the evidence. In our model, the granularity of the rating is the e-evidence (the transaction). Each party involved in a case can process its own FCR.

The rating is calculated thanks to a taxonomy, primarily proposed by the security researcher Bruce Schneier. In [14], he defines a taxonomy of social networking data, and that taxonomy can be extended to fit any digital artifact in or out cyberspace.

Bruce Schneier defines six data types in the framework of social networking and we added three more data types to capture a broader set of e-evidences.

The first six data types defined in [14] are the following:

(1) *Service data* is the data you give to a social networking site in order to use it.

(2) *Disclosed data* is what you post on your own pages, or social media.

(3) *Entrusted data* is what you post on other people's pages.

(4) *Incidental data* is what other people post about you.

(5) *Behavioral data* is data the site collects about your habits by recording what you do and who you do it with.

(6) *Derived data* is data about you that is derived from all the other data.

The three additional data types are the following:

(7) *System data* is produced when a system is recording an action (or lack of action).

(8) *Private data* is basically any data that you don't want to disclose, or only if the person you disclose the data to is at your highest degree of confidence.

(9) *Leaked data* is a mutation of all others types of data (and chiefly private data), over time, to disclosed data. With the added particularity that this data was not supposed to be disclosed by its rightful owner.

Each data type has a confidence rating that is primarily applied to the e-evidence. This confidence rating may evolve, especially when additional elements of the investigation appear. The e-evidence itself may change its type over time (and thus, the associated rating). For instance, a service data can be disclosed and therefore, its initial rating will evolve.

A. *Service data*

Service Data may have a high degree of confidence since it is usually checked against public records, or via state issued documents. For instance, a bitcoin trading platform will require an official identification document to prove who you are, and to conform to bank-related laws. Of course, identification documents can be forged, but the suspect is then under the threat of not being accepted on the platform. And the suspect needs to use the platform. Eventually, some information is necessary correct, in order to cash money or to receive goods.

An example of a real case can be found in [15]. In this US district court of Northern District of California, a federal court order allows the Internal Revenue Service (IRS) to order Bitcoin exchange platform Coinbase to give up their customers' identities.

Note that getting access to service data might be hindered by applicable laws, depending of the ruling courts where the service is registered.

B. *Disclosed data*

Disclosed data is heavily used by law enforcement or intelligence bodies. Since the data is meant to be read by other people or machines, every disclosed data of an individual is scrutinized by law enforcement when he/she becomes a suspect. Law enforcement is not the only body interested in disclosed data: lawyers, journalists, insurance companies, activists or common people can access the information.

An example of disclosed data is the bitcoin blockchain, where every transaction is recorded. Provided that you know the public bitcoin addresses used by a person, all his/her transactions can be scrutinized.

Another example, from a real case, is taken from [16] and [17], where a comment left on MySpace is recognized as a central evidence:

M. Clark was found guilty of murdering a two-year-old girl left in his care and was sentenced to life in prison without parole. On appeal, Clark argued that the trial court improperly admitted evidence from his MySpace account in violation of Ind. R. Evid. 404(b). Taking up the "novel question" of the propriety of admitting such evidence, the Supreme Court of Indiana ruled that the trial court did not err in admitting the evidence, particularly where Clark's own testimony made his character a "central issue" of his defense. The verdict and sentence were therefore affirmed.

The confidence rate that can be associated to disclosed data is not as high as with service data.

C. *Entrusted data*

The suspect has intentionally left data at someone else social media. In our opinion, this data type is exactly the same, from a forensics point of view, as disclosed data. The only exception being that data is left privately at someone else social media, who in turn makes it public. For instance, a Twitter direct message that is sent to the public tweet list.

As a matter of fact, US courts make strong distinction between private and non-public data [18] (although most of the cases are civil litigations, the concept is the same for criminal justice):

Litigants continue to believe that messages sent and posts made on their Facebook pages are "private" and should not be subject to discovery during litigation. In support of this, litigants claim that their Facebook pages are not publicly available but, instead, are available only to a limited number of designated Facebook "friends." Courts consistently reject this argument, however. Instead, courts generally find that "private" is not necessarily the same as "not public." By sharing the content with others - even if only a limited number of specially selected friends - the litigant has no reasonable expectation of privacy with respect to the shared content. Thus, the very purpose of social media - to share content with others - precludes the finding of an objectively reasonable expectation that content will remain "private."

Provided that law enforcement knows the social media identification code of the suspect, all entrusted data is searchable and may produce an interesting confidence rate.

D. *Incidental data*

If one considers Incidental data, the order of confidence should be lower than service data, since people can lie or make assumptions, mistakes or alternative truth as it is now called. It can, however, color the suspect profile. For instance, think about a political forum arguing about a candidate to some election.

The rating may not be very high in the first place, but a photo where a suspect appears, or a list of a meeting attendees including the suspect's name, all this constitutes incidental data that may rise the confidence rating.

E. *Behavioral data*

For Bruce Schneier, who designed this taxonomy in the context of social media, behavioral data is the information, at large, that a social media can affix to your identity. If you get past the context of social media, we can consider behavioral data as any data that is produced by a legitimate user (human or machine) at any cyber service.

For instance, the list of whatsapp calls you made is behavioral data. Getting access to behavioral data might be hindered by applicable laws, depending of the ruling courts where the service is registered.

When behavioral data is recorded by an autonomous machine, it can possess a high confidence rating.

F. *Derived data*

Derived data is what is produced by artificial intelligence, big data, data mining and such services. Derived data is also more prosaically a basic oriented graphs showing relations among people or machines. This data is much more in demand in investigations, even for intelligence gathering.

The confidence rating is of course function of the confidence rating associated to the original data.

G. *System data*

System data is produced when a system is recording a suspect action (or lack of action). For instance, a suspect enters a room and a sensor is triggered: this is a system data. Another example is when a suspect accesses his/her email box: a system data is produced. Another example is when the suspect phone is Bluetooth enabled and it comes close to another Bluetooth enabled phone: both phones may produce system data with the Bluetooth address and name.

In our example of TABLE VIII, the e-evidence is taken from a registry hive and is considered of System data type. It can be associated with a high confidence since few people can modify registry keys on a Windows System.

TABLE XI shows the confidence rating associated to our example. The first e-evidence has no user name associated with the USB key description, so we can lower its confidence rating since we don't know which user inserted the USB key. Thus, creating a new record in the FCR, as shown in TABLE XII.

TABLE XI. EXAMPLE OF AN EVIDENCE RATING

TransactionID	Time of Rating	Rating
b8fc4661d646	25.12.2017 01:00:00	High
9f423577e25a	25.12.2017 01:00:00	High
645bad667141	25.12.2017 01:00:00	High

TABLE XII. EXAMPLE OF A RATING ALTERATION

TransactionID	Time of Rating	Rating
b8fc4661d646	25.12.2017 01:00:00	High
9f423577e25a	25.12.2017 01:00:00	High
645bad667141	25.12.2017 01:00:00	High
b8fc4661d646	25.12.2017 12:30:00	Medium

H. Private data

Private data is basically any data that you don't want to disclose, or only if the person you disclose the data to is at the highest degree of confidence. For instance, medical records, intimate diary, bank account credentials. During an investigation, all this data might be disclosed to forensics investigators, even if they are not case-related.

The confidence rating of the case related data is usually unknown.

I. Leaked data

Leaked data is a mutation of all others types of data (and chiefly private data), over time, to disclosed data. With the added particularity that this data was not supposed to be disclosed by its rightful owner.

It is a special case of data because not all courts accept stolen or leaked data as evidence. A good example of these kind of data is the banking records stolen at some financial institutions.

Even derived data can be leaked. Since a tier is involved in the leakage, leaked data can be tampered with before they are released. Therefore, the confidence rating should be low.

VI. Conclusion and future works

In this paper, we proposed a preliminary framework for building a fact-based confidence rating of e-evidence.

First, e-evidence is collected inside an immutable e-evidence blockchain, the Digital Evidence Inventory (DEI). This blockchain enforces the "chain of evidence": *who* obtained the evidence; *where* and *when* the evidence was obtained; *who* secured the evidence; *who* had control or possession of the evidence. In this manner, the consistency of the evidence can be traced back and validated.

Every party in a trial can have access to the DEI that provides also traceability. Therefore, the DEI is a verification tool that can be activated by any party, depending on the legal system.

Then e-evidence is categorized into basic data types and each data type is associated with a measure of the certainty and relevance of the e-evidence. This measure can evolve through time, as well as the categorization. That is the reason why the measures are kept in an external data structure, the Forensics Confidence Rating (FCR), linked to the DEI. All the rating modifications are recorded. Each party involved in a trial can have its own FCR, with its own ratings, depending on its own view of the trial.

And finally, a Global Digital Timeline (GDT), also linked to the DEI, is created. It is our experience that magistrates are primarily interested in the chain of events that led to a specific crime, or action. This chain of events gives the landscape of the actions taking place before, during, and after a crime is committed.

In overall, this framework, called '*Aldiana'* allows for a better confidence rating of e-evidence, both for the forensics investigators and the courts. The framework is a valuable documentation tool for the forensics investigators, that can be cross-examined.

Future works include a finer tuning of the blockchain protocol, a semi-automated tool for the building of the GDT and a more precise confidence rating by adding error rate probabilities and relevance.

Acknowledgment

Our sincere thanks to HES-SO for supporting this research partially through grant ISNet 74354.

The author thanks the anonymous reviewers for their helpful comments.

References

[1] D. Billard, « Weighted forensics evidence using blockchain », présenté à International Conference on Computing and Data Engineering, Shanghai, China, 2018.

[2] R. McKemmish et A. I. of Criminology, *What is Forensic Computing?* Australian Institute of Criminology, 1999.

[3] E. Casey, « Error, uncertainty, and loss in digital evidence », *Int. J. Digit. Evid.*, vol. 1, nº 2, p. 1-45, 2002.

[4] Chief Justice Rehnquist, *Daubert v. Merrell Dow Pharmaceuticals, Inc.* 1993.

[5] United States District Court, E.D. Pennsylvania, *UNITED STATES of America, v. Carlos Ivan LLERA PLAZA, Wilfredo Martinez Acosta, and Victor Rodriguez*. 2002.

[6] S. L. Garfinkel, « Digital forensics research: The next 10 years », *Digit. Investig.*, vol. 7, p. S64-S73, 2010.

[7] Strong, J. W., « McCormick on Evidence », *4th Ed*, vol. section 294, 1992.

[8] Open Provenance Model, « Open provenance model ». .

[9] P. Turner, « Digital provenance – interpretation, verification and corroboration », *Digit. Investig.*, vol. 2, nº 1, p. 45-49, févr. 2005.

[10] U. Mukhopadhyay *et al.*, *The Scrybe Provenance Framework: Scalable Secure Data Provenance Using Blockchain Technology*. 2017.

[11] E. Casey, S. Barnum, R. Griffith, J. Snyder, H. van Beek, et A. Nelson, « Advancing coordinated cyber-investigations and tool interoperability using a community developed specification language », *Digit Investig*, vol. 22, nº C, p. 14-45, 2017.

[12] E. Casey, G. Back, et S. Barnum, « Leveraging CybOX^{TM} to standardize representation and exchange of digital forensic information », *Digit. Investig.*, vol. 12, p. S102-S110, 2015.

[13] Satoshi Nakamoto, « Bitcoin: A Peer-to-Peer Electronic Cash System ». 24-mars-2009.

[14] B. Schneier, « A Taxonomy of Social Networking Data », *IEEE Security & Privacy*, vol. 8, nº 4, p. 88, 2010.

[15] *John Doe vs USA*. 2016.

[16] *Ian J. CLARK v. STATE of Indiana*. 2009.

[17] Electronic Discovery Law, « Indiana Supreme Court Rules Trial Court Properly Admitted Evidence of Defendant's MySpace Page in Murder Trial ». 23-oct-2009.

[18] Margaret (Molly) DiBianca, « Discovery and Preservation of Social Media Evidence ». 02-janv-2014.

David Billard was born in France, 1968. Mr Billard received a PhD in computer science from University of Montpellier, France, in 1995.

He worked at the University of Geneva, Switzerland, from 1995 to 2000 as a research fellow, then he headed the University software developments until 2008. Since 2008 he is associate professor at the University of Applied Sciences in Western Switzerland in Geneva. Since 1999, He is a sworn expert to the courts in France and Switzerland (French speaking cantons) and participated to more than 150 criminal investigations and 50 civil litigations. He publishes regularly about digital forensics and privacy.

Prof. David Billard is involved in several program committees, like DFRWS EU or Forensics International.