Preface

Cloud computing has emerged not only as a new concept, but also as a new paradigm for applications of information and communications technologies. The last few years has witnessed Information and Communication Technology (ICT) providers investing heavily in developing technologies and enormous server facilities which allow global end users to access Web-based applications and store their data off-site. Cloud systems essentially comprise a network of shared pool of resources (e.g. networks, servers, storage, applications, and services) ready to be used on-demand from clients. When someone uses a cloud facility, data is distributed and processed in a network of resources around the world.

Optimization of resource utilization is achieved through sharing in an open environment. This new computing paradigm has triggered a new wave of ICT revolution, which will undoubtedly bring about enormous opportunities for the world economy and exciting possibilities in every walk of the modern society. Willingly or reluctantly, we are all now immersed in such a cyberspace, full of opportunities and possibilities, and permeated with rich multimedia and information. However, this type of close and strong interweaving also causes concerns and poses threats. When exploited with malign intentions, the same technologies provide means for doing harm at a colossal scale. The £1.3 billion loss of the Swiss banking group, UBS, due to unauthorized trading by a rogue trader in its investment bank in 2011 is just one of many typical examples. If this incident is not significant enough to cause concern, then the statistics provided in the 2012 Norton Cybercrime Report would probably raise the eyebrows of the most optimistic. According to the report, the scale of consumer cybercrime is one-and-a-half million victims daily (i.e., 18 victims per second) and the global price tag of consumer cybercrime is US$110 billion annually. These statistics and the scales of the damages due to security incidents create anxiety and uncertainty about our privacy, the reality of the information and business we deal with, and the security of the information infrastructures we are relying on today. The recent growth of insurance against losses due to information security breach is the best manifesto of the need for protecting IT systems and data, and ensuring business continuation.

The widespread of social networks and smart devices with image acquisition capability provide unprecedented ease for multimedia sharing. However, this also gives rise to increased amount of pirated content and sharing of illegal media. Moreover, powerful multimedia processing tools also allow multimedia content to be manipulated at ease. In the forensic and legal contexts when multimedia is involved, investigators need to analyze the integrity of the content and to ensure that the content they
have collected or analyzed is not contaminated such that its admissibility at the court of law is rejected. In the meantime, technologies need to be available to assert the integrity of authentic content when its admissibility as evidence is challenged at the court of law.

Due to the rise of digital crime and the acute need for methods of fighting these forms of criminal activities, there is an increasing awareness of the importance of digital forensics and investigations. As a result, the last decade has also seen the emergence of the new interdisciplinary research field of digital forensics and investigations, 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 evidential 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 research, use, design, and development of techniques related to digital forensics and investigation. This book is divided into two main sections according to the thematic areas covered by the contributed chapters.

- **Section 1**: Multimedia Forensics and Security
- **Section 2**: Digital Evidence

**Section 1**: Multimedia Forensics and Security consists of the first 8 chapters, which concerns with the forensics and security issues surrounding the use of multimedia, such as audios, images, videos, and websites. Multimedia has long been used as tools for aiding forensic investigations or as evidence in legal proceedings. This usually entails the use of signal processing and pattern recognition techniques either in the collection of evidence or in the analysis regarding the admissibility of evidence. For example, images in question may be manipulated (e.g., object addition and deletion) to mislead forensic investigation. The integrity of multimedia used as evidence at the court of law may be challenged. To prevent from being misled by manipulated content or to assert the integrity of the multimedia at the court of law both require the application of multimedia signal processing and pattern recognition techniques. Their role is becoming increasingly important in law enforcement given the pace of the advances of electronic devices and ICT. Biometrics, the art of using human physiological or behavioral traits for people identification and identity verification, also requires extensive use of pattern recognition and digital signal processing techniques to analyze the sensed data. Section 1 will touch on human fingerprint liveness detection, content integrity, copyright protection, steganography for covert communication, and Spam detection.

**Section 2**: Digital Evidence comprises the last 8 chapters of the book and cover a wide ranges of issues surrounding digital evidence, investigation tools and models, and standardization. Maintaining the chain-of-custody for evidence is of paramount importance in civil and criminal legal cases. To ensure the admissibility of evidence in the court of law, technical measures applied in the digital forensic investiga-
tion 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 usually 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.

Recent deployment of cloud computing techniques poses new challenges as many existing digital forensic guidelines are not adequate to deal with the more sophisticatedly distributed and shared IT facilities. Therefore, two of the thematic topics of this book are the discussions of the emerging challenges of cloud computing in digital forensics and international standardization.

**SECTION 1: MULTIMEDIA FORENSICS AND SECURITY**

In Chapter 1, “Fingerprint Liveness Detection Based on Fake Finger Characteristics,” Marcialis, Coli, and Roli point out that the robustness of a fingerprint verification system under fraudulent attempts through fake fingers is an acute issue to be addressed. It has been shown that many fingerprint sensors can be deceived by the attacker submitting a “gummy” finger. The majority of the methods for prevent this type of fraudulent attempts is based on the use of hardware embedded in the sensor, which can detect the vitality, or liveness, of the finger, e.g., through a blood pressure measurement. Another class of approaches is based on the extraction of features that can discriminate between live and fake fingerprints by processing images acquired by the sensor. This class of software-based solutions is obviously less intrusive and cheaper than the hardware-based ones. Therefore, this is the approach the authors of this chapter take. The vitality detection of fingerprints is currently acknowledged as a serious issue for personal identity verification systems. This problem, raised some years ago, is related to the fact that the 3D shape pattern of a fingerprint can be reproduced using artificial materials. An image quite similar to that of true, alive, fingerprint is derived if such “fake fingers” are submitted to an electronic scanner (optical and capacitive). Since introducing hardware dedicated to liveness detection in scanners is expensive, software-based solutions, based on image processing algorithms, have been proposed as alternative. So far, proposed approaches are based on features exploiting characteristics of a live finger (e.g., features characterizing finger perspiration). Such features can be named live-based or vitality-based features. In this paper, first the authors propose and motivate the use of a novel kind of features exploiting characteristics noticed in the reproduction of fake fingers, that they name fake-based features. They propose a possible implementation of these kinds of features based on the power spectrum of the fingerprint image. Their proposal is compared and integrated with several live-based features at the state-of-the-art, and shows
very good liveness detection performances. Experiments are carried out on a dataset much larger than commonly adopted ones, containing images from three different optical sensors.

Aiming at the identification of Natural Images (NI) and Computer-Generated Graphics (CG), Peng, Liu, and Long propose a novel method based on hybrid features extracted from the content in Chapter 2, “Identification of Natural Images and Computer- Generated Graphics Based on Hybrid Features.” A common practice in image forgery is to cut one area from one image and paste the area in another to create fake image. Such a cut-and-paste operation may involve natural images and computer graphics. Since the image acquisition pipelines are different, there exist some differences in statistical, visual, and noise characteristics between natural images and computer-generated graphics. Exploiting these discriminative characteristics, the authors first extract the mean, variance, kurtosis, skewness, and median of the histograms of images to serve as statistical features. Secondly, the fractal dimensions of the images are extracted to serve as visual features. Photo-Response Non-Uniformity (PRNU) left by imaging device sensors in the content, which is lacking in computer-generated graphics is also used as features. Based on these three sets of features, a Support Vector Machine (SVM) classifier is used to differentiate natural images from computer-generated graphics. The authors have observed that while their approach is capable of classifying computer-generated graphics and natural images, it does perform well in differentiating natural images. To improve the applicability of their method, the authors have pointed out the possibility of 1) including more precise methods for extracting the PRNU and more effective formulation of features extracted from PRNU (such as porosity and multi-fractal spectrum of PRNU) and 2) including Color Filter Array (CFA) features in images.

Chapter 3, “An Image Region Description Method Based on Step Sector Statistics and Its Application in Image Copy-Rotate/Flip-Move Forgery Detection,” deals with the same image forgery problems as described in Chapter 1 in a different approach. The authors, Chen, Lu, and Ni, start with a comparative description of active and passive image forgery detection. Because active image forgery detection (such as digital watermarking) requires the embedding of extrinsic data in the content, which limits its applicability, the authors review a number of methodologies that make use of intrinsic information in the content (i.e., passive image forgery detection) for detecting copy-move forgery. During the copy-move process, the duplicated regions may undergo geometrical modifications such as rotation, scaling, and/or illumination adjustment in order to blend the region smoothly into the new background. The tampered images may also be blurred, noised, or compressed so as to conceal the trace of manipulations. Thus, a good forgery detection algorithm should take these operations into account. In the light of the limitations of the reviewed methods, the authors propose a new approach about step sector statistics as features for describing image regions. The means and the standard deviations along the radial direction of the circular image region are extracted through the sector masks, and the rearrangement of these statistics makes this image region description method rotation-robust. The proposed method has been applied to the detection of copy-rotate-move forgery, and it can even detect the exact rotation angle between the duplicate regions. With minor extension, the proposed description method can also be applied to the detection of copy-flip-move forgery.

In Chapter 4, Sun, Lu, Yu, and Shen propose a “Daubechies Wavelets Based Robust Audio Fingerprinting for Content-Based Audio Retrieval.” As the authors described, audio fingerprinting is the process of obtaining a compact content-based signature that summarizes the essence of the audio clip in question. Such audio fingerprints can be utilized for asserting copyright ownership. The authors first review the
related methods in a taxonomical manner by dividing them into three categories, namely time-domain based, transform-domain based, and compressed-domain based approaches. Informed by the pros and cons of the reviewed approaches, the authors present a novel framework for content-based audio retrieval based on the audio fingerprinting scheme that is robust against large linear speed changes. In the proposed scheme comprising a fingerprint extraction algorithm and a fingerprint retrieval algorithm, 8 levels Daubechies wavelet decomposition is adopted for extracting time-frequency features. Experimental results show that the proposed scheme not only has good robustness to content-preserving operations and Additive White Gaussian Noise (AWGN) but also shows high robustness against big linear speed changes. Furthermore, it can reduce the storage space and computational costs.

Due to the popularity of camcorders and smart phones, digital videos are widely used in our everyday life. The ever-increasing video sharing sites on the Internet makes the distribution of digital video easy and fast. Apart from the many benefits this new set of technologies has to offer, the forged videos can also be easily created with a variety of video editing software, which certainly have forensic and legal implications. Chapter 5, “An Improved Fingerprinting Algorithm for Detection of Video Frame Duplication Forgery,” proposed by Hu, Li, Wang, and Liu, is intended to deal with video frame duplication. Frame duplication is a common way of digital video forgeries. State-of-the-art approaches to duplication detection usually suffer from heavy computational load. This chapter deals with a new algorithm for detecting duplicated video frames based on video sub-sequence fingerprints. The authors argued that the detection time can be shortened when the temporal information and the spatial information are integrated into one feature. A possible way of integrating the temporal information and the spatial information is to consider a video as a three-dimensional matrix and to perform Three-Dimensional Discrete Cosine Transform (3D-DCT) on it. However, the complexity of computing 3D-DCT is still known to be very high. The authors propose to extract the fingerprints from the DCT coefficients of the Temporally Informative Representative Images (TIRIs) of the sub-sequences and introduce a simple metric for the matching of video sub-sequences. Experimental results show that the proposed algorithm overall outperforms three related duplication forgery detection algorithms in terms of computational efficiency, detection accuracy, and robustness against common video operations like compression and brightness change.

Steganography is the technique for covert communications. Minimizing the embedding impact on the stego-object is an important aspect in designing steganographic systems as it reduces the probability of drawing steganalysts’ attention, thus, increasing the systems’ security. Wang, Ni, Wang, and Zhang argue in Chapter 6, “An Adaptive JPEG Steganographic Scheme Based on the Block Entropy of DCT Coefficients,” that by virtue of this strategy, the development of steganographic systems can be formulated as the construction of distortion profile reflecting the embedding impact and the design of syndrome coding based on a certain code. They propose a new JPEG steganographic scheme based on a new distortion profile exploiting both the block complexity and the distortion effect due to flipping and rounding errors. The distortion profile is then incorporated in the framework of Syndrome Trellis Coding (STC), which forms the core of their steganographic system. The STC provides multiple candidate solutions to embed messages to a block of coefficients while the constructed content-adaptive distortion profile guides the determination of the best solution with minimal distortion effect. In so doing, the total embedding distortion and impact is significantly reduced, thus attributing to less detectability of the hidden message. They demonstrate through a series of experiments that the proposed JPEG steganographic scheme greatly increases the embedding capacity without sacrificing imperceptibility and shows significant superiority over some existing JPEG steganographic schemes.
Yu, Shi, Zhao, Ni, and Cao conduct “A Study on Embedding Efficiency of Matrix Encoding” when applied to steganography in Chapter 7. Steganography is another important area in information forensics and security. The most important requirement of steganography is undetectability (security), i.e., malicious unintended recipients should not be able to distinguish between cover and stego objects with success better than a random guess. Steganographic techniques can be divided into two main categories depending on the ways they enforce the security of the schemes. Techniques of the first category embed information into host signals in such a way that no image features are significantly perturbed during the embedding, while techniques of the other category embed data in such a way that it distorts the steganalysis’s estimate of the cover image statistics. Nowadays, the most popular way is to seek the lowest possible rate of modification to the cover signal or the highest possible embedding capacity at a given distortion level, which falls in the first category. In this chapter, the authors focus on embedding efficiency, which influences the most concerned performance of steganography, security, directly. Embedding efficiency is a quantified measurement of the embedding capacity at a given distortion level, which is defined as the number of random message bits embedded per one embedding change. The authors make two main contributions in this chapter. First, embedding changes are not only evaluated on the number of changed coefficients but also on the varying magnitude of changed coefficients. Second, embedding efficiency of matrix embedding with different radixes is formalized and investigated. They concluded that ternary matrix embedding can achieve the highest embedding efficiency.

In Chapter 8, “Spam 2.0: State of the Art,” Hayati and Potdar defined Spam 2.0 as the propagation of unsolicited, anonymous, mass content to infiltrate legitimate Web 2.0 applications. A fake eye-catching profile in social networking websites, a promotional review, a response to a thread in online forums with unsolicited content or a manipulated Wiki page, are examples of Spam 2.0. This chapter provides a comprehensive survey of the state-of-the-art, detection-based methods (i.e. content-based, linked-based, and rule-based), prevention-based, and early-detection-based Spam 2.0 filtering methods. A detection-based Spam 2.0 filtering method relies on filtering spam by analyzing and evaluating content posted on Web 2.0 platforms. Methods in this category search for spam keywords, templates, attachments, etc. within the content. Prevention-based Spam 2.0 filtering strives to limit automated access to Web systems by blocking or slowing down spam distribution. By introducing computation, time, and economic challenges, prevention methods make Spam 2.0 distribution difficult and allow the genuine user to enter the system while keeping the spammer out of the system. The early-detection-based Spam 2.0 filtering methods try to discover spam behavior inside the system prior to or just after spam content is distributed. These methods typically examine the nature and behavior of spam generation and spam origin, such as spambots. By studying the behavior of content distribution and investigating behavioral features, early-detection methods are able to distinguish spam from legitimate content. Behavioral features vary depending on the different platforms. The state-of-the-art surveyed in this chapter will provide a guide to future research in this field.

SECTION 2: DIGITAL EVIDENCE

Chapter 9, contributed by Lempereur, Merabti, and Shi, covers “Pypette: A Platform for the Evaluation of Live Digital Forensics.” Live digital forensics presents unique challenges with respect to maintaining forensic soundness, but also offers the ability to examine information that is unavailable to quiescent
analysis. Any perturbation of a live operating system by a forensic examiner will have far-reaching effects on the state of the system being analysed. Numerous approaches to live digital forensic evidence acquisition have been proposed in the literature, but relatively little attention has been paid to the problem of identifying how the effects of these approaches, and their improvements over other techniques, can be evaluated and quantified. In this chapter, the authors present Pypette, a novel platform enabling the automated, repeatable analysis of live digital forensic acquisition techniques. While existing approaches have focused on evaluation based on a percentage of memory change before and after acquiring live forensic evidence, the authors of this chapter consider the accuracy and effects of methods in terms of the artefacts forensic examiners actually need to extract from systems, and the mechanisms they use for achieving this. Initial results have shown that their platform is capable of conducting repeatable experiments and generating consistent results. Their method provides a novel approach to:

- Evaluating the effects of live digital forensic techniques.
- Gathering meaningful results that can be correlated across platforms and approaches.
- An automated, consistent, technique that reduces the chance of errors in the analysis and evaluation process.

Ishihara’s studies in Chapter 10, “Probabilistic Evaluation of SMS Messages as Forensic Evidence: Likelihood Ratio-Based Approach with Lexical Features,” one of the first likelihood ratio-based forensic text comparison studies in forensic authorship analysis. The likelihood-ratio-based evaluation of scientific evidence has started being adopted in many disciplines of forensic evidence comparison sciences, such as DNA, fingerprints, handwriting, etc., and it is largely accepted that this is the way to ensure the maximum accountability and transparency of the process. Due to its convenience and low cost, Short Message Service (SMS) has been a very popular medium of communication for quite some time. Unfortunately, however, SMS messages are sometimes used for reprehensible purposes, e.g. communication between drug dealers and buyers, or in illicit acts such as extortion, fraud, scams, hoaxes, and false reports of terrorist threats. In his chapter, Ishihara performs a likelihood-ratio-based forensic text comparison of SMS messages focusing on lexical features. The Likelihood Ratios (LRs) are calculated using multivariate kernel density procedure, and are calibrated. The validity of the system is assessed based on the magnitude of the LRs using the log-likelihood-ratio cost. Ishihara demonstrates that the system with lexical features performed better than the one with N-grams. However, we pointed out that many of the derived LRs (calibrated) are weak in their strength as evidence, providing only limited support for either hypothesis.

In Chapter 11, “A Model for Hybrid Evidence Investigation,” Vlachopoulos, Magkos, and Chris-sikopoulos discuss that with the advent of Information and Communication Technologies, the means of committing a crime and the crime itself are constantly evolving. In addition, the boundaries between traditional crime and cybercrime are vague: a crime may not have a defined traditional or digital form since digital and physical evidence may coexist in a crime scene. They review a selection of investigation models for physical/digital evidence. Because various items found in a crime scene may worth be examined as both physical and digital evidence, they introduced the term hybrid evidence and propose a model for investigating such crime scenes with hybrid evidence. This model unifies the procedures related to digital and physical evidence collection and examination, taking into consideration the unique characteristics of each form of evidence. This model can also be implemented in cases where only digital or physical evidence exist in a crime scene.
An event is an occurrence within a computer system that converses with other systems or users. Computer and network systems contain event logs that hold enormous amounts of data. These event logs hold records of any behaviors or actions a network device performs. Events may also involve illegal activities such as malicious attacks or unexpected data movement. Argano, Gidwani, Issa, Betham, and Yan present in Chapter 12, “A Comprehensive Survey of Event Analytics,” an important area of study that avoids further incidents or risks after the events have occurred. Event analytics is a complex scheme; therefore, the authors have divided this into four major components: computer event surveillance, computer event monitoring, computer event forensic, and computer event security. Argano et al. combine computer event surveillance and computer event monitoring together and computer event forensic and computer event security together. An extensive study is conducted throughout these subcategories.

Chapter 13, “Hypothesis Generation and Testing in Event Profiling for Digital Forensic Investigations,” by Batten, Pan, and Khan discusses the need for an automated approach to forensic digital investigation and reviews a number of related attempts which have been made in the past decade. The aim of this chapter is to assist the forensic investigator with the generation and testing of hypotheses in the analysis phase. In doing so, they present a new architecture that facilitates the move to automation of the investigative process; this new architecture draws together several important components of the literature on question and answer methodologies including the concept of “pivot” word and sentence ranking. Their architecture is supported by a detailed case study that demonstrates the use of hypothesis generation and hypothesis ranking during a digital forensic investigation where many assumptions and decisions are made by investigators. In particular, they show an example of applying the proposed architecture to speed up the investigative process while retaining the benefits of using hypothesis generation and relationship building techniques. Their new approach moves beyond relationship building, already used by several authors, to focus on hypotheses generation and analysis. With the use of a ranking scheme, the investigator can quickly identify those important relations and hypotheses about which he/she is confident and is able to eliminate less convincing statements.

Haggerty et al. discuss in Chapter 14, “A Framework for the Forensic Analysis of User Interaction with Social Media,” that the increasing use of social media, applications, or platforms that allow users to interact online ensures that this environment will provide a useful source of evidence for the forensics examiner. They point out that current tools for the examination of digital evidence find this data problematic, as they are not designed for the collection and analysis of online data. This chapter presents a framework for the forensic analysis of user interaction with social media. In particular, it presents an inter-disciplinary approach for the quantitative analysis of user engagement to identify relational and temporal dimensions of evidence relevant to an investigation. This framework enables the analysis of large data sets from which a (much smaller) group of individuals of interest can be identified. In this way, it may be used to support the identification of individuals who might be “instigators” of a criminal event orchestrated via social media, or a means of potentially identifying those who might be involved in the “peaks” of activity. In order to demonstrate the applicability of the framework, this paper applies it to a case study of actors posting to a social media website.

As Grispos, Storer, and Glisson discuss in Chapter 15, “Calm Before the Storm: The Challenges of Cloud Computing in Digital Forensics,” cloud computing is a rapidly evolving Information Technology (IT) phenomenon. Rather than procure, deploy, and manage a physical IT infrastructure to host their software applications, organizations are increasingly deploying their infrastructure into remote, virtualized environments, often hosted and managed by third parties. This development has significant implications for digital forensic investigators, equipment vendors, law enforcement, as well as corporate
compliance and audit departments (among others). Many guidelines for digital forensic investigations were developed prior to the advent of cloud technologies and largely assume that the investigator has physical access and control over the target system or device, and in particular, its storage media. This assumption is likely to be invalidated when investigating activity in a cloud environment. This chapter summarizes the key aspects of cloud computing and analyses the challenges raised by cloud computing with respect to existing models of digital forensic investigations, such as the DFRW Investigative Process (DIP) model and the ACPO principles and guidelines. DIP model provides a comprehensive review of the stages employed in the digital forensic process, and so is convenient for analyzing the impact of cloud forensics on this process. Other models are also referenced where appropriate, in particular the Association of Chief Police Officers’ principles and guidelines. Several new research challenges addressing this changing context are also identified and discussed.

Chapter 16 is concerned with “The Need for Digital Evidence Standardization.” Grobler discussed in this chapter that continuous developments in forensic processes and tools have aided in elevating the positioning of digital forensics within the legal system. The equally continuous developments in technology and electronic advances, however, are making it more difficult to match forensic processes and tools with the advanced technology. It is therefore necessary to create and maintain internationally accepted standards to control the use and application of digital forensic processes. This chapter addresses this need and touches on the motivation for such internationally recognized standards on digital evidence. It also looks at current work in and progress towards the establishment of digital evidence related documents addressing all phases of the digital forensic process, namely pre-incident, during-incident, and post-incident. However, devising internationally acceptable standardization is by no means trivial. This chapter discusses the hurdles to international digital evidence standards from the aspects of environmental differences, technological differences, judicial differences, and training and certification differences. Environmental differences relate to the increasing number of users of computers, in combination with other environmental variables. Grobler notes that, although very relevant to the digital forensics discipline, technical aspects (e.g., storage devices, smart phones, and operating systems) are generally not discussed in the International Standards, in order to future-proof the content of these documents. According to Grobler, jurisdiction refers to the authority of a specific court to make decisions regarding a specific person or a certain matter in order to administer justice. Different legislations apply in different jurisdictions, and these are not always compatible. The differences in levels of technical expertise and a global skills shortage is another major concern discussed in this chapter.

Chang-Tsun Li
University of Warwick, UK