Title	Maintaining hard disk integrity with digital legal professional privilege (LPP) data
Author(s)	Jiang, ZL; Fang, JB; Law, YW; Lai, PKY; Ieong, RSC; Kwan, YK; Chow, KP; Hui, CK; Yiu, SM
Citation	IEEE Transactions on Information Forensics and Security, 2013, v. 8 n. 5, p. 821-828
Issued Date	2013
URL	http://hdl.handle.net/10722/190317
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Maintaining Hard Disk Integrity With Digital Legal Professional Privilege (LPP) Data

Zoe L. Jiang, Junbin Fang, Frank Y. W. Law, Pierre K. Y. Lai, Ricci S. C. Ieong, Michael Y. K. Kwan, K. P. Chow, Lucas C. K. Hui, S. M. Yiu, and K. H. Pun

***Abstract*—The concept of legal professional privilege (LPP) in the Common Law is to enable a client to make full disclosure to his legal advisor for seeking advice without worrying that anything so disclosed will be used against him. Thus, some of the communications and documents between a legal advisor and his client can be excluded as evidence for prosecution. Protection of LPP information in the physical world is well addressed and proper procedures for handling LPP documents have been established. However, there does not exist a forensically sound procedure for protecting digital LPP information. In this correspondence, motivated by a real case of a commercial crime investigation, we introduce the LPP data integrity problem. While finding an ideal solution to solve the problem is difficult, we propose a practical solution that was adopted to solve the real case investigation. We also analyze the performance of our solution based on simulated data.**

***Index Terms*—Computer forensics, data integrity, LPP document.**

I. Introduction

DIGITAL crime investigation replicates many legal practices in real world crime investigation. For example, taking cryptographic hash is used to preserve the integrity of digital evidence, resembling taking a snapshot of the scene of crime in real world crime investigation. However, not all real world legal practices have been incorporated into the digital forensics investigation framework. One issue, that is essential to proper administration of justice but rarely addressed, is the maintenance of hard disk integrity with privilege documents in digital world.

The rationale behind having this legal professional privilege (LPP) in the Common Law is to enable a client who may not have enough legal knowledge to fully disclose everything to his legal advisor for seeking of advice without worrying that anything disclosed for this purpose will be used against him. Thus, the accused is protected from being prosecuted based on private or unintentional discussion [3] and he has the right to prevent a particular document from presenting as evidence against him if it is classified as privileged document [1].

Nowadays, with the advancement in information technology, communications are no longer limited to paper or telephone conversations and have extended to various kinds of electronic communications such as e-mails, instant messaging chats, VoIP phones and digital video conferences. Furthermore, the documents prepared for legal proceedings may not exist as hard copies but may be stored in a computer as Word documents or Excel spreadsheet files. All these have an impact on the traditional investigation approach as crime investigators are required to handle LPP documents in digital format. This is a new area required to be addressed and a systematic approach is needed to assist investigators in handling privileged digital documents properly and effectively.

In the physical world, LPP documents are handled with extreme caution because the disclosure of any LPP information may jeopardize the legal proceedings involved. Under the Common Law, investigators are not allowed to inspect any documents for which privilege is claimed. When a document is claimed to be privileged, but an investigator thinks that a document is likely to be related to a criminal case, he can seize it and put it in a sealed container, such as an exhibit envelope, and later submit to the court for determining if the document can be used for prosecution [8].

When it comes to the digital world, the situation is not as simple. Usually, these documents and numerous other digital files are stored in a single physical hard drive. The law enforcement office will seize the hard drive. There is no well-defined and forensically sound procedure to deal with LPP documents. Simple solutions as in the following may not work. For example, putting the whole hard drive that contains the claimed LPP files into an exhibit envelope and letting the court decide the proper way to handle the digital files stored in there is certainly not appropriate. This may result in a very lengthy process as the court may not have the expertise for the task. The non-LPP files that are in the same drive are needed for investigation. This will delay the investigation a lot.

Manuscript received March 27, 2011; revised September 15, 2011, March 19, 2012, and March 07, 2013; accepted March 16, 2013. Date of publication April 04, 2013; date of current version April 25, 2013. This work was supported in part by two grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project HKU 7136/04E and HKU 7132/06E), one grant from the National Natural Science Foundation of China (Grant 61240011), and one grant from Shenzhen Strategic Emerging Industries Program (Grant ZDSY20120613125016389). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Yong Guan.

Z. L. Jiang was with the Department of Computer Science, The University of Hong Kong, Hong Kong. She is now with the Shenzhen Key Laboratory of Internet Information Collaboration, and also with Harbin Institute of Technology Shenzhen Graduate School, Shenzhen 518055, China.

J. Fang is with the Key Laboratory of Optoelectronic Information and Sensing Technologies of Guangdong Higher Education Institutes and the Department of Optoelectronic Engineering, Jinan University, Guangzhou 510632, China, and also with the Department of Computer Science, The University of Hong Kong, Hong Kong.

F. Y. W. Law, P. K. Y. Lai, R. S. C. Ieong, M. Y. K. Kwan, K. P. Chow, L. C. K. Hui, S. M. Yiu, and K. H. Pun are with the Department of Computer Science, The University of Hong Kong, Hong Kong, China (e-mail: smyiu@cs.hku.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2013.2256784

In fact, the majority of existing digital forensics investigation models or procedures do not incorporate a procedure for supporting legal professional privilege data protection. For instance, in the DFRWS framework [2], digital investigation covers the Identification, Preservation, Collection, Examination, Analysis and Presentation of digital evidence. It focuses on the technical aspects in collecting, examining and explaining the hypothesis of incidents without fully incorporating the proper practice from the legal perspective [4]. How to handle legal privilege documents was not adequately addressed. Thus, the protection of these documents can only rely on individual practitioners and is handled differently in a case by case manner. In this paper, we identify two issues of handling privileged documents: (1) the lack of a systematic procedure and (2) the LPP data integrity problem. crime investigation case in a law enforcement department of Hong Kong. It is difficult to have an ideal solution for the LPP data integrity problem. In this correspondence, we provide a practical solution that was adopted to solve the real case investigation. The rest of the paper is organized as follows. Based on a real crime investigation case, Section II describes the problems we want to address for handling digital LPP documents. Our proposed solution and the analysis of the solution are presented in Sections III and IV. Section V concludes the paper. We remark that although the technical part of the solution may not be very innovative and our proposed solution still cannot solve the problem completely, the solution as a whole is shown to be feasible and serves as a starting point for further research.

II. The Problems—Motivated by a Real Criminal Investigation Case

We first describe a real criminal investigation case and show that the two common approaches for handling digital forensic data, namely clone and erase, selective cloning, cannot handle this case as it involves LPP documents.

A. The Background

In October 2008, a Hong Kong-listed company announced that the company had bet wrongly on massive foreign-currency target-redemption forward contracts and daily accrual (accumulator) contracts for Australian and European dollars, causing a loss of 800 million Hong Kong Dollars on these contracts from 1 July to 17 October 2008. Since the contracts could only be ended until 2010, there was a provisional loss of more than 15 billion Hong Kong Dollars for these contracts. The announcement caused a 55% drop in the stock price of the company. The Government Securities and Future Commission had initiated an investigation against the company. The case was subsequently forwarded to the Commercial Crime Bureau of Hong Kong Police to investigate any criminal aspect on leveraging the foreign exchange transactions by the company.

In April 2009, the Police raided the head office of the company with a search warrant and requested its directors to provide information with regard to the foreign exchange contracts entered into in 2007 and 2008 and announcements made by the Company from July 2007 to March 2009. The warrant related to an investigation of alleged offences, including (i) false statements by company directors; and/or (ii) conspiracy to defraud under the common law. The Police had seized a wide range of documents as well as computers and clones of servers that held almost the entire soft records of the company kept in the head office for the purpose of its investigation into the allegation. The seized documents and computer exhibits included approximately 2 terabytes of digital data in total. However, the data was widely spread in multiple PCs, computer servers and external storage devices of the company.

Since some of the digital data contained information that is subject to legal professional privilege (LPP), the legal representatives of the company subsequently lodged an application to the High Court of Hong Kong claiming LPP in respect of some of the exhibits. With the amount of digital data involved and the absence of official protocols in handling digital LPP data, the company formed a special board committee consisting of directors new to the board and external legal counsels to discuss with the police on how the materials seized should be handled. The Technology Crime Division (TCD) of Commercial Crime Bureau was the dedicated unit to deal with the digital data involving LPP information.

B. Existing Practices and Shortcomings

There is no standardized, forensically sound procedure to protect LPP information. Two existing practices, namely clone and erase, selective cloning, for handling digital information were explored in the case meeting between TCD and the company.

1) Clone and Erase: Clone and erase is a straightforward approach. It first prepares a cloned image, that is a bit stream image acquired from the target digital storage media, and then erases the LPP documents from it. The sanitized disk image would be used for later investigation. This process should be carried out in front of both parties so as to prevent potential evidence in non-LPP files from being removed intentionally.

One obvious problem is that, the image would replicate all the digital data, including the deleted LPP documents, which exist in the storage media. With the assistance of standard computer forensic tools, one could easily inspect all data including logical files, deleted files or fragmented file data that exist in unallocated space within the image. This is an obstacle to the proper protection of LLP documents as deleted LPP information may still be accessible in the context of investigation if the erasing is not thorough. Also, the data owner may take a very long time to view and segregate LPP documents from the enormous digital data stored in the storage media, and the LPP identification may be error-prone under a stressful environment. For better accuracy and efficiency, the process that requires face-to-face interaction should be kept minimal.

2) Selective Cloning: To avoid any duplication of sensitive LPP documents, selective cloning tries to conduct a selective data copy [6], [8] instead of cloning the whole storage medium. This is, to copy the non-LPP files from the source hard drive to another storage media for later investigation. The examiner may firstly connect to the target storage media with a write blocker device and then selectively extract digital data, excluding the LPP materials, that are relevant to the investigation. To avoid any dispute about unauthorized access to LPP data, the whole process should be carried out in the presence of the data owner, who is responsible to identify any LPP materials and monitor the

actions being performed by the examiner. This method offers the best protection to the LPP files. However, as the copying is performed with a logical view, deleted files or fragmented file data in unallocated spaces, which are invisible in the logical file system view, may be missed out. It can be very unfavorable for the investigation. Even though the examiner could utilize computer forensic tool to search for deleted or relevant data in unallocated spaces, the process would be very time-consuming and is not practical to be conducted at the scene of crime. Philip Turner suggested utilizing a selective imaging approach [7] to perform a selective acquisition of data on a hard drive using the concept of digital evidence bags [9]. This method is obviously in contrast to the traditional bit stream cloning but derives a way to properly handle large amount of digital information in a forensically sound manner. But, it still involves a very lengthy face-to-face process.

C. The Problems

Besides other disadvantages, both approaches described in the above require a lengthy face-to-face process and assume that the parties involved can identify all LPP documents at the spot. This assumption is not realistic. To summarize, there are two issues not adequately addressed in the current practice and research community. The first one is a standardized procedure for handling digital media with LPP documents inside. The second one is the LPP data integrity problem. It is a normal practice for the investigation officers to clone the hard disk once it is seized (even before the defendant can claim any LPP documents). The original hard disk will then be sealed and will not be accessed except a few exceptional cases in order not to corrupt the data in the original hard disk. The investigation will be carried out on the cloned copy. An integrity scheme will usually be enforced during the cloning to ensure that any data found in the cloned copy is exactly the same as it is inside the original copy.

Now, with the LPP documents, the situation becomes more complicated. After the cloning, the defendant is allowed to delete the agreed LPP documents from the cloned copy before the investigation officers can work on it. Thus, in the ideal case, we need an integrity scheme so that even if some of the data are deleted, it can still verify the integrity of the remaining data. Obviously, designing such an ideal solution is very difficult. Thus, we propose a solution so that there will be a high chance that this can be done. Then, supplement it with a "dirty" trick, we show a practical solution to solve the problem which has been adopted to the real case (see the discussion in the concluding section).

III. Proposed Approach

In this section, we design a high-level procedure to handle LPP documents while keeping the evidence of hard disk, followed by discussing the integrity issue in this procedure and proposing our hashing scheme.

A. High-Level Procedure

In order to comply with ordinary principles of computer forensics, it is observed that a bit stream image of the entire storage device should be taken whenever practicable. To avoid replicated LPP documents in cloned image being accessed in the context of computer forensic examination, the acquired image should not be examined until the content has been sanitized. For the removal of the LPP documents, the data owner would be invited to identify the privileged data from the image. The LPP documents would then be selected and an assessment would be made to determine if the documents are legally privileged or not. This assessment may be conducted by a trusted third party, for example, other investigator who is not involved in the case or any independent person who is not involved in the investigation. After the confirmation of the nature of the document, the LPP documents are destroyed from the data image. The original digital storage media would then be sealed whilst the sanitized image would be ready for computer forensics examination [8].

Taking all these factors into account, we propose the following procedure for handling LPP information:

1) A bit-stream image (cloning) is obtained from the original hard disk under the supervision of both parties to prevent any dispute of unauthorized access to LPP document.
2) Upon the completion of cloning, the original storage media will be sealed and kept in a safe location by the prosecution, whilst the acquired image will be given to the data owner for a reasonable period of time to remove any LPP documents that exist therein by forensically sound methods. This can be done by zeroing out the data content or replacing the content by some easily identifiable characters like "This is LPP data" in its storage sector(s).
3) After the removal of LPP documents, the data owner will return a sanitized image and a list of removed files to the prosecution for verification and record.
4) If there is any dispute or error in the context of LPP data removal, either party could still recover the erased data from the original media.

Note that in the above proposed procedure, the data owner will be given enough time to examine the image, without the presence of the examiner, in order to identify all LPP files (including those logically deleted, but still exist in the image) to be removed. This provides a more feasible solution to the problem as the volume of storage media is getting bigger and bigger. Requiring face-to-face interaction during the examination process will soon be impractical. This procedure also solves the problem of having the examiner look at the LPP files (as in the case of selective cloning).

On the other hand, to realize the above procedure, we need an effective integrity scheme that is done in the cloning step which can check the integrity of undeleted sectors in the sanitized disk as the deletion is done without the presence of the examiner. A bit-by-bit checking is not practical as it requires the use of the sealed original storage media and also, it requires the presence of both parties again to overlook the lengthy checking process.

B. The Integrity Issue

To ensure the integrity of a data item, cryptographic hash value is a common technique. The concept is as follows. Given any given data item, we can compute a unique hash value based on a mathematical formula as the signature of the data. The hash value has the properties that even one bit of the data is modified, the computed hash value will be different and it is difficult to

have two different files with the same hash value. It is easy to check whether the file has been modified if we have stored its hash value.

For our problem, one can compute a single hash value for the whole hard disk during the cloning process. But then, it is not possible to verify the integrity of individual sectors in the sanitized disk as some of LPP-related sectors have been deleted. Another simple method is to compute one hash value for one sector. Thus, the undeleted sector can still be verified by its corresponding hash value in the sanitized image. Nevertheless, the hash set is huge. For example, more than 1 billion hash values need to be stored and computed for a 500 GB hard disk with sector size of 512 bytes.

Note that the above approach devours large space although it offers the best accuracy among the other schemes. With the rapid changes in technology, the size of hard drive becomes larger and larger and the number of hashes required would be substantially increased. This methodology may no longer be competent to handle the task practically. To enhance the efficiency, and at the same time maintain the accuracy of the process, we show look for other efficient schemes.

C. $k - \mathcal{D}$ Hashing Scheme

In [10], [11], the authors proposed a $k - \mathcal{D}$ hashing scheme to address the issue for checking the integrity of a hard drive even if some of the sectors become bad sectors without storing the hash value for each sector. We develop our scheme based on this idea. We use $k = 3$, each sector $s_m (0 \leq m \leq N - 1)$ can be represented by three coordinates, $s_{x,y,z}$ where $1 \leq x, y, z \leq N^{1/3}$, and N is the total number of sectors. How to map each sector to a 3-coordinate point is given in Section III-D. Before we present the mapping scheme, we provide the framework on how this 3-$\mathcal{D}$ scheme works.

Instead of computing one hash value for each sector, by using this method, for each fixed (y, z) pair, it generates a hash value for X dimension using the sectors $s_{x,y,z}$ for all x from 1 to $N^{1/3}$; for each fixed (x, z) pair, it generates a hash value for Y dimension using the sectors $s_{x,y,z}$ for all y from 1 to $N^{1/3}$; for each fixed (x, y) pair, it generates a hash value for Z dimension using the sectors $s_{x,y,z}$ for all z from 1 to $N^{1/3}$. The size of each chain is about $N^{1/3}$ and we compute one hash value for each sector chain. All the hash values are stored in a secure place for later comparison. After some time when hard disk integrity checking is required, all the hash values will be recalculated using the same 3-$\mathcal{D}$ scheme on the hard disk sectors, then compared to the originally stored ones.

Each sector $s_{x,y,z}$ has been used to compute three hash values. The integrity can be verified by comparing any one of the three corresponding hash values. In our case, however, it is possible that a sector cannot be verified if there exist deleted sectors in all three chains. We will address this issue in the later part of the paper. One advantage of the 3-$\mathcal{D}$ scheme is to greatly decrease the number of hash values needed to be stored compared to N, especially when N is large. Using the same example of a 500 GB hard disk, the number of hash values to be stored decreases from 1 billion to about 3 millions.

Generally, the number of hash values of the $k - \mathcal{D}$ scheme can be calculated as: $N_{hash} = k \times N^{((k-1)/k)}$ since there are $N^{(k-1)/k}$ hash values in each dimension. The number of hash values for $k - \mathcal{D}$ scheme varying with the number of total sectors is shown in Fig. 1. As for the computational cost, the number of hash calculations for the $k - \mathcal{D}$ scheme is $k \times N$ since every sector exists in k chains. A hard disk with capacity of 250 GB (total number of sectors $N = 488,392,065$) was tested with the 3-$\mathcal{D}$ scheme to evaluate the scheme's practical performance. Experimental results including absolute time (T_{abs}), computational time (T_{compu}) and the actual storage of hash values (S_{hash}) are shown in Table I. Here absolute time refers to the total time for generating integrity information with the 3-$\mathcal{D}$ scheme, including the time for reading all sectors and computing hash values. We also check the time for reading all sectors of a hard disk, denoted as T_{read}. The time required for computations as $T_{compu} = (T_{abs} - T_{read})$. It takes 9,459 seconds for reading the hard disk in our test. The size of each hash value is 128 bits as MD5 is used in this experiment.

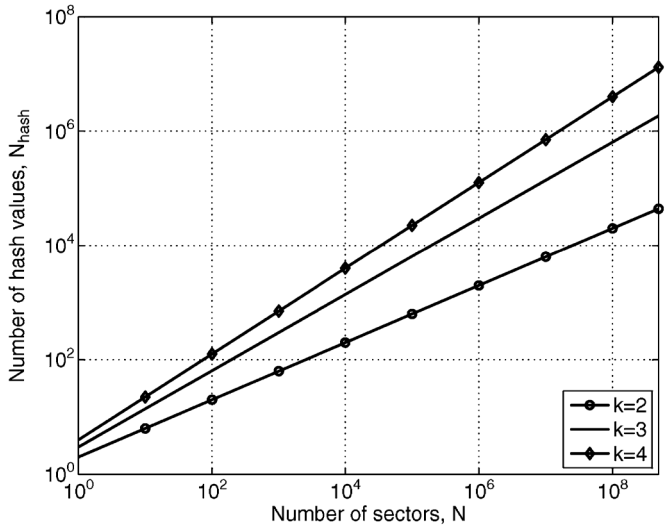


Fig. 1. Number of hash values for $k - \mathcal{D}$ scheme versus the number of total sectors N.

Table I
TIME AND STORAGE NEEDED FOR 3-$\mathcal{D}$ SCHEME

	$T_{abs}(Sec)$	$T_{compu}(Sec)$	$S_{hash}(MB)$
Pure reading	9,459		
3-$\mathcal{D}$	16,245	6,786	29.8

However, there is a problem with the scheme as mentioned in [10], [11], the scheme fails to verify the integrity of a sector when there is at least one bad sector on each chain to which the sector belongs. The sector which cannot be verified due to the tainted chain is named as affected sector. For example, in Fig. 4, the scheme fails to verify the integrity of sector $s_(x, y, z)$ (affected sector) for there is a bad sector s_{i_0,j_0,k_0} with $(j_0 = y, k_0 = z)$ in X dimension, a bad sector s_{i_1,j_1,k_1} with $(i_1 = x, k_1 = z)$ in Y dimension, and a bad sector s_{i_2,j_2,k_2} with $(j_2 = y, k_2 = z)$ in Z dimension.

Since Word document is a quite common digital format that is used by ordinary users frequently, it is a typical and important form of LPP data stored on a hard disk. As a preliminary work, we erase multiple Word documents to investigate the number of affected sectors using 3-$\mathcal{D}$ scheme. Our experiments are based on tens of the hard disks using NTFS file system, whose storage sizes are all 250 GB with 488,392,065 sectors. We search for

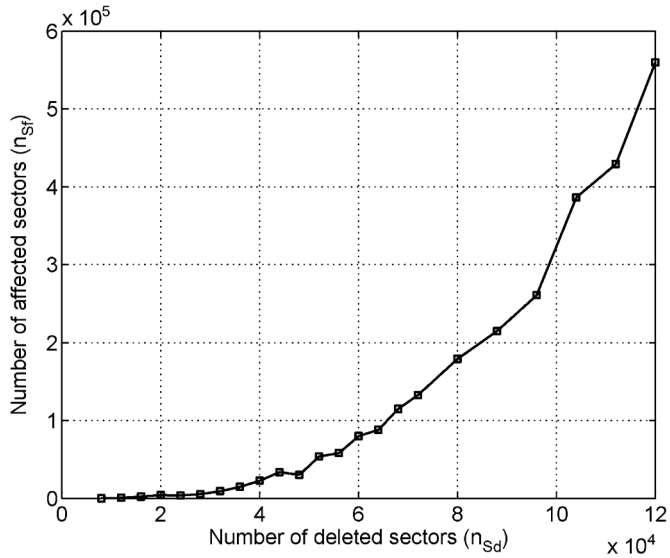


Fig. 2. Number of affected sectors in 3-$\mathcal{D}$ scheme varies with the number of deleted sectors.

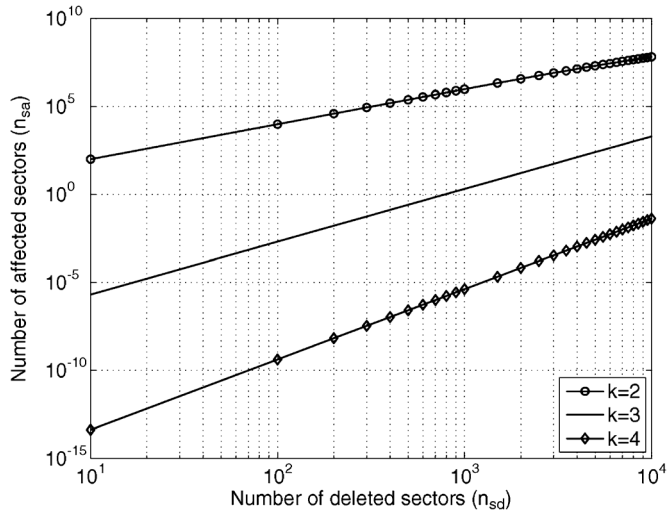


Fig. 3. Estimated number of affected sectors in $k-\mathcal{D}$ scheme varies with the number of deleted sectors.

all Word documents with "doc" and "docx" as file extension from normal users' hard disks, and find that on average there are about 3000 Word documents on a hard disk with an average size of 200 K Bytes (i.e. 400 sectors or 50 clusters). The experimental results for 3-$\mathcal{D}$ scheme are shown in Fig. 2. 3000 Word documents are prepared for this experiment and each document occupies 400 continuous sectors in the hard disk on average. When the number of deleted sectors for LPP documents is increased from 8000 to 120000, the number of affected sectors for 3-$\mathcal{D}$ scheme is increased from 100 to 551036 on average, i.e., the number of affected sectors is increased 5510 times. This result indicates that when the number of deleted sectors increases, the chance that an undeleted sector can be verified becomes lower.

The probability of a sector which becomes unverifiable can be computed as follows. Assume that there is a hard disk drive containing N sectors and the sectors are distributed on k dimensions. The length of each dimension should be $R = N^{1/k}$. If there is only one sector deleted, the probability for each chain in $d1$ dimension which will become unverifiable due to the deleted sectors will be $1/N^{(k-1)/k}$ because there are totally $N^{(k-1)/k}$ chains in $d1$ dimension, and the probability for each chain remains unaffected will be $(1 - 1/N^{(k-1)/k})$. Therefore, if there are n_{sd} sectors deleted, the probability for each chain remains unaffected will be $(1 - 1/N^{(k-1)/k})^{n_{sd}}$, and the probability for each chain becomes unverifiable will be $1-(1-1/N^{(k-1)/k})^{n_{sd}}$. The probability of a sector which will become unverifiable in $d1$ dimension is equivalent to the probability for each chain becomes unverifiable due to the deleted sectors, denoted as $p_{d1} = 1 - (1 - 1/N^{(k-1)/k})^{n_{sd}}$. Similarly, the probability for the other dimensions can be denoted as $p_{dn} = 1 - (1 - 1/N^{(k-1)/k})^{n_{sd}}$, where $dn = d1, d2, \ldots k$. A sector in the $k - \mathcal{D}$ space will be affected only when all of those chains across the sector become unverifiable. Thus, the probability of a sector becomes unverifiable will be $p_{sf} = p_{d1} \times p_{d2} \times \ldots \times p_{dk}$. Assuming that the affected sectors are independent, then we can calculate the upper bound of the estimated number of affected sectors due to deleted LPP sectors as: $n_{sf} = (N - n_{sd}) \times p_{d1} \times p_{d2} \times \ldots \times p_{dk}$, as shown in Fig. 3. Note that the assumption of independence may not hold in real cases, the formula only provides a rough estimation for reference only.

D. Our Scheme

Recall that we want to solve the following problem. Denote the total N sectors in the hard disk as a set $S_1 = \{s_m | 0 \leq m \leq N-1\}$. Conceptually, each sector s_m can be mapped to a unique point, $s_{x,y,z}$, in a 3-dimensional space as follows. Roughly speaking, we order the sectors along the Z dimension, then Y, then X dimension.

$$x = \left\lfloor \frac{m}{N^{\frac{2}{3}}} \right\rfloor ; y = \left\lfloor \frac{m \bmod N^{\frac{2}{3}}}{N^{\frac{1}{3}}} \right\rfloor ;$$
$$z = (m \bmod N^{\frac{2}{3}}) \bmod N^{\frac{1}{3}}.$$

Then each sector s_m can be rewritten as $s_{x,y,z}$ and the set of all sectors S_1 can be represented as $S_3 = \{s_{x,y,z} | 0 \leq x, y, z \leq N^{1/3}\}$. For simplicity, we consider the case that N is an integer to the power of 3. For the other cases, the scheme can be easily extended to handle them. Let l be the number of sectors (containing LPP documents) to be deleted. We want to verify the integrity of the remaining $N - l$ sectors without doing a bit-by-bit comparison with the "sealed" original hard disk.

For convenience, the set of l sectors (the ones that have been crossed in Fig. 4) to be deleted is denoted as

$$Sd_3 = \{s_{i,j,k}\} \subset S_3,$$

and the remaining $N - l$ sectors denoted as

$$Snd_3 = \{s_{x,y,z}\} = S_3 - Sd_3.$$

So, the problem can be described as how to verify the integrity of all sectors in Snd_3 after sectors in Sd_3 have been deleted.

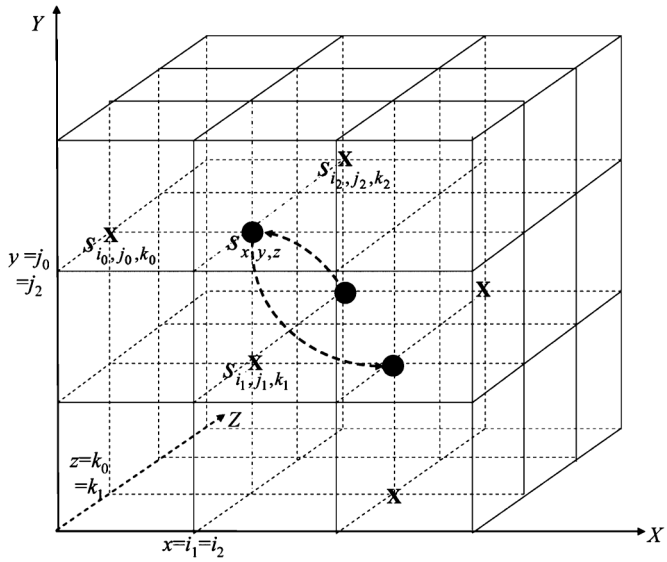


Fig. 4. Three-dimensional structure.

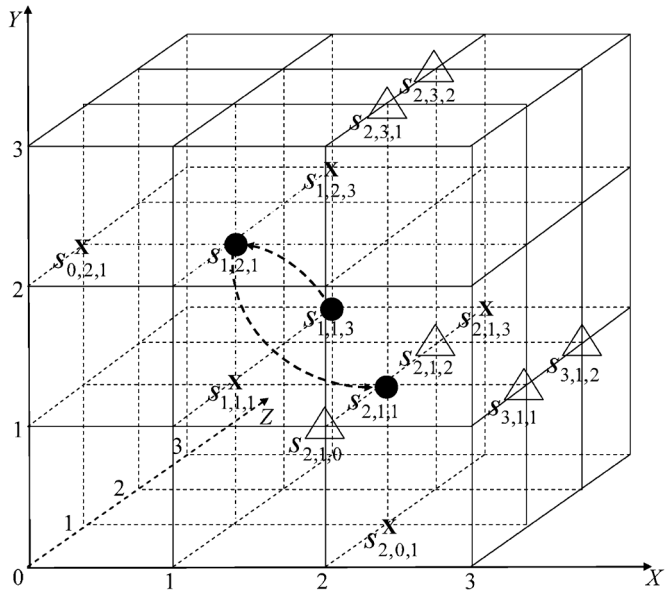


Fig. 5. Three-dimensional example.

Let the hash values for sector chains in three dimensions be denoted, respectively, as follows.

$$VX_{y,z} = Hash\left(s_{1,y,z} \| s_{2,y,z} \| \cdots \| s_{N^{\frac{1}{3}},y,z}\right),$$
$$VY_{x,z} = Hash\left(s_{x,1,z} \| s_{x,2,z} \| \cdots \| s_{x,N^{\frac{1}{3}},z}\right),$$
$$VZ_{x,y} = Hash\left(s_{x,y,1} \| s_{x,y,2} \| \cdots \| s_{x,y,N^{\frac{1}{3}}}\right).$$

These hash values should be precomputed and stored in a secure place. After the deletion of l sectors including LPP documents, we compare the new hash values for the sector chains with the precomputed hash values. The integrity of a sector $s_{x,y,z} \in Snd_3$ can still be verified if no sectors $\in Sd_3$ in at least one of its corresponding three sector chains have been deleted. If that is the case, to verify whether $s_{x,y,z} \in Snd_3$ is changed or not, we do the following.

VERIFICATION(1): We recalculate the three hash values of sector chain containing $s_{x,y,z}$, $VX'_{y,z}$, $VY'_{x,z}$, and $VZ'_{x,y}$, respectively. Obviously if at least one of the following holds, we can guarantee the integrity of $s_{x,y,z}$: $VX_{y,z} = VX'_{y,z}$ or $VX_{x,z} = VX'_{x,z}$ or $VX_{x,y} = VX'_{x,y}$.

However, there may exist cases (undeleted sectors) in which the above does not hold if there exists deleted sectors in its three chains.

1) A Practical Issue: In our practical cases, if the sectors containing the evidence are all verifiable using the hash values, then it is perfect. However, this may not be true. To handle this case, we make an extra hash value of the affected sectors chain after the LPP-related sectors are known. Note that this step is undesirable as it requires the access of the original sealed hard disk. First, this may not be needed. Second, the affected sectors are expected to be small compared to the total number of sectors, so hopefully the chance of damaging the original copy is low by accessing this limited number of sectors. Of course, since we will access the sealed copy, there is no need to compute the hash value of the additional chain, we can directly compare the bit stream of the affected sectors. However, this would increase the number of access to the original copy since in practice, we may find out evidence in different stages which may imply multiple access to the original hard disk.

The modified 3-$\mathcal{D}$ hash scheme is as follows. Let $Sf_3 \subset Snd_3$ be the sectors that fail the verification. For example, all the black sectors illustrated in Fig. 5 are the affected sectors by Sd_3. To verify these sectors, we construct an extra hash value for the chain of sectors in Sf_3 (see the dashed line in Fig. 5):

$$VE = \mathrm{Hash}(Sf_3).$$

It is easy to define an order for the sectors in Sf_3 for computing VE. For example, we can sort the sectors according to its first index, then its second index and so on.

VERIFICATION(2): To verify the sectors in Sf_3, we compute VE' as the hash value for the sectors in Sf_3 and check if $VE = VE'$.

Fig. 5 illustrates a concrete example with $N = 64$, where $Sd_3 = \{s_{0,2,1}, s_{1,1,1}, s_{1,2,3}, s_{2,0,1}, s_{2,1,3}\}$, $Sf_3 = \{s_{1,1,3}, s_{1,2,1}, s_{2,1,1}\}$, and the extra sector chain is the dash line with arrows showing the order of the sectors in this chain.

Note that to apply this scheme, we need to modify the "Procedure for Handling LPP information" given in Section III-A as follows. Note that the 3-$\mathcal{D}$ hash values are computed at Step 1. We need to add Step 5 which is to compute the additional hash value for the affected sectors chain if needed. Note that both parties should keep a copy of all hash values computed in the process. We expect that l may not be a very big number compared to the total number of sectors in the hard drive. Therefore, this computation can be done efficiently. After all hash values have been computed, there is no need to refer to the sealed original hard drive for integrity verification.

E. Identification of Modified or Wrongly Deleted Sectors

In this scheme, based on the verification process, it is possible to identify the set of sectors that may have been wrongly modified or deleted when there are mismatched hash values that do not satisfy the verification (1).

After the verification, let the set of hash values VF be the hash values that fail the verification (1). Then calculate a set, Sh_3, including all the intersections of every three hash value of sector chains $\in VF$. Denote $Sp_3 = Sh_3 - Sd_3 - Sf_3$. Sp_3 is exactly those problematic sectors which are wrongly deleted.

For example in Fig. 5, Assume that we got $VF = \{VX_{0,1}, VX_{1,0}, VX_{1,1}, VX_{1,2}, VX_{1,3}, VX_{2,1}, VX_{2,3}, VX_{3,1}, VX_{3,2}, VY_{0,1}, VY_{1,1}, VY_{1,3}, VY_{2,0}, VY_{2,1}, VY_{2,2}, VY_{2,3}, VY_{3,1}, VY_{3,2}, VZ_{0,2}, VZ_{1,1,}, VZ_{1,2}, VZ_{2,0}, VZ_{2,1}, VZ_{2,3}, VZ_{3,1}\}$. Then all the intersections of every three sector chains corresponding to hash values $\in VF$ form a set Sh_3. Finally, we can identify the wrongly deleted sectors $Sp_3 = Sh_3 - Sd_3 - Sf_3 = \{s_{2,3,2}, s_{2,1,0}, s_{2,1,2}, s_{3,1,2}, s_{3,1,1}, s_{2,3,1}\}$, which are denoted by triangles in Fig. 5.

IV. Discussion and Conclusions

The solution proposed by the paper has been adopted to resolve the case described in Section II. The following is the actual procedure (slightly different from what we described in Section II) used in the real case. Note that the time taken for the defendant to identify LPP documents is actually very long which matches our observation that a face-to-face meeting for identifying LPP documents is not feasible. 1) The Police clones the original image making two copies—one evidence copy and one screening copy; 2) The evidence copy is retained and not reviewed; 3) The screening copy is provided to the company representatives involved in the privilege issue to identify and redact privileged information; 4) This screening routinely takes a very long time; 5) Once the screening and redaction is done, the material is returned to the Police; 6) It is then digitally compared against the original evidence copy, and the redactions assessed for relevance; 7) Only then can the digital forensic examinations begin.

In Step 1, they use the hashing approach suggested in the paper (with $k = 3$). The approach facilitates enough times at both sides to conduct necessary data identification and searching, resulting in a more effective, efficient and legally justified approach to deal with LPP data. Luckily, we do not need to access the original sealed copy to compute a hash value for the affected sector as the sectors containing evidence can be verified by the hash values computed in Step 1. Note that in this real practical, instead of having only one cloned copy, the police has created two to increase the resilience. As a remark, the bottleneck still lies on Step 4 (the screening of files by the defendant to identify LPP documents). Since there is no proper way to prescreen the material to validate the claim, it seems that the proposed approach is still a feasible one and has a good balance between investigation and justice. Because the LPP data identification phase took very long time in this real case, the screening and redaction are only completed for about 30% of the digital data. Nevertheless, some of the redacted data have been returned to the police for forensic examination. The approach was proved to be effective in regard to the current case progress.

To summarize, in this paper, we address an important, but not adequately addressed in the community, issue for protecting digital legal professional privilege (LPP) information during forensics investigation. We highlighted the differences of digital LPP information and physical LPP information and discussed the difficulties of handling digital LPP data. We also investigated the current practices for handling this kind of information and concluded that these practices cannot guarantee the protection of LLP data or may create obstacles for forensics investigation. Also, both practices rely on the face-to-face interaction between the examiner and the data owner during the identification of LPP information in the target storage media. As the volume of storage device become larger, this involves a lengthy process and it is not practical to require face-to-face interaction.

We then propose a feasible solution to solve this problem. In our proposed approach, there is no need for both parties to get together to identify the LPP information, thus providing a better protection to the privacy of the LPP information and avoiding the lengthy and impractical face-to-face interaction between both parties. We show the effectiveness of our solution on a real case. Future research directions include the following. The integrity scheme proposed in the paper, of course, is not the only solution nor an ideal solution to the problem. Finding a better scheme and procedure to solve this digital LPP protection problem is always desirable. As this problem is still new to the law enforcement parties, the number of real cases is limited. When more real cases exist, a more comprehensive study should be conducted. The bottleneck process of identifying LPP documents should be further investigated and see if a more efficient, but forensically sound approach could be developed. We hope that this paper can catch the attention of the community to help developing a better solution to solve this problem.

References

[1] P. E. Nygh and P. Butt, *Butterworths Concise Australian Legal Dictionary*. Sydney, Australia: Butterworths, 1997.

[2] DFRWS, Report from the First Digital Forensic Research Workshop. DTR-T001-01 FINAL. A Road Map for Digital Forensic Research, Nov. 6, 2001.

[3] *Cavendish Lawcards Series—Evidence*. London: Cavendish Publishing Ltd., 2004.

[4] R. S. C. Ieong, "FORZA—Digital forensics investigation framework that incorporate legal issues," in *Proc. Digital Forensics Research Workshop (DFRWS)*, 2006, pp. 29–36.

[5] F. Y. W. Law, P. K. Y. Lai, Z. L. Jiang, R. S. C. Ieong, M. Y. K. Kwan, K. P. Chow, L. C. K. Hui, S. M. Yiu, and C. F. Chong, "Protecting digital legal professional privilege (LPP) data," in *Proc. Third Int. Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE2008)*, California, USA, 2008, pp. 91–101.

[6] Association of Chief Police Officers (ACPO), Good Practice Guide for Computer Based Electronic Evidence [Online]. Available: http://www.dataclinic.co.uk/ACPO%20Guide%20v3.0.pdf accessed on Jan. 31, 2008

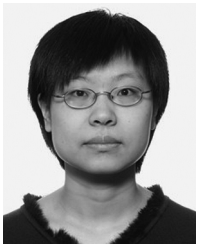
[7] P. Turner, "Selective and intelligent imaging using digital evidence bags," *Digital Investigat.*, vol. 3, no. 1, pp. 59–64, 2006.

[8] Anti Cartel Enforcement Manual International Competition Network, April 2006 [Online]. Available: http://www.internationalcompetitionnetwork.org/media/library/conference 5th capetown 2006/DigitalEvidenceGathering.pdf, accessed on Jan. 30, 2008

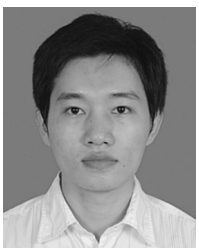
[9] P. Turner, "Unification of evidence from disparate sources (digital evidence bags)," in *Proc. Digital Forensic Research Workshop (DFRWS)*, 2005, pp. 223–228.

[10] Z. L. Jiang, L. C. K. Hui, K. P. Chow, S. M. Yiu, and P. K. Y. Lai, "Improving disk sector integrity using 3-dimension hashing scheme," in *Proc. 2007 Int. Workshop on Forensics for Future Generation Communication*, 2007, vol. 2, pp. 141–145.

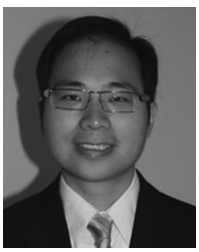
[11] Z. L. Jiang, L. C. K. Hui, and S. M. Yiu, "Improving disk sector integrity using k-dimension hashing," in *Proc. 4th Ann. IFIP WG 11.9 Int. Conf. Digital Forensics*, 2008, vol. 285, no. 14, pp. 87–98.



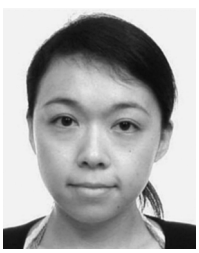
Zoe L. Jiang is a postdoctoral researcher with the Shenzhen Graduate School, Harbin Institute of Technology. She received the Ph.D. degree from The University of Hong Kong in 2010. Her research interests include digital forensics and applied cryptography.



Junbin Fang is an associate professor with the Department of Optoelectronic Engineering, Jinan University, Guangzhou, China. He is also a Visiting Scholar with the Department of Computer Science, The University of Hong Kong, Hong Kong, China. His research interests include digital forensics, information security, and quantum cryptography.



Frank Y. W. Law has been working with the Hong Kong Police Force since 1998, and has been involved in technology-crime-related policing since 2001. He is currently the Head of Cyber Security of the Hong Kong Police Force, in charge of three teams concerning the areas of cyber attack responses. His research interests include live systems forensics, digital forensics, and digital timestamp analysis.



Pierre K. Y. Lai is a guest lecturer with the Department of Computer Science, University of Hong Kong, Hong Kong, China. Her research interests include cryptography, peer-to-peer networks, and digital forensics.

Ricci S. C. Ieong is a Ph.D. candidate in the Department of Computer Science, The University of Hong Kong, specialized in peer-to-peer, cloud, and network forensics.

Michael Y. K. Kwan is an appointed Honorary Assistant Professor with the Department of Computer Science, University of Hong Kong. His research is on digital forensic analysis.

K. P. Chow is an Associate Professor with the Department of Computer Science and the Associate Director of the Center for Information Security and Cryptography (CISC), The University of Hong Kong. His areas of research interest are computer forensics, cryptography, computer security, Internet surveillance, and privacy.

In the past few years, Dr. Chow has been invited to be a computer forensic expert to assist the Court in Hong Kong.

Lucas C. K. Hui is the founder and Honorary Director of the Center for Information Security and Cryptography (CISC), and concurrently an associate professor with the Department of Computer Science, The University of Hong Kong. His research interests include different and diversified areas in information security, privacy protection in e-commerce, Internet security, mobile device security, smart grid security, and cryptography.

S. M. Yiu is currently an Associate Professor with the Department of Computer Science, the University of Hong Kong. His research interests include computer security, cryptography, digital forensics, and bioinformatics.

K. H. Pun is an Associate Professor of Computer Science, a Codirector of the Law and Technology Centre, and the Director of the Hong Kong Legal Information Institute (HKLII), University of Hong Kong. He is also a barrister in Hong Kong practising intellectual property, information technology law, and other civil matters.