Preface

The last few decades have witnessed the unprecedented development and convergence of information and communication technology (ICT), computational hardware and multimedia techniques. At the personal level, these techniques have revolutionized the ways we exchange information, learn, work, interact with others and go about our daily life. At the organisational level, these techniques have enabled a wide spectrum of services through e-commerce, e-business and e-governance. This wave of ICT revolution has undoubtedly brought about enormous opportunities for the world economy and exciting possibilities for every sector of the modern societies. Willingly or reluctantly, directly or indirectly, we are all now immersed in some ways in the cyberspace, full of ‘e-opportunities’ and ‘e-possibilities’, and permeated with data and information. However, this type of close and strong interweaving also poses concerns and threats. When exploited with malign or intentions, the same technologies provide means for doing harms at colossal scale. These concerns create anxiety and uncertainty about the reality of the information and business we deal with, the security the information infrastructures we are relying on today and our privacy. Due to the rise of digital crime and the pressing need for methods of combating these forms of criminal activities, there is an increasing awareness of the importance of digital forensics and investigation. As a result, the last decade has also seen the emergence of the new interdisciplinary research field of digital forensics and investigation, which aims at pooling expertise in various areas to combat the abuses of the ICT facilities and computer technologies.

The primary objective of this book is to provide a media for advancing research and the development of theory and practice of digital crime prevention and forensics. This book embraces a broad range of digital crime and forensics disciplines that use electronic devices and software for crime prevention and investigation, and addresses legal issues. It encompasses a wide variety of aspects of the related subject areas and provides a scientifically and scholarly sound treatment of state-of-the-art techniques to students, researchers, academics, personnel of law enforcement and IT/multimedia practitioners, who are interested or involved in the study, research, use, design and development of techniques related to digital forensics and investigation. This book is divided into four main sections according to the thematic areas covered by the contributed chapters.

- Section 1. Digital Evidence
- Section 2. Combating Internet-Based Crime
- Section 3. Content Protection through the Use of Extrinsic Data
- Section 4. Applications of Pattern Recognition and Signal Processing Techniques to Digital Forensics
However, it should be noted that these four sections are closely related, e.g., Chapter 6 of Section 2 fits in Section 4 quite well because of its use of pattern recognition and image processing techniques. Such a division is only meant to provide a structural organisation of the book to smooth the flow of thoughts and to aid the readability, rather than proposing taxonomy of the study of digital forensics.

**Section 1: Digital Evidence**

Maintaining the chain-of-custody for evidence is of paramount importance in civil and criminal legal cases. To ensure the admissibility of evidence, technical measures, including hardware and software, applied in the digital forensic investigation procedures are required to assure not only that evidence is not tampered with or manipulated due to their application, but that malicious attacks aiming at hiding or manipulating evidence are effectively detected. To serve these purposes, like physical world forensic investigation, digital forensic investigation has to follow three main steps:

1) *Event preservation* which entails, for example, the need for a bit-by-bit duplication of the volatile memory or file systems;
2) *Evidence search* which aims at collecting forensic information such as making timelines of system and file activities, device fingerprint (e.g., sensor pattern noise of digital cameras), keywords, contraband media, telecommunication data, steganography;
3) *Event Reconstruction* which is about interpreting the collected information/evidence in order to establish what have happened and who has involved in what.

To address the need for maintaining the chain-of-custody for evidence, the first section of this book covers four chapters concerning with technical and legal issues surrounding the chain-of-custody for evidence. In Chapter 1 - *Providing Cryptographic Security and Evidentiary Chain-of-Custody with the Advanced Forensic Format, Library, and Tools*, Simson Garfinkel proposes a method for assuring the integrity of digital evidence collected from data storage media based on novel public key cryptography. Garfinkel points out two potential problems the current practice of manually recording hash codes and including them in investigative reports has, i.e., 1) it is difficult to perform automated processing and validation because the hash codes are recorded in a free-format report narrative, and 2) If the disk image becomes damaged, the hash code will only indicate that it no longer matches, but does not allow the damage to be detected or corrected. He also discusses the vulnerabilities of the commonly used hashing algorithm MD5 and presents his recent work for addressing these problems. In addition to its capability of assuring the integrity of digital evidence, his method can also allow for digital documentation of evidential transfer, reconstruction of damaged evidence, recovery of partial evidence when reconstructing damaged evidence is infeasible, encryption with both symmetrical and asymmetrical cryptosystems.

With the emergence of VoIP technology as one of the important form for telecommunications, Chapter 2 - *Voice Over IP: Privacy and Forensic Implications*, presented by Jill Slay and Matthew Simon provides the authors’ insight into the issues of VoIP and Security, such as denial of service, exploitations of server and protocol vulnerability, call surveillance and hijacking, identity theft, tampering of audio streams, and VoIP and crime, such as the possibilities available to criminals and terrorists to communicate through the decentralised systems that use strong encryption algorithms and require no verification of users’ details. The authors also conduct experiments on the use of several techniques in the process of recovering VoIP evidence in an after the event Win XP to Win XP communication scenario and report
their findings surrounding the issues of privacy, telecommunication interceptions and digital evidence preservation. They also advise as to whether the VoIP technology should be used in restricted environments from a technical point of view.

Most enterprises conducting e-business and e-commerce rely on mission-critical computer systems to guarantee continuity of their service even when the system is under attack or forensic investigation. In Chapter 3 - *Volatile Memory Collection and Analysis for Windows Mission-Critical Computer Systems*, Antonio Savoldi and Paolo Gubian propose a live forensic technology for collecting the full state of mission-critical systems under investigation without interrupting services. Savoldi and Gubian start with the review of the state-of-the-art techniques related to data collection from the volatile memory with three factors considered, namely fidelity, atomicity and integrity. Fidelity refers to the capability of collecting admissible bit-by-bit content of the volatile memory. Atomicity refers to the capability of collecting a snapshot of the volatile memory without altering it. Integrity is about preventing malicious tampering of memory snapshot, e.g., subversive actions conducted by kernel root kits. The authors then present a method for virtual memory space reconstruction and a possible way of collecting the page file on Windows-based mission-critical systems.

While the previous three chapters are contributed by computer scientists, Chapter 4 - *Evidentiary Implications of Potential Security Weaknesses in Forensic Software*, on the other hand, contributed by Chris Ridder from a law practitioner’s perspective, is concerned with the vulnerabilities that could be exploited by an attacker to hide, add or change data without being detected so as to prevent forensic software from collecting admissible evidence or mislead investigators and courts. He discuss a number of legal doctrines designed to ensure that the chain-of-custody of the evidence presented to the courts of law is preserved, but thus far the courts law have not enforced them due to the concerns over the possibility of security weaknesses in forensic software. He also studies how the courts of law may react to such claims, and recommends approaches that attorneys and courts can adopt to ensure that digital evidence presented in court is both fair to litigants and admissible.

**Section 2: Combating Internet-Based Crime**

Computer networks are one of the main avenues for various forms of malicious activities and cybercrime. Section 2 of this book focuses on some forms of these malicious activities, aiming at providing insight into their characteristics, how they work and how to detect or prevent them. For most, if not all, email users, unsolicited email spam constitutes the majority of the messages they received in their inbox, mainly for advertising purpose. These inundating adverts do represent a significant waste of resources, including the abuse of systems and the time the recipients spend on deleting them and avoiding over looking non-spam emails. However, the number of fraudulent emails intended for collecting recipient identities or credit card details (i.e., phishing schemes) is on the rise and the consequences of their success is far greater than the advertising email spam can incur. To address these issues surrounding email spam, in Chapter 5 - *Methods to Identify Spammers*, Tobias Eggendorfer discusses how spammers work, including email addresses acquisition through address traders, spam delivery (e.g., the use of botnets, which are responsible for the delivery of over 80% of email spam), and the operation of anonymous online shops and payment mechanism. Although spam filtering has provided some degree of relief, it has also given rise to a range of problems, such as high rates of false positives (i.e., mistaking authentic emails as spam) and false negatives (i.e., mistaking spam emails as authentic), increased system overheads and risk of security breaches. With these problems in mind, Eggendorfer proposes some alternatives
for tracking spammers, including email analysis, observing botnets, surveillance of purchases, partner
shops, payment process and server owners and methods for indentifying address traders through email
address identification and distributed tar pit networks.

In order to help with the identification of spam clusters and phishing groups, in Chapter 6 - *Spam
Image Clustering for Identifying Common Sources of Unsolicited Emails*, Chengcui Zhang et al. propose
a spam image clustering method based on data mining techniques to analyse the image attachments of
spam emails. This chapter is of particular interest because, unlike most spam filers which rely on textual
analysis, it presents a new pictorial content based approach which is rarely attempted. Given a large set
of emails with attached images, they first model attached images based on their visual features, such as
foreground text layout, foreground picture illustration and background texture analysis. An unsupervised
clustering algorithm is then applied to cluster the emails into groups according to the similarity of the
images’ visual features. Since there is no *a priori* knowledge about the actual sources of the spam emails,
visual validation is conducted to evaluate the clustering results. Different feature matching techniques
such as Scale Invariant Feature Transform (SIFT) similarity metrics, are applied to measure the feature
similarity of different visual features extracted from the images.

In many digital investigations, attributing events to specific people or sources is crucial and, in a
networked environment, attribution depends much on the identification of which computers were using
what IP-addresses at what points in time. If the IP-addresses are dynamically assigned, the originating
computers can only be identified if logs of the usage of the addresses can be acquired, and the points in
time of the relevant events can be accurately established. However, the use of timestamps (records of
specific moments in time) as evidence may not be admissible in a networked environment due to the
inconsistent settings of clocks at various nodes. In Chapter 7 - *A Model Based Approach to Timestamp
Evidence Interpretation*, Svein Willassen addresses this challenge by formulating historical clock settings
as a clock hypothesis and tests the hypothesis for consistency with timestamp evidence by modelling
actions affecting timestamps in the system under investigation. Acceptance of a clock hypothesis with
timestamp evidence can help establish when events occurred in civil time and can be used to correlate
timestamp evidence from various sources, including detecting correct originators.

The introduction of a wireless gateway to the automobile in-vehicle networks enables remote diag-
nostics and firmware updates, thus reducing costs considerably. Unfortunately, insufficiently protected
wireless gateways are not immune from cyber attacks and proper means for detecting and investigating
security-related events are yet to be developed. In Chapter 8 - *Conducting Forensic Investigations of
Cyber Attacks on Automobile In-Vehicle Networks*, Dennis Nilsson and Ulf Larson carry out an analysis
of the current features of in-vehicle networks and develop an attacker model to help devising countermea-
sures. Based on their attacker model and informed by a set of commonly practised forensic investigation
principles, they also derive a set of requirements for event detection, forensic data collection and event
reconstruction. They then use the Integrated Digital Investigation Process (IDIP) proposed by Carrier
and Spafford in 2004 as a template to illustrate the impact of the adoption of the derived requirements
on digital investigations.

The emergence of online games as an avenue for providing entertainment and a vehicle for generating
revenue has pave the way for trading and exchange of virtual properties to become another type of cyber
financial activity. Virtual properties can attain their physical value and be traded for physical currency.
However, disputes among the players are also foreseeable and measures for resolving them are necessary,
but not in place yet. Chapter 9 - *Dealing with Multiple Truths in Online Virtual Worlds*, Jan Sablatnig et
al analyse the system requirements needed to prevent or resolve the problem of simultaneous existence
of multiple truths and cheating. They start the chapter with a discussion on the real value of virtual property and the future of the online game in a virtual environment, including object complexity and scalability of the number of players. They then discuss the potential forms of fraud and cheating. They also studied ways for simplifying end user support in order to reduce the cost of running online games.

Section 3: Content Protection through the Use of Extrinsic Data

This section is concerned with multimedia content protection and authentication through cryptography and the embedding of extrinsic data. Cryptography is a mature scientific discipline and has a long application history in protecting digital content. Content protection through the use of extrinsic data, a relatively younger discipline, is about protecting the value of digital content or verifying the integrity and authenticity by embedding secret data in the host media and matching the hidden secret data against the original version at a later stage. Digital watermarking is a typical example of content protection and authentication based on extrinsic data and has been an active research area in the past 15 years. This set of digital watermarking techniques have found their applications in copyright protection, such as ownership identification, transaction tracking/traitor tracing and copy control, which is of great interest to the multimedia and movie industry, and content integrity verification and authentication, which is of high interest to the security sector, medical community, legal systems, etc. To ensure the security of these content protection schemes, the sophistication of countermeasures and attack models have to be in the mind of the developers of the protection schemes.

The greater the amount of extrinsic data (i.e., payload or capacity) is embedded, the better in the covert communication applications (i.e., steganography). High payload also strengthens security of digital watermarking schemes. However, hiding extrinsic data in host media inflict distortion, making imperceptibility an issue for the scheme designers to contend with. Results stemmed from recent research on information theory of steganography indicate that the detect ability of payload in a stego-object (a piece of media carrying hidden messages) is proportional to the square of the number of changes made during the embedding process. In Chapter 10 - Locally Square Distortion and Batch Steganographic Capacity, Andrew Ker investigates the implication when a payload is to be spread amongst multiple cover objects, and give asymptotic estimations about the maximum secure payload. Kerr studied two embedding scenarios, namely embedding in a fixed finite batch of covers and continuous embedding in an infinite stream and observed that the steganographic capacity, as a function of the number of objects, is sub-linear and strictly asymptotically lower in the second scenario.

Broadcast encryption is widely used for protecting the content of recordable and pre-recordable media. The encryption of the media key is stored in the header of media and used to encrypt the content stored after the header. The media key is repeatedly encrypted using all the chosen device keys to form a Media Key Block (MKB), which is then sent along with the content when the content is distributed. Proposed in Chapter 11 - Efficient Forensic Analysis for Anonymous Attack in Secure Content Distribution, due to Hongxia Jin, is a forensic technology aiming at preventing piracy and tracing traitors in the context of secure multimedia content distribution, with a specific focus on defending against anonymous attack where the attacker can re-broadcast the per-content encryption key or decrypted plain content. Jin also discusses the lack of some practical considerations in existing systems and points out four requirements that need to be implemented in the design of the future systems.

Picture archiving and communication systems (PACS) are typical information systems, which may be undermined by unauthorized users who have illegal access to the systems. In Chapter 12 - Protection
Li et al. propose a role-based access control framework comprising two main components—a content-based steganographic module, called *Repetitive Index Modulation* (RIM), and a reversible watermarking module, to protect mammograms on PACSs. Within this framework, the content-based steganographic module is to hide patients’ textual information into mammograms without having to change the important details of the pictorial contents and to verify the authenticity and integrity of the mammograms. The reversible watermarking module, capable of masking the contents of mammograms, is for preventing unauthorized users from viewing the contents of the mammograms. The scheme is compatible with mammogram transmission and storage on PACSs.

Digital watermarking schemes introduce extrinsic data into the host media, which distort the fidelity of the content of the host media. In some applications where even incidental distortion is not acceptable, reversibility (i.e., the capability of the scheme’s being able to restore the original version of the watermarked media after verification or authentication) is desirable. Moreover, many watermarking schemes also require the availability of original version of the host media at the verification stage, thus reducing the scheme applicability in various scenarios, e.g., multimedia database applications. Chapter 13 - *Reversible and Blind Database Watermarking Using Difference Expansion*, presented by Gaurav Gupta and Josef Pieprzyk, aims at addressing these two issues so as to provide a reversible and oblivious (also called blind) method for multimedia database watermarking. By incorporating the popular *Difference Expansion* technique in their scheme, the scheme provides reversibility to high quality media, effective identification of rightful owner, resistance against secondary watermarking attacks, and does not require a duplicated secure database of the original media for verification purpose.

In Chapter 14 - *Medical Images Authentication through Repetitive Index Modulation Based Watermarking*, Li and Li proposed a RIM-based digital watermarking scheme for authentication and integrity verification of medical images. By exploiting the fact that many types of medical images have significant background areas and medically meaningful Regions of Interest (ROI), which represent the actual contents of the images, the authors separate the ROIs from the background and scramble a little amount of information extracted from the contents of the ROI under the control a secret key to create a content-dependent watermark. The watermark is then embedded in the background areas, using the same Repetitive Index Modulation (RIM) scheme proposed in Chapter XII. On the verification side, the same operations are performed under the control of the same key as the one used in the embedding side to create the original watermark. This newly calculated watermark is compared against the watermark embedded in the background area. If both versions are the same, the received image is deemed authentic. Otherwise it is not trustworthy. By doing so, when any pixel of the ROI is attacked, the watermark embedded in the background areas will be different from the watermark calculated according to the attacked contents, thus raising alarm to alert that the image in question is inauthentic. Because the creation of the watermark is content-dependent and the watermark is only embedded in the background areas, the proposed scheme can actually protect the content/ROI without distorting it.

**Section 4: Applications of Pattern Recognition and Signal Processing Techniques to Digital Forensics**

Pattern recognition and digital signal processing techniques have been in use by expert witnesses in forensic investigations for decades. Section 4 of this book deals with methods that harness these two sets of techniques for biometric applications and multimedia forensics. Identity authentication is of
paramount importance in many aspects of our everyday life and business. While traditional authentication measures, such as PINs and passwords, may be forgotten, stolen or cracked, biometrics provides authentication mechanisms based on unique human physiological and behavioural characteristics that cannot be easily duplicated or forged.

When audio, image and video are presented in the court of law as evidence, their authenticity became an immediate concern and expert witnesses have to be called in for assistance. The following four chapters focus on the use of signal processing techniques in aiding digital forensic investigations. Matthew Sorell presented an interesting case in Chapter 15 - *Unexpected Artifacts in a Digital Photograph* surrounding the appearance of an unexpected logo of a non-sponsoring sports company on the jersey of a famous footballer in just one of a sequence of images of tournament. With deliberate video tampering as a cause eliminated, Sorell proposes a hypothetical sequence of circumstances, concerning optical pre-processing, infrared sensitivity, colour filtering and micro lens, exposure and lighting conditions, image acquisition, processing and enhancement and JPEG compression. The hypotheses are tested using a digital SLR camera. The investigation is of interest in a forensic context as a possible explanation as to why such a phenomenon has occurred.

Quantisation tables in a JPEG images have previously been shown to be an effective discriminator of the manufacturers and model series of digital cameras. There have been reports on using quantisation tables as device model signature for identifying source digital cameras. However, JPEG compressed images may be further compressed for making better use of transmission bandwidth or storage space. The secondary quantization during the further compression process will undoubtedly erase or damage the signature of the initial quantization, making source device identification based on JPEG compression difficult, if not impossible. Matthew Sorell points out in Chapter 16 - *Conditions for Effective Detection and Identification of Primary Quantisation of Re-Quantized JPEG Images*, that it is possible to identify initial quantisation artefacts in the image coefficients, provided that certain image and quantisation conditions are met. This chapter studies the conditions under which primary quantisation coefficients can be detected, and hence can be used image source identification. Forensic applications include matching a small range of potential source cameras to an image.

Although signature verification is an ancient authentication technique, which has been practised for centuries, because of the simplicity in acquiring signatures and the cost-effectiveness in verification, it is still in intensive use today, even in a networked environment. The main challenges of signature verification lie in detecting forgeries while accepting certain degree of variations in the authentic signatures of the same person and in the inadequacy of training samples. Both problems may lead to high rates of false positives and false negatives. To address these problems, Yan Chen, Xiaqing Ding and Patrick Wang propose a dynamic structural statistic model based online signature verification algorithm in Chapter 17 - *Dynamic Structural Statistical Model Based Online Signature Verification*. Dynamic time warping is adopted to match two signature sequences in order to extract a correspondent characteristic point pair from the matching result. A multi-variable statistical probability distribution is then employed to describe the variations of a characteristic point. Three methods, namely point dependent distribution model, point independent distribution model and point cluster distribution model, for estimating the statistical distribution parameters are evaluated. With this dynamic structural statistical methodology and based on the criterion of minimum potential risk, a discriminate function is derived to decide as to whether a signature in question is genuine or not.
Chapter 18 - Efficient Image Matching Using Local Invariant Features for Copy Detection, is concerned with the retrieval based approach, which has recently emerged as an attractive option for image copy detection. The Content Based Copy Detection (CBCD) can be treated as a restricted case of near duplicate image detection. Near duplicate images can be: (1) perceptually identical images (e.g. allowing for change in color balance, change in brightness, compression artifacts, contrast adjustment, rotation, cropping, filtering, scaling etc.), (ii) images of the same 3D scene (from different viewpoints). In this chapter, H.R. Chennamma et al., focus on image matching strategy based on local invariant features that will assist in the detection of forged (copy-paste forgery) images. They also introduce a novel strategy for pattern matching of key point distributions for copy detection. Typical experiments conducted on real case images demonstrate the success in near duplicate image retrieval for the application of digital image forensics. Efficiency of the proposed method is corroborated by comparison, with contemporary methods.

The manual review of videos captured by surveillance cameras is an inefficient and error-prone process, but of significant forensic value. In Chapter 19 - Reliable Motion Detection, Location and Audit in Surveillance Video, the authors, Samaan Poursoltan and Matthew Sorell discuss two key challenges often encountered in such a tedious video review task: 1) ensuring that all motion events are detected for analysis and 2) demonstrating that all motion events have been detected so that the evidence survives challenges in the court of law. The authors have demonstrated in one of their previous works that tracking the average brightness of video frames can provide a more robust measurement of motion than other commonly hypothesised motion metrics. This chapter extends that work by setting automatic localised motion detection thresholds, maintaining a frame-by-frame single parameter normalised motion metrics, and locating regions of motion events within the footage. A tracking filter approach is utilised to analyse localised motion, which adapts to localised background motion or noise within each image segment. After motion detection, location and size are estimated and used to describe motion events.

As identity theft is a big concern and the number of unique physiological and behavioral characteristics for each person is limited (e.g., each person has one face, two eyes, ten fingerprints, one signing style, etc.), re-issuing biometric traits for identifying the same person can be problematic. To resolve this non-reissuable problem and to thwart identity theft, researches on cancellable biometrics is gaining more momentum. Cancellable biometrics is the technique that utilizes non-invertible transformation functions to operate on the same original biometric sample in order to generate multiple variants/templates to represent the same person. The idea of this approach is to issue a transform template for authentication and verification purpose, and should the template is stolen or copied, the stolen or copied template is revoked with no fear of having the original sample obtained by the attacker through inverse transformation. The non-invertibility is to make it computationally infeasible to obtain the original sample from the stolen transformed template. Since the features of transformed variants of the same original template are similar in the feature space, a new transformed template can be issued after the stolen one is revoked/cancelled from the system. Contributed by Emanuele Maiorana, Patrizio Campisiand Alessandro Neri, Chapter 20 - Cancellable Biometrics for On-Line Signature Recognition is concerned with the renewability of a protected online signature based on this idea. They use a Dynamic Time warping (DTW) strategy to compare the transformed templates and carry out several experiments on the public MCYT signature database to evaluate their methodology.

The afore-mentioned six chapters provide technical treatment of various forensics aspects that require the use of pattern recognition and signal processing techniques. However, techniques alone should not
be expected to provide resolutions to all legal cases. The reader is referred to the J. Tibbitts and Y. Lu’s work entitled *Forensic Applications of Signal Processing*, IEEE Signal Processing Magazine, March 2009, for detailed discussions on issues such as scientific interpretation of reasonable doubt, lack of understanding of scientific principles, lack understanding of signal processing, and in particular, lack of understanding of speech and image processing.

*Chang-Tsun Li*
*University of Warwick, UK*

*Anthony T. S. Ho*
*University of Surrey, UK*