Verifiable Limited Disclosure: Reporting and Handling Digital Evidence in Police Investigations

Thein Tun[1] Blaine Price[1] Arosha Bandara[1] Yijun Yu[1] Bashar Nuseibeh[1,2]

[1]The Open University, UK
{firstname.lastname}@open.ac.uk

[2]Lero, Ireland
firstname.lastname@lero.ie

Abstract—**Police investigations involving digital evidence tend to focus on forensic examination of storage units on personal electronic devices (laptops, smartphones, etc). However, a number of factors are making digital forensic tools increasingly ineffective: (i) storage capacities of electronic devices have increased, and so has the amount of personal information held on them; (ii) cyber crimes are increasingly committed on social media, and evidence of crimes are held on social media platforms, not necessarily on personal devices; (iii) there is a greater need for protecting digital privacy, especially when examining digital evidence from witnesses and victims of cyber crimes. These factors pose a number of practical challenges for both law enforcement agencies and citizens when disclosing and handling the digital evidence. This paper defines and illustrates the key challenges, and proposes the concept of *verifiable limited disclosure*, which defines a communication protocol to ensure privacy, continuity and integrity of digital evidence. More specifically, the protocol allows (i) citizens to decide what evidence to disclose to law enforcement agencies and (ii) any of the two parties to be able to prove any tampering of the disclosed evidence. The paper discusses methods for implementing the communication protocol using standard security and privacy tools and presents a pathway to evaluating their effectiveness.**

Index Terms—**Digital Forensic, Social Media, Cyber Crimes, Digital Privacy**

I. Background and Context

Police investigations involving digital evidence tend to focus on forensic examination of digital storage units partly because evidence for many offences can be gathered from web browsing history, emails, chats, as well as photos and other files stored on those devices. In this "image and search" approach, the law enforcement agencies take cryptographically secure disk copies of storage units and search for evidence on the disk copies so that the integrity of evidence is preserved.

A YouGov online survey between 16 and 17 March 2015 shows that the average British household owns 7.4 internet devices such as smartphones and laptops [1]. Similarly, a survey conducted by Pew Research Center between 17th March and 12th April 2015 shows that about 1 in 3 Americans own three devices: a smartphone, a computer and a tablet [2]. With increasing availability of high capacity storage units (100s of gigabytes, or terabytes are increasingly common), together with high volumes of data people store on them, this "image and search" approach is increasingly inefficient. Even the process of triaging, i.e, the initial assessment process where cases are filtered and prioritised by digital forensic experts, is coming under increasing pressure. For example, London Metropolitan Police services receive over 38,000 digital devices a year for digital forensic examination, a situation described as "resource overload" [3].

Beside the challenges associated with the current digital forensic technologies, online crime patterns are also changing. Criminal offences on social media, primarily Facebook and Twitter platforms, now constitute the bulk of all criminal cases investigated by the UK police forces [4]. This emerging class of social media offences include "cyber bullying", "trolling" and "virtual mobbing".

When investigating such offences, the traditional digital forensic method has limited effectiveness for two main reasons. First, it is the social media organisations, not necessarily the devices used by citizens, that may hold the evidence relating to offences committed on the social media. For example, Facebook has the entire history of a user's activities, but the user's smartphone may not. However, social media organisations sometimes decline requests for information. For example, in the second half of 2015, Twitter received 427 requests for information (the highest since 2012) from law enforcement agencies in the United Kingdom[1]. Twitter complied partially or fully in 76% of requests relating to 956 Twitter, Periscope, and Vine accounts. In contrast, in the second half of 2012, it received 25 requests for information, and only in 4% of the cases relating to 27 accounts yielded some information. Although UK law enforcement agencies are making many more requests for information, and are more successful, some requests for information are still resisted by social media companies such as Twitter[2].

Second, electronic devices sent for forensic examination are typically seized during arrest of suspects. A suspect under police investigation may not have strong claims for digital privacy. Moreover, with social media offences, a person may come forward voluntarily to give evidence as a witness and/or a victim. In such cases, protecting the digital privacy of the person could encourage cooperation and increase public trust. Since anonymous crime reporting has been a common practice, those who report cyber crimes should also be able to maintain their digital privacy.

[1] https://transparency.twitter.com/country/uk

[2] Tweets are normally public, and therefore requests for information must relate to non-public information such as user's identity, location, and deleted tweets. Recent new features of Twitter allow tweets within a smaller group and between individuals, which could be the subject of future requests.

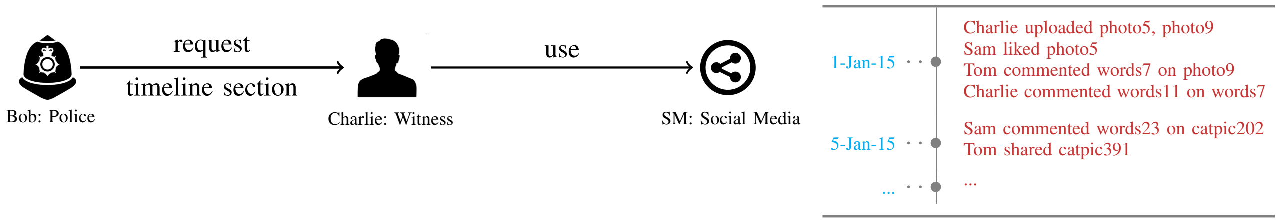

request
timeline section
use
Bob: Police
Charlie: Witness
SM: Social Media
1-Jan-15
Charlie uploaded photo5, photo9
Sam liked photo5
Tom commented words7 on photo9
Charlie commented words11 on words7
5-Jan-15
Sam commented words23 on catpic202
Tom shared catpic391
...
...

Fig. 1. Collecting Digital Evidence on Social Media

However, evidence collection and handling methods for social media crimes lack technical sophistication. For example, Facebook gives the following advice for evidence collection in case of cyber bullying on its platform[3].

> Is someone at risk of physical harm? Contact your local law enforcement immediately if you're worried about the safety of the person reaching out to you. Take screenshots of all bullying posts, and print them out to save.

Screenshots and printouts are not reliable digital evidence, and evaluating their continuity and integrity later by law enforcement agencies could be a slow and laborious process. What is needed is an evidence collection process that is forensically sound (irrespective of whether it is for a suspect, a victim or a witness), so that there can be no question about mishandling of the evidence by any party involved.

In summary, when collecting evidence for social media offences, the "image and search" approach could be slow, ineffective, and potentially inappropriate in terms of digital privacy. The current "print screen and print out" approach also suffers from the same deficiencies.

II. Problem Statement

This section presents a scenario for evidence collection for social media offences, and discusses the requirements for potential solutions.

A. *Scenario*

Charlie uses the social media platform SM (see Fig. 1). Typically, his activities on SM include creating, reading and viewing content (messages, photos, videos, comments, likes, rating, etc). These activities can be ordered chronologically along a timeline. All the information is hosted on the servers of SM, and Charlie may use multiple devices at different times to access SM. On 1-Jan-15, after Charlie uploaded two photos (photo5 and photo9), both Sam and Tom commented on them. Charlie considers Tom's comment words7 in breach of criminal laws. Charlie contacts the police officer Bob, who asks for information about all his activities between 1-Jan-15 and 4-Jan-15. Charlie wants to comply with Bob's request.

B. *Possible Approaches*

There are a number of approaches Charlie can use when reporting evidence to Bob, including:

(1) Charlie hands over his smartphone to Bob who performs the "image and search" approach on storage units.
(2) Charlie takes a screenshot, and prints out the relevant part of the time line (removing what he thinks is irrelevant/private), and gives it to Bob.
(3) Bob asks SM to show the relevant parts of Charlie's timeline with his consent (assuming that SM complies with all requests from Bob).

Charlie may not want to hand over his device (1), or give Bob direct access to his entire timeline (3) due to privacy concerns. The fact that Charlie is not in control of what Bob can learn from his device and his SM activities may discourage him from reporting crimes. If however Charlie has full control over the evidence he discloses (1) and (2), then Bob has to check with SM to ensure that the evidence he receives from Charlie is accurate. This could be a resource expensive process for Bob.

C. *Requirements*

Practical evidence collection methods for social media crimes have to address the issues identified in Section II-B. We now formulate these issues as requirements for a software system that supports an evidence collection method for social media crimes. We identify three main requirements as discussed below.

Requirement 1: Charlie does not over-disclose: Charlie never has to reveal more information than Bob's request, to which Charlie has agreed. In other words, Charlie has some ways of maintaining his digital privacy.

Requirement 2: Charlie and Bob cannot lie: Charlie can prove to the world that the evidence he gives to Bob has not been modified by him. Likewise, Bob can prove to the world that the evidence he uses in his investigation has not been modified by him.

Requirement 3: Charlie and Bob cannot conceal: Charlie can prove to the world that he has not withheld evidence relevant to Bob's request. Likewise, Bob can prove to the world that he has not withheld evidence revealed to him by Charlie.

These three requirements are about maintaining privacy, continuity and integrity of evidence. In some sense, these requirements are an elaboration of the phrase "the truth, the whole truth and nothing but the *relevant* truth". The first requirement deals with "the relevant truth", the third requirement with "the whole truth" and the second requirement with "nothing but the truth".

[3] https://www.facebook.com/help/420576171311103/

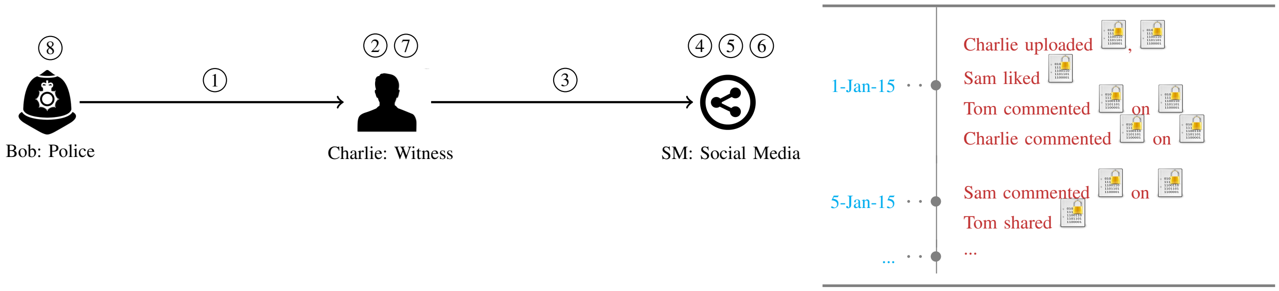


Fig. 2. Communication Protocol for Verifiable Limited Disclosure

In addition to these requirements, there are two auxiliary requirements, namely: (i) *automation*: much of the tasks involved should be done automatically, therefore quickly, with little or no human involvement, (ii) *usability*: the system should be usable with little or no training of the users. These requirements deal with the problem of "resource overload".

III. Verifiable Limited Disclosure

This paper proposes the concept of *verifiable limited disclosure* in order to satisfy the requirements described in Section II-C.

- We use the term "disclosure" to mean a witness, a victim, or a suspect of a cyber crime giving evidence to law enforcement agencies.
- "Limited disclosure" means that (i) the evidence disclosed must be relevant to a potential crime, and (ii) reasonable concerns for digital privacy must be respected.
- "Verifiable" means if evidence is mishandled either deliberately or accidentally, then it is possible to prove that evidence has been tampered with.

A. Protocol

Verifiable limited disclosure can be achieved by means of a protocol that uses two main security tools: cryptographic hash function and public key encryption. The key actors in this protocol and their communication are shown in Fig. 2. The main steps in the protocol are as follows:

① Bob requests section s of Charlie's timeline (s could be 1-Jan-15 to 4-Jan-15)

② Charlie announces his public key ($PK_{Charlie}$) to the world

③ Charlie requests encrypted timeline from FB

④ FB encrypts every object in Charlie's timeline using $PK_{Charlie}$

⑤ FB computes the cryptographic hash value (CHV_k) of all encrypted objects along Charlie's timeline

⑥ FB announces CHV_k and its timestamp to the world

⑦ Charlie decrypts a section of his timeline and gives the entire timeline to Bob.

⑧ Bob encrypts the decrypted section of Charlie's timeline, computes the hash value and compares it to the one announced by SM in Step ⑥

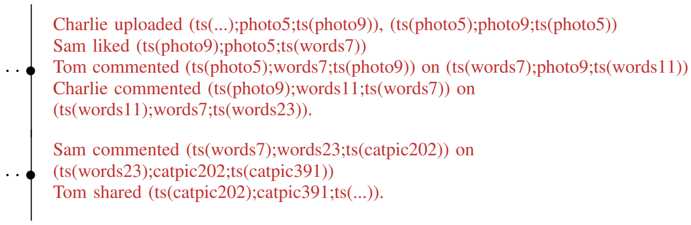
This protocol satisfies (a) *Requirement 1* because Charlie chooses the section of the timeline he decrypts and Bob cannot read anything else on his timeline and (b) *Requirement 2* because if Charlie modifies anything in his timeline, Bob can prove the modification by means of a mismatched hash value. The protocol does not satisfy the Requirement 3 because if Charlie hides anything, such as his comments words11, then Bob has no way of knowing about the concealment. Similarly, if Bob withholds evidence revealed to him by Charlie, Charlie cannot prove that Bob is withholding evidence (because Bob can claim that the evidence is in the encrypted part).

Therefore, in order to handle *Requirement 3*, it is necessary to disclose the metadata indicating the activity counts on each day. For example, the metadata can state that there are 4 items on Charlie's timeline on 1-Jan-15, where Charlie first uploaded two pictures, followed by Sam's like, Tom's comment, and Charlie's comment, in that order. Obviously disclosing all of Charlie's metadata will not satisfy *Requirement 1*. Alternatively, Bob can check with SM to see if Charlie's response contains any concealment. This creates extra work for both Bob and SM.

B. Handling Metadata

A different way to handle the meta-data is to partially encode it into the timeline data. When SM encrypts, every encrypted object contains the timestamp of the previous and the next object along Charlie's timeline, as shown in Tab. I. The second photo Charlie uploaded, photo9, now contains the timestamp of the previous object photo5, and the timestamp of the next object, photo5. This scheme only focuses on timestamps of objects but it is easy to extend it to actions, as well as relations between actions and objects (such as 'like'). For reasons of space, we omit those details.

TABLE I
FULLY-SEQUENCED OBJECTS IN CHARLIE'S TIMELINE


Charlie uploaded (ts(...);photo5;ts(photo9)), (ts(photo5);photo9;ts(photo5))
Sam liked (ts(photo9);photo5;ts(words7))
Tom commented (ts(photo5);words7;ts(photo9)) on (ts(words7);photo9;ts(words11))
Charlie commented (ts(photo9);words11;ts(words7)) on (ts(words11);words7;ts(words23)).

Sam commented (ts(words7);words23;ts(catpic202)) on (ts(words23);catpic202;ts(catpic391))
Tom shared (ts(catpic202);catpic391;ts(...)).

Returning to the problem of Charlie withholding words11, it is clear that with this new data structure, Charlie will be found out because decrypting words7 will reveal that the timestamp of the next object is within the date range Bob is interested in. In order for the timstamps to join up, Charlie cannot do anything but reveal all the objects Bob has requested to see. Therefore, Charlie has to reveal everything until the timestamps fall outside the range, without hiding anything in between. Concealment of relevant information by Bob can be proved in the same way (but less directly). When applied to this revised data structure, the protocol will satisfy all three requirements discussed in Section II-C.

C. Discussion

In addition to the three requirements, the protocol also satisfies the automation requirement because all verification tasks are performed by the hashing and encryption algorithms. For example, Bob can automatically check the integrity of Charlie's evidence. Evaluation of the usability requirement depends on particular implementation and interaction design choices, and therefore will be examined in our future work.

The protocol design can also be used in other applications as well. Recently there has been some public discussions around the use of forensic evidence from smartphones in case of traffic accidents in New York [5] (potentially on the basis of "implied consent", because searching a smartphone without a warrant is illegal there). A protocol for verifiable limited disclosure can be used to disclose for example, only the texting activities on the phone within a time range.

Having said that, the current protocol design has some limitations. First of all, it assumes that the information is chronologically ordered (along a timeline) and Bob's requests are for specific time ranges. However, if Bob has requests that do not fit the timeline, such as all comments posted to a specific friend, then the request cannot be handled without violating *Requirement 3*. Secondly, the protocol uses public-key encryption where key management can be a difficult issue.

There is currently limited software support for "open-source evidence collection" especially for social media offences. Existing tools rarely provide the degree of integrity needed in forensic processes. For example, Huber et. al. [6] developed data harvesting methods for online social networks without addressing the issues of privacy, continuity and integrity of evidence. Therefore, such tools are only useful for initial investigations.

The use of steganographic techniques to disguise illegal communications on social media is an important issue. Chee [7] presents an evaluation of steganographic techniques on a number of popular social neworks.

IV. Conclusions

Traditional digital forensic tools are increasingly ineffective due to the vast amounts of data stored on digital devices. Moreover, when investigating crimes on the social media, the evidence is usually held on external servers, rather than on personal devices.

We have proposed a protocol to implement the notion of verifiable limited disclosure that satisfies a number of requirements for privacy, continuity and integrity of digital evidence. The protocol uses the public-key encryption and cryptographic hashing function.

In future, we plan to develop a working prototype of this protocol and evaluate it with real data. Initially, we plan to develop new methods for encoding meta-data in order to support arbitrary queries. We also plan to investigate alternative architectures to simplify the interactions between Bob, Charlie and the social media platform, so that they can communicate securely but not necessarily using the public-key infrastructure.

The current protocol requires a degree of cooperation from social media companies in Steps ④, ⑤ and ⑥, namely encrypting and computing and publishing hash values. Although these operations are not computationally expensive, there is currently no legal and business framework within which this cooperation might take place. In our view, the advantages of using the proposed protocol outweigh the costs for all parties involved.

Although evidence collection is treated as a reactive process, we will explore how it can be done more proactively using adaptive approaches [8].

Acknowledgement

The work is supported by The Open University Centre for Policing Research and Learning, with funding from the Police Knowledge Fund, and by SFI grants 10/CE/I1855 and 13/RC/2094, and ERC Advanced Grant no. 291652 (ASAP). We thank the members of the police forces which work closely with the Centre for valuable input, in particular, DS Steve Fenn from Gwent Police, Danny Inglis, Dave Smith and colleagues from Greater Manchester Police, and DI Eric Halford from Lancashire Constabulary. We also thank our colleague Mu Yang for comments and suggestions on an earlier draft.

References

[1] Press Association. (2015, Apr.) Online all the time average British household owns 7.4 internet devices. [Online]. Available: https://www.theguardian.com/technology/2015/apr/09/online-all-the-time-average-british-household-owns-74-internet-devices

[2] M. Anderson. (2015, Nov.) Smartphone, computer or tablet? 36% of Americans own all three. [Online]. Available: http://www.pewresearch.org/fact-tank/2015/11/25/device-ownership/

[3] R. E. Overill, J. A. Silomon, and K. A. Roscoe, "Triage template pipelines in digital forensic investigations," *Digital Investigation*, vol. 10, no. 2, pp. 168 – 174, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1742287613000261

[4] Keith Moore. (2014, Jun.) Social media 'at least half' of calls passed to front-line police. BBC Radio 4's Law in Action. [Online]. Available: http://www.bbc.co.uk/news/uk-27949674

[5] M. Richtel. (2016, Apr.) Texting and driving? Watch out for the Textalyzer. [Online]. Available: http://www.nytimes.com/2016/04/28/science/driving-texting-safety-textalyzer.html

[6] M. Huber, M. Mulazzani, M. Leithner, S. Schrittwieser, G. Wondracek, and E. Weippl, "Social snapshots: Digital forensics for online social networks," in *Proceedings of the 27th Annual Computer Security Applications Conference*. New York, NY, USA: ACM, 2011, pp. 113–122. [Online]. Available: http://doi.acm.org/10.1145/2076732.2076748

[7] A. Chee, "Steganographic techniques on social media: Investigation guidelines," Master's thesis, School of Computing and Mathematical Sciences, Auckland University of Technology, 2013, https://aut.researchgateway.ac.nz/handle/10292/5577.

[8] L. Pasquale, Y. Yu, M. Salehie, L. Cavallaro, T. T. Tun, and B. Nuseibeh, "Requirements-driven adaptive digital forensics," in *21st IEEE International Requirements Engineering Conference*. IEEE Computer Society, 2013, pp. 340–341. [Online]. Available: http://dx.doi.org/10.1109/RE.2013.6636745