A Digital Forensic Investigation and Verification Model for Industrial Espionage

Jieun Dokko[1,2(✉)] and Michael Shin[1]

[1] Department of Computer Science,
Texas Tech University, Lubbock, TX 19409, USA
{jieun.dokko,michael.shin}@ttu.edu
[2] Supreme Prosecutors' Office, Seoul 06590, Republic of Korea

Abstract. This paper describes a digital forensic investigation and verification model for industrial espionage (DEIV-IE) focusing on insider data thefts at the company level. This model aims to advance the state-of practice in forensic investigation and to verify evidence sufficiency of industrial espionage cases by incorporating the crime specific features and analysis techniques of digital evidence. The model is structured with six phases: file reduction, file classification, crime feature identification, evidence mapping, evidence sufficiency verification, and documentations. In particular, we focus on characterizing crime features that have multiple aspects of commonalities in crime patterns in industrial espionage; and the evidence sufficiency verification that is a verification procedure for digital evidence sufficiency for court decision using these crime features. This model has been developed based on analysis of five industrial espionage cases and the literature review, being validated with three additional cases in terms of the effectiveness of the model.

Keywords: Digital forensic investigation · Digital evidence verification
Evidence prioritization · Behavioral evidence analysis · Digital forensics triage
Industrial espionage

1 Introduction

Historically, digital forensic researches have focused on quickly obtaining digital evidence in a crime scene in a limited time and analyzing digital evidence in technical manners. Within this framework, various digital forensic investigation models have been developed to support analytical techniques for digital forensics [1]. These researches enabled a digital forensic investigator to follow the procedures defined for the techniques, thereby discovering the evidence. However, legal practitioners, such as prosecutors or case investigators, have always faced the challenge of understanding the technical complexity of the evidence [2] because existing digital forensic investigation models are too procedural and technical. Moreover, digital forensic investigators cannot fully understand each type of crime and what information should be examined for the successful prosecution of the case [3]. Also, legal practitioners cannot always advise forensic investigators as to what information should be searched during the

F. Breitinger and I. Baggili (Eds.): ICDF2C 2018, LNICST 259, pp. 128–146, 2019.
https://doi.org/10.1007/978-3-030-05487-8_7

entire investigation process. Thus, there always exists the gap between digital forensic investigators and legal practitioners [4] and some meaningful evidence for court decision can be often overlooked. There is, however, little research examining each crime's features and the application of the features to the digital forensic investigation, particularly in the case of industrial espionage.

As the use of information technologies has become an ever growing factor in successful business, industrial espionage has been growing, thereby producing a high volume of digital data to be investigated [5]. Plus, industrial espionage cases usually require discovering indirect evidence that is related to certain circumstances implicated in the plots and might be connected to other factual information, which can lead to a probable conclusion in the case. Therefore, without a reasonable analysis and verification method for industrial espionage cases, the investigation can become multifaceted and result in overlooked evidence.

To fill this gap, this paper proposes a digital forensic investigation and verification model for industrial espionage (DEIV-IE), which identifies crime features and specifies available digital evidence from the crime features. Also the model verifies the sufficiency of evidential findings. The main contribution of this work is twofold. First, it advances the crime specific investigation practice of industrial espionage cases. Secondly, it provides a digital evidence verification practice in the investigation, which is necessary to establish factual information for court decision.

This paper is organized as follows. Section 2 presents the relevant work related to crime specific investigation and verification methods. Section 3 describes the overview of this model. Section 4 presents the file classification with the characteristics related to the legal recognition and the usages of digital evidence. Section 5 characterizes evidentiary crime features and maps them to evidence. Section 6 describes the evidence sufficiency verification. Section 7 validates the efficiency of the model and Sect. 8 concludes this paper.

2 Related Work

2.1 Crime Specific Investigation

Some studies have focused on developing the crime specific investigation models that can compensate forensic investigators for lack of domain knowledge of a crime. Among them, the author in [8] describes a semi-automated and crime specific triage model, which helps an investigator prioritize and lead the examination process. Even though it discusses crime specific features such as live forensics, computer profile, and crime potential, it is still a high level framework that should be adjusted to a specific type of crime. The research in [6, 7] presents a mobile forensic triage model relying upon manually prioritized crime features in child pornography and copyright infringement cases. This model is enhanced by going one step further than the previous one because it discusses where to search for evidence focusing on child pornography and copyright infringement. The authors in [6, 7] verified their models based on mathematical theories, probabilities, and comparison of algorithms in digital forensic studies. The authors in [9] have studied a digital forensic investigator's decision

process model for child exploitation cases, which helps investigators exclude non-relevant media from further in-depth analysis.

Authors in [35] study forensic investigation of cyberstalking cases using behavioral evidence analysis (BEA). Unlike other crime specific investigation models that find digital evidence by using specified crime features, the study in [35] reversely finds the specific crime features of cyberstalking (such as traits of offenders or means of committing the crime) by using digital evidence detected in actual cyberstalking cases. The research in [36] proposes a behavioral digital evidence analysis approach applied to 15 actual child pornography cases using P2P networks. As in other BEA analysis, the study in [36] aims to profile offender characteristics and behaviors by analyzing the potential evidential files. However, this work simply identifies potential evidential files using video or image files, and identifies the location of these files in the Download folder or the Program Files directory where P2P software and downloaded child pornography files are stored by default. Even though this work addresses the pool of potential evidential files, it lacks analytic approaches because such crime doesn't require the in-depth analysis to identify the potential evidential files in nature.

2.2 Digital Evidence Verification

Investigators need to verify that the digital evidence presented to courts is qualified enough to prove the crime, as analysis of digital evidence plays a key role in crime solving [10]. To answer this need, many researches have proposed various frameworks for verifying digital evidence and enhancing the investigation process for a legal argument. The study in [2] suggests the need of a validation stage in the investigation domain that creates a chain of evidence, and clarifies or nullifies an additional assertion derived from the succeeding evidence, bridging the gap between digital forensic experts and legal practitioners. However, this work mostly focused on the legal side, but not actual examination of a specific crime. The authors in [11] present a framework to help investigators to assess the validity, weight and admissibility of evidence with less effort using an interrogative approach. This work evaluates potential evidence using a relationship between other potential evidences, which enhances the presentation and interpretation of digital evidence in a legal process. Similarly, the research in [10] proposes a framework for preparing a qualified report where the traceability of digital evidence has also been enhanced by the proof of digital evidence's origin and history. The authors in [12] describe a traceability model based on a scenario for a digital forensic investigation process, which can help digital forensic investigators to identify the origin of the incident as well as the evidence itself. The study in [13] explains a genetically traceable framework highlighting that the identity, history and origin of extracted digital evidence should be verifiable through scientifically accepted manners in a legal argument. Research in [1] assesses the eleven existing digital forensic investigation process models (DFIPM) against the five criteria of 'Daubert Test' to decide a level of reliability of models. The research shows that no one DFIPM can take the most scientific approach during the investigation because each model has developed based on personal experience and on an ad-hoc basis.

3 Digital Forensic Investigation and Verification Model

The digital forensic investigation and verification model for industrial espionage (DEIV-IE) is designed to examine a stand-alone computer where a windows operating system is installed, sometimes connected with external devices. We assume that the forensic image files captured from seized devices have always been acquired in a forensically sound manner before the investigation process starts, and the acquisition of the seized devices has not been altered since they were acquired. Thus, this model does not address the digital evidence integrity and the environment verification including application, operating system and hardware platform. Figure 1 depicts the overview of DEIV-IE consisting of file reduction, file classification, crime feature identification, evidence mapping, evidence sufficiency verification, and documentations.

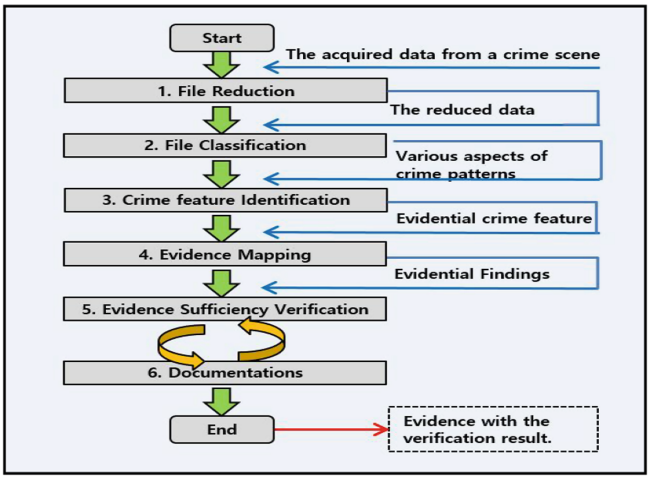

Fig. 1. Overview of the digital forensic investigation and verification model.

File Reduction. This phase eliminates the files created by a non-user from all the files that are acquired from a targeted computer. We assume that a file signature analysis is performed using forensic software and non-user created files are reduced automatically by means of hash techniques [14] before DEIV-IE proceeds to the next phase. For example, the files created by an installation program or operating system can be reduced because they are types of files unlikely to be created and accessed by users [15].

File Classification. In this phase, the reduced files in the previous phase can be categorized into five individual groups based on the expertise of an investigator: user files, communication files, user input files, system files, and less-identified files by considering both the way of creating a file, and a court's different determination regarding the admissibility for each type of file as evidence.

Crime Feature Identification. This phase identifies evidentiary crime features, which are used to guide investigation priority setting to discover the evidence to determine

whether an industrial espionage has occurred. The crime features are characterized on account of the multiple aspects of the crime patterns [16] such as behaviors and characteristics of the offender(s) and victims, the property taken or stolen by the crime, and the time the crime occurred.

Evidence Mapping. This phase finds the relevant files that can be used for solving the questions derived from each crime feature by systematically mapping between the crime features and the categorized file groups.

Sufficiency Verification and Documentations. This phase serves as a type of checklist to verify whether an investigator has found sufficient evidence to establish factual information for court decision on an industrial espionage case. This phase carries out the five interrogative questions and their answering, and the process documentations.

4 File Classification

File classification, commonly carried out regardless of crime types, has been implemented in various digital forensic tools by sorting and grouping the files based on technical criteria, such as file extension, size, and metadata [17]. DEIV-IE, however, classifies files into five groups based on the experience and expertise of investigators, in particular, considering practical criteria such as 'who creates the file', 'why the file is created' and 'how the file is created' and also by considering various legal viewpoints in the file groups that are shown in Table 1. Thus, this grouping is adjustable and the list is not complete, it can be added to or altered based on investigators' expertise.

Table 1. File classification

File Group	Sub Group	Files (e.g., of its extension)
User file	Document	doc(x), ppt(x), xls(x), hwp, pdf, txt, cvs, xml
	Image	jpg, jpeg, tiff, bmp, png, gif
	Multimedia	mp3, mp4, avi, wma, wmv, rm, ram, mpg
	Program code	java, c, cpp, py, dll, css, asp, sys
Communication file	Email	pst, dbx, ost, eml, edb, msg, idx, nsf
	SMS or MMS	db, splite, im
User input file	Financial document Business activity	dbf, dbk, db, fp, md, mda, sql, xls(x)
System file	Internet	html, htm, dat, db, exb, sql, url, png
	Registry	registry files
	Log	log, txt, db, md, sql, spl, shd
	Shortcut	lnk, yak
	Installation	ext, bat, dll, sys, vxd, bin, java, c, cpp, py
Less-identified file	Compressed, backup	zip, tar, pak, cab, jar, tgz, rar, alz, backup
	Embedded, Encrypted	tmp, EMF, sav

User Files. User files are manually created by a user and classified into 4 subgroups: document, image, multimedia, and program code file group. Some files in these groups are generated by the system; image files less than 1 MB can be web cache or temporarily downloaded webpage files, whereas most of the user created image files tend to be larger than 1 MB. User files are rarely accepted in court [18] even if they link to a crime directly because they are regarded as the same as a statement and can be highly modified by a user. Such files can fall into 'hearsay', which cannot be used as evidence in most courts [19].

Communication Files. Communication files are mainly used to communicate and share information among people using Internet, and classified into 2 subgroups: emails, instant or text message file groups. They are likely accepted in court although they are created by a user, because these files' metadata (e.g., the sent or received timestamp) is created by software, and rarely rejected by court [20]. Also a file in the sender side can be cross-checked against the file in the receiver side in terms of the file's integrity and authenticity.

User Input Files. User input files are created by computer processing with a user input, being related to either regularly conducted business activity or financial records, and classified into financial document and business activity file groups. These files are admissible in court, because a court accepts regularly conducted activity information based on the inherent reliability of business records [21]. But, DEIV-IE excludes this group from evidence mapping phase because such files are not related to user's intentional activities and scarcely noticed in industrial espionage cases.

System Files. System files are automatically created and utilized by the system to manage operations, and classified into 5 subgroups: Internet, registry, log, shortcut, and installation. Internet file group mainly demonstrates a user's Internet activities or system's Internet usages. Registry shows information about the system and its devices as well as the activities generated by a user or a system. Log files record program activities created by the system. Shortcut files are created to facilitate repeated access swiftly when a user accesses a file, an external device and a shared network, so previous activities of a user or a system are traceable [22]. Installation files including executable files is created by system according to its pre-scripted procedure. These files are credible in court because they are not hearsay and created solely by a computer without any human interaction [23].

Less-identified Files. A file falling into a less-identified file group can contain another file(s) or latent pieces of information embedded somewhere inside the file, thus the challenge of examining and processing evidential information out of them remains. Less-identified files in this model are classified into two subgroups: compressed or backup, and embedded or encrypted. A compressed file may contain multiple individual files, and an encrypted file cannot be read without its decryption.

The file classification can be beneficial to preparing for trials because the authentication and reliability of files in each group are differently accepted by court from legal viewpoints. The evidence corroboration can be carried out with the combination of different evidence(s) in each file group, which can be an essential tool for the successful prosecution of a case. These evidences can support or corroborate each other so that

they confirm the proposition and enhance the reliability of other evidence. For example, a document file that describes a plot, motive, and criminal behaviors of a suspect cannot be used alone as evidence in a court due to hearsay. However, the document file can be assumed to be accurate if reliable evidence (e.g., log, registry, emails) can link this document to the suspect and prove the crime.

5 Crime Feature Identification and Evidence Mapping

5.1 Industrial Espionage

70% of a company's assets lies in market-sensitive information [24] such as client lists, supplier agreements, personal records, research documents, and prototype plans for a new product or service. Industrial espionage makes use of many different methods such as computer hacking, dumpster diving, electronic surveillance, and reverse engineering [25], but almost 85% of espionage cases originate from insiders within an organization [26] who cooperate with a criminal authority outside the organization. Our model deals with the theft of trade secrets and intellectual property of a company, focusing on an insider threat. This model has been established using a typical scenario of industrial espionage cases revealed by the review of digital forensic analysis reports of five cases as follows:

A discontented employee in a company conspires with a rival company who seeks a competitor's assets e.g., patents, inventions, or trade secrets. The employee covertly takes the company's confidential information, and either joins a rival company with a promotion agreement or receives compensation for handing over the information.

Five Reference Cases. The digital forensic cases that were reviewed for developing DEIV-IE in this paper were examined by 7 digital forensic investigators with 2 to 8 years' analysis experience in the prosecutors' office in Korea 2008 to 2014. The summaries of each case are as follows.

Case 1. A suspect working for a company that provides industrial maintenance products and services for cleaning waste water was suspected of stealing the company's protected document files related to marketing strategy (e.g., loan annex) using his office computer before his resignation.
Case 2. A suspect working as a marketing manager for a manufacturer of Radio-Frequency Identification (RFID) devices was accused of revealing the company's secret files of marketing and products to a rival company, where the suspect subsequently gained employment.
Case 3. A suspect working as a salesperson in a company of eco-friendly products, e.g., electronic railway trolleys and oil flushing equipment, was suspected of stealing protected computer-aided design (CAD) files before his resignation.
Case 4. A suspect who worked at a light aircraft manufacturer was charged with stealing protected information about agricultural aircraft development and marketing.
Case 5. A researcher in an electronic technology institute was suspected of releasing document files related to nuclear power energy to a person outside the company.

5.2 Crime Features with Evidence

The crime features in this paper are defined in accordance with the principle of 4W1H questions that should always be asked during the entire investigation of a crime; 'When' for the time period that the crime occurred, 'who' for the suspect(s), 'whom' for the accomplice(s), 'what' for the files or data misappropriated or stolen, 'how' for the criminal behaviors or activities that occurred. However, this model excludes 'where' and 'why' because the computer to be examined is where for the place the crime occurred [27] which is the scope the model is limited to and the motivation of the crime can be acknowledged indirectly through the investigation. This model finds the files relevant to each crime feature by systematically mapping crime features to classified file groups. It is not necessary for the model to identify specific evidential files, but it narrows down the range of potential evidential files. This mapping result can also be used to guide investigation priority setting. The evidence mapping in DEIV-IE is hierarchically depicted in Fig. 2, which categorizes the 4W1H questions, the crime features, the classified file groups, and the sub-file groups. The relationships between the categories (Fig. 2) are represented by means of generalization and specialization, question and answer, and the whole and its parts, which are described with the notation in Table 2.

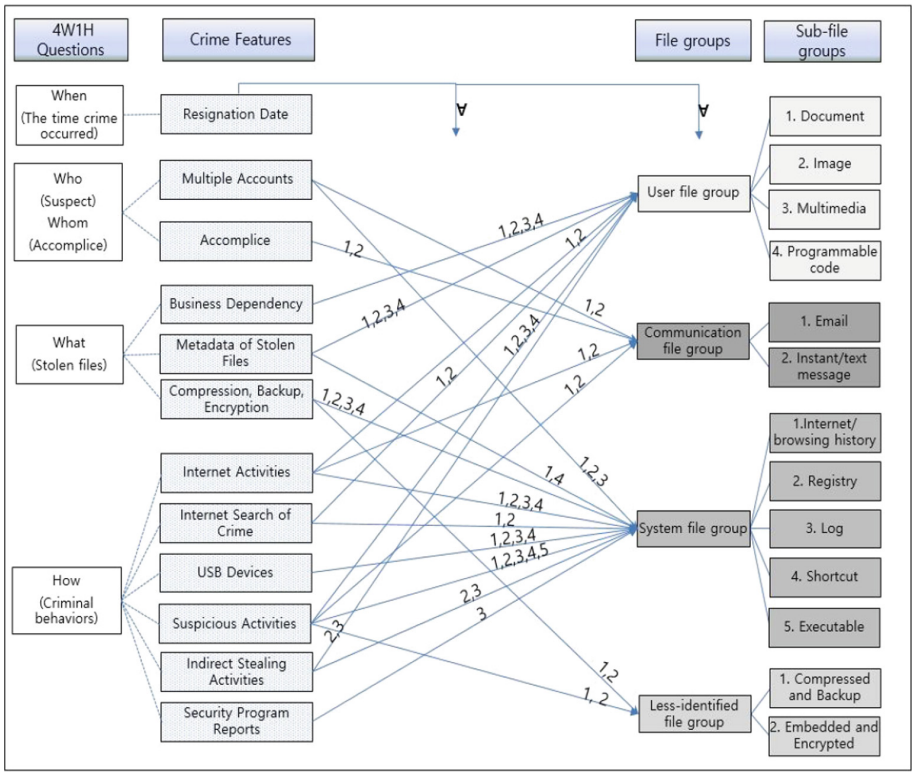

Fig. 2. Outline of mapping between crime features and potential evidence in industrial espionage

Table 2. Relationship between in the model.

Relationship	Description	Notation
Generalization, Specialization	A component of a higher layer is semantically the general question of component(s) of the lower layer, expanded question(s) for more specified inquiry	
Question & Answer	A component of a higher layer can be proved by file(s) in the file group(s) pointed out by an arrow from the component (One-directional relationship). The number(s) on an arrow indicates the reference number of its sub file group(s) in its lower layer	
Whole & Parts	A whole component of a higher layer consists of its distinct components of a lower layer	

Resignation Date (When). Accurate estimation of the period of suspected criminal activities is a crucial skill in the crime investigation. The majority of suspects in industrial espionage cases tend to plan their activities in advance. Five cases of the digital forensic analysis reports show that the most critical evidences are revealed in the period four months prior and up until a suspect leaves a company. This feature should be always prioritized in the entire investigation and examined in conjunction with other crime features. An investigator can narrow down the scope of investigation and detect a suspect's doubtable activities when he/she concentrates on the activities that have occurred for four months before a suspect resigns. The resignation date feature is mapped to all file groups because each file has the timestamps in terms of creation, deletion, modification or copy, thereby being relevant to determining the period of criminal activities. The resignation date feature is expressed as '∀' above the top two arrows in Fig. 2, which are applied to all other arrows and all file groups.

Multiple Accounts (Who). The investigation needs to identify who has used the computer during the alleged time period. In industrial espionage, the range of suspects tends to be limited to the employees who used the suspected computer and who have rights to access the targeted information. However, a suspected computer might be used by multiple users who were not the owner of the computer in an organization. Also, a suspect can create an unauthorized backdoor account or use other person's account to steal secret information [26]. Case 3 shows that several employees used the suspected computer during the alleged time period. This feature should not be neglected because several, unexpected employees can be involved in this crime, even if in some cases it can be insignificant because most of the Hi-Tech companies tend to have their own policies that prohibit employees from shared computer usage.

This feature can be mapped to the registry, Internet history, and log files in the system file group and communication file group to identify who has used the computer (Fig. 2). Registry files contain each user's profile including registered accounts, computer name, the last login account with timestamps, and the registered user to a certain service or software [22]. The Internet history files (e.g., index.dat, webcachev01.dat or cookie files) enable an investigator to trace each account and the timestamps for the user's Internet use and browsing history. A log file (e.g., a security

event log for window auditing) can be used to track user accounts with their activities. Emails and Instant messages (IM) in the communication file group contain the email addresses, and the sender and receiver names. Cross-checking of different files can help in finding the suspect because a system or software creates or updates different types of files (e.g., registry, log, Internet history, email, and IM files) resulting from just one action.

Accomplice (Whom). Industrial espionages tend to be committed in cooperation with accomplices [28]. This feature cannot only provide us with the accomplice but also reveal additional detailed clues for a case, such as suspect, means, motives, alibi, where a suspect communicates with the accomplices to conspire the crime. This feature can be mapped to emails or IMs in the communication file group (Fig. 2). In the cases 2, 4 and 5 in the analysis reports, accomplices outside the company are identified on email investigation. Especially in case 5, the suspect has discussed her new job offer and sent her resume to the accomplice.

Business Dependency (What). Identifying stolen files is the first step of investigation on industrial espionages. However, stolen files may not be easily identified due to the diversity of valuable information in a computer. Most stolen files can be dependent upon the computer programs required to run businesses. Source code, design blueprint, and prototypes are valuable to Hi-Tech manufacturers, while sales forecasts, financial and customer information are valuable in merchandising business [24].

To identify stolen files, this feature can be mapped to user file groups (Fig. 2) but the types of stolen files are different depending on the types of business. In the analysis reports, the computer-aided design (CAD) files are generally targeted for the technical data theft, whereas database and spread sheets files are dominant for the marketing data crime, and document files (e.g., Microsoft Word or Adobe Acrobat files) are quite commonly used for most of the trade secrets espionage regardless of the types of business. The CAD files were stolen in case 3, the spreadsheet files in cases 1 and 2, the image files in cases 1 and 4, and the document files were stolen in cases 2, 4, and 5. However, the fact that the stolen files were discovered on the suspected computer might not be sufficient evidence to prove a criminal activity because the stolen files might be originally supposed to be on the computer. Thus a court might not accept the stolen files alone as evidence without proving that the suspect misused the files.

Metadata of Stolen Files (What). Each file has its own metadata that describes the characteristics of a file, which includes the file name, owner, created, accessed and modified timestamps, and directory path where the file is stored [29]. Without opening a file to look at its contents, the metadata of a file is helpful for proving that the file existed at a specific location and time. The status of a file's metadata is updated whenever the file is accessed, modified, deleted or moved, thus the trace of a stolen file's metadata can reveal the changes of the file. Although the file has been deleted, the change history of a stolen file allows an investigator to ascertain whether the file was stored in a certain location in a computer.

This feature can be answered by mapping to the metadata of the user file groups (Fig. 2). Also the metadata of a stolen file can be obtained from a browsing history (e.g., index.dat and webcachev01.dat) or a shortcut file (e.g., a file with the lnk

extension) in the system file group (Fig. 2) if the file has been recently opened in the system. The new metadata is recorded in a new shortcut file if the file is reopened after the location, timestamps or other characteristics of the file are changed. Thus, by learning the difference between the various versions of shortcut files, an investigator can prove the trail of the stolen file that has been moved, changed or copied to other devices in the computer. In the analysis reports for all 5 cases, we observed that the metadata of targeted files were examined to trace the stolen files.

Compression, Backup, Encryption of Stolen Files (What). A stolen file might be compressed, backed-up, or encrypted by organizations to protect the contents in accordance with their security policy, or by a suspect to hide the crime. Suspects often encrypt the stolen files to transmit them to outsiders covertly [26], back-up and compress the stolen files to copy them to USB, especially in the alleged time period. In this way, many protected files can be transmitted or copied without releasing their name and contents. This feature is mapped to the less-identified file group and mapped to the system file group because the shortcut, browsing history, registry and log files can show that a stolen file has been compressed, backed-up, or encrypted (Fig. 2). A shortcut file and a browsing history file can be created when a compressed file is opened, whereas a registry file can contain the mounted information about an encrypted or backed-up volume if the volume has been mounted to the computer. In case 4, the stolen files were found to be compressed during the alleged time period.

Internet Activities (How). The Internet is often used by a suspect to communicate with an outsider [30] and to send the stolen files to the outsider covertly. In particular, email is the most important object of investigation because nearly 45% of data theft is sent to an outsider using email [24] and the emails between offenders (such as a suspect and an accomplice) may reveal their intent for a crime, especially, the motivation of the crime. Although web-hard and cloud services can also be used to share the targeted files with outsiders [31], they are beyond the scope of this model that is confined to the investigation of a computer.

The Internet activities feature is mapped to Internet history, registry, log, and shortcut files in the system file group, emails and IMs in the communication file group, and webpage files with web cache in the user file group (Fig. 2). The Internet history files store the Internet access record with timestamps, user accounts, paths of visited URLs, and the number of visits to a specific website or a web-hard. The registry files store the URLs the user visited, the user's start page, and, in some cases, the keywords the user searched for [22]. The log files can contain a user's activities; for instance, cloud services (e.g., Dropbox, OneDrive, and iCloud) generate certain log files containing the path, name, size, and timestamp of a downloaded or uploaded file along with user accounts or email addresses [32]. The shortcut files (e.g., bookmark or favorite files) can record user's favorite and bookmarked web-pages, which can make it possible to assume that the user recognized and often visited these web-pages. The temporarily stored webpage files with web cache can be useful for rebuilding an original webpage the suspect browsed. The analysis reports in cases 2, 4 and 5 show that the stolen files have been sent to outsiders via email. Specifically in case 5, the criminal conspiracy is obviously incorporated in the email messages and the web-hard service is used for sharing the stolen files with an accomplice.

Internet Search of Crime (How). A suspect tends to search the Internet for information related to industrial espionage just before or after the crime. A suspect might take a look at a company to which he/she moves, or search for the information about the crime such as a range of potential penalties, or the latest news and updates on similar crimes. This feature gives circumstances of a crime to an investigator, thus, the investigator can infer that the suspect has committed the crime. This feature is mapped to the web-pages with web caches in the user file group, and the Internet history and registry files in the system file group (Fig. 2). In case 4, a suspect's internet search activities are examined in the Internet history and registry files that contain the searched keywords and typed URLs.

USB Devices (How). It is highly plausible for a suspect to copy a secret file to a USB device covertly during the alleged time period. An investigator can identify the model, volume serial number, or labeled name of the USB that was attached to the suspected computer. The identified USB can also reveal the names of stolen files that were copied on the USB. But, this feature has become less effective; only less than 10% of data theft has taken place using USBs [24], as many companies disable USB devices on employees' systems in order to protect their secret data. To identify the suspicious USB usage, this feature is mapped to registry, shortcut, browsing history, and log files in the system file group (Fig. 2). Registry files contain external device information from the sub-keys, e.g., Mounted device, USBSTOR, and USB. Shortcut and browsing history files can provide a clue as to when or where (which USB) the stolen files were copied. Log files can contain USB installation, along with the timestamps (e.g., setupapi.dev. log), and many monitoring software can create a log file containing a USB and the name of copied files on the USB [22]. In cases 1, 2, and 4, the analysis reports detected that suspect used USBs during the alleged time period.

Suspicious Activities (How). An investigator should discern a suspect's unusual, irregular or suspicious activities that deviate from the ordinary activities. When a suspect emails a particular person, he might use an outside email account rather than his company email account. During the suspected dates, a suspect may exchange too many emails with a particular person, or access a strange web-server for uploading files. A suspect may email the secret files to himself to elude a company's email checking policy. New external devices might be attached to the computer so that a bunch of files could be backed up and copied on them especially during the suspected dates. Also unknown programs might run on the computer prior to his resignation.

To detect suspicious activities, an investigator can search the Internet related files in the system file group, the emails and IMs in the communication file group, and the files in the user file group, and compressed or backed-up files in the less-identified file group (Fig. 2). An investigator needs to cross-check the different versions of files' metadata (especially timestamps and directory paths of the stolen files) so that he/she can find different types of suspicious activities (e.g., moving, deleting, or copying many files). The metadata of the files are stored not only inside the files, but also in the browsing history and shortcut files in the system file group. An investigator also needs to detect some suspicious login attempts or insecure USB connections to the system or protected storage using the registry and log files [29] in the system file group. In cases 1, 2, 4, and 5, the analysis reports show that the suspects accessed most of the targeted files and

attached several USB devices to the computers a few days before a suspect left the company. In case 4, the suspect had 7 email addresses and used one of those emails during the alleged time period.

Indirect Stealing Activities (How). As an alternative to directly stealing the targeted files (e.g., emailing the files to an outsider), a suspect often obtains the targeted information indirectly by means of printing, capturing, drawing, or taking photos of the targeted file without notice. As a company installs and operates security software that restricts the use of networks and external media, it is difficult to transfer data without permissions [24]. Thus indirect stealing activities feature has become more important in the investigation of industrial espionage. With a smart phone, a suspect can easily take photos of the secret information, and send emails or messages with the photos to outsiders; but a smart phone is beyond the scope of our model.

To find indirect stealing activities, this feature can be mapped to the system file group, and user file group (Fig. 2). Printing or capturing a target file can be detected in the system file group using both the log files of the print program, if any, the registry files relating to program activities, and files containing the printed images of a file and the information for each printing job (e.g., print spool files) [33]. The stolen files created or changed through indirect stealing activities (e.g., screenshot or printing) should be detected using the image and multimedia files in the user file group. In case 4, after analyzing print spool files, an investigator found that a suspect printed the targeted files using a local and a network printer during the suspected dates. The use of a capturing program was discovered in a registry file in case 1.

Security Program Reports (How). Most of the Hi-Tech companies run security programs: logging, security monitoring or data loss prevention (DLP) software for the security of files, valuable data, and the detection of leaked information. If the company has a security program in place, the log files of the program on the suspected dates need to be analyzed [26]. This feature is mapped to the log files in the system file group (Fig. 2). The log files of a security program (e.g., 'Activity Monitor') can include the typed keystrokes, records of switching between programs with timestamps, application path, and window names, visited web sites, passwords, email, chat conversation, USB, and printing usages. The security program reports feature identified a suspect in case 2 via an investigation of the log files in the system file group.

6 Evidence Sufficiency Verification and Documentation

6.1 Evidence Sufficiency Verification Process

After identifying all the supporting potential evidential files, DEIV-IE creates a more solid reconstruction of the crime using the evidence sufficiency verification process. This process consists of a series of questions and answers that should be proved to persuade a court for industrial espionage. The process begins with the identification of when a crime occurred, followed by who was a suspect, with whom a suspect committed the crime, what were the stolen files, and how the suspect committed the crime. The following describes each step of this process.

When did a Crime Occur? The first step of the process identifies the time period when the crime occurred. This step checks if the resignation date feature identifies the crime dates or not. If it is 'Yes', the result is documented. However, if it is 'No', the investigator inquires within the victimized company for a plausible range of time. However, if the company cannot provide information about the crime time, this process assumes that the crime time frame is within four months preceding resignation.

Who was a Suspect? The second step of the process identifies the suspect(s). This step checks if the multiple accounts feature identifies a suspect or not. If the suspect(s) is named, it is documented with more specified crime dates. But, if the suspect(s) is not identified, the investigator needs to inquire within the company for a range of possible suspect. If the company fails to identify the suspect(s), the range of suspect(s) narrows down to the employees who used the suspected computer or who had a right to access the stolen files.

With whom did a Suspect(s) Commit the Crime? The third step of the process identifies the accomplice(s) using the accomplice feature. If the accomplice(s) is identified, the result is documented with an updated conclusion of evidence. However, if the accomplice(s) is not identified, this step is pending until additional findings in the subsequent steps help to identify the accomplice. If no accomplice is associated with the crime or no additional findings help to identify the accomplice, this step fails.

What was the Stolen File(s)? This step identifies stolen files by checking if the business dependency, metadata, compression, backup, or encryption of stolen files feature identifies the stolen file(s). If it is 'Yes', the result is documented with updated previous findings. If stolen files are not identified, the investigator inquires within the company for a range of the possible stolen files. Otherwise, this model assumes that the range of the stolen files is the user file group.

How did the Suspect Commit the Crime? The final step of the process finds criminal behaviors. This step checks if the internet activities, internet search of the crime, USB devices, suspicious activities, indirect stealing activities, and security program reports features identify criminal activities. If it is 'Yes', the result is documented with updated previous findings. If the final findings contain the evidence to answer all the questions such as the crime dates, suspect(s), accomplice(s) if any, stolen file(s), and criminal activities, it concludes that this investigation has sufficient evidence to convict the suspect for the crime committed. But the final findings might be insufficient to convict under the legislation when they only establish partial facts from the existing evidence. Lack of evidence cannot provide proof of comprehensive evidence validation, and it can make the exiting evidence unreliable for court decision [2].

6.2 Documentations

Whenever a file or information is identified and updated as evidence, the result is documented. The documentation is not accumulative, but it is created individually, so that the updates or changes of preceding evidential information are newly documented along with a new identified evidence. This is a procedure to specify and narrow the findings. After completion of the entire process, the final report is created.

7 Validation

DEIV-IE was validated with three industrial espionage cases; two cases of the prosecutors' office in Korea in 2010 and 2014, and one of 'M57-Jean' (Corpora, 2008) that posted on a website for use in computer forensics education research. An ex-digital forensic investigator was involved in this validation, and the tool 'Encase' was allowed for the application of this model to the cases. Each case is described as follows:

Case 1: A suspect who is a son of the owner of a rival company joined the chemical manufacturer and was suspected of ex-filtrating secret files to the rival company on the day before his resignation.
Case 2: An accomplice was suspected of involvement in trade secret theft of nuclear power and the computer of the accomplice was investigated.
Case 3: (M57-Jean): The protected information of a company was posted on the Internet. A suspect's laptop in the workplace was examined for the crime investigation. The case is originally designed as a case study for the document ex-filtration from a corporation rather than industrial espionage. However, the case is utilized in our validation due to the similarity of its criminal activities such as stealing and revealing information of a company.

In this validation, we use two sets of assessment criteria – main assessment criteria (Table 3) and sub-criteria (Table 3). The main assessment criteria focus on whether the crime features described in chapter 5 can detect sufficient evidence to answer the '4W1H' five questions (Table 3) when they are mapped to the relevant file groups defined in chapter 4. This is because the answers to the five questions are typically considered the basic information that determines whether a person is guilty. One point has been given to the score of each case if one of five main assessment criteria is answered by investigating any files or data categorized as sub-criteria (Table 3). This assessment records the score of cases on a scale of 0 to 5, with 0 being the no sufficient information and with 5 being the sufficient information, in terms of the main assessment criteria. All five questions are answered for cases 2 and 3 (Table 3), each of which is given five points. In addition, this validation uses the score quantified by the sub-criteria (Table 3) which demonstrates how much main assessment criteria is supported by the sub-criteria. When a main assessment criteria is substantiated with more than one file, the fact proved by the main assessment criteria is more valid than the fact proved using just one file. We give one point to each case if a file of the sub-criteria is used to indicate the fact proving the main assessment criteria. The number of files used for each case are counted and represented as a score along with the score for main assessment criteria. The case 2 has the score 5(13), with 5 being the number answered by the main assessment criteria and with 13 being the number of files used to prove the main assessment criteria.

The DEIV-IE model has discovered sufficient evidence to prove the crime in cases. In our validation, the case 2 fulfills the five main assessment criteria with five points supported by 13 different types of evidential files, denoted by 5 (13) in Table 3. The case 2 has the eight points of sub-criteria and five points of the main assessment criteria. The court acquires all the five answers to the main assessment criteria so that it

Table 3. A score obtained by DEIV-IE in the verification cases.

Main Assessment Criteria (4W1H)	Score (# of findings)			Crime feature	Sub-Criteria (Files or data used to detect evidence)		
	Casel	Case2	Case3		Casel	Case2	Case3
When was crime dates	1(1)	1(1)	1(1)	Resignation date	Timestamps	Timestamps	Timestamps
Who was the suspect	1(1)	1(1)	1(1)	Multiple accounts	Registry	Registry	Registry
Who was the accomplice	0(0)	1(2)	1(1)	Accomplice		Email, IM	Email
What files were stolen	1(4)	1(2)	1(2)	Business dependency	Document file	Document file	Document file
				Metadata of stolen files	Metadata	Metadata	Metadata
				Compression, backup or encryption	Shortcut, compressed file		
How the crime was committed	1(5)	1(7)	1(3)	Internet activities		Email, Internet history, IM	Email, attachment
				Internet search of crime		IM	Email
				USB devices	Registry, shortcut, browsing history		
				Suspicious activities	Shortcut, deleted stolen file	Browsing history, registry, stolen file	
				Indirect stealing activities			
				Security program reports			
Total Score	4(11)	5(13)	5 (8)				

can determine if a criminal is guilty. The case 1 answers the four questions of main assessment criteria with its sub-scoring 11 points (4 (11) in Table 3). Our model cannot prove the accomplice of case 1, but it discovers sufficient evidence for court decision if we assume that case 1 doesn't have an accomplice. This validation approach is not a conclusive validation, but a score-based assessment to verify whether the crime features with their file groups can reveal sufficient evidence of each case.

8 Conclusion and Discussion

This paper has described a digital forensic investigation and verification model (DEIV-IE), which finds digital evidence using crime features and verifies the sufficiency of the evidence to establish factual information on industrial espionage for court decision. This study used a deductive, crime-based approach by examining digital forensic analysis reports of five actual industrial espionage cases, and studying the state of literature of this crime. In DEIV-IE, files have been classified to their file groups from the different legal perspectives and notable crime patterns have been streamlined to the crime features. Then DEIV-IE identifies specific crime features mapped to its relevant file groups, so that an investigator examines the files in the file groups for each feature.

However, this model is limited in scope only addressing the espionage cases of insider data thefts at corporate level where are only stand-alone computers running windows operating system, and external devices that are examined. As a result, this model does not address other common attack vectors in industrial espionage and has several limitations. First, it does not discuss user endpoint malware infection, and server vulnerability exploitation for exfiltration. Second, it does not address the decryption issues to find the targeted files even though a company's sensitive files tend to be encrypted due to the company's security policy. Third, it does not support comprehensive analysis between computers and other digital devices such as badge readers, CCTV footage, in particular, smart phones even if they have become major tools for industrial espionage cases recently. Forth, this model assumes that evidence is a file or exists in a file, so it may not be applicable to a situation where a criminal tampers with the evidence. Criminals can remove targeted files to interrupt the business of the victimized company deliberately. This case requires additional processes such as file carving from fragments and file reconstruction from the tempered evidence.

DEIV-IE, despite these weaknesses, still contributes to industrial espionage investigations. To date, there are no digital forensic investigation models for industrial espionage that have incorporated the crime features into digital forensic analysis techniques. In addition, the crime features in the model do not serve as standard practice but serve as an example of essential building blocks in the crime investigation, which will be inevitably expanded as different attack vectors are involved in the crime. The framework approach in this paper is easily adjustable based on new activities, the environments (e.g., Linux, smartphone) where it is applied, the expertise of an investigator and also open to adjustment to develop the DEIV model for other crimes.

This research will be extended to developing a tool supporting DEIV-IE, so that it extracts the information required for each crime feature from the proper files and verifies the findings. It will also involve defining the relationships among the crime features so that investigators use the connection between the features in the investigation.

References

1. Montasari, R.: Review and assessment of the existing digital forensic investigation process models. Int. J. Comput. Appl. **147**, 7 (2016)
2. Boddington, R., Hobbs, V., Mann, G.: Validating digital evidence for legal argument. In: Australian Digital Forensics Conference (2008)
3. Karie, N.M., Venter, H.S.: Towards a framework for enhancing potential digital evidence presentation. In: Information Security for South Africa. IEEE (2013)
4. Ieong, R.S.C.: FORZA–digital forensics investigation framework that incorporate legal issues. Digit. Investig. **3**, 29–36 (2006)
5. Søilen, K.S.: Economic and industrial espionage at the start of the 21st century–Status quaestionis. J. Intell. Stud. Bus. **6**, 3 (2016)
6. Marturana, F., et al.: A quantitative approach to triaging in mobile forensics. In: IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). IEEE (2011)
7. McClelland, D., Marturana, F.: A digital forensics triage methodology based on feature manipulation techniques. In: IEEE International Conference on Communications Workshops (ICC). IEEE (2014)
8. Cantrell, G., et al.: Research toward a partially-automated, and crime specific digital triage process model. Comput. Inf. Sci. **5**(2), 29 (2012)
9. James, J.I., Gladyshev, P.: A survey of digital forensic investigator decision processes and measurement of decisions based on enhanced preview. Digit. Invest. **10**(2), 148–157 (2013)
10. Karie, N., Venter, H.: A generic framework for enhancing the quality digital evidence reports. In: 13th European Conference on Cyber Warfare and Security ECCWS-2014 the University of Piraeus Piraeus, Greece (2014)
11. Karie, N.M., Venter, H.S.: Towards a framework for enhancing potential digital evidence presentation. In: Information Security for South Africa 2013. IEEE (2013)
12. Mohamed, I.A., Manaf, A.B.: An enhancement of traceability model based-on scenario for digital forensic investigation process. In: Third International Conference on Cyber Security, Cyber Warfare and Digital Forensic (CyberSec). IEEE (2014)
13. Karie, N., Kebande, V., Venter, H.: A generic framework for digital evidence traceability. In: European Conference on Cyber Warfare and Security. Academic Conferences International Limited (2016)
14. National Institute of Standards and Technology (NIST) (2002). The National Software Reference Library (NSRL). https://www.nist.gov/software-quality-group/nsrl-introduction. Accessed 24 Jan 2018
15. Holt, T.J., Bossler, A.M., Seigfried-Spellar, K.C.: Cybercrime and Digital Forensics: An Introduction. Routledge, Abingdon (2015)
16. Bruce, C., Santos, R.B.: Crime Pattern Definitions for Tactical Analysis (2011)
17. Raghavan, S., Raghavan, S.V.: A study of forensic & analysis tools. In: Eighth International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE). IEEE (2013)
18. Teppler, S.W.: Testable reliability: a modernized approach to ESI admissibility. Ave Maria L. Rev. **12**, 213 (2014)
19. Legal Information Institute (Hearsay 2017). https://www.law.cornell.edu/wex/hearsay. Accessed May 2017
20. United States v. Hamilton, 413 F.3d 1138, 1142 (10th Cir. 2005)
21. Records of Regularly Conducted Activity, Rule 803(6), Federal Rule of Evidence
22. Carvey, H.: Windows forensic analysis DVD toolkit. Syngress, Amsterdam (2009)

23. United States v. Washington, 498 F.3d 225, 233 (4th Cir. 2007)
24. Casey, E.: Error, uncertainty, and loss in digital evidence. Int. J. Digit. Evid. **1**(2), 1–45 (2002)
25. Sinha, S.: Understanding industrial espionage for greater technological and economic security. IEEE Potentials **31**(3), 37–41 (2012)
26. Wright, L.: People, risk, and security: How to Prevent Your Greatest Asset from Becoming your Greatest Liability. Springer, London (2017). https://doi.org/10.1057/978-1-349-95093-5
27. EC-Council: Computer Forensics: Investigating Network Intrusions and Cyber Crime. Nelson Education (2009)
28. Carrier, B., Spafford, E.H.: An event-based digital forensic investigation framework. In: Digital Forensic Research Workshop (2004)
29. Bhatti, H.J., Alymenko, A.: A Literature Review: Industrial Espionage (2017)
30. EC-Council: Computer Forensics: Hard disk and Operating Systems. Nelson Education (2009)
31. Hultquist, J.: Distinguishing cyber espionage activity to prioritize threats. In: 13th European Conference on Cyber Warfare and Security ECCWS-2014, The University of Piraeus Piraeus, Greece (2014)
32. Tun, T., et al.: Verifiable limited disclosure: reporting and handling digital evidence in police investigations. In: IEEE International Conference on Requirements Engineering Conference Workshops (REW). IEEE (2016)
33. Chung, H., et al.: Digital forensic investigation of cloud storage services. Digit. Investig. **9**(2), 81–95 (2012)
34. Sammons, J.: The Basics of Digital Forensics: The Primer for Getting Started in Digital Forensics. Elsevier, Waltham (2012)
35. Al Mutawa, N., et al.: Forensic investigation of cyberstalking cases using behavioural evidence analysis. Digit. Investig. **16**, S96–S103 (2016)
36. Al Mutawa, N., et al.: Behavioural evidence analysis applied to digital forensics: an empirical analysis of child pornography cases using P2P networks. In: 10th International Conference on Availability, Reliability and Security (ARES) 2015. IEEE (2015)