A Framework for Digital Forensics and Investigations

Article *in* International Journal of Digital Crime and Forensics · August 2015

DOI: 10.4018/jdcf.2013040101

CITATIONS
0

READS
2,417

3 authors:

Benjamin Aziz
University of Portsmouth
142 PUBLICATIONS **713** CITATIONS
SEE PROFILE

Clive Blackwell
Royal Holloway, University of London
24 PUBLICATIONS **72** CITATIONS
SEE PROFILE

Shareeful Islam
University of East London
50 PUBLICATIONS **992** CITATIONS
SEE PROFILE

IGI PUBLISHING
WWW.IGI-GLOBAL.COM

CALL FOR ARTICLES

International Journal of Digital Crime and Forensics

An official publication of the Information Resources Management Association

MISSION:

The mission of the **International Journal of Digital Crime and Forensics (IJDCF)** is to provide and foster a forum for advancing research and development of the theory and practice of digital crime prevention and forensics. IJDCF addresses a broad range of digital crimes and forensic disciplines that use electronic devices and software for crime prevention and investigation. This journal informs a broad cross-sectional and multi-disciplinary readership ranging from the academic and professional research communities, to industry consultants and practitioners. IJDCF publishes a balanced mix of high quality theoretical and empirical research articles, case studies, book reviews, tutorials, and editorials.

COVERAGE/MAJOR TOPICS:

- Computational approaches to digital crime preventions
- Computer virology
- Crime scene imaging
- Criminal investigative criteria and standard of procedure on computer crime
- Cryptological techniques and tools for crime investigation
- Data carving and recovery
- Digital document examination
- Digital evidence
- Digital signal processing techniques for crime investigations
- Identity theft and biometrics
- Information warfare
- Machine learning, data mining, and information retrieval for crime prevention and forensics
- Malicious codes
- Network access control and intrusion detection
- Policy, standards, protocols, accreditation, certification, and ethical issues related to digital crime and forensics
- Practical case studies and reports, legislative developments, and limitations, law enforcement
- Small digital device forensics (cell phones, smartphone, PDAs, audio/video devices, cameras, flash drives, gaming devices, GPS devices, etc.)
- Steganography and steganalysis
- Terrorism knowledge portals and databases
- Terrorism related analytical methodologies and software tools
- Terrorist incident chronology databases
- Watermarking for digital forensics

ISSN 1941-6210
eISSN 1941-6229
Published quarterly

All submissions should be mailed to:
Chang-Tsun Li, Editor-in-Chief
ijdcf@dcs.warwick.ac.uk

Ideas for Special Theme Issues may be submitted to the Editor-in-Chief.

Please recommend this publication to your librarian. For a convenient easy-to-use library recommendation for. please visit:
http://www.igi-global.com/IJDCF

International Journal of Digital Crime and Forensics

April-June 2013, Vol. 5, No. 2

Table of Contents

Research Articles

Book Review

The *International Journal of Digital Crime and Forensics* is currently listed or indexed in: Applied Social Sciences Index & Abstracts (ASSIA); Bacon's Media Directory; Cabell's Directories; Compendex (Elsevier Engineering Index); DBLP; GetCited; Google Scholar; INSPEC; JournalTOCs; Library & Information Science Abstracts (LISA); MediaFinder; Norwegian Social Science Data Services (NSD); SCOPUS; The Index of Information Systems Journals; The Standard Periodical Directory; Ulrich's Periodicals Directory

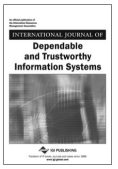
Related Journals

International Journal of Cyber Warfare and Terrorism

Matthew Warren (Deakin University, Australia)

ISSN: 1947-3435; Established 2011; Published Quarterly

Publishes original innovative findings on ethical, political, legal, and social issues relating to security and cybernetic wars. This journal focuses on cyber warfare and terrorism using examples from around the world.



International Journal of Dependable and Trustworthy Information Systems

A.F. Salam (The University of North Carolina at Greensboro, USA) and Sandra A. Vannoy (Appalachian State University, USA)

ISSN: 1947-9050; Established 2010; Published Quarterly

Publishes research on the broad and comprehensive field of dependable and trustworthy computing and information systems.

International Journal of Information Security and Privacy

Hamid Nemati (The University of North Carolina at Greensboro, USA)

ISSN: 1930-1650; Established 2007; Published Quarterly

Creates and fosters a forum where research in the theory and practice of information security and privacy is advanced.

International Journal of Monitoring and Surveillance Technologies Research

Nikolaos Bourbakis (Wright State University, USA), Konstantina S. Nikita (National Technical University of Athens, Greece), and Ming Yang (Southern Polytechnic State University, USA)

ISSN: 2166-7241; Established 2013; Published Quarterly

Publishes cutting-edge research on monitoring and surveillance technologies.

International Journal of Risk and Contingency Management

Kenneth David Strang (State University of New York, USA and APPC Research, Australia)

ISSN: 2160-9624; Established 2012; Published Quarterly

Publishes interdisciplinary research papers, reviews, and case studies that examine risk, uncertainty, and contingency.

A Framework for Digital Forensics and Investigations:
The Goal-Driven Approach

Benjamin Aziz, School of Computing, University of Portsmouth, Portsmouth, UK

Clive Blackwell, Department of Computing and Communication Technologies, Oxford Brookes University, Oxford, UK

Shareeful Islam, School of Architecture, Computing and Engineering, University of East London, London, UK

ABSTRACT

Digital forensics investigations are an important task for collecting evidence based on the artifacts left in computer systems for computer related crimes. The requirements of such investigations are often a neglected aspect in most of the existing models of digital investigations. Therefore, a formal and systematic approach is needed to provide a framework for modeling and reasoning about the requirements of digital investigations. In addition, anti-forensics situations make the forensic investigation process challenging by contaminating any stage of the investigation process, its requirements, or by destroying the evidence. Therefore, successful forensic investigations require understanding the possible anti-forensic issues during the investigation. In this paper, the authors present a new method for guiding digital forensics investigations considering the anti-forensics based on goal-driven requirements engineering methodologies, in particular KAOS. Methodologies like KAOS facilitate modeling and reasoning about goals, requirements and obstacles, as well as their operationalization and responsibility assignments. The authors believe that this new method will lead in the future to better management and organization of the various steps of forensics investigations in cyberspace as well as provide more robust grounds for reasoning about forensic evidence.

Keywords: *Anti-Forensics, Digital Forensics, Investigative Methodologies, KAOS, Requirements Engineering*

INTRODUCTION

Digital forensics is a complex and important field emerging because of the increasing nature and complexity of modern day cybercrime and the ever-increasing utilization of computer systems and digital media in real world crimes. The likelihood of becoming a target of cybercrime is a fear of almost every computer user. Therefore, cybercrime is a significant challenging problem that could cause severe financial damage. Digital forensics is a craft-based discipline

DOI: 10.4018/jdcf.2013040101

that has grown out of the need to enforce law and justice in cyberspace bringing together the whole body of knowledge in computer sciences to the legal system.

Generally cyber criminals leave evidence, which is correlated and analyzed by forensics investigators to understand who, what, why, when, where and how a crime was committed. Forensic evidence should be admissible, authentic, complete, reliable and believable by the legal system to prosecute the criminals (Brezinski & Killalea, 2002). However, anti-forensics methods have recently gained popularity by criminals who aim to interfere with the forensic processes by destroying digital evidence using different methods and tools or increasing the examiners' overall investigation time and cost. According to various international reports, the usage of anti-forensics has recently risen to over one third of cybercrime cases in recent years (Verizon Business, 2009). Therefore, a reliable framework for digital forensics investigations in terms of tools and methods is needed while at the same time addressing anti-forensic methods, particularly when time, cost and resources are critical constraints in an investigation.

Digital forensics investigation models have remained at an informal level of expressivity and there are very few attempts in literature that aim at the formalization of what a digital forensics investigation is (Leigland & Krings, 2004). For example, Carrier (2006) showed that the concept of digital forensics investigations could be mapped onto computing concepts by demonstrating that a particular program created some file, and Gladyshev (2004) analyzed a printer queue to show who printed a particular document. However, these attempts are detailed analyses of single pieces of evidence. Blackwell (2009) systematically analyzed credit card fraud using attack trees, which could also be applied to forensic investigations, and would benefit from using a more formal and systematic methodology.

According to Leigland and Krings (2004), such formalization might have several benefits, which can be classified as follows:

- **Procedural:** By reducing the amount of data and their management;
- **Technical:** By allowing digital forensic investigations to be modified to take account of the technological changes underlying them;
- **Social:** In that the capabilities of an attack are captured within the social as well as technical dimension, and finally;
- **Legal:** In that it allows the expression of the legal requirements in an investigation.

In this paper, we develop a framework to support digital forensics investigations considering possible anti-forensic situations. We use a goal-driven formal requirements engineering methodology called KAOS (van Lamsweerde, 2009) in formalizing the goals, obstacles, procedures and responsibilities involved in any digital forensics investigation. Therefore, we map the KAOS concepts such as goals, obstacles and agents with concepts used in typical digital forensics investigations.

The main contributions of this paper therefore are:

- By providing a structured framework for describing the requirements of forensic processes, including anti-forensic measures, the forensic investigator will be better equipped and guided in dealing with ad-hoc crime scenarios;
- The framework aids the forensic analyst in analyzing the value of evidence collected from the crime scene;
- The use of the concept of goals and requirements along with their operationalization and responsibility assignment captures what is in our view an important aspect of any digital investigation process that is

the current state of the investigation. This concept distinguishes our approach from other existing approaches;
- The inclusion of anti-forensic reasoning in the framework allows better coverage of technical measures for dealing with anti-forensic techniques deployed by criminals.

A main product of this approach would be to establish a pattern library for various investigations and anti-forensic models, in a similar fashion to the work of Fernandez et al. (2007) in establishing attack patterns. This library can then be used to the benefit of the investigators in guiding an instance of a digital investigation based on the main goal of the investigation, providing suggestions for ways to implement the requirements of the investigation, and assigning responsibilities to the qualified personnel or automated systems.

We believe that forensic investigations can and probably should make their goals and requirements more explicit. We think our approach is generally applicable because (at worst) it still provides a framework for collecting and analyzing evidence making it easier for investigators to determine the state of the investigation and its limitations even if the resulting goal tree is not entirely accurate. The KAOS model has been developed over the last 20 years for all aspects of requirements engineering, and is a very detailed, refined and flexible framework. We believe the framework can make investigative difficulties clear and therefore aid their discussion and planning with remedial forensic actions, but that does not mean they can necessarily be overcome just because they have been made explicit.

The rest of the paper is organized as follows: In the second section, we discuss existing works on digital forensics and anti-forensic models. In the third section, we provide a brief overview of the KAOS requirements engineering methodology. In the fourth section, we demonstrate how the various elements underlying a digital forensics investigation can be expressed in KAOS model elements. In the fifth section, we give an example of this mapping applied to digital forensics. In the sixth section, we illustrate the use of our framework with the infamous Ceglia versus Zuckerberg case, before concluding the paper in the seventh section and giving some insight into future research pathways for our approach.

RELATED WORKS

The work in this paper extends similar idea proposed recently (Aziz, 2012), which presents the first attempt in literature in utilizing KAOS as a methodology for capturing the requirements of digital forensics investigations. The current paper extends that work by including the anti-forensics dimension, which is naturally represented in KAOS using the concept of obstacle. It also demonstrates how the KAOS-based framework is used in a real world case involving document forgery.

There are several works, which consider the forensic investigation process, its challenges and tool support. This section presents some of the works that are relevant to ours. Various models, for example (Beebe & Clark, 2005; Carrier & Spafford, 2004; Casey & Rose, 2010; Ó Ciardhuáin, 2004; Cohen, 2009; Ieong, 2008; Palmer, 2001; Reith, Carr & Gunsch, 2002), have been proposed in the past to capture the process of a digital forensics investigation, which have the purpose of managing and organizing a digital investigation rather than dictating its specific steps and procedures.

Cohen (2009) provides a multistage model with several stages, including transport, storage and destruction of evidence that are typically incorporated into other stages of, or omitted altogether from other models. Ó Ciardhuáin (2004) combines existing models and proposes an extended model of cybercrime investigations, which is able to capture information flows in a full investigation. Awareness is the first step, which triggers an investigation. Authorization is obtained then from internal and external authoritative entities. Planning is considered to involve strategies and policies, whereas dissemination is used for guiding future inves-

tigations and procedures. Again, it is a staged model containing all the main investigative steps, but incorporating some aspects missing from other models.

Beebe and Clark (2005) argue the need for an objective-based framework for digital forensics owing to the uniqueness of every forensic investigation. They propose a two-tiered hierarchical framework decomposing the investigation process into seven stages further divided into sub-stages with particular goals. However, the focus of this framework is to maintain evidence integrity at all stages of an investigation. Reith et al. (2002) define an abstract model of the digital forensics procedure, which provides a mechanism to support future digital technologies and aid understanding by non-technical observers. Palmer (2001) sets out six phases to define what a digital investigative process is. These are defined around the concept of *digital evidence* and they include the identification, preservation, collection, examination, analysis and presentation of evidence. Finally, Casey (2010) also proposes a four-step model of the investigative process, which involves recognition, preservation, classification and reconstruction of digital evidence.

The above models all decompose incidents into stages, whereas the two following models consider other investigative concerns. Ieong (2008) defines a model called FORZA, which defines the differing investigative concerns for all the involved stakeholders. The second dimension of the model asks the questions who, why, what, how, when and where for each of the stakeholders, which aligns itself to our KAOS-based questioning approach, even though it does not follow a refinement process. Carrier and Spafford (2004) consider physical crime scene procedures for the forensic investigation process. The framework focuses on preserving the state of as many digital objects as possible whilst incorporating the computer crime scene within the overall crime scene. Table 1 below summarizes the models discussed above.

Our model aids the existing frameworks rather than providing a completely new one, since the other forensic models focus on capturing the forensic process itself, rather than the state of each stage in the forensic process and its requirements. More specifically, advantages of our goal-driven approach over the above models include:

- Aiding the explicit determination of the investigative goals and potential methods of achievement;
- Helping to plan a strategy, as the goal tree keeps track of the state of the investigation, and helps to determine and overcome its deficiencies;
- Allows traceability by linking the goals to the evidence and vice versa aiding the construction of valid arguments and exposing any weaknesses and assumptions;

Table 1. Summary of comparison of existing digital forensics models

Model/Framework	Advantages	Disadvantages
Multistage models (Beebe & Clark, 2005; Casey & Rose, 2010; Ó Ciardhuáin, 2004; Cohen, 2009; Palmer, 2001; Reith, Carr & Gunsch, 2002)	Decomposition into stages helps to provide a systematic approach to the investigation. They also consider the activities necessary to protect the evidence and verify its authenticity	Do not consider specific reasoning about the place of evidence in the investigation. Also, do not consider the evidence needed for presentation in court or to progress between stages
Forza (2008)	Provides a comprehensive list of concerns for the investigation	Does not provide any process to answer the investigative concerns
Carrier & Spafford (2004)	Considers computer evidence within the context of the overall investigation	States what needs to be considered without giving the detailed steps

- Allowing decomposition of complex cases into simpler and more tractable parts that can be separately analyzed;
- A completed goal tree can be used in the presentation stage to marshal the arguments to provide a coherent structure for explaining the conclusion that links to the investigative goals and explains any weaknesses and assumptions.

The limitations of the above models are that most of them consider the progression through the various stages with their different purposes rather than reason about the evidence. Most of these models do not explicitly capture the state of the investigation, the relationships between the evidence to the goals, or the legal issues because they focus mainly on technological concerns. Conversely, our model provides explicit reasoning about the evidence within a structured framework that aids the understanding and management of the state of the investigation and its weaknesses. Goal refinement makes clear how the evidence supports goal satisfaction helping us to find alternatives for missing evidence, expose our assumptions, overcome obstacles and examine unsatisfied goals. In our model, a goal tree changes during the different stages showing the progression of the investigation and potential obstacles, and finishing with the completed tree in the presentation stage.

On the other hand, there are works, which focus on the anti-forensics tactics and methods used to thwart the forensic investigation (Dahbur & Mohammad, 2011; Harris, 2006; Kessler, 2007; Rekhis & Boudriga, 2012). Most of these techniques are used to obstruct the reliability of the crime evidence, forensic investigation processes and tools. Forensic software vendors do not develop the forensic tools to function in a hostile environment with flaws as stack overflows, improper management of memory pages, and unsafe exception handling leakage. Therefore, a forensic investigation may not be successful without understanding possible anti-forensic issues for the crime.

Harris (2006) defines categories of techniques related to anti-forensics such as destroying evidence, hiding evidence and eliminating evidence sources. Several recommendations are proposed to control the anti-forensic problems. These include the necessity to control human elements such as investigator educational level, real-world experience and willingness to think in new directions that could all affect the detection of anti-forensics, and dependency on the tools, as they are not immune to attack. Several challenges imposed by anti-forensics for the forensic investigation, such as forensic constraints (i.e., time, cost and money), vulnerabilities of the forensic software, victim privacy, and nature of the digital evidence, are presented in (Dahbur & Mohammad, 2011). The work categorizes anti-forensics based on attacked target, techniques and tactics (traditional and non–traditional) and functionality.

Rekhis and Boudriga (2012) propose a theoretical approach for digital forensics investigations using awareness of anti-forensic attacks. The formal investigation process of the approach generates potential attack scenarios and formalizes the attacks based on the evidence collected from the different sources. Finally, Kessler (2007) emphasizes a time-sensitive anti-forensics concept, where the length of time that the anti-forensics method can protect data against discovery is greater than the length of time to identify, acquire, examine and analyze the data.

All the works described above are important, but they indicate general issues that need addressing or consider incidents from the perpetrator's viewpoint. Anti-forensics techniques and tools continue to provide difficulties and challenges to the overall forensic investigation process. Therefore, it is important to consider

possible anti-forensic scenarios during the forensic investigation. However, in current literature, there is limited work combining both forensics and anti-forensics into a single systematic reasoning framework. This work contributes to this direction. Table 2 provides a comparison of the above existing works on anti-forensics.

Our model again does not provide an alternative to the above anti-forensic works and models, but is also compatible with them. Our focus is on the goals of the investigation and the forensic actions to collect the necessary evidence and overcome obstacles, independent of their specific technical achievement.

BRIEF OVERVIEW OF KAOS

Knowledge Acquisition in autOmated Specification or *Keep All Objects Satisfied (KAOS)* is a generic methodology based on capturing, structuring and precise formulation of system goals (van Lamsweerde, 2009). A goal is a prescriptive description of system properties, formulated in non-operational terms. A system includes not only the software to be developed but also its environment. Goals are refined and operationalized in a top-down manner as the system is designed, or in a bottom-up approach while reengineering existing systems. The approach also supports adverse environments, containing possibly malicious agents trying to undermine the system goals rather than to collaborate in goal fulfillment.

A KAOS model is composed of a number of sub-models. We mention here four such sub-models that are most relevant to our approach:

- The *goal model* captures and structures the assumed and required properties of a system by formalizing a property as a top-level goal that is then refined to intermediate sub-goals and finally to low-level requirements that can be operationalized. Goals are organized in AND/OR refinement/abstraction hierarchies, where higher-level goals are generally strategic, coarse-grained and involve multiple agents, whereas lower-level goals are technical, fine-grained and involve fewer agents. AND-refinement

Table 2. Summary of comparison of existing anti-forensics works

Anti-Forensics Works	Advantages	Disadvantages
Provides an overview of anti-forensic types, techniques and issues that could reduce the effectiveness of anti-forensic types (Harris, 2006).	Different anti-forensic methods are discussed. Analyzes the root causes of anti-forensic methods from human, tools, and physical perspectives.	The anti-forensic types do not directly link with the forensic issues such as process and evidence. Moreover, consequences of these anti-forensic types within the overall investigation are also missing.
Classifies computer anti-forensic challenges for the forensic investigation (Dahbur & Mohammad, 2011).	Lists of challenges are presented relating to investigation constraints, software, attacks, privacy and nature of evidence.	This work provides an overview of anti-forensic challenges; however, it does not describe how such challenges obstruct the investigation process.
Defines a model of digital forensics investigations, which is aware of anti-forensic attacks (Rekhis & Boudriga, 2012).	A digital investigation process is proposed that includes analyzing the anti-forensic attacks by modeling attack scenarios and security solutions and their formal presentation.	Less emphasis on analyzing the collected evidence. Lacks reasoning facilities within the investigation process.
Defines categories of anti-forensic methods and different anti-forensic aspects (Kessler, 2007).	Time-sensitive anti-forensics concept is defined besides the common anti-forensic concepts.	The overview of the anti-forensic concepts does not link with the forensic issues; in particular, how the anti-forensic methods obstruct the overall investigation process.

links relate a goal to a set of sub-goals possibly conjoined with domain properties or environment assumptions; this means that satisfying all sub-goals in the refinement is sufficient for satisfying the goal. OR-refinement links relate a goal to a set of alternative refinements; this means that satisfying one sub-goal in the refinement is a sufficient condition for satisfying the goal. KAOS also provides the means for expressing a set of desirable properties of the AND/OR refinements, which includes the completeness, consistency and minimality of the refined sub-goals. This is more systematic than the simpler attack trees that simply decompose attacks into their different means of achievement using AND/OR refinement without considering the context or agents involved (Schneier, 1999);

- The *obstacle model* is a linking from obstacles to goals that the obstacles obstruct. The top-level obstacle is the root obstacle, which is refined through AND/OR refinements showing how sub-obstacles can satisfy the root obstacle, in a similar manner to goal refinement. KAOS presents different types of obstacles such as hazard, threat, dissatisfaction, misinformation, inaccuracy and non-usability depending on the nature of the goal and associated domain-specific properties. An anti-goal is an obstacle caused by a threat actor who deliberately interferes with goal satisfaction;
- The *agent model* assigns goals to agents in a realizable way. Discovering the responsible agents is the criterion to stop a goal-refinement process;
- The *operation model* details, at a state-transition level, the actions an agent has to perform to reach the goals for which it is responsible.

It is worth noting that the rigor of the KAOS methodology stems from the fact that any concepts defined within its sub-models incorporate formal definitions using Linear Temporal Logic (LTL) (Vardi, 2001) formulae. In this paper, we do not delve into LTL encodings of the forensic requirements; however, these have the potential for establishing formally verified pieces of evidence due to the current support from technologies such as model checking, theorem proving and static analysis.

KAOS GOAL-DRIVEN CONCEPTS FOR DIGITAL FORENSICS INVESTIGATIONS

Our work focuses on the KAOS modeling language for digital forensics investigations. In this section, we demonstrate how elements of a digital forensic investigation can be mapped to KAOS model elements. We assume that a digital forensic investigation consists of processes and their anti-forensics, actions and the personnel including law enforcement agencies and the active systems used by the personnel. These entities map to KAOS concepts such as goals, obstacles, agents and operations.

Conceptual View

A goal is the center of our approach. Goals in our case are the objectives, expectations and constraints stemming from the forensic investigation processes, tools and digital evidence. These goals need to be satisfied for a forensic investigation to be successful. Obstacles are the negation of goals (Islam & Houmb, 2010). In particular, they refer to what could go wrong during a forensic investigation specifically relating to the evidence collection and analysis within the forensic process.

We follow the anti-forensic methods and tools to identify the obstacles that directly reduce the reliability of the digital evidence. These obstacles need to be addressed, which we do through determining and performing suitable forensic actions. Suitable agents undertake the responsibility to execute the actions so that goals can be attained. Our approach defines an agent as an expert human investigator, a law enforcement agency or an active system component

such as a forensic computer system, tool or software. Figure 1 shows the conceptual view of the proposed approach. Goals are derived from the forensic process, tools and evidence, and obstacles are derived from anti-forensic situations, tools and obstruction of the digital evidence.

Levels of Abstraction

The proposed goal-driven digital forensics investigation model supports different levels of abstraction from goals to obstacles and finally to forensics actions. Figure 2 gives an overview of the different levels of abstraction using exemplary questions that symbolize the characteristics of the method (Islam, 2011). We divide the levels of abstraction into three main areas. These levels build the bridge from the goals of a forensics investigation to the obstacles, which obstruct the goals to be satisfied, and finally to the actions that oppose the obstacles in order to achieve the goals. On the top level, there are the goals, i.e. objectives, expectations, and constraints for the forensics investigation, which map to the processes involved within the forensics investigation. The root goal is refined into sub-goals through AND/OR refinement links.

In the middle level, obstacles obstruct the goals of a forensic investigation by destroying, hiding, manipulating and preventing the creation of evidence. Obstacles are the different types of anti-forensics, which confront the availability and usefulness of the evidence for the forensics investigation. Therefore, for a successful forensics investigation, we need to address the obstacles related to anti-forensics; in particular, it is necessary to completely prevent, or if not feasible then reduce the effectiveness of the anti-forensic methods. An identified obstacle is refined by AND/OR refinements through the obstruction of crime evidence or the forensic investigation process.

In the bottom level, forensic actions support goal fulfillment and obstacle opposition. These actions operationalize the requirements for the

Figure 1. Conceptual view of digital forensics investigation using the goal-driven approach

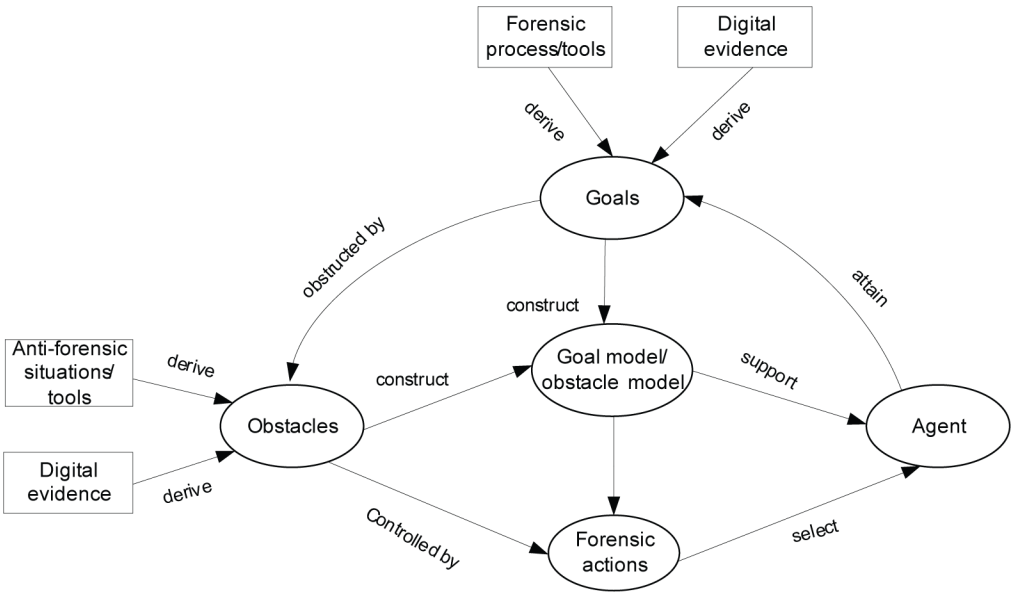
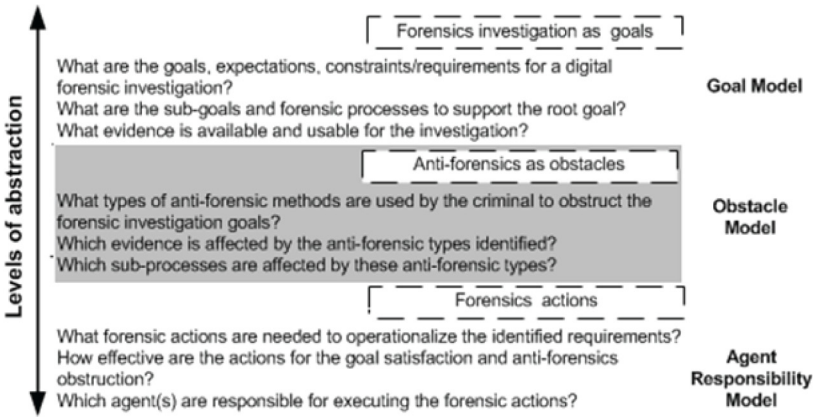


Figure 2. Levels of abstraction



forensic investigation and provide countermeasures to obstruct the obstacles. These actions are linked with the agent, such as software, system, law enforcement officer or expert, responsible for performing the action.

A goal model refines higher-level goals to lower level finely grained sub-goals through AND or OR refinement so that sub-goals contribute to the main goals. We consider that digital forensic investigation is the parent goal based on the context of the crime event. This goal is refined to sub-goals by different activities of the forensic investigation process. The refinement provides a goal model of the forensic investigation.

An obstacle model analyses the anti-forensic situations of the cybercrime. In KAOS, obstacles are the negation of a goal's fulfillment. We map the anti-forensic type as the root obstacle and obstruction of evidence and the forensic process as a sub-obstacle. Refinement from root obstacle to sub-obstacle produces the obstacle model and includes an obstruction link from obstacle to goal. Agent responsibility determines who are responsible for executing the forensic actions for goal satisfaction and obstacle negation. Therefore, an agent has a role and needs certain capabilities for goal satisfaction.

GOAL-DRIVEN FORENSIC AND ANTI-FORENSIC INVESTIGATIONS

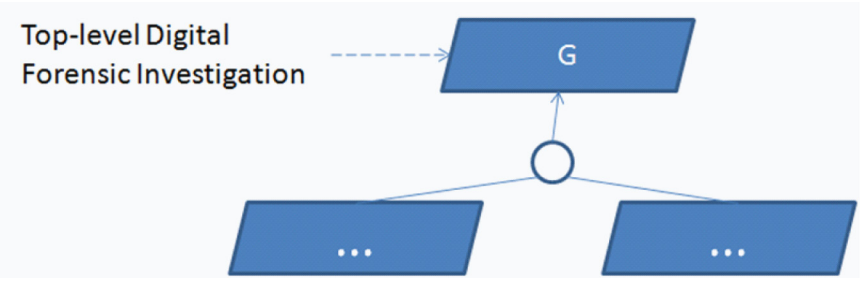
In this section, we demonstrate how forensic processes and their anti-forensic obstacles can be modeled in terms of the KAOS goal engineering methodology. We shall use the Ceglia versus Zuckerberg case (Stroz Friedberg, 2012) as our working example throughout this section, and give more details later on in the case analysis in the sixth section.

Forensic Investigations as the Root Goal

The starting point is to map a forensic investigation into the KAOS goal model. This is done by defining the root goal of a KAOS goal tree, G, as a representation of the top-level digital forensic investigation, as shown in Figure 3.

In this sense, a digital forensic investigation is seen as the top-level goal that the responsible law enforcement agency aims to achieve, and is therefore the root of the KAOS-based goal tree. Examples of such a goal would be to *investigate fraud case in some bank, investigate a case of identity theft,* or *investigate online criminal activity.* However, it is important to

Figure 3. Mapping digital forensics investigations



keep in mind that such a root goal is necessarily abstract at this stage, with the next step aiming at its refinement in terms of more concrete goals.

Forensic Processes as Sub-Goals

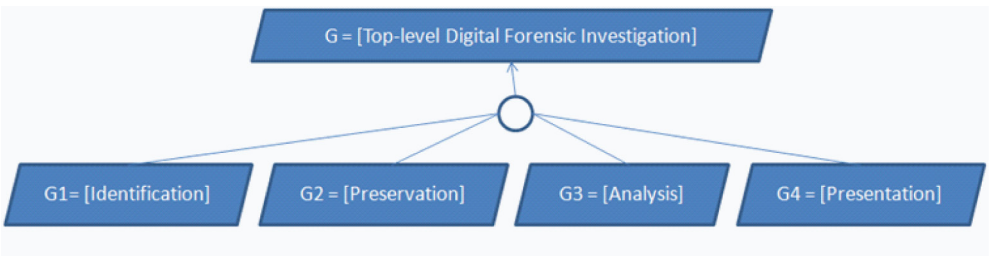
The next mapping involves the various steps of the investigative process. For simplicity, we adopt McKemmish's (1999) early and simple model, which defines a digital forensic process as "*The process of identifying, preserving, analyzing and presenting digital evidence in a manner that is legally acceptable.*" This definition implies the activities of *identification, preservation, analysis* and *presentation*. These are modeled in terms of the sub-goals of the root goal representing the top-level investigation, as shown in Figure 4. Note that our model is general and can be used to capture any other definition of the digital forensic process.

The specific steps involved in each of these activities will represent further sub-goals that must be achieved in order for the activity to be achieved. Such sub-goals may be linked by either the AND or the OR refinement relations. The refinement of Figure 4 is an example of an AND-refinement, which implies that unless all the sub-goals (identification, preservation, analysis and presentation) are satisfied, the main goal of the investigation will not have been achieved. In the analysis of the image of a hard disk acquired from Ceglia (Stroz Friedberg, 2012), the investigators may need to achieve both analysis of files' internal content and system log files (AND refinement).

A different type of refinement is the OR-refinement, which implies that any of the sub-goals is sufficient for the achievement of the parent goal. However they may have more than one choice in one of the two analyses; for example, analyzing the alleged contract by examining the original or a deleted version (OR refinement).

Each of the above main activities will be refined until one arrives at the lowest possible requirements corresponding to specific elements of each activity that cannot be refined (detailed)

Figure 4. Mapping activities of the digital forensic process



any further. A requirement is considered a leaf in the goal tree of the main investigation. Once all the requirements have been identified, it is necessary to a) operationalize them by means of appropriate forensic actions and b) assign the actions to the responsible agents. If it proves impossible to operationalize all goals, then this indicates that the investigation may not be successful.

For example, if critical evidence cannot be collected and there are no alternatives, then it may be possible to detect this early before resources are wasted on fruitless investigation paths. In the case study, Ceglia did not provide the original electronic contract to Zuckerberg's expert witnesses, and so file carving was required to recover the deleted versions from a hard disk that showed numerous inconsistencies with his account, including timestamps that appeared to show that the contract was created several years after Ceglia claims. In this case, exhaustive search discovered several other inconsistent draft versions on Ceglia's hard disk.

Anti-Forensics as Obstacles

As stated, the main objective of a digital forensics investigation process is to systematically obtain, examine and analyze digital information for evidence in a court of law. Anti-forensics obstructs the forensics goals; in particular, it hinders the efforts of forensics experts using techniques and tactics to destroy, alter or hide data leading to a lack of good evidence.

Evidence analysis is essential for the forensics investigation process. It is not possible initially to work out the complete goal tree because there may be unavailable evidence or the discovery of new obstacles. However, the goal tree can help case management to discover and plan how to overcome potential missing evidence and obstacles in advance.

Evidence should be:

- **Admissible:** Must be able to be used in court, so must be relevant and not prejudicial;
- **Authentic:** Original and unchanged;
- **Reliable:** Correct and accurate, and can be used to make sound inferences;
- **Complete:** Important evidence is available and entire; and
- **Believable:** Easy to understand by and credible to a jury (Brezinski & Killalea, 2002).

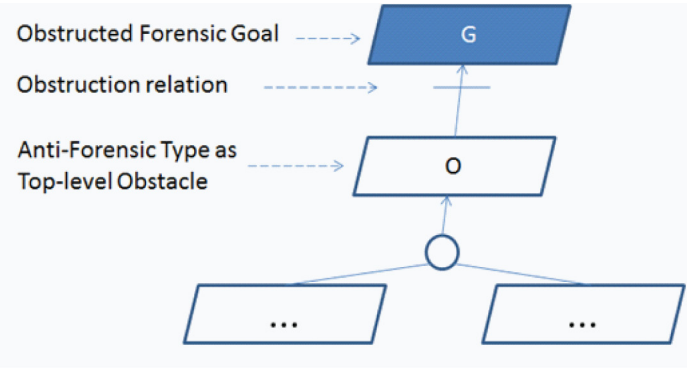
Anti-forensic methods focus on obstructing the objective of obtaining evidence in a digital forensic investigation, one that exhibits the above properties. Depending on the context, criminals use various types of obstructive approaches to forensics investigation including hiding, wiping or corrupting the evidentiary artifacts or overwriting data (Guidance Software, 2007; Rekhis & Boudriga, 2012).

As Figure 5 shows, we consider an anti-forensic type as the main obstacle, O, which directly obstructs specific goals or sub-goals of the forensic investigation. An obstruction link from the anti-forensics type to the goal provides traceability regarding the specific goal being obstructed. Therefore, we know the potential effect on the investigation if a particular goal is obstructed, and so the goal tree helps to understand and manage the state of the investigation by giving ways to defeat the obstacle or achieve alternative goals if available.

The main obstacle identified previously is the type of anti-forensic attack. The obstacle needs refinement eventually into anti-requirements to provide detail about the anti-forensic attacks used during the crime. The obstacle anti-requirements bear the same relationship to obstacles as do requirements to goals, in that they are specific enough to determine actions that can be carried out successfully to cause an obstacle to a goal. This refinement supports proper understanding of the crime and identification of forensic actions for dealing with the crime.

Considering potential obstacles helps planning to overcome them in advance, such as when the opposing party fails to supply incriminating evidence as required by warrant. For example, Ceglia did not supply any of the original evidence to Zuckerberg's expert, which

Figure 5. Mapping anti-forensics



can be decomposed into failing to provide computers, disks, files, documents and email at least. He did not supply complete computers only hard disks, several storage devices such as USB sticks were missing, the original contract had been deleted from a supplied disk, and the supporting email had been cut-and-pasted into a Word document. Anticipating these obstacles, their refinement into anti-requirements and operationalization with anti-forensic actions leads to predicting their effect on the investigation. It also helps to consider other possible sources of evidence, such as third parties that had a copy of a contract that differed from the one provided by Ceglia, and system and file metadata that shows manipulation of the alleged contract and supporting email respectively.

We consider two types of obstacle refinement involving obstruction of the forensic investigation process or evidence or both. The investigation can be led astray by creating false leads, or the opposition can waste the investigator's time by failing to supply evidential items such as email accounts, or by providing a mass of irrelevant data such as the provision of over 1,000 data CDs by Ceglia.

The digital forensics investigation process we adopted here consists of the four main activities of identification, preservation, analysis and presentation, each of which can be obstructed. For example, the failure of Ceglia to provide the original evidence caused issues for each of the four stages:

- He failed to identify and supply all relevant computer systems, storage media and email accounts that he owned, possessed or used;
- He failed, amongst other things, to preserve the originals of the alleged contract and the supporting email between himself and Zuckerberg;
- He hindered analysis of the contract and email by apparently making changes to the file and email metadata so that they appeared legitimate;
- He delayed presentation in court with legal manoeuvring and several changes of lawyers.

An obstacle can oppose the integrity, completeness, reproducibility, timeliness and believability of the activity and outputs produced by the activity. Therefore, obstruction of any of these properties is an obstacle for the digital forensics investigation. Criminals can also undermine the forensic tools such as Encase, FTK or WinHex, so that the produced evidence is incorrect, incomplete or unreliable (Guidance Software, 2007; Kessler, 2007). As stated previously, evidence for legal prosecution should achieve all of the five different rules: admissibility, authenticity, completeness, reliability and believability.

We consider five different evidence violation rules for refinement with examples from the case study:

- **Inadmissible:** Evidence lacks relevance or is too prejudicial, such as introducing previous crimes into court proceedings. Ceglia has previously been arrested for grand larceny for his company's failure to provide goods, but this evidence was admitted into court even though it might be prejudicial because it was relevant;
- **Unauthentic:** Evidence is not original or does not come from the correct source. Ceglia did not provide the originals of the alleged contract and email he exchanged with Zuckerberg. However, deleted drafts of the contract were discovered and copies of the email that were cut-and-pasted into a Word document;
- **Incomplete:** Evidence is missing or partial. Ceglia's evidence was incomplete, because he failed to provide complete computers only hard disks, several storage media including at least six USB sticks, or even an electronic version of the alleged contract;
- **Unreliable:** Lack of correctness of evidence, or invalid inferences that extend from permissible interpretation of the evidence into speculation. Zuckerberg's experts (Broom, 2012) argued that the email timestamps in the headers of the email saved in Word mentioned above were often one hour out, indicating that they were created at a different time of year when daylight saving hours were in operation. However, email timestamps are often incorrect because accounts are misconfigured;
- **Unbelievable:** Unclear or implausible evidence. The apparently legitimate alternative contract agreed between Ceglia and Zuckerberg, which did not mention Facebook, was found on a server belonging to Ceglia's lawyer and in the Outbox of the email account of Ceglia's parents. Broom's expert report (2012) for Ceglia claimed that this was a forgery and proposed a convoluted method of how the contract could have been fabricated and placed in these locations by Zuckerberg or his accomplices. However, there was convincing evidence that Ceglia's lawyer had acted upon the contract back in 2004, so the claim would be that Zuckerberg faked the contract before 2004 in anticipation of a possible case six years later in 2010.

A criminal can use data wiping techniques, such as strong encryption or strong deletion techniques, to destroy the evidence by using some wiping tools such as BC wipe, Window Washer or Secure Clean. Data wiping as an obstacle is a refinement of the "non-authenticity of evidence" higher anti-forensic goal since the evidence can be corrupted by wiping parts of it leading to the loss of its authenticity. The alleged paper contract supplied by Ceglia apparently had attached a fake page 1 to the legitimate page 2 of the contract he actually signed with Zuckerberg, whose lack of authenticity was shown by the inconsistency between the paper, toner and ink of the two pages.

Data wiping also obstructs the identification of evidence by removing it completely and hence is a refinement of the "incompleteness of evidence" anti-forensic goal. Ceglia had deleted draft versions of the alleged contract, but they were recoverable using file carving, and there was strong evidence that they had been created several years after Ceglia claimed the contract was agreed from the inconsistency of the file timestamps. Figure 6 shows these concepts.

Forensic Actions as Operations

Once the goal tree has been sufficiently specified starting from the top-level digital investigation goal and ending with the leaves corresponding to the low-level requirements, these requirements (and consequently the goal tree) need to be *operationalized* by necessary and sufficient forensic actions satisfying the requirements and thereby eventually leading to the satisfaction of the main goal of the investigation.

Figure 7 illustrates how the sub-goal Identification can be first refined to the three requirements, R1, R2 and R3, of authorized legal seizures, seizures of suspected PCs and seizures of suspected mobile devices, respec-

Figure 6. The obstacle model with data wiping

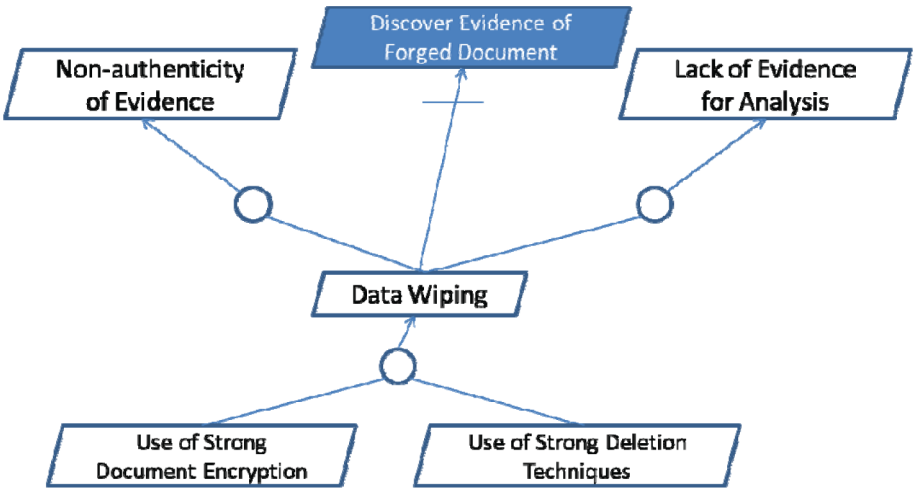
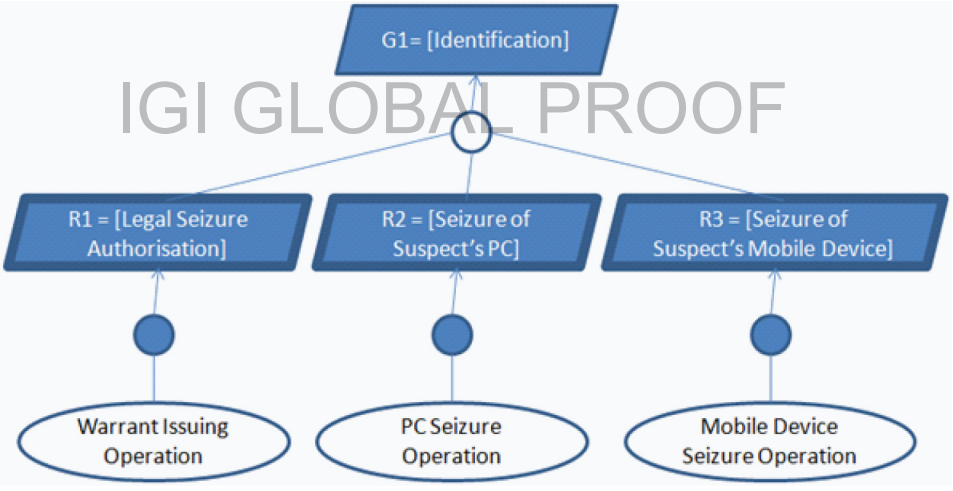


Figure 7. Requirements operationalization



tively. These then are operationalized in terms of the three operations: warrant issuing, PC seizure and mobile device seizure.

In general, the assignment of the requirements to the operations is crucial as it ensures that every requirement of the forensic investigation can be satisfied through the application of some appropriate operation (or method or activity). If it is determined that some operations cannot be carried out adequately, remedial measures can be taken to overcome the problem another way, or to the early abandonment of the investigation if the issues cannot be surmounted. Ideally, potential operational difficulties can be planned for and alternative actions placed in the goal tree in advance, discovered from patterns of previous similar cases.

Law Enforcement Agencies, Personnel and Systems as Agents

The final step in our modeling approach is to model law enforcement agencies and personnel as well as their active systems and tools in terms of KAOS agents. For example, Figure 8 illustrates a Forensic Scene Investigator agent in relation to the requirements and operations of the Identification sub-goal. The dashed line represents the relation that the Forensic Investigator applies the forensic actions operationalizing those requirements, whereas the solid line represents the relation that the Investigator is indeed responsible for the satisfaction of the requirement to which it is related.

CASE STUDY: INVESTIGATING DIGITAL MEDIA FORGERY

In this section, we show one example of a case study involving the application of the KAOS method to modeling the goals, obstacles and operations of a digital forensics investigation. The investigation is focused on alleged document forgery in a legal case involving Paul Ceglia and Facebook's Mark Zuckerberg. The digital aspects of the investigation were carried out by a digital risk management company called Stroz Friedberg, who submitted the report to the court (Stroz Friedberg, 2012). A more detailed KAOS-based analysis of this case has recently been shown (Blackwell, Islam, & Aziz, 2013).

Paul Ceglia filed a legal complaint in 2010 against Facebook and Mark Zuckerberg claiming that the latter sold him half of Facebook's assets in exchange for $2000, as part of a disputed contract. It was agreed that there was a contract covering some work that Zuckerberg did for Ceglia in a different (pre-Facebook) project called StreetFax, but Ceglia provided a contract giving him half of Facebook, whereas Zuckerberg's version did not mention Facebook. Based on the outcome of the report (Stroz Friedberg, 2012), we generated the goal model of Figure 9 to capture the main goals and requirements involved in an investigation of document forgery.

As usual, we refer to a goal or a sub-goal with the symbol *G* and to its requirements with the symbol *R*. We only consider the sub-goal of identification (of evidence or the lack of it) in the case, and leave out the other sub-goals related to preservation, analysis and presentation.

Figure 8. Responsibility assignment

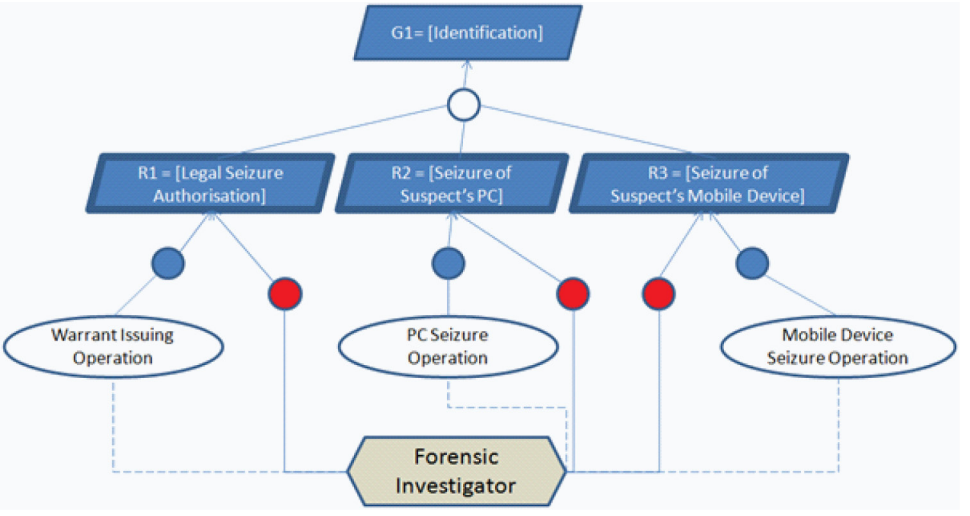
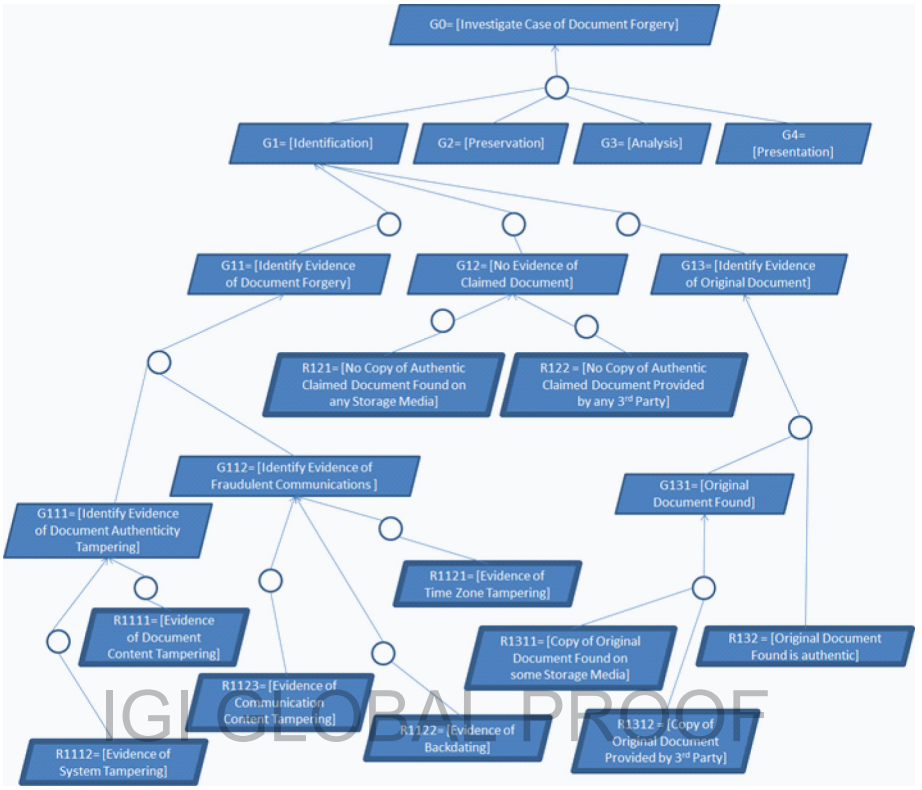


Figure 9. Modeling the goal of the investigation: Investigating a case of document forgery



The overall goal for the forensics investigator (Stroz Friedberg, 2012) was to prove the alleged contract a forgery, which would cause Ceglia's claim for part ownership of Facebook to fail, as it is the only supplied evidence capable of proving his version of events. Proving the contract a forgery is sufficient to defeat Ceglia's claim, but in complex cases, proving document forgery may only be a sub-goal supporting the overall investigation.

We create a structured goal tree including the general goals and obstacles in proving document forgery. An initial generic goal tree developed from previous similar cases can help determine an initial approach that focuses attention on the likely evidence and its potential locations. In addition, we need to determine the specific issues for each new case of document forgery to help decompose the goal tree as the investigation progresses to take account of new evidence and overcome proffered obstacles.

As the investigation proceeded, this was broken down in terms of three main sub-goals:

- Identifying evidence of contract document forgery by Ceglia;
- Showing that no evidence existed regarding the presence of an authentic version of the claimed contract; and finally
- Identifying the presence of the original contract document signed for project StreetFax.

Two of the branches relate to proving the forgery of the alleged contract, whereas the third relates to the existence of the alternative

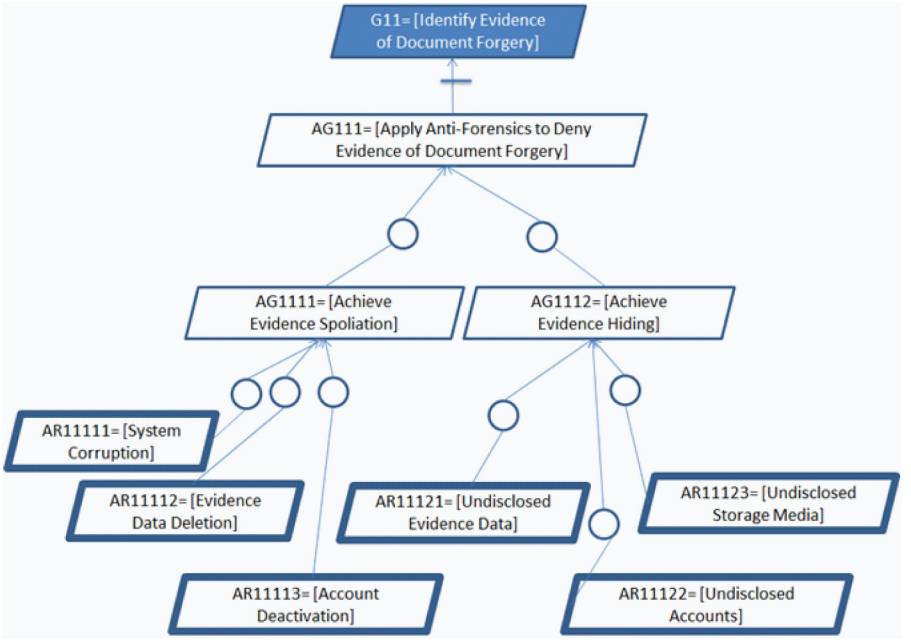
StreetFax contract that did not mention Facebook. The first sub-goal directly aims to prove the contract a forgery, whereas the second shows there is a lack of evidence for the validity of the contract. Proving the second goal is easier to satisfy and suffices, as Ceglia did not provide any other compelling evidence. However, it is a weaker goal that is not definitive as the opposition may obstruct it by providing stronger evidence for the contract's existence.

There is also the particular sub-goal in this case of proving the alternative StreetFax contract is valid. It was agreed between the parties that only one contract existed between them, so that proving the authenticity of the StreetFax contract amounted to proving the claimed contract a forgery. It is important to include the domain assumption of the existence of only one contract, as it is external evidence outside the remit of the forensic investigation, and the argument for the third goal is rendered null and void if the assumption is found to be invalid.

These are OR-refinements of the main investigation goal, since any of these is sufficient to achieve the goal. However, although it is sufficient to prove forgery in one way only, we might consider proving forgery in multiple ways to handle unanticipated obstacles. All these three sub-goals are then broken down in terms of other sub-goals and requirements, and can be related to the specific parts of the Stroz Friedberg report (2012, Sections VI-XI).

The report also highlighted (Stroz Friedberg, 2012, Sections XIII and XIV) some examples of the use of anti-forensic methods to deter or obstruct the investigators from showing the non-authenticity of the claimed documents. The anti-forensic methods (goals and requirements) given in the report are summarized in Figure 10, and are generally divided into evidence spoliation (interferes with authenticity

Figure 10. An anti-goal model obstructing the document forgery sub-goal



and reliability) and evidence hiding (interferes with completeness) anti-goals. Similar to the standard requirements, the operationalization of anti-requirements leads to the specification of the actual tools and operations needed to achieve them. Recall that an anti-goal is simply an obstacle that someone causes deliberately to satisfy their needs. We shall refer to an anti-goal as *AG* and to its refined anti-requirements as *AR*.

Figure 11 next shows the operationalization of the requirements for discovery of document authenticity tampering (G111 in Figure 9) and fraudulent communications (G112 in Figure 9).

The next Figure 12 shows the operationalization of the requirements for finding the original of the alleged contract (G13 in Figure 9) and the lack of authenticity copy of claimed contract (G12 in Figure 9). It would be expected that if the alleged contract was legitimate, it would be found on the media provided by Ceglia or be provided by third parties. However, goal G12 should not be relied upon unless there is no alternative, as the discovery of an apparently legitimate contract later could provide an insurmountable obstacle.

Finally, Figure 13 demonstrates the operationalization of the anti-requirements depicted in Figure 10 with the operationalization of evidence spoliation (AG1111) above and of evidence hiding (AG1112) below.

CONCLUSION AND FUTURE WORK

The execution of a digital forensic investigation can be a complex and disorganized exercise, often leading to trust in invalid pieces of evidence or failure in the process leading to the reconstruction of such evidence. Therefore, we agree with Eoghan Casey that the use of a formal methodology in describing the process of a digital forensics investigation "encourages a complete, rigorous investigation, ensures proper

Figure 11. The operationalization of requirements for the discovery of document authenticity tampering (G111 above) and fraudulent communications (G112 below)

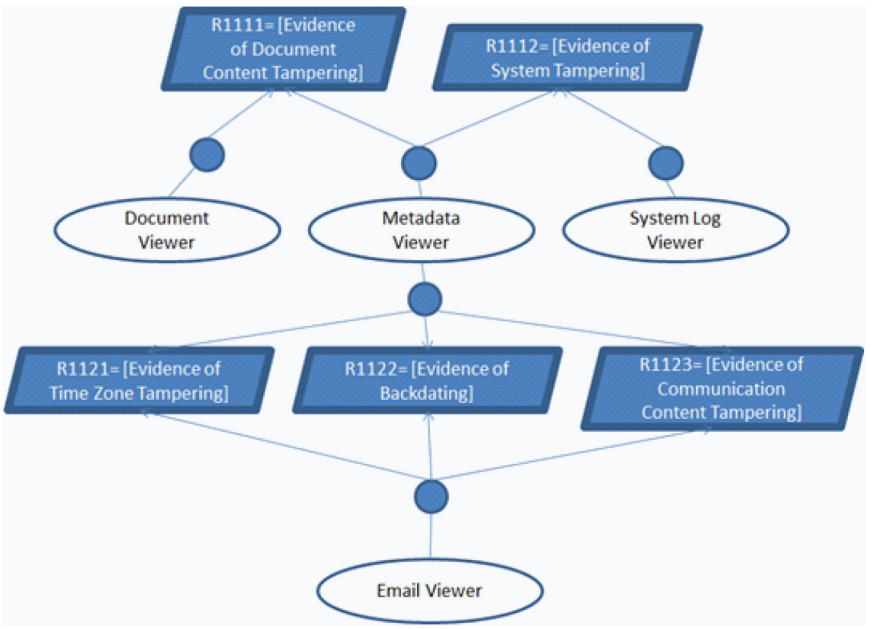


Figure 12. The operationalization of requirements for the discovery of the alleged contract (G13 below) and the lack of authenticity of the claimed contract (G12 above)

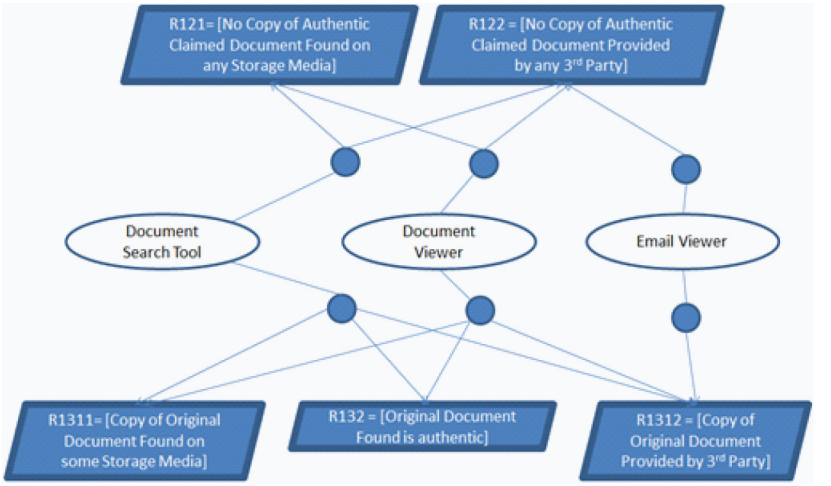
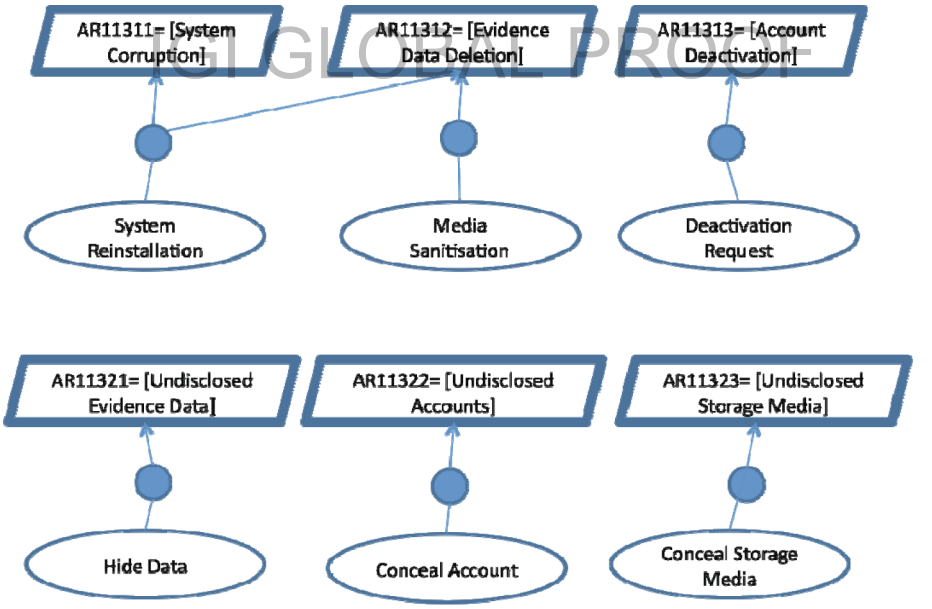


Figure 13. The operationalization of anti-requirements for evidence spoliation (AG1111 top) and evidence hiding (AG1112 bottom)



evidence handling and reduces the chance of mistakes created by preconceived theories, time pressures and other potential pitfalls." (Casey, 2011). In this paper, we defined one such systematic approach based on a well-established and rigorous requirements engineering methodology, namely KAOS. In addition to its systematic approach, KAOS also provides a formal model for describing its various entities (requirements, agents, operations etc.), which is based on LTL

logic. Such LTL-based semantics could be used in the future to provide more robust reasoning about the worthiness of the evidence, specifically when temporal properties are concerned, such as formalizing timeline analysis.

Our approach focuses on the concept of the state of each stage in the forensic investigative process. This facilitates applications to the inherent evidence collection and analysis issues in investigations relating to the availability, integrity and authenticity of evidence that may never have been collected or cannot be shown to be reliable. In addition, evidential issues can be caused by benign agents such as system administrators trying to be helpful, victims to avoid distress or embarrassment, witnesses who detect and correct suspicious changes, and other system users who may inadvertently alter or destroy evidence. This is known as evidence dynamics, which refers to events that may change, relocate, obscure or obliterate evidence after the incident (Casey & Rose, 2010).

Goal modeling also helps to discover that sufficient evidence cannot be collected if it is not possible to operationalize all goals. Goal refinement may help us to detect failure early before resources are wasted on fruitless investigation. One aspect pertaining to the rigor of this approach is that it ensures that all goals are acted upon, as the crucial obstacles and their potential effects on the investigations are exposed.

Of the limitations of our framework is the difficulty with which goals/requirements and other KAOS concepts can be specified when dealing with crimes or incidents that are ill defined or poorly understood. At present, our focus is on types of crime for which computer evidence is fundamental such as document forgery or online fraud. To analyze complex crime or incidents such as the Advanced Persistent Threat (APT), other techniques would be required. The results could also be problematic if there were unexpected issues with the investigation, caused by unusual variations in the crime or further undiscovered offenses, which requires continuous questioning of the assumptions, evidence, obstacles and so forth. This is not a problem with the framework per se, but of an unthinking acceptance of its results without question. In particular, the model helps examine alternative hypotheses within the goal tree helping to avoid tunnel vision. The model can also help isolate and make explicit any problems allowing subsequent feedback taking account of the new information with the addition of branches to the generic goal tree for the type of case. Some of these questions will be the focus of our continuing research.

In addition, this work can be extended in many other directions. First, we plan to provide a general goal library of some of the most common patterns of digital forensics investigations and their requirements. The Common Attack Pattern Enumeration and Classification (CAPEC) (MITRE Corporation, 2013) schema has over 400 attack patterns that can possibly be adapted for use in digital forensics. Second, we also plan to investigate other aspects of KAOS, such as the anti-goal model (van Lamsweerde, 2004), in guiding models of attacks in the cyber world and their relationship to real world crime. Finally, we plan to conduct real-world case studies to investigate the pros and cons of this method in enhancing the digital forensics investigation process.

REFERENCES

Aziz, B. (2012). Towards goal-driven digital forensics investigations. In *Proceedings of the 2nd International Conference on Cybercrime, Security and Digital Forensics (Cyfor-12)*, London, UK.

Beebe, N. L., & Clark, J. G. (2005). A hierarchical, objectives-based framework for the digital investigations process. *Digital Investigation*, *2*(2), 146–166. doi:10.1016/j.diin.2005.04.002.

Blackwell, C. (2009). A reasoning agent for credit card fraud on the internet using the event calculus. *International Journal of Electronic Security and Digital Forensics, 2*(1), Inderscience.

Blackwell, C., Islam, S., & Aziz, A. (2013). Implementation of digital forensics investigations using a goal-driven approach for a questioned contract. In *Proceedings of the 9th Annual IFIP WG 11.9 International Conference on Digital Forensics*, Orlando, FL.

Brezinski, D., & Killalea, T. (2002). *Guidelines for evidence collection and archiving*. RFC 3227.

Broom, N. (2012). *Declaration of Neil Broom, Ceglia v. Zuckerberg and Facebook, Inc. No. 1:10-cv-569-RJA-LGF*, Technical Resource Center, Inc.

Carrier, B. (2006). *A hypothesis-based approach to digital forensic investigations*. PhD thesis, CERIAS, Purdue University, CERIAS Tech Report 2006-06.

Carrier, B. D., & Spafford, E. H. (2004). An event-based digital forensic investigation framework. In *Proceedings of the 2004 Digital Forensics Research Workshop*, Baltimore, MD.

Casey, E. (2011). *Digital evidence and computer crime – Forensic science, computers and the internet* (3rd ed.). Elsevier.

Casey, E., & Rose, C. (2010). *Forensic discovery, handbook of digital forensics and investigation*. Academic Press.

Ciardhuáin, Ó, S. (2004). An extended model of cybercrime investigations. *International Journal of Digital Evidence*, *3*(1).

Cohen, F. (2009). *Digital forensic evidence examination*. Fred Cohen & Associates.

Dahbur, K., & Mohammad, B. (2011). The anti-forensics challenge. In *Proceedings of the 2011 International Conference on Intelligent Semantic Web-Services and Applications*, ACM.

Fernandez, E., Pelaez, J., & Larrondo-Petrie, M. (2007). Attack patterns: A new forensic and design tool. In *Proceedings of the 3rd Annual IFIP WG 11.9 International Conference on Digital Forensics, Springer Advances in Digital Forensics III* (pp 345-357), Orlando, FL.

Gladyshev, P. (2004). *Formalising event reconstruction in digital investigations*. Unpublished PhD thesis, Department of Computer Science, University College Dublin.

Guidance Software. (2007). *Guidance software response to iSECxReport*. Retrieved June 1, 2012, from http://www.securityfocus.com/archive/1/474727

Harris, R. (2006). Arriving at an anti-forensics consensus: Examining how to define and control the anti-forensics problem. In *Proceedings of the 6th Annual Digital Forensic Research Workshop (DFRWS06)*, Elsevier.

Ieong, R. S. C. (2008). FORZA – Digital forensics investigation framework that incorporates legal issues. In *Proceedings of the 8th Digital Forensic Research Workshop*, Baltimore, MD.

Islam, S. (2011). *Software development risk management model – A goal-driven approach*. Unpublished PhD thesis, Technische Universität München, Germany.

Islam, S., & Houmb, S. H. (2010). Integrating risk management activities into requirements engineering. In *Proceedings of the 4th IEEE Research International Conference on Research Challenges in Information Science (RCIS2010)*, France.

Kessler, G. C. (2007). Anti-forensics and the digital investigator. In *Proceedings of the 5th Australian Digital Forensics Conference*, Perth, Australia.

Leigland, R., & Krings, A. W. (2004). a formalization of digital forensics. *International Journal of Digital Evidence*, *3*(2).

McKemmish, R. (1999). What is forensic computing? *Trends and Issues in Crime and Criminal Justice, 118*.

MITRE Corporation. (n.d.). *Common attack pattern enumeration and classification (CAPEC)*. Retrieved February 1, 2013, from http://capec.mitre.org

Palmer, G. (2001). *A road map for digital forensic research* (DFRWS Technical Report T001-01). Retrieved February 1, 2013, from http://www.dfrws.org/2001/dfrws-rm-final.pdf

Reith, M., Carr, C., & Gunsch, G. (2002). An examination of digital forensic models. *International Journal of Digital Evidence*, *1*(3).

Rekhis, S., & Boudriga, N. (2012). A system for formal digital forensic investigation aware of anti-forensic attacks. *IEEE Transactions on Information Forensics and Security, 7*(2).

Schneier, B. (1999). Attack trees: Modeling security threats. *Dr. Dobb's Journal*, *24*(12), 21–29.

Stroz Friedberg. (2012). Report of digital forensic analysis. In Paul D. Ceglia v. Mark Elliot Zuckerberg (Eds.), *Individually, and Facebook, Inc. Civil Action No: 1:10-cv-00569-RJA*. Retrieved March 26, 2012, from http://www.wired.com/images_blogs/threatlevel/2012/03/celiginvestigation.pdf

van Lamsweerde, A. (2004). Elaborating security requirements by construction of intentional anti-models. In *Proceedings of the 26th ACM-IEEE International Conference on Software Engineering (ICSE'04)* (pp. 148-157). IEEE Press, Edinburgh, U.K.

van Lamsweerde, A. (2009). *Requirements engineering: From system goals to UML models to software specifications*. Wiley.

Vardi, M. Y. (2001). Branching vs. linear time: Final showdown. In *Proceedings of the 7th International Conference On Tools and Algorithms for the Construction and Analysis of Systems (TACAS 2001)*, Springer Lecture Notes in Computer Science 2031 (pp 1–22), Springer, Genoa, Italy.

Verizon Business (2009). *2009 data breach investigations report*. A study conducted by the Verizon RISK Team in cooperation with the United States Secret Service.

Benjamin Aziz is a senior lecturer at the School of Computing, University of Portsmouth. Benjamin has research experience in the field of computer and information security spanning 15 years where he worked in the past at Rutherford Appleton Laboratory and Imperial College, and has published more than 70 articles and book chapters in areas related to the security of large-scale systems, formal security, requirements engineering and digital forensics. He is on board program committees for several conferences and working groups in relevant areas, including the Cloud Computing Security Alliance, ERCIM Formal Methods for Industrial and Critical Systems, ERCIM Security and Trust Management and IFIP WG 11.3 on Data and Application Security and Privacy. Benjamin is a certified forensic investigation practitioner and a certified wireless security analyst.

Clive Blackwell is currently a research fellow at Oxford Brookes University, where his main area of research is in cybersecurity and digital forensics. He studied for a PhD at Royal Holloway, University of London, with his main field of research in security architecture, where he developed a three-layer security architecture to model complex systems such as the Internet and critical infrastructure. Clive holds undergraduate degrees in Mathematics from Warwick University, and in Computer Science from London University where he finished top in his class, and an MSc in Information Security also from London University. He has more than 50 conference publications and presentations to his name. He was the co-chair of the Cyberpatterns workshop and regularly presents to industry both at home and abroad. His is also interested in the science of digital forensics, and the use of abduction and argumentation to reason about fraud and digital forensics.

Shareeful Islam is currently working at the School of Architecture, Computing and Engineering (ACE), University of East London, UK. He was awarded his PhD for a thesis on a Software Risk Management Model using a goal-driven approach by the Chair of Software & Systems Engineering (I4), Technische Universität München, Germany. He received M.Sc. in Information Communication System Security from the Royal Institute of Technology (KTH), Sweden and MSc in CS and BSc (Hons) in APE from the University of Dhaka, Bangladesh. He is a Fellow of the British Higher Education Academy (HEA) and has published more than 40 referred papers in high-quality journals and international conferences. He participated in EU, industry, KTP projects. His research interests and fields of expertise are risk management, requirements engineering, security, privacy, trust, and cloud computing.

Collision Analysis and Improvement of a Parallel Hash Function Based on Chaotic Maps with Changeable Parameters

Min Long, School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha, China

Hao Wang, School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha, China

ABSTRACT

Recently, a parallel hash function based on chaotic maps with changeable parameters was proposed by Li et al (2011, pp.1305-1312). In this paper, the security of it is analyzed and the weakness of the architecture is pointed out. It is found that the main limitations are the error using of floor, round and exclusive OR operations in the algorithm. In order to counterstrike these, some improvements are done to strength its security. Theoretical analysis and experimental results illustrate that the improved Hash function is more secure and practical than the original one.

Keywords: *Changeable Parameters, Chaos, Collision, Hash Function, Security Analysis*

1. INTRODUCTION

One-way hash function is a fundamental technique for information security, and it is usually applied for integrity protection, digital signatures and message authentication. In the past few years, chaos has been found that it has great potential to be used in the construction of hash function due to its sensitivity to initial conditions and system parameters, ergodicity and random like behavior. Thus, many works have been done on the chaos-based hash functions (Akhavan & Samsudin, 2009; Ren & Wang, 2009; Xiao & Liao, 2008; Zhang &Wang, 2007). Among them, wide attention has been paid to parallel hash function, where the sub-blocks of a message are processed in a parallel mode with high efficiency (Xiao & Liao, 2008). However, cryptanalysis of chaos-based hash functions is also developed very fast. Some chaos-based hash functions also been ben proved to be insecure (Li & Li 2006;

DOI: 10.4018/jdcf.2013040102

Guo & Wang, 2009; Wang & Li, 2012; Wang & Wang, 2008; Wang & Xu, 2010; Wang & Zhao, 2010). Collisions and flaws exist if two or more distinct messages or keys are found to obtain a same hash value, which can be implemented by adversary to fabricate fake messages. For this reason, collision resistance is a basic requirement for a secure hash function.

Recently, a parallel hash function based on chaotic maps with changeable parameters is proposed by Li et al. (2011). Detailed analysis is performed to it, and it is found that it is vulnerable to collision attacks, thus, some measurements of how to improve its security are proposed in this paper.

The rest of the paper is organized as follows. The original algorithm is described and analyzed in the second Section 2, and some improvements are made in Section 3. In Section 4, the experiments and analysis are performed to evaluate the performance of the improved hash function. Finally, some conclusions are drawn in the Section 5.

2. ORIGINAL HASH FUNCTION AND ITS SECURITY ANALYSIS

2.1. Description of the Chaos-Based Parallel Hash Function

In the parallel hash function, two chaotic maps are used. One is tent map defined in Equation (1), where $0 < \alpha < 1$, and $0 \le x_i \le 1$. The other is piecewise linear maps defined in Equation (2) (Box 1), where $0 < P < 0.5$, and $0 \le X(t) \le 1$:

$$x_{i+1} = \begin{cases} \dfrac{x_i}{\alpha}, & if(0 \le x_i \le \alpha) \\ \dfrac{1-x_i}{1-\alpha}, & if(\alpha < x_i \le 1) \end{cases} \quad (1)$$

The hash function is composed of the following three steps:

Step 1: Message expansion. The message is first padded with 64 bits, which represents the length of the original message, and then padded another s bits $(1010\ldots10)_2$ so that the padded message can be partitioned into n blocks. Each block consists of 1016 bits, and the padded message can be expressed by a matrix M:

$$M = \begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,127} & r_2 \\ m_{2,1} & m_{2,2} & \cdots & m_{2,127} & r_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ m_{n-1,1} & m_{n-1,2} & \cdots & m_{n-1,127} & r_n \\ c_{127} & c_{126} & \cdots & c_1 & r_1 \end{bmatrix} \quad (3)$$

Box 1.

$$X(t+1) = F_P(X(t)) = \begin{cases} X(t)/P, & 0 \le X(t) < P \\ (X(t)-P)/(0.5-P), & P \le X(t) < 0.5 \\ (1-X(t)-P)/(0.5-P), & 0.5 \le X(t) < 1-P \\ (1-X(t))/P, & 1-P \le X(t) \le 1 \end{cases} \quad (2)$$

where elements c_j $(j = 1,2,...,127)$ and r_i $(i = 1,2,...,n)$ are defined in Equation (4) and (5), respectively.

$$c_j = \bigoplus_{i=1}^{n}(m_{i,j} \oplus q_i) + q^j \quad (j = 1,2,...,127) \tag{4}$$

$$r_i = \begin{cases} \sum_{j=1}^{127}((m_{i,j} + q^j) \oplus q_i) & i = 1,2,...,n-1 \\ \sum_{j=1}^{127}((col_j + q^j) \oplus q_i) & i = n \end{cases} \tag{5}$$

where q^j $(j = 1,2,...,127)$ and q_i $(i = 1,2,...,n)$ are the (n+127) successive values multiplied by 2^8, which are obtained by iterating chaotic asymmetric tent map with initial values of $\alpha = 0.6$ and $x_0 = 0.7654$, respectively. "$\oplus$" represents bitwise exclusive OR operation, "+" represents addition modulo 2^8, and "$\sum$" represents summation modulo 2^8:

Step 2: Parallel processing. For each message block M_i, set $M = (M_1, M_2, \cdots, M_n) = (m_{i,j})$, $(i = 1,2,\cdots,n; j = 1,2,\cdots,128)$, where M_i denotes the *i*th row elements of the matrix M. Firstly, the current state value $x_{atm}(m_{i,j})$ is generated by iterating tent map $\lfloor j / Hl \times m_{i,j} \rfloor$ times with an initial value $x_{plm}(m_{i,j-1})$ and a changeable parameter $\alpha = (i/n + j/Hl)/2$ when $j = 1$ holds. Otherwise, α remains unchanged. Here, $x_{plm}(m_{i,j-1})$ is the last iterated value of the piece-wise linear map. Secondly, the current value $x_{plm}(m_{i,j})$ is generated by iterating the piece-wise linear map $\lfloor (1 - j / Hl) \times m_{i,j} \rfloor$ times with a changeable parameter $\beta = \alpha / 2$. Thirdly, the value of $x_{plm}(m_{i,j-1})$ is rounded to 0 or 1. Until all sub-blocks are proposed, *Hl* numbers of 0 or 1 will be obtained. The intermediate *Hl*-bit hash value $H(i)$ is generated by cascading *Hl* numbers of 0 or 1:

Step 3: Hash value generation. After all plaintext blocks have been processed, the hash value $H(M) = H(1) \oplus H(2) \oplus \cdots \oplus H(i) \oplus \cdots \oplus H(n)$ is obtained.

2.2. Security Analysis of the Hash Function

In the original hash function scheme, floor operations $\lfloor j / Hl \times m_{i,j} \rfloor$ and $\lfloor (1 - j / Hl) \times m_{i,j} \rfloor$ are used to generate the iteration times of the chaotic maps. However, it is not sensitive to tiny difference in messages. Moreover, intermediate hash value $H(i)$ is composed of 0 or 1, which is generated from the rounding results of $x_{plm}(m_{i,j-1})$. However, analysis found that the probability of $round(x_{plm}(m_{i,j-1})) = round(x_{plm}(m'_{i,j-1}))$ is 50%, where $m_{i,j-1} \neq m'_{i,j-1}$. It means that it has great probability to obtain a same final hash value corresponding with two different messages. The collision analysis is generalized as the following 3 propositions:

2.2.1. Proposition 1

When the length of original message is less than 127, given an original message matrix $Me = [m_{i,j}]$, where $i = 1; j = 1,2,\cdots,n$. After padding, Matrix M denotes the padded message, and M is shown in Equation (3). If there exist $\lfloor j \times m_{1,j} \rfloor = 128Z$ $(Z \in \{1,2,3,\cdots,n\})$, j=64, and $m_{i,j}$ is even, the probability of a collision is 33.35%:

- **Proof:** Construct another message matrix $Me^{'} = [m'_{i,j}]$, where $i = 1; j = 1,2,\cdots,n$, $m'_{i,j} = m_{i,j}$ ($j \neq 64$) and $m'_{i,j} = m_{i,j} + 1$ ($j = 64$). After padding, it can be expressed as M'. Here, three situations are considered:

  i. It can be found that there exist $\lfloor j \times m_{1,j} / Hl \rfloor = \lfloor j \times m'_{1,j} / Hl \rfloor$ and $\lfloor (1\text{-}j/Hl) \times m_{1,j} \rfloor = \lfloor (1\text{-}j/Hl) \times m'_{1,j} \rfloor$ ($j = 1,\cdots,127$) when $Hl = 128$;
  ii. Computer exhaustive searching shows that $\lfloor j \times c_{128-j} / Hl \rfloor = \lfloor j \times c'_{128-j} / Hl \rfloor$ and $\lfloor (1\text{-}j/Hl) \times c_{128-j} \rfloor = \lfloor (1\text{-}j/Hl) \times c'_{128-j} \rfloor$ ($j = 64$) hold with a same probability of 66.7%. Here, Computer exhaustive searching is performed by setting $c_{128-j} = 1,2,3\cdots254,255$ and $c'_{128-j} = 2,3,4\cdots255,1$ successively;
  iii. When $r'_1 \neq r_1$ and $r'_2 \neq r_2$, only the last bit of intermediate hash value is influenced, i.e., $H(1') = H(1)$ and $H(2') = H(2)$ hold with a same probability of 50%.

According to the above analysis, it can be inferred that the probability of $H(M') = H(M)$ is 33.35%.Given two different messages Message1 and Message2 as followings:

Message1: fgqtauvrsokcwreybcdabpemoydztyeH;
Message2: fgqtauvrsokcwreybcdabpemoydztyeI.

It can be found that they have a same hash value: 0x6FBD75CCD84AE834473D-0B0E00FA 3220.

2.2.2. Proposition 2

When the length of the original message is less than 127, give an original message matrix $Me = [m_{i,j}]$, where $i = 1; j = 1,2\cdots$. If there exist $\lfloor k_1 \times m_{1,k_1} \rfloor = 128Z_1, \lfloor k_2 \times m_{1,k_2} \rfloor = 128Z_2$ ($Z_1, Z_2 \in \{1,2,3\cdots\}$), and $m_{1,k_1} = m_{1,k_2} + 1$, the probability of a collision is 22.22%:

- **Proof:** Construct another message matrix $Me^{'} = [m'_{i,j}]$, where $m'_{i,j} = m_{i,j}$ ($j \neq k_1, k_2$), $m'_{i,k_1} = m_{i,k_2}$ and $m'_{i,k_2} = m_{i,k_1}$. After padding, it can be expressed as M'. Here, three situations are considered:

  i. It can be found that:
  $\lfloor k_1 \times m_{1,k_1} / Hl \rfloor = \lfloor k_1 \times m'_{1,k_1} / Hl \rfloor$,
  $\lfloor k_2 \times m_{1,k_2} / Hl \rfloor = \lfloor k_2 \times m'_{1,k_2} / Hl \rfloor$,
  $\lfloor (1 - k_1 / Hl) \times m_{1,k_1} \rfloor = \lfloor (1 - k_1 / Hl) \times m'_{1,k_1} \rfloor$
  and:
  $\lfloor (1 - k_2 / Hl) \times m_{1,k_2} \rfloor = \lfloor (1 - k_2 / Hl) \times m'_{1,k_2} \rfloor$;
  ii. Computer exhaustive searching shows that:
  $\lfloor k_2 \times c_{128-k_2} / Hl \rfloor = \lfloor k_2 \times c'_{128-k_2} / Hl \rfloor$,
  $\lfloor (1 - k_2 / Hl) \times c_{128-k_2} \rfloor = \lfloor (1\text{-}k_2 / Hl) \times c'_{128-k_2} \rfloor$,
  $\lfloor k_1 \times c_{128-k_1} / Hl \rfloor = \lfloor k_1 \times c'_{128-k_1} / Hl \rfloor$,
  and:
  $\lfloor (1 - k_1 / Hl) \times c_{128-k_1} \rfloor = \lfloor (1\text{-}k_1 / Hl) \times c'_{128-k_1} \rfloor$
  hold with a same probability of 66.7%, and the computer exhaustive searching is performed by setting $c_{128-k_1} = 1,2,3\cdots254,255$ and $c_{128-k1} = 1,2,3\cdots255$ successively;
  iii. When $r'_1 \neq r_1$, $r'_2 \neq r_2$ hold, $H(1') = H(1)$ and $H(2') = H(2)$ hold with a same probability of 50%.

According to the above analysis, it can be inferred that the probability of $H(M') = H(M)$ is 22.22%. Two different messages Message1 and Message2 are given as followings:

Message1: k@iyiejskejegaltexcmniuqcjncqpqA;

Message2: kAiyiejskejegaltexcmniuqcjncqpq@.

After calculation, it can be found that they have a same hash value: 0xF9642C04591833D-0473D0B0FD2ED E857.

2.2.3. Proposition 3

When the length of original message is more than 127, give an original message $Me = [m_{i,j}]$, where $i = 1,2,\cdots,n-1; j = 1,2,\cdots 127$ .Ifthere exist $\left\lfloor k_1 \times m_{a,k_1} \right\rfloor = 128Z_1, \left\lfloor k_2 \times m_{a,k_2} \right\rfloor = 128Z_2$ ($Z_1, Z_2 \in \{1,2,3\cdots\}$), $m_{a,k_1} = m_{a,k_2} + 1$, $m_{a,k_1} = m_{b,k_2}$ and $m_{a,k_2} = m_{b,k_1}$, the probability of a collision is 50%:

- **Proof:** Construct another message matrix $Me' = [m'_{i,j}]$, where $m'_{i,j} = m_{i,j}$ (except $m'_{a,k_1} = m_{a,k_2}, m'_{a,k_2} = m_{a,k_1}, m'_{b,k_1} = m_{b,k_2}$ and $m'_{b,k_2} = m_{b,k_1}$), and the padded messages and original of that are denoted as M' and M, respectively. It can be found:

$$\left\lfloor (k_2 / 128) \times m_{a,k_2} \right\rfloor = \left\lfloor (k_2 / 128) \times m'_{a,k_2} \right\rfloor = \left\lfloor (k_2 / 128) \times m_{a,k_1} \right\rfloor = Z_1$$

$$\left\lfloor (1 - k_2 / 128) \times m_{a,k_2} \right\rfloor = \left\lfloor (1 - k_2 / 128) \times m'_{a,k_2} \right\rfloor = m_{a,k_2} - Z_1$$

$$\left\lfloor (k_1 / 128) \times m_{a,k_1} \right\rfloor = \left\lfloor (k_2 / 128) \times m'_{a,k_1} \right\rfloor = Z_2$$

and:

$$\left\lfloor (1 - k_1 / 128) \times m_{a,k_1} \right\rfloor = \left\lfloor (1 - k_2 / 128) \times m'_{a,k_1} \right\rfloor = m_{a,k_1} - Z_2$$

$$\left\lfloor k_2 \times c_{128-k_2} / Hl \right\rfloor = \left\lfloor k_2 \times c'_{128-k_2} / Hl \right\rfloor$$

$$\left\lfloor (1 - k_2 / Hl) \times c_{128-k_2} \right\rfloor = \left\lfloor (1\text{-}k_2 / Hl) \times c'_{128-k_2} \right\rfloor$$

$$\left\lfloor k_1 \times c_{128-k_1} / Hl \right\rfloor = \left\lfloor k_1 \times c'_{128-k_1} / Hl \right\rfloor$$

and:

$$\left\lfloor (1 - k_1 / Hl) \times c_{128-k_1} \right\rfloor = \left\lfloor (1\text{-}k_1 / Hl) \times c'_{128-k_1} \right\rfloor$$

hold, respectively, where:

$$c_k = c'_k,\ k = 1,2,\cdots,127$$

When $r'_a \neq r_a$ and $r'_b \neq r_b$, only the last bit of intermediate hash value is influenced. i.e., $H(a') = H(a)$ and $H(b') = H(b)$ hold with a same probability of 50%.

According to the above analysis, it can be inferred that the probability of $H(M') = H(M)$ is 50%.Give two different messages Message1 and Message2 as followings:

Message1: akfjmvoiagkbvJd9dfkafjds-ftldlfk8jfkdjkjjkjaklfhiokhjdhfjahfoe-idfkankjgjagfpfhelajl kvmkjoqiuioreyeiu poanfcmcvnankewrihqyuoedkn,vmcxnv bhahkfjdad8jeijkjakdjfkdjf9;

Message2: akfjmvoiagkbvJd8dfkafjds-ftldlfk9jfkdjkjjkjaklfhiokhjdhfjahfoe-idfkankjgjagfpfhelajl kvmkjoqiuioreyeiu poanfcmcvnankewrihqyuoedkn,vmcxnv bhahkfjdad9jeijkjakdjfkdjf8.

It can be found that they have a same hash value: 0xE423E99D470DACAA7D4D39CFC-CA6C 57B.

In addition, there is no measure to avoid the weak key $\alpha = (i / n + j / Hl) / 2 = 0.5$ in Equation (1).

3. IMPROVEMENTS TO THE ORIGINAL SCHEME

Here, some improvements are done to enhance the security of the hash function scheme by Li et al (2011). For the convenience of comparison and description, only the different parts between the original scheme and the improved version are described. They are described as follows:

1. The Equation (4) is changed to:

$$c_j = \sum_{i=1}^{n} (m_{i,j} + q_i) \oplus q^j \quad (j = 1,2,...,127) \tag{6}$$

Where "$\oplus$" represents bitwise exclusive OR operation, "+" represents addition modulo 2^8, and "$\sum$" represents summation modulo 2^8. Since "$\oplus$" in Equation (4) follows the commutative principle, i.e., exchange of two elements in different rows has no influence on the result c_j, which will lead to a collision. After the change, as seen from Equation (6), the operation of addition modulo "+" does not follow the commutative principle. Even if two messages meet the requirements of proposition 1 proposition 2 or proposition 3, c_j is still different from c'_j, Therefore, the two different messages will have different iteration times:

2. α, β and the initial value in the *step* 2 are changed to:

$$\alpha = (i/(n+1) + j/(Hl+1) + (m_{i,j} + 0.8)/256)/3 \tag{7}$$

$$\beta = (i/(n+1) + j/(Hl+1) + (m_{i,j} + 0.8)/256)/7 \tag{8}$$

$$x'_{i0} = (0.8 + ((\sum_{j=1}^{127} m_{i,j}) \bmod 256))/256 \tag{9}$$

where x'_{i0} is the initial value of the asymmetric tent map for M_i.

As α and β defined in the original algorithm is a main reason for leading to same results in the symmetric position in the matrix M, it is needed to redefine α and β. As seen in Equation (7)-(9), α, β and x'_{i0} are associated with the message $m_{i,j}$. Thus, different α, β will be generated from different $m_{i,j}$:

3. The iteration times of the asymmetric tent map and the piecewise liner map in *step* 2 are changed into $\lfloor j/Hl \times m_{i,j} \rfloor + m_{i,j}$ and $\lfloor (1-j/Hl) \times m_{i,j} \rfloor + m_{i,j}$ respectively.

Analysis results indicated that the iteration time in the original algorithm eliminates the subtle differences between messages. Thus, it is redefined to enhance the sensitivity to the change of message.

4. Operation "$\oplus$" in *step* 3 is changed to the equation shown in Box 2.

In the improved algorithm, complete different matrix *M*, iteration time and parameters in the chaotic map can be obtained from messages with little differences. Even there exist

Box 2.

$$\oplus = \begin{cases} XOR & if\ i \bmod 2 = 1 \\ XOR\ after\ H(i)\ is\ circle\ shifted\ 64\ bits & otherwise \end{cases} \tag{10}$$

two messages fulfilling the requirements of proposition 1, proposition 2 and proposition 3, it can be found that:

$$\left\lfloor j \times m_{i,j} \,/\, Hl \right\rfloor \neq \left\lfloor j \times m'_{i,j} \,/\, Hl \right\rfloor$$

$$\left\lfloor (1\text{-}j/Hl) \times m_{i,j} \right\rfloor \neq \left\lfloor (1\text{-}j/Hl) \times m'_{i,j} \right\rfloor$$

$$\left\lfloor j \times c_{128-j} \,/\, Hl \right\rfloor \neq \left\lfloor j \times c'_{128-j} \,/\, Hl \right\rfloor$$

and:

$$\left\lfloor (1\text{-}j/Hl) \times c_{128-j} \right\rfloor \neq \left\lfloor (1\text{-}j/Hl) \times c'_{128-j} \right\rfloor$$

Thus, the collision can be avoided.

4. PERFORMANCE ANALYSIS OF THE IMPROVED SCHME

4.1. Distribution of the Hash Value

A secure hash function should achieve a uniform distribution of hash value. Experiments are performed to the improved scheme to compute the hash value with a message *M*:

The Internet of Things is a technological revolution that represents the future of computing and communications, and its development needs the support from some innovational technologies One of the biggest breakthroughs of the Internet of Things is making the physical world and information world together. Sensors play very important role to bridge the gap between the physical world and information world. Sensors collect data from their environment, generating information raising awareness.

The distribution of the ASCII codes of the original message is shown in Figure 1, which is localized within a small and stable area. At the same time, the hash value in hexadecimal spreads around uniformly, as seen in Figure 2. The results indicate that the improved scheme has a good diffusion and confusion.

4.2. Statistical Analysis

Here, a same statistics analysis method described by Li et al (2011, pp.1310-1311) is used. Four statistics are defined as follows:

- Mean changed bit number:

$$\bar{B} = \frac{1}{J}\sum_{i=1}^{N} B_i$$

- Mean changed probability:

$$P = (\bar{B} \,/\, H_l) \times 100\%$$

- Standard deviation of the changed bit number:

$$\Delta B = \sqrt{\frac{1}{J-1}\sum_{i=1}^{N}(B_i - \bar{B})^2}$$

- Standard deviation:

$$\Delta P = \sqrt{\frac{1}{J-1}\sum_{i=1}^{N}(B_i / H_l - P)^2} \times 100\%$$

Where J is the total number of tests, B_i denotes changed 1 bit number in the i^{th} test, and H_l is the length of Hash value.

The statistical analysis results are listed in Table 1.

As seen From Table 1, it can be found that the mean changed bit number $\bar{B}$ and the mean changed probability P are both very close to the ideal value 64 bits and 50%, respectively, all index are very near to the ideal level. Compared with the original hash function, standard deviation of the changed bit number and standard deviation is smaller in the improved one. The mean changed bit number in hash value with 1 bit changed in message is illustrated in Figure 3. The tests are performed *J*-time, where *J*=2048. It can be seen that the changed bit

Figure 1. Distribution of the original message in ASCII

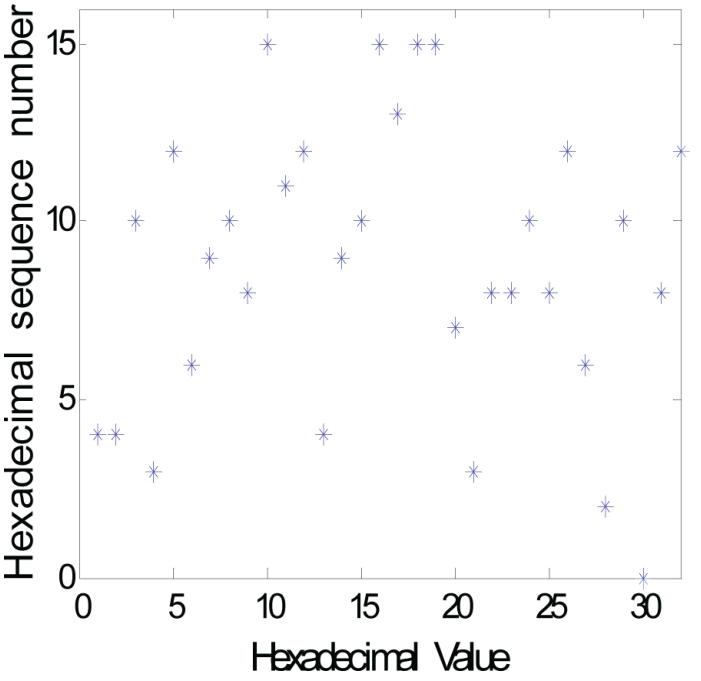


Figure 2. Distribution of the final hash values in hexadecimal

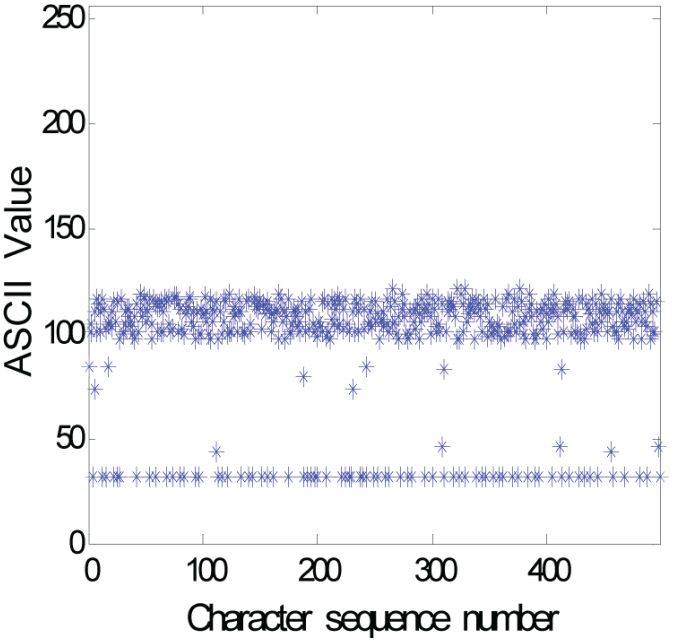
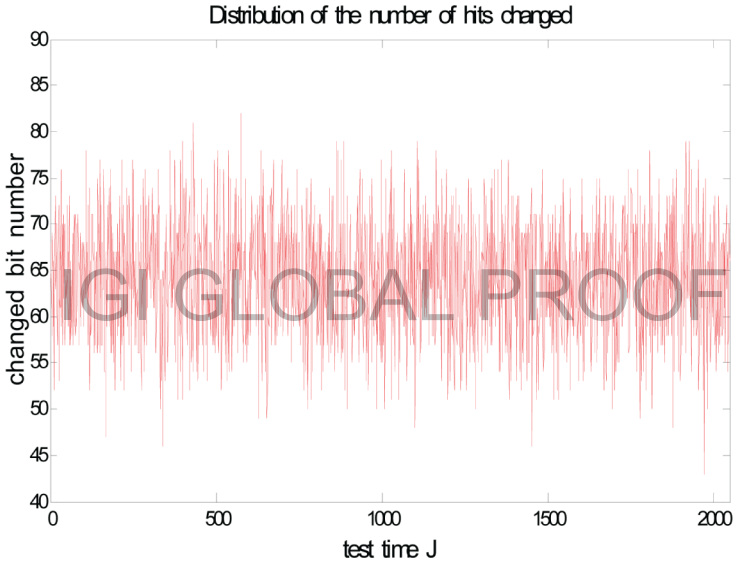


Table 1. Statistics of number of changed bit

J	*J=256*	*J=512*	*J=1024*	*J=2048*	*Mean*
$\overline{B}$	64.2890	64.2988	64.2783	64.0488	64.2287
$P(\%)$	50.22	50.23	50.21	50.03	50.17
ΔB	5.6895	5.9250	5.9055	5.8284	5.8371
$\Delta P(\%)$	4.44	4.62	4.61	4.55	4.56

Figure 3. Distribution of the number of bits changed in improved algorithm



number corresponding to 1 bit changed message concentrates around the ideal changed bit number 64 bits. The above results indicate that the improved scheme has a stronger capability in diffusion and confusion than that of the original hash function scheme.

4.3. Analysis of Collision

Here collision analysis is conducted with the same message. Following the method described by LI et al (2011), $d = \sum_{i=1}^{N} \left| t\left(e_i\right) - t\left(e_i^{'}\right) \right|$ is used for the evaluation, where e_i and $e_i^{'}$ are the i^{th} ASCII character of the original and the new hash value, respectively, and function $t(\cdot)$ converts the entries to their equivalent decimal values. The collision test is performed with 2048 times, including 1927 tests with no hit, 118 tests with once hit, and 3 tests with twice hits. The distribution of the number of ASCII characters with the same value at the same location in the hash value is shown in Figure 4.

Figure 4. Distribution of the number of ASCII characters with the same value at the same location in the hash value

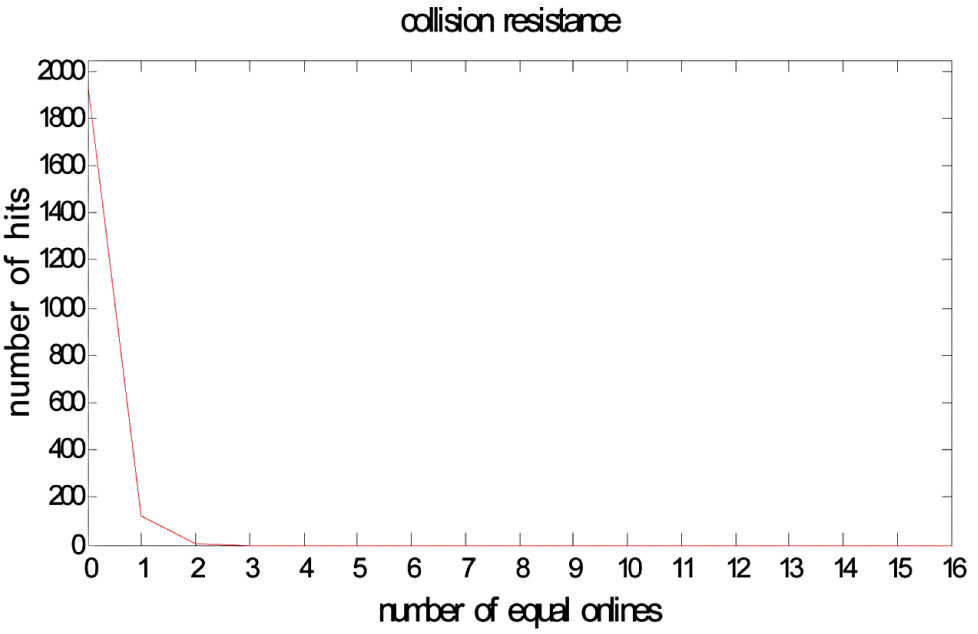


Table 2. Comparison of the two algorithms on the collision test

Scheme	Message	Hash Value (128bits)
The original scheme	M1	0x: A9F7A21C65D0BD493B0E78D0815DF9D0
	M2	0x: A9F7A21C65D0BD493B0E78D0815DF9D0
	M3	0x: 6283066216279E04D3F992057A2B160F
	M4	0x: 6283066216279E04D3F992057A2B160F
	M5	0x: 62D9D387DF4464C7CF87BDA97DF8B7DC
	M6	0x: 62D9D387DF4464C7CF87BDA97DF8B7DC
The improved scheme	M1	0x: 019CA5A69B388F931B6E16ABC606FA97
	M2	0x: FC958DD0809D878A6946498F0AAA64E6
	M3	0x: C4A218B3096E5DF92CA425666AEBD888
	M4	0x: D195A02C2D889AAA83CFEC4A80AA7A5C
	M5	0x: 71E7B475702A3CC6F6860E5BCEA926D7
	M6	0x: 67CFB899339EDB3830458BA0C937ABF2

The original algorithm and the improved version are conducted tests with the following six cases. *M* is the message described in Section 4.1:

- **M1:** *"abcdefgopqrstuvwxyzabcdefglmoqrhabcd";*
- **M2:** *Change the 32nd character "h" to "i" in M1;*
- **M3:** *Change the 8th and 247th character in M to "p", and change the 120th and 135th character in M to "q";*
- **M4:** *Change the 8th and 247th character in M to "q", and change the 120th and 135th character in M to "p";*
- **M5:** *Change the 8th and 143rd character in M to "0", change the 16th and 135th character in M to "1", Change the 32nd and 175th character in M to "X", and change the 48th and 159th character in M to "Y";*
- **M6:** *Change the 8th and 143rd character in M to "0", change the 16th and 135th character in M to "1", Change the 32nd and 175th character in M to "Y", and change the 48th and 159th character in M to "X".*

From Table 2, it can be seen that it is easy to find a collision in the original algorithm, while the improved one has effectively overcome this weakness. The results illustrate the good collision resistance capability of the improved hash function scheme.

Compare to another parallel keyed hash function (Xiao & Liao, 2008), this algorithm iterates the chaotic asymmetric tent map and the chaotic piecewise linear map with changeable parameters, which are dynamically obtained from the position index of the corresponding message blocks. At the same time, it employs an operation addition modulo "+", which does not follow the commutative principle. Therefore, it can resist forgery attack.

5. CONCLUSION

In this paper, the security of a recent proposed chaos-based parallel hash function has been investigated in details. Analysis results show that it is not secure enough to resist collision attacks. Improved measures are proposed to enhance its security. Experimental results and analysis also illustrated that the improved chaos-based hash function scheme has stronger confusion and diffusion capability, and better collision resistance.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (Grant No. 61001004), the Education Department Foundation of Hunan Province (Grant No.11B002), and the Overseas Teacher Project of Hunan Province (Grant No.2012007).

REFERENCES

Akhavan, A., Samsudin, A., & Akhshani, A. (2009). Hash function based on piecewise nonlinear chaotic map. *Chaos, Solitons, and Fractals*, *42*(2), 1046–1053. doi:10.1016/j.chaos.2009.02.044.

Guo, W., Wang, X., & He, D. (2009). Cryptanalysis on a parallel keyed hash function based on chaotic maps. *Physics Letters. [Part A]*, *373*(36), 3201–3206. doi:10.1016/j.physleta.2009.07.016.

Li, C., Li, S., & Lou, D. C. (2006). On the security of the Yen-Guo's domino signal encryption algorithm. *Journal of Systems and Software*, *79*(2), 253–258. doi:10.1016/j.jss.2005.04.021.

Li, Y., Xiao, D., & Deng, S. J. (2011). Parallel hash function construction based on chaotic maps with changeable parameters. *Neural Computing & Applications*, *20*(8), 1305–1312. doi:10.1007/s00521-011-0543-4.

Ren, H., Wang, Y., & Xie, Q. (2009). A novel method for one-way hash function construction based on spatiotemporal chaos. *Chaos, Solitons, and Fractals*, *42*(4), 2014–2022. doi:10.1016/j.chaos.2009.03.168.

Wang, J., Wang, M., & Wang, Y. (2008). The collision of one keyed hash function based on chaotic map and analysis. *Acta Physica Sinica*, *57*(5), 2737–2743.

Wang, J., Xu, S., & Tian, M. (2010). The analysis for a chaos-based one-way hash algorithm. In *Proceedings of the 2010 International Conference on Electrical and Control Engineering* (pp. 25-27).

Wang, S., Li, D., & Zhou, H. (2012). Collision analysis of a chaos-based hash function with both modification detection and localization capability. *Communications in Nonlinear Science and Numerical Simulation*, *17*(2), 780–784. doi:10.1016/j.cnsns.2011.06.017.

Wang, X., & Zhao, J. (2010). Cryptanalysis on a parallel keyed hash function based on chaotic neural network. *Neurocomputing*, *73*(16-18), 3224–3228. doi:10.1016/j.neucom.2010.05.011.

Xiao, D., Liao, X. F., & Deng, S. J. (2008). Parallel keyed hash function construction based on chaotic maps. *Physics Letters. [Part A]*, *37*(3), 4682–4688. doi:10.1016/j.physleta.2008.04.060.

Zhang, J., Wang, X., & Zhang, W. (2007). Chaotic keyed hash function based on feedforward–feedback. *Physics Letters. [Part A]*, *362*(5-7), 439–448. doi:10.1016/j.physleta.2006.10.052.

Min Long received the PhD degree in circuits and systems from the South China University of Science and Technology, Guangzhou, China, in 2006. She was a visiting fellow with the Department of Computer Science at U.K. University of Warwick from 2009 to 2010. Currently, she is an associate professor with the College of Computer and Communication, Changsha University of Science and Technology, Changsha, China. Her areas of interest include digital watermarking, chaos-based secure communication, and UWB secure communication.

Hao Wang is a master's candidate of the College of Computer and Communication, Changsha University of Science and Technology, Changsha, China. His areas of interest include chaos-based secure communication and information security analysis.

An Effective Selective Encryption Scheme for H.264 Video Based on Chaotic Qi System

Fei Peng, School of Information Science and Engineering, Hunan University, Changsha, China

Xiao-wen Zhu, School of Information Science and Engineering, Hunan University, Changsha, China

Min Long, College of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha, China

ABSTRACT

With the wide use of H.264 in Internet and wireless network, many concerns have been made to the security of it. Aiming at providing an effective content protection method for H.264 video, a novel selective encryption scheme based on chaotic Qi system is proposed in this paper. The unpredictability of chaotic system is implemented to construct a pseudo-random number generator based on the 3-dimensional chaotic Qi system, and then some key data such as intra-prediction modes, residual coefficients and MVD are encrypted by the generated key stream. Experimental results and analysis show that the proposed scheme can achieve a good encryption result, a low computational complexity, little impact on compression ratio, and good format compatibility. It has a great potential to be applied in some real-time applications.

Keywords: *Chaotic Encryption, H.264, Qi System, Selective Encryption, Video Encryption*

INTRODUCTION

H.264 is a video coding standard jointly developed by ITU and ISO/IEC (JVT, 2005). It can provide a higher compression performance than H.263 and MPEG-4, and save about 30%~50% bitrate meanwhile guarantee the quality of image coding. It plays an important role in digital television broadcasting, real-time video communication, and network streaming media delivery due to its significant compression performance. However, the vulnerability of the network increases the risk of the disclosure of some sensitive content of H.264 video, which

DOI: 10.4018/jdcf.2013040103

may result in serious consequence for the society. For this reason, many concerns have been made on the research of the encryption of H.264 video.

Encryption of video content is proved to be an effective way to protect videos. Due to the bulk data and real-time requirement, traditional encryption algorithms such as DES, AES or RSA cannot be applied to H.264 video directly (Stutz T. & Uhl A.,2012). To maintain the compression performance and avoid the computational overhead, many researchers have paid much attention to selective video encryption, which only some important parts are encrypted (Wu & Kuo, 2001; Ahn, Shim, Jeon, & Choi, 2004). At the same time, format (or syntax) compliance also attracts researchers' attentions because the encrypted video can be decoded correctly with no destroy in syntax, which can achieve many advantages, such as maintaining synchronization and error resiliency. However, most of the existing video encryption methods are proposed for MPEG standard, and cannot achieve a good balance between security and compression performance, which is difficult to meet the requirements of real application (Stutz & Uhl, 2012). herefore, it still needs more works to be done for the protection of the content of H.264 video.

In this paper, an effective selective encryption scheme for H.264 video based on chaotic Qi system is proposed. Firstly, a pseudo-random number generator is designed based on a 3-dimensional chaotic Qi system, and then some key data such as intra-prediction mode, residual coefficients and MVD are encrypted by the generated key stream. Experimental results and analysis show that the proposed scheme can obtain a good balance between security, computational complexity, and compression ratio. At the same time, it can maintain the format-compliance to the H.264 standard decoder.

The remaining content of this paper is organized as follows: the related work is introduced in the second section; the preliminary knowledge about chaos and H.264 is presented in the third section; the selective video encryption scheme is described in the fourth section in details; experiments and analysis are performed in the fifth section; finally, some conclusions are drawn in the last section.

RELATED WORKS

Encryption is a basic mean for the protection of the content of video data. The existed methods can be classified into the following two categories according to the range of the content to be encrypted:

1. Full encryption of video data

Full encryption of video data can achieve the highest security. The video data is simply regarded as a binary sequence and is encrypted as a whole with traditional encryption algorithms, such as DES, AES and RSA. VEA (Video Encryption Algorithm) is proposed to encrypt video data (Qiao & Nahrstedt, 1997). It divides plaintext block into odd-numbered bytes and even-numbered bytes to form two new byte streams called as odd list and even list. One part of cipher-text is achieved by encrypting the odd list with DES, and the other is the XOR results between the even list and the encrypted odd list. The cryptographic complexity is reduced to almost half of the original one. The complexity is further reduced to one fourth of the original one by re-dividing the odd list into two parts (Tosun & Feng, 2001). This kind of encryption methods takes advantage of high security of the traditional cryptography, but it is not format-compliance, and the computational efficiency is low, which cannot be applied in some real-time applications (Stutz & Uhl, 2012):

2. Selective encryption of video data

The characteristics of the video data and the requirements of video compression standard are considered in selective encryption of video data, and only some key data in the video are selected for encryption. An encryption scheme for MPEG-1 named as SECMPEG is proposed (Meyer & Gadegast, 1997). It combines

selective encryption with additional header information to achieve four levels of security with DES and RSA. However, the encrypted video is not compatible with MPEG-1 standard. A scrambling method is proposed for the encryption of video (Tang, 1996). A random list instead of "Zigzag" sequence is implemented to scramble the DCT coefficients. It is suggested to use random key and multiple random tables to improve its security. However, the random permutation list within MPEG compression process will reduce the compression ratio, and unable to resist the known plaintext attacks. After that, an improvement is done (Tosun & Feng, 2000), 64 DCT coefficients are divided into 3 layers according to the frequency band, and only the lower two layers are encrypted. It can obtain a higher compression ratio and an acceptable security, but there is still a significant impact on compression ratio. A new selective and scalable encryption (SSE) method for intra dyadic scalable coding framework based on wavelet/sub-band (DWTSB) for H.264/AVC is put forward (Rukhin & Soto, 2001). It is accomplished by scrambling quantized transform coefficients (QTCs) in all the sub-bands of DWTSB.

An encryption algorithm based on the scrambling of the intra-prediction modes is proposed (Ahn et al., 2004). However, the security is limited because the scrambling space is limited; meanwhile, the use of fixed-length pseudo-random sequence generator may vulnerable to some attacks. After that, some improvements have been made to it (Jiang et al., 2009), but it still cannot provide enough security alone. Video encryption is integrated in the H.264/AVC/SVC syntax (Stutz & Uhl, 2008). It preserves format-compliance and keeps the computational complexity low. However, it is not suitable for perceptual/ transparent encryption for plain H.264/AVC. An improved sign bit encryption of motion vector based on H.264/AVC is proposed (Wang et al., 2012). It can provide a good scrambling effect, and keep format-compliance and the compression ratio unchanged. It is also recommended to encrypt intra prediction modes to further enhance the security.

Several video encryption methods based on statistical models are put forward (Wu & Kuo, 2001). As for Huffman codec, the encryption is carried out by choosing different entropy statistical models under the control of a secret key. Since the number of Huffman tables is limited, it cannot resist brute-force attack and known-plaintext attack. As for QM codec, the encryption is performed by choosing multiple state indices under the control of a secret key. However, it is vulnerable to chosen-plaintext attack and increases the computational cost. Selective encryption (SE) is applied to carefully selected codewords and bin-strings of CAVLC and CABAC of H.264/SVC (Shahid, Chaumont, & Puech, 2011; Asghar, Ghanbari, & Reed, 2012). Owing to no escalation in bitrate and maintaining the full bitstream compliance, the algorithm is well suited for real-time multimedia streaming. However, the limited syntax elements make its security questionable. A video encryption scheme in H,264 compressed domain is presented (Mao, Zhuo, Zhang, & Li, 2012). Only the most significant bits are extracted and encrypted. It optimizes the tradeoff between security and computational complexity, furthermore, and it maintains the format-compliance to the H.264 standard decoder and has no impact on the compression efficiency. However, the using of AES results in a high computational complexity.

Privacy preservation is also a concern in video encryption. The signs of selected transform coefficients and some bits of the code stream are pseudo-randomly flipped to conceal regions of interest (ROIs) based on transform-domain or code-stream-domain scrambling (Dufaux & Ebrahimi, 2008). The private data in ROIs is successfully hided while the scene remains comprehensible. The impacts on coding efficiency and computational complexity are negligible. However, it cannot resist error-concealment attacks. After that, an efficient

MPEG video encryption scheme is presented based on two simple chaotic maps (Shang, Sun, & Cai, 2008). It uses chaotic stream cipher to encrypt fixed length codeword and then shuffles macro-blocks of each frame. However, the macro-block is not secure because it has a high level of visual content. For low-dimensional chaotic systems exist apparent inadequacy, an encryption algorithm which combined the process of video compression with encryption based on hyper-chaos system is proposed (Chen & Zhang, 2012). However, there is still a gap between theory and its practical application.

Although some improvements have been made to video encryption, the existed methods still have limitations such as significant impact on compression ratio, format-compliance, increment of encoding/decoding time and vulnerability in resisting known attacks. To counterstrike these weaknesses, an effective selective video encryption scheme based on chaotic Qi system is proposed in this paper.

PRELIMINARY KNOWLEDGE

Chaos Theory and Qi System

Chaos is generally defined as unpredictable states of motion in a deterministic system, and it is a complicated system, which is widely existed in nature.

There are some characteristics in a chaotic system, such as the sensitivity to the initial conditions, broadband frequency spectrum, difficulty of prediction in a long term, exponential amplification of errors, and local instability versus global stability. These characteristics indicate the possibility of its implementation in the field of cryptography. C.E. Shannon has proposed two basic principles for the design of cryptography: diffusion and confusion (Shannon, 1949). The mixing of chaotic orbit is corresponded to the diffusion of traditional encryption system, and the randomness and the sensitivity to system parameters are corresponded to the confusion. Obviously, the excellent mixing characteristic guarantees the diffusion and the confusion of chaotic encryption systems.

Currently, some encryption algorithms based on chaos have been put forward. The existed method can be classified into chaos-based block cipher (Chen, Mao, & Chui, 2004) and chaos-based stream cipher (Li, Li, Halang, & Chen, 2007). Nevertheless, there still exist some limitations such as finite-precision effect and degradation of dynamic characteristics of chaotic system. How to improve the randomness of digitalized chaotic signal is becoming a research hot spot. Here, a chaotic stream cipher based on a 3-dimensional Qi system is proposed and implemented to protect the content of H.264 video.

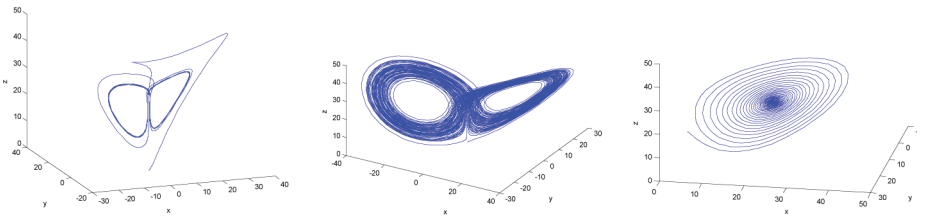
Qi system is modified from Lorenz system (Sparrow, 1982), in which each equation contains a single quadratic cross-product term. According to some non-linear dynamic analysis (Chen, 2003), a cross-product nonlinear term is added to the first equation of the Lorenz system and obtained a new system named as Qi system. Compared with Lorenz system, the new system has some distinct differences such as five equilibria, some larger chaotic regions, and more complex bifurcation behaviors. Notably, each equation has one single cross-product term, and the Qi system is described as follows:

$$\begin{cases} \dot{x} = a(y - x) + yz, \\ \dot{y} = cx - y - xz, \\ \dot{z} = xy - bz. \end{cases} \tag{1}$$

where a, b and c are the control parameters of the system. When a=35, $b \in$(2.25, 11.5] and c=25 hold, the system is chaotic with a positive Lyapunov exponent. In this paper, the value of b is fixed as b=7, and its phase portrait is shown in Figure 1(b).

Actually, there is no clear boundary between the periodic solution and chaos. When $b \in$(0, 2.25], the Equation (1) has a

Figure 1. Phase portraits of Qi system, a=35, c=25. (a) 3D view on the x-y-z space, b=1.5, (b) 3D view on the x-y-z space, b=7 and (c) 3D view on the x-y-z space, b=13.



(a) (b) (c)

periodic orbit, and it is shown in Figure 1(a) when b=1.5 exists. Through calculation, there exist three real equilibria ($S1$=[0,0,0], $S2$=[7.8838,4.6445,24.4109], $S3$=[-7.8838,-4.6445, 24.4109]) and two complex equilibria. When b>11.5 holds, the maximum Lyapunov exponent becomes negative, which implies the orbit of Equation (1) converges to an equilibrium, as shown in Figure 1(c).

Actually, when a=35, b=8/3, and $c \in$ (17, 189] hold, the system is chaotic. For example, when the value of c is 80, the chaotic attractor of Equation (1) is shown in Figure 2.

Bit-Stream Syntax Structure of H.264

In order to obtain a better compression efficiency, H.264 has introduced some new features which make it more effective than previous codec and can be implemented in various network environments (JVT, 2005):

1. Intra-prediction coding

In H.264, the encoded macroblocks are used to predict the pixel values of the current macroblock, and encoding is only performed to the predicted errors. For a 4×4 luminance block, there are 9 IPMs (Intra-prediction Modes) (JVT, 2005). However, it will occupy a large number of bits if each 4×4 prediction mode is transmitted, thus, the correlation of prediction modes between neighboring blocks is implemented for further compression of data.

For example, given encoded 4×4 luminance blocks A and B, they are on the above and on the left of the current block C, respectively. If the prediction mode of A and B are both mode 2, the prediction mode of C is most likely mode 2. Therefore, if the best prediction mode of C is mode 2, the syntactic element *prev_intra4×4_pred_mode_flag* will be set as 1; otherwise, it will be set as 0. After that, the prediction mode of the current block will be the output. After obtaining the best prediction mode, the prediction mode, the difference between the predicted value and the actual value will be encoded by entropy coder. This drastically reduces the bit rate compared to direct encoding of frames.

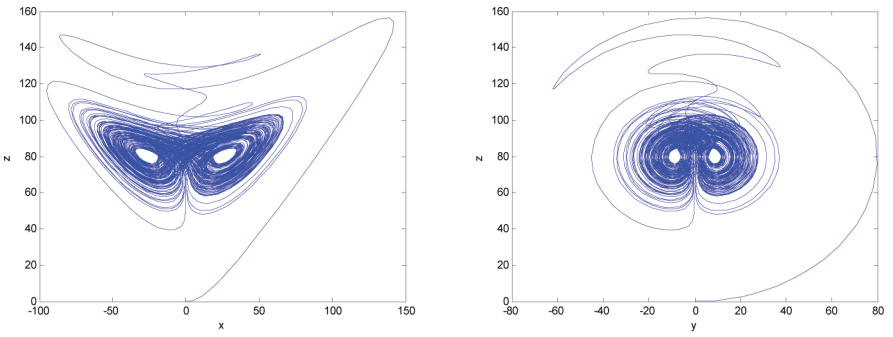
For some flat areas with very little spatial information in an image, H.264 also supports the intra prediction coding for 16×16 luminance block:

2. Inter-prediction coding

Since there exists correlation between consecutive images in video, inter-prediction can be implemented to reduce redundancy on the time by the usage of motion compensation coding. Some new technologies and methods including variable size block motion compensation, one-fourth pixel accuracy motion estimation and multiple referenced frames are used in the inter-prediction in H.264 (JVT, 2005).

During the process of inter prediction in H.264, each macroblock can be split into 4 forms, they are 1 sub-macroblock with a size of 16×16, 2 sub-macroblocks with a size of

Figure 2. Chaotic attractor of Equation (1), a=35, b=8/3, c=80. (a) Projection on x-z plane and (b) projection on y-z plane.



(a) (b)

16×8, 2 sub-macroblocks with a size of 8×16, or 4 sub-macroblocks with a size of 8×8. For a sub-macroblock with a size of 8×8, it can be further split into 1 sub-macroblock with a size of 8×8, 2 sub-macroblocks with a size of 4×8, 2 sub-macroblocks with a size of 8×4, or 4 sub-macroblocks with a size of 4×4. Each sub-macroblock has independent motion compensation. For each motion vector (MV) needs a considerable number of bits for coding, in order to reduce the transmitted bits, the adjacent encoded MV is used to predict the current MV due to the strong correlation between adjacent MVs. The predictive vector is calculated based on MV and MVD (Motion Vector Difference), and they will be encoded and transmitted later.

If the signs or amplitudes of the MVD have been destroyed, errors will be generated in motion estimation and motion compensation in the remaining macroblocks of the coding slice, which will result in a great distortion in the quality of video:

3. Entropy coding

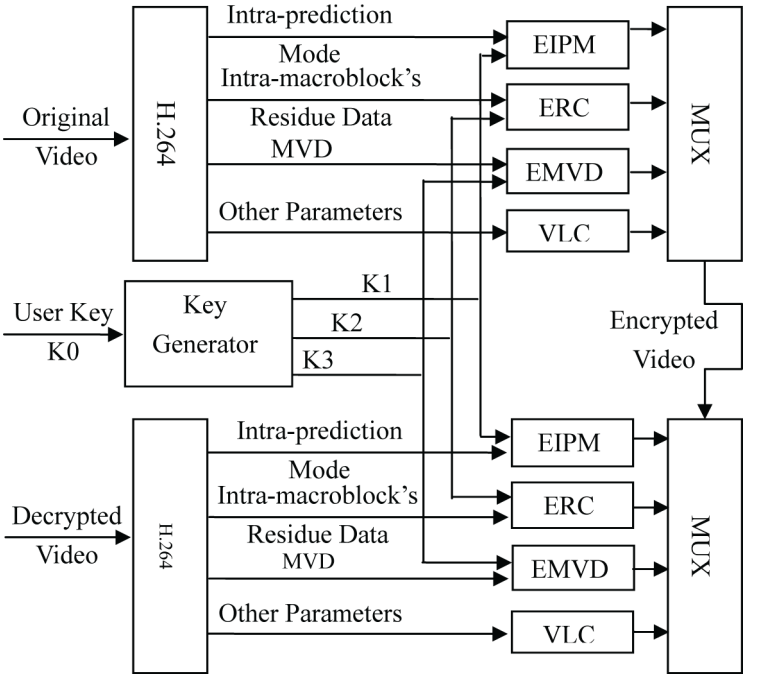
CAVLC is used for the coding of the residual data of luminance and chrominance. It can achieve a very high compression ratio through dynamically updating the tables according to the encoded syntactic elements.

After transformation and quantization, residual coefficients show the following characteristics: the non-zero coefficients are mainly concentrated in the low frequency field while the high frequency coefficients are 0; the non-zero coefficient values near the DC coefficients are large, and the most of non-zero values are ± 1 in the high frequency area after Zigzag scanning; the number of non-zero coefficients of adjacent 4×4 block is related. CAVLC takes full advantage of these characteristics to further reduce the redundant information in the compression process, and forms the basis for enhancement of H.264 encoding efficiency.

THE PROPOSED ENCRYPTION SCHEME

Based on the analysis of the bit-stream syntax structure of H.264, IPM, MVD, and residual data are selected for encryption in this paper. As for intra-prediction mode, EIPM is used to encrypt the code words; as for MVD, its sign and amplitude code words are both encrypted; as for residual data, only the signs of the non-zero coefficients are encrypted in order to reduce the encryption data volume. The decryption is a reverse of the encryption, and the whole process is illustrated in Figure 3:

Figure 3. Diagram of encryption and decryption processes



1. Encryption of IPM (EIPM)

In a video sequence, intra-prediction mode can be easily modified. It represents the prediction information between the pixels in a frame, and it is a basis for the reconstruction of image. Besides, error can be propagated during the intra prediction coding. Assuming intra-prediction is used for the first frame and the inter-prediction mode is used for the followed frames, the first frame will be referenced by the followed frames. Thus, if the first frame is encrypted, the error in the first frame will result in the distortion in the followed frames with inter-prediction.

H.264 provides intra_4×4 and intra_16×16 prediction modes. Actually, intra_4×4 prediction mode is used by most of the video frames, while video frames using intra_16×16 prediction mode are coded with CBP (Code Block Pattern) information, which is difficult to be encrypted. For this reason, only intra_4×4 prediction modes are considered for encryption in this paper.

There are 9 intra_4×4 prediction modes in H.264. When the current IPM is the best mode, a bit "1" is used to represent it, otherwise, 4 bits are used to represent it, where the first bit is "0", the followed 3 bits are used to encode the current intra-prediction mode.

The encryption process for intra_4×4 prediction modes is described in the following:

Step 1: Extract 16 IPMs from a macroblock, and put them into a 4×4 matrix;
Step 2: Arnold transformation is done to the matrix;
Step 3: Assign the transformed IPMs to the original positions of the macroblock;

Step 4: When the current intra-prediction mode is not the best mode, XOR operation is done to the 3 bits represented the IPM and a key stream, which is described in Equation (2).

$$en\_IPM = \begin{cases} IPM \oplus k_1, & flag = 0 \\ IPM, & flag = 1 \end{cases} \quad (2)$$

where $\oplus$ represent XOR operation, k_1 is bit stream with a length of 3 bits, which is acquired from a pseudo-random binary sequence generated from a chaotic system, and *en_IPM* represent the encrypted IPM:

2. Encryption of MVD(EMVD)

In H.264 standard, Exp-Golomb is used to encode the MVD coding parameters in *P* and *B* frames. The codeword can be represented as [*Mzeros*][1][*INFO*], where prefix and information bits are included. If the length of *INFO* is *M* bits, the prefix code is composed of *M* bits of "0" and 1 bit of "1". Its codeword structure with an input from 0 to 8 is listed in Table 1.

From Table 1, given an input *Code_Num*, the length of the information *M* and the informaiton bits *INFO* can be obtained according to:

$$M = floor(log_2(Code\_Num)+1) \quad (3)$$

$$INFO = Code\_Num + 1 - 2^{M} \quad (4)$$

where *floor*(·) represents a floor function. Here, *M* bits of key stream are used to encrypt *INFO*, as seen in Figure 4 and Equation (5):

$$en\_INFO = INFO \oplus k_2 \quad (5)$$

where k_2 is a key stream, which is also acquired from a pseudo-random binary sequence generated from a chaotic system, *INFO* and *en_INFO* represented the information bits before and after encryption, respectively:

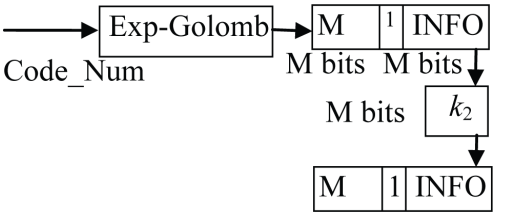
3. Encryption of residual coefficients(ERC)

According to the process of CAVLC entropy coding, the data before CAVLC coding can be represented by *pLevel* and *pRun* after Zigzag scanning is done to luma and hue chroma data in H.264, which is corresponded to the number of non-zero coefficients and the number of zeros before each non-zero coefficient, i.e., *pLevel* and *pRun* (an array with variable length less than 16) represent the positions of the coefficients of a 4×4 block. Thus, it is possible to encrypt the residual data by scrambling *pLevel* and *pRun*.

Table 1. The codeword structure of Exp-Golomb

Code_Num	Codeword Structure
0	1
1	010
2	011
3	00100
4	00101
5	00110
6	00111
7	0001000
8	0001001
…	…

Figure 4. Encryption operation for Exp-Golomb



According to the decoding rules, it is easy to obtain the number of trailing ±1 and non-zero coefficients. If the syntax elements are encrypted, the decoder cannot perform normal decoding operation to all coefficients. In order to guarantee that the encrypted data still follow the H.264 standard and achieve a better encryption effect with fewer encrypted data, the sign of non-zero coefficients is selected for encryption.

The process of the encryption of residual data is described as follows:

Step 1: Arnold transformation is done to *pLevel* and *pRun*, respectively. For most non-zero coefficients in high frequency positions are ±1, in order to reduce the influence of order of energy in Zigzag scanning, the scrambling operation in *pLevel* is not done to trailing coefficients;

Step 2: The signs of non-zero coefficients are encrypted according to:

$$en\_level\_signs = level\_signs \oplus k_3 \quad (6)$$

where k_3 is the key stream generated by the chaotic system, *level_signs* and *en_level_signs* represent the signs of levels before and after encryption, respectively:

4. Generation of Chaotic Sequence

Chaotic sequence is a kind of non-linear and pseudo-random sequence. It is complex and difficult to be analyzed and predicted, which is suitable to be implemented in cryptography. At the same time, chaotic stream cipher is length variable, high computation speed, and limited error propagation, which can be implemented for the encryption of video stream (Chen & Zhang, 2012; Qi, Chen, Du, Chen, & Yuan, 2005).

Since Qi system is in chaos when a=35, b=7 and c=25 hold (Qi, Chen, Du, Chen, & Yuan, 2005), i.e., the sequence $\{(x_n, y_n, z_n) \mid n = 0,1,2,\cdots\}$ generated from Qi system using 4-rank Runge-Kutter method is non-periodic, non-convergent, and sensitive to initial values. For the chaotic sequence is generally acquired values in real, binarization operation needs to be done to it.

Here, based on Qi system, a chaotic key stream generator is constructed as follows:

Step 1: The initial values of Qi are set as x_0=0.01, y_0=0.01 and z_0=0.01;

Step 2: Qi system is iterated $T_1=t_{num}+len/2$ times with 4-rank Runge-Kutter method, where $t_{num}\in[2000, 5000]$, and *len* is the length of the binary sequence. The t_{num}^{th} iteration results are the initial values of the next iteration;

Step 3: Obtain 5 continuous numbers after decimal of z_n, and an integer Z_n is formed from these numbers. Modulo operation is done to this integer, as seen in Equation (7):

$$S_n = 1 + (Z_n \% 10) \quad (7)$$

where:

$$Z_n = floor((|z_n| - floor(|z_n|)) \times 10^5)$$

Step 4: Obtain the S_n th number from x_n and y_n, and extract their least significant bit as key stream, as seen in Equation (8):

$$key = [B(X_0)B(Y_0)B(X_1)B(Y_1) \cdots B(X_{n/2})B(Y_{n/2})] \quad (8)$$

where $B(X)$ represents the least significant bit of X, and:

$$\begin{cases} X_n = floor((|x_n| \times 10^{s_n - 1} - floor(|x_n| \times 10^{s_n - 1})) \times 10) \\ Y_n = floor((|y_n| \times 10^{s_n - 1} - floor(|y_n| \times 10^{s_n - 1})) \times 10) \end{cases}$$

Since it is random in the transformation of chaotic sequence from real to integer, the chaotic sequence generated by this method can guarantee a good randomness. Besides, two bits are generated in each step of iteration, which greatly improve the efficiency of key stream generation.

When the key stream is used to encrypt the video, the distribution of keys is based on a slice. For each slice, 256 bits of key sequence will be generated by the key stream generation system, and the distribution process is synchronized with the encoding of slice and encryption process. These two operations are performed dependently, and the key sequence can be repeatedly distributed until the encryption of the slice is finished.

EXPERIMENTAL RESULTS AND ANALYSIS

Experiments are performed on JM10.2. Baseline profile is used in the coding, and the parameters for the configuration file are: encode 100 frames, the periodic of I-frame is 15, the quantization parameter of *I* and *P* frame is 30, entropy coding method is CAVLC, and QCIF (176×144) sequences are used for experiments. The key stream used in the experiments is generated by the chaotic key stream generator.

Experimental Results

Here, two QCIF sequences including Foreman and Salesman are used in the experiments, and IPMs, MVDs and residual coefficients are selected as encryption/decryption data. The results are shown in Figure 5.

As seen in Figure 5, the proposed method can achieve a good encryption/decryption performance.

Key Spaces Analysis

In order to evaluate the key space, analysis is done to key stream generation method based on chaotic systems. Here, the initial values and t_{num} of the chaotic system are regarded as keys.

Assuming the data precision of the initial value (x_0, y_0, z_0) is double, x_0, y_0, z_0 can acquire values from the range of x, y, z of Qi system, and the value of t_{num} is generally acquired from [2000,5000]. The key space S_k can be calculated as:

$$\begin{aligned} S_k &= Card\{x_0\} \cdot Card\{y_0\} \cdot Card\{z_0\} \cdot Card\{t_{num}\} \\ &= (20 \times 10^{15}) \cdot (30 \times 10^{15}) \cdot (40 \times 10^{15}) \times 10^3 \\ &= 2.4 \times 10^{52} \approx 2^{174} \end{aligned} \quad (9)$$

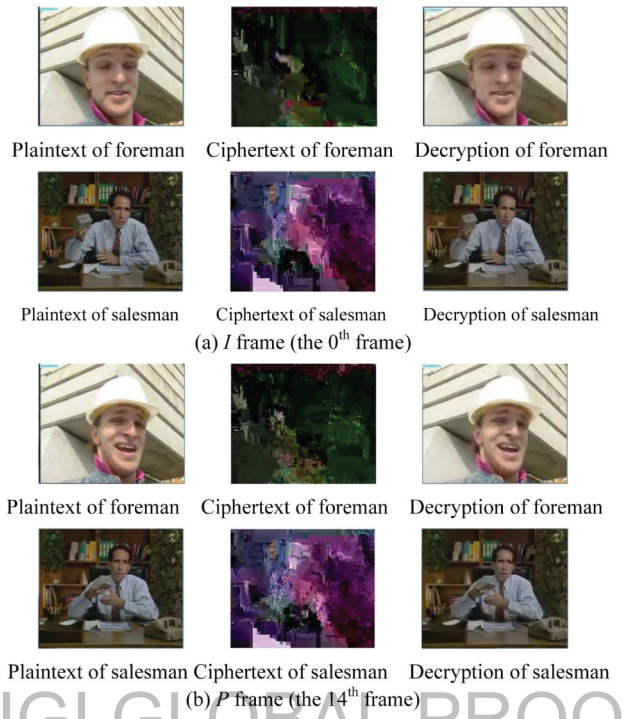
where $Card\{\cdot\}$ represents the cardinality of a set.

As seen in Equation (9), the key space is about 2.4×10^{52}, and it equals to a key with a length of 174 bits, which can resist brute force attacks.

Analysis of the Randomness of the Key Stream

In order to test the randomness of the generated chaotic key stream, NIST (National Institute of Standards and Technology) test suite is used to evaluate 16 statistics (Rukhin & Soto, 2001). Each evaluation result is represented as *P*-value. According to the standard, given a significant level $\alpha \in [0.001, 0.01]$, if all $P\text{-value} \geq \alpha$ holds, it

Figure 5. The encryption and decryption results of foreman and salesman



means that the sequence has passed the test and it is regarded as random, where the confidence is (1-α)×100%; Otherwise, it means that the sequence has not passed the test.

In the test, the initial value for the chaotic system is (0.01,0.01,0.01), the length is 10^6 (it is recommended by NIST that the length of the testing length N for the sequence is between 10^3~10^7), and α is 0.01. The testing results are listed in Table 2.

As seen in Table 2, it can be found that the *P-Value* of each item is greater than α, which indicates the good randomness of the key stream generated by the key stream generator.

Security Analysis

The capability of resisting cryptanalysis is greatly improved by encrypting the intra prediction modes and residual data. Assuming the attacker has guessed out the key stream bits for the encryption of intra prediction modes, he cannot successfully validate it because the scrambling operation confuse the positions of the intra prediction modes, which will result in a failed attack.

At the same time, experiments are also done to analyze the ability of resisting error-concealment based attacks (ECA) (Wen, Severa, & Zeng, 2002). ECA are based on statistical information and knowledge of the video format, and try to conceal the resultant quality degradation by treating unbreakable data as lost and then attempting to minimize the impact on quality as a result of loss. The format compliance of the encryption scheme makes it possible for the attacker to guess the values of some data elements separately in ciphertext-only attacks. As for the encrypted frame of Foreman in Figure 5(a), ECA is done to it. As seen in Figure 6, the IPMs are set as 2 in (a), all DC coefficients are set as 1 in (b), and the hue coefficients are set as 1 in (c). The experimental results are shown in Figure 6. Apparently, the attacker cannot deduce the encrypted data from unencrypted data without

Table 2. NIST randomness test results

Statistical Test	P-Value	State (+,-)
Frequency	0.415645	+
BlockFrequency	0.845546	+
CumulativeSums	*	+
Runs	0.959594	+
LongestRun	0.210799	+
Rank	0.439381	+
FFT	0.897775	+
NonOverlappingTemplate	*	+
OverlappingTemplate	0.807859	+
Universal	0.772788	+
ApproximateEntropy	0.542316	+
RandomExcursions	*	+
RandomExcursionsVariant	*	+
Serial	*	+
LinearComplexity	0.847701	+

Note: + and – represent passed or not, * represents there are more than one value.

Figure 6. The experimental results of ECA

(a)All IPMs are set as 2. (b) All DCs are set as 1. (c) All luma are set as 1

key, so the attacked image still can protect the information in the video, which illustrates the good ability of resisting the ECA.

Analysis of Computation Complexity

Generally, the computation complexity of encryption algorithm is depended on the data volume and encryption operations.

In the proposed scheme, the data for encryption includes IPMs, MVDs and residual coefficients. In I-frame, IPMs and residual coefficients are selected for encryption. As for IPMs, there are about 50% blocks are implemented for encryption when the quantization parameter is 30 and the coding mode is intra_4×4 prediction mode. The data volume for encryption is about $50\% \times W \times H/(4\times4)$, where W and H represent the width and height of the frame in QCIF. As for residual coefficients, only the signs of non-zero coefficients are encrypted, therefore, the data volume is about $W \times H \times 16/(4\times4)$. As for P and B frame, every block with a size of 4×4 has one motion vector, so the data volume of MVD for

encryption is $W\times H/(4\times4)$. Take the encryption of Foreman ($W\times H$=176×144) for example, the encrypted data is at most $W\times H\times16/(4\times4)+W\times H/(4\times4)\approx2.8\times10^4$.

Since the encryption of IPMs, MVDs and residual coefficients are based on XOR operation, it is efficient. So the computation complexity is very low.

At the same time, experiments are also done to evaluate the time efficiency. The results are shown in Table 3 and Table 4, respectively.

As seen in Table 3, comparing with direct compression coding, the time used in the coding of the video with encryption is only have an overhead less than 5%, which indicate the low complexity of the proposed scheme. For the overhead in the decoder, as seen in Table 4, it is also less than 5%, which can meet the needs of real time application.

Analysis of Coding Compression Performance

According to the encryption process, the length of bit stream before and after encryption is the same when the signs of residual coefficients are encrypted. Nevertheless, the scrambling of non-zero coefficients in 4×4 block changes the positions of the coefficients, and it will destroy the statistical characteristics of the non-zero coefficients in the block, which will impose influence to the compression ratio. At the same time, the encryption of IPMs and MVDs both have no influence to the length of bit stream, so the whole influence to the compression performance is very limited.

Here, the changed rate of compression ratio Δr is defined in Equation (10) to evaluate the compression performance as following:

$$\Delta r = \frac{(r_2 - r_1)}{r_1} \times 100\% \tag{10}$$

where r_1, r_2 represents the data volume of the video before and after encryption, respectively.

The results are listed in Table 5.From Table 5, the bit stream has little increase for the video after encryption, and the increased data is mainly due to scrambling of non-zero coefficients in 4×4 block. Nevertheless, since the percentage of the increase of bit stream is very tiny, the proposed scheme still can obtain a good coding compression performance.

Analysis of Operability

In the proposed scheme, there is no change to the format and control information of the video, so the encryption has no negative influence to the error robustness. For the key stream is chaotic

Table 3. Change of encoding time (s)

Sequence	Before Encryption	After Encryption	Overhead(%)
Foreman	206.615	209.535	1.41
Salesman	201.356	204.533	1.58

Table 4. Change of decoding time (s)

Sequence	Before Encryption	After Encryption	Overhead(%)
Foreman	6.178	6.333	2.51
Salesman	5.899	6.045	2.47

Table 5. Change of compression ratio (bit)

Sequence	r_1	r_2	Δr(%)
Foreman	429280	430488	0.28
Salesman	271072	272456	0.51

sequence, the control precision for bit rate can be retained in bit, so the error in the bit stream will not bring error propagation, which also strengthen the error robustness.

For the encrypted bit stream still follow the H.264 standard, there is no influence to the operability of the encrypted bit stream, and it can be normally decoded with the existed standard codec.

CONCLUSION

In this paper, a selective encryption scheme for H.264 video based on chaos is proposed. It combines chaotic encryption and selective encryption, and takes the advantages from chaotic encryption and selective encryption such as good randomness, low computation complexity, real time, and little influence to compression ratio. Experimental results and analysis show that the pseudo-random key steam sequence generated from the chaotic Qi system has good randomness, and the proposed encryption scheme can strike a good balance between computation complexity and security. It provides a novel method to combine chaotic encryption and selective encryption, and has great potential in the protection of the content of H.264 video.

ACKNOWLEDGMENT

This work was supported in part by project supported by National Natural Science Foundation of China (Grant No. 61070195, 61001004), project supported by Hunan Provincial Natural Science Foundation of China (Grant No.12JJA006), project supported by Youth Growth Plan of Hunan University (53107040055), and project supported by the Education Department Foundation of Hunan Province (Grant No.11B002).

REFERENCES

Ahn, J., Shim, H. J., Jeon, B., & Choi, I. (2004). Digital video scrambling method using intra prediction mode. In *Proceedings of the 5th Pacific Rim conference on Advances in Multimedia Information Processing - Volume Part III,* Tokyo, Japan.

Asghar, M. N., Ghanbari, M., & Reed, M. J. (2012). Sufficient encryption with codewords and bin-strings of H.264/SVC. In *Proceedings of 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)* (pp. 443-450).

Chen, G. (2003). Chaotification via feedback control: Theories, methods, and applications. In *Proceedings of 2003 International Conference on Physics and Control* (pp. 468-474).

Chen, G., Mao, Y., & Chui, C. (2004). a symmetric image encryption scheme based on 3d chaotic cat maps. *Chaos, Solitons, and Fractals*, *21*(3), 749–761. doi:10.1016/j.chaos.2003.12.022.

Chen, Q., & Zhang, Z. (2012). New video compression and encryption algorithm based on hyper-chaos. In *Proceedings of 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)* (pp. 1869-1873).

Dufaux, F., & Ebrahimi, T. (2008). Scrambling for privacy protection in video surveillance systems. *IEEE Transactions on Circuits and Systems for Video Technology*, *18*(8), 1168–1174. doi:10.1109/TCSVT.2008.928225.

Jiang, J., Xing, S., & Qi, M. (2009). An intra prediction mode-based video encryption algorithm in H.264. In *Proceedings of International Conference on Multimedia Information Networking and Security (MINES '09)* (pp. 478-482).

Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-G050. (2005). *Draft ITU-T recommendation and final draft international standard of joint video specification* (ITU-T Rec. H.264|ISO/IEC 14496-10 AVC).

Li, P., Li, Z., Halang, W. A., & Chen, G. (2007). A stream cipher based on a spatiotemporal chaotic system. *Chaos, Solitons, and Fractals*, *32*(5), 1867–1876. doi:10.1016/j.chaos.2005.12.021.

Mao, N., Zhuo, L., Zhang, J., & Li, X. (2012). Fast compression domain video encryption scheme for H.264/AVC streaming. In *Proceedings of 14th International Conference on Advanced Communication Technology (ICACT)* (pp. 425-429).

Meyer, J., & Gadegast, F. (1995). *Security mechanisms for multimedia data with the example MPEG-1 video. Project Description of SECMPEG*. Germany: Technical University of Berlin.

Qi, G., Chen, G., Du, S., Chen, Z., & Yuan, Z. (2005). Analysis of a new chaotic system. *Physica A: Statistical Mechanics and its Applications, 352*(2-4), 295-308.

Qiao, L. T., & Nahrstedt, K. (1997). A new algorithm for MPEG video encryption. In *Proceeding of The First International Conference on Imaging Science, Systems, and Technology (CISST' 97)* (pp. 21-29).

Rukhin, A., & Soto, J. (2001). A statistical test suite for random and pseudorandom number generators for cryptographic application. *NIST Special Publication 800-22*.

Shahid, Z., Chaumont, M., & Puech, W. (2009). Selective and scalable encryption of enhancement layers for dyadic scalable H.264/AVC by scrambling of scan patterns. In *Proceedings of 16th IEEE International Conference on Image Processing (ICIP)* (pp. 1273-1276).

Shahid, Z., Chaumont, M., & Puech, W. (2011). Fast protection of H.264/AVC by selective encryption of CAVLC and CABAC for I and P frames. *IEEE Transactions on Circuits and Systems for Video Technology*, *21*(5), 565–576. doi:10.1109/TCSVT.2011.2129090.

Shang, F., Sun, K., & Cai, Y. (2008). An efficient MPEG video encryption scheme based on chaotic cipher. In *Proceedings of Congress on Image and Signal Processing (CISP '08)* (pp. 12-16).

Shannon, C. E. (1949). Communication theory of secrecy systems. *The Bell System Technical Journal*, *28*(10), 656–715.

Sparrow, C. (1982). *The Lorenz equations: Bifurcations, chaos, and strange attractors*. New York, NY: Springer-Verlag. doi:10.1007/978-1-4612-5767-7.

Stutz, T., & Uhl, A. (2008). Format-compliant encryption of H.264/AVC and SVC. In *Proceedings of Tenth IEEE International Symposium on Multimedia* (pp. 446-451).

Stutz, T., & Uhl, A. (2012). A survey of H.264 AVC/SVC encryption. *IEEE Transactions on Circuits and Systems for Video Technology*, *22*(3), 325–339. doi:10.1109/TCSVT.2011.2162290.

Tang, L. (1996). Methods for encrypting and decrypting MPEG video data efficiently. In *Proceedings of the Fourth ACM International Conference on Multimedia,* Boston, MA.

Tosun, A. S., & Feng, W. C. (2000). Efficient multi-layer coding and encryption of MPEG video streams. In *Proceedings of IEEE International Conference on Multimedia and Expo(ICME 2000)* (pp. 119-122).

Tosun, A. S., & Feng, W. C. (2001). Lightweight security mechanisms for wireless video transmission. In *Proceedings of the International Conference on Information Technology: Coding and Computing* (pp. 157-161).

Wang, Y., O'Neill, M., & Kurugollu, F. (2012). The improved sign bit encryption of motion vectors for H.264/AVC. In *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)* (pp. 1752-1756).

Wen, J. T., Severa, M., & Zeng, W. J. (2002). A format-compliant configurable encryption framework for access control of video. *IEEE Transactions on Circuits and Systems for Video Technology*, *12*(6), 545–557. doi:10.1109/TCSVT.2002.800321.

Wu, C. P., & Kuo, C. C. J. (2001). Efficient multimedia encryption via entropy codec design. In Proceedings of SPIE 4314, Security and Watermarking of Multimedia Contents III (pp. 128-138).

E-Behaviour Trends and Patterns among Malaysian Pre-Adolescents and Adolescents

Selvi Salome Gnasigamoney, School of Business Infrastructure, Infrastructure University Kuala Lumpur, Kajang, Malaysia

Manjit Singh Sidhu, Department of Graphics and Multimedia, College of Information Technology, University Tenaga Nasional (UNITEN), Selangor, Malaysia

ABSTRACT

The threat of cyber-related crimes due to excessive usage of Internet and current e-behaviour amongst the younger children is not new in this new millennium but stays as an issue for consideration. This paper provide a general pattern of online related behaviours that seem to be taking place among Malaysian pre-adolescents and adolescents and its possible impact on their behaviours leading towards cyber-related crimes. Facts and finding from various researches conducted from different parts of the world, including Malaysia were reviewed. The results from various studies reveal that a great concern and strategies have to be put into place as the age group using the Internet has reduced and the routine activity of pre-adolescence and adolescence are changing and are based on Internet. Non-awareness of their current online behaviours and its possible link to cyber-related crimes may lead these young children to a greater threat when using e-Commerce or any other Internet dependent activities in the future. This paper focuses on the facts collected from various studies to justify the importance of having future research on this phenomenon.

Keywords: *Cyber Related Crimes, E-Behaviour, Internet Usage, Malaysian Adolescence, Malaysian Pre-Adolescence, Malaysian Youths, Online Behaviours*

INTRODUCTION

The Internet can be considered as a great source of information provider for almost all ages and areas such as science, medicine, engineering, education etc. and a virtual place for people to share ideas, build communities, promote businesses and socialise. Although the usage of Internet had become popular in other parts of the world since its existence, the Internet age in Malaysia began in 1995 (John and Jackie, 2001). The growth in the number of Internet hosts in Malaysia began around 1996. The country's first search engine and web portal company, Cari Internet, was also founded in that year (Sreejit, 2001). According to the first Malaysian Internet survey conducted from October to November 1995 by MIMOS and Beta

DOI: 10.4018/jdcf.2013040104

Interactive Services in (1996), one out of every thousand Malaysians had access to the Internet (20,000 Internet users out of a population of 20 million). In 1998, this number grew to 2.6% of the population. The total number of computer units sold, which was 467,000 in 1998 and 701,000 in 2000 indicate an increasing growth (Lee, 2000). Today, Internet usage or access is not limited via a personal computer with LAN connection but it could be accessed by using a 3G mobile phone.

INTERNET USAGE TREND

Internet usage among individuals such as children, adults and professionals are not a new phenomenon in this new millennium. The world statistics on the Internet usage recorded as of 31st March 2011 compared to 31st December 2000 shows a tremendous increase (Miniwatts Marketing Group (20/12/2011), from 360,985,492 to 2,095,006,005 and latest report as per recorded in June 30, 2012 (refer to Table 1) appears to be at 2,405,518,376. On the whole in June 2012, worldwide Internet usage statistic shows an increase by 14% from March 2011 and 566.38% from December 2000.

As for Malaysia, the Internet usage has increased 41% in 2010 which has increase 15% over the previous year, based on The Nielsen Company's Mobile Insights Survey (Nielsenwire, 2011) wherelse Internet usage statistics in terms of overall population was reported as 56.62% as shown in Figure 1. Economist Intelligence Unit (EIU) estimated that the Internet usage in Malaysia would be up to 62% in 2011 nevertheless the report from the World Bank Group (2012) records at 60% and by 2015, this rate is expected to reach 77%. These percentages are an indication of progressive increase in Internet usage among Malaysian.

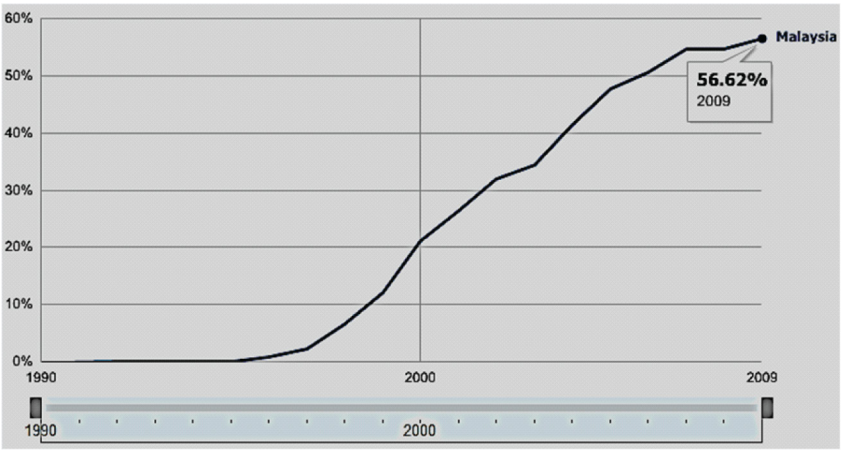
Worldwide Internet usage report which demonstrates an increase in the pattern of Internet use (Internet World Stats, 2012; NAS Recruitment Communications, 2009) is in line with Malaysia's Internet usage as detailed by Digital Media Across Asia, (2010), however the age of the Internet users are reducing in many countries (Australian Communications & Media Authority, 2010; pg. 42) which in due course experiences changes among millennium pre-adolescent and adolescent online behaviour.

This paper is therefore aimed to recognise and draw attention to the age group and behavioural pattern of the pre-adolescent and

Table 1. Internet usage statistics: World internet users and population stats (modified from Miniwatts Marketing Group (17/02/2013)

World Internet Usage & Population Statistics June 30, 2012			
World Regions	**Population (2012 Est.)**	**Internet Users (Dec. 31, 2000)**	**Internet Users Latest Data**
Africa	1,073,380,925	4,514,400	167,335,676
Asia	3,922,066,987	114,304,000	1,076,681,059
Europe	820,918,446	105096093	518,512,109
Middle East	223,608,203	3,284,800	90,000,455
North America	348,280,154	108,096,800	273,785,413
Latin America/Carib.	593,688,638	18,068,919	254,915,745
Oceania/Australia	35,903,569	7,620,480	24,287,919
World Total	7,017,846,922	360,985,492	2,405,518,376

Figure 1. Internet usage as percentage of population (http://www.google.com.my/publicdata/ - Internet 12/30/2012)



adolescent on the Internet use in many parts of the world. Understanding these concerns can support research to be more focused on the trends and patterns of activities among Malaysian pre-adolescents and adolescents on their e-behaviour.

PRE-ADOLESCENCE AND ADOLESCENCE AND INTERNET USAGE

Individual undergoes various stages of life development during their life time (Cherry, 2012). It begins from infancy and may end in very old age. The stages can be differentiated according to the developmental stage and age group (Cherry, 2012; Newman & Newman, 2009).

Pre-adolescence is an age category between 8 – 12 years old, in addition it is an age where a child would develop a life outside home circle (Hall, 2006), adolescence by comparison will be those aged between 12 – 24 years old, as exemplified by Newman and Newman (2009). Adolescent represents, "*the period when a person's identity is formed, and is a common period for experiencing confusion and frustration*." (Chapman, 2006 – 2012; Kim et al., 2005). Both pre-adolescent and adolescent are age categories which are crucial as formations of behavior are taking place.

Internet users are not targeted to just adults aged above 21 years but reducing to young children as well. Most of the young children (pre-adolescent and adolescent) in this new millennium are 'wired' and to a greater extent influence the behavior formation (Ybarra & Mitchell, 2004) and lifestyle.

Philips (2010) a senior analyst for eMarketers.com reported that in America, 12 – 24 years old represents a major component of the Internet users at 51.7 million (23.4% of the total). In addition, Philips (2010) estimated that in 2011 there will be 20.2 million children under 11 years old going online. Further it noted that this number may increase to 24.9 million American kids by 2014 (refer to Table 2).

Likewise in United Kingdom, European Travel Commission (2013) in its website recited the Office for National Statistic in August 2011 and pointed that the largest Internet users for United Kingdom were in the 16 – 24 age group, which is at 98.8% and that it represents 7.19 million people.

Table 2. US internet users by age, 2008 – 2014 (abstracted from Philips, 2010)

	2008	2009	2010	2011	2012	2013	2014
0-11	15.6	17.0	18.6	20.2	21.8	23.3	24.9
12-17	23.3	23.5	23.8	23.9	24.1	24.3	24.5
18-24	26.6	27.3	27.9	28.4	28.7	28.9	29.0
25-34	34.2	35.5	36.5	37.4	38.2	38.9	39.6
35-44	33.7	34.2	34.5	34.9	35.6	36.2	36.7
45-54	32.7	33.9	35.1	36.1	36.9	37.5	37.9
55-64	22.2	23.9	26.0	27.7	29.0	30.4	31.7
65+	14.9	16.3	18.6	20.5	22.6	24.6	26.4
Total	**203.2**	**211.7**	**221.0**	**229.2**	**236.9**	**244.1**	**250.7**

Note: an Internet user is a person who uses the Internet from any location at least once per month
Source: eMarketer, February 2010

Similarly, a survey done in Malaysia confirms the high usage of Internet among the age group between 15 – 19 years old (refer to Table 3). If fact, Household Use of the Internet Survey made between 2005 on a sample size of 4,925 Internet users in private households, identified back then that, "*below 15 categories are the youngest age groups and they already accounted for 42.3% of all users*" (Household Use of the Internet Survey, 2005).

Likewise the latest survey made in 2009 and presented by Koay Hock Eng on behalf of Malaysian Communication and Multimedia Commission in the Asia-Pacific Internet Research Alliance, 7th International Conference, New Delhi, India (2010) displayed a table of figures that has made the significance for research to be conducted on Malaysian pre-adolescent and adolescent age group between 10 to 19 years old (refer to Table 4).

Table 3. Percentage share of household user base on age group (obtained from Household Use of the Internet survey, 2005)

Age category	Percentage
Below 15	6.5%
15 – 19	18.6%
20 – 24	17.2%
25 – 29	12.5%
30 – 34	12.2%
35 – 39	9.9%
40 – 44	9.6%
45 – 49	5.1%
Above 50	8.4%

42.3

Single largest age group, 15 – 19 years has 18.6%

Table 4. Percentage share of household user base on age group (Malaysian Communication and Multimedia Commission, 2010)

Percentage Share of Household User Base on Age Group				
Age Group	**2005**	**2006**	**2008**	**2009**
Below 15	6.5	7.3	6.8	8.1
15 – 19	18.6	18.7	17.9	19.2
20 – 24	17.2	16.3	15.7	14.2
25 – 29	12.5	11.3	11.9	12.9
30 – 34	12.2	12.3	11.7	11.4
35 – 39	9.9	10.4	11.2	9.5
40 – 44	9.6	10.6	9.3	9.4
45 – 49	5.1	6.1	6.1	5.1
Above 50	8.4	7.1	9.4	10.2

Despite the fact that 19 years old age and below are just 27.3% from the overall age group in Malaysia, awareness and emphasis on their online behavior should be researched as these age group would turn to be the adults using the Internet in the future. There have been great concerns raised on age group classified as pre-adolescent and adolescent online behaviour.

PRE-ADOLESCENCE AND ADOLESCENCE E-BEHAVIOUR

In a study conducted by the Observatory from Greek Information Society (2008), 73% of children ages between 10 and 15 appear to be the highest users and in addition, they are well aware of the potential Internet dangers. Further to this, during the first six month, the cybercrime such as child pornography (27%), Spam-Phishing (27%) and Financial Fraud (18%) was recorded (Christodoulaki & Fragopoulou, 2010). Consequently, the Internet usage among pre-adolescent and adolescent have an impact on risky online behaviour (e-Behaviour) leading into cybercrime.

A general characteristic of pre-adolescents and adolescents and the possible risky online behaviors is summarized in Table 5. The table clearly displays how the various bands of pre-adolescents and adolescents may be posed towards great dangers if no strategies and action plans are put forth at the initial stage of their Internet use. This may also indirectly assist in understanding of their possible reasons for e-behavior or online behavior and indirect involvement and victimization towards cyber related crimes. As such, basic research should be in place to identify the possibilities that have been described by Newman and Newman (2009) in Table 5 among Malaysian pre-adolescents and adolescents.

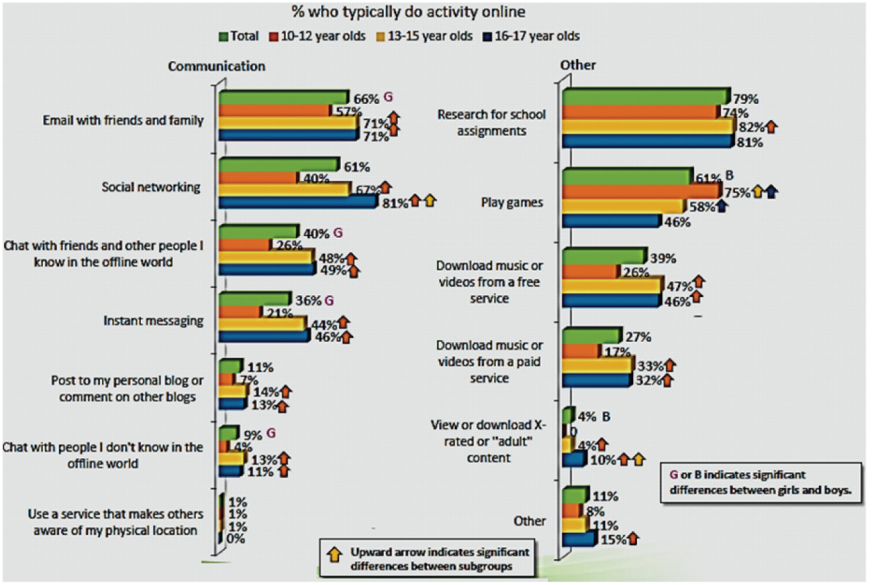
In order to identify some of the e-Behaviours among pre-adolescents and adolescents, this article uses a summary of the research conducted by Pieters and Krupin (2010) on U.S. respondents in the Harris Poll Online (HPOL) opt-in panel (refer to Figure 2). This research was conducted on 1357 sample size from the age range of 10 – 17 years old (606 males, 751 females; 402 pre-adolescent age 10 – 12, 593 adolescent age 13 – 15 and 362 adolescent age 16 – 17).

From the above mentioned research adolescents were identified to be more likely to engage into communicating online (the breakdown can be found in Figure 2). Girls (55.34%) are more likely than boys (44.66%) to be involved in online behaviour surrounding communication.

Table 5. Life stage, age group, development tasks and summary findings from Newman & Newman, 2009

Life Stage	Age Group	Development Tasks	Description/Summary Abstracted Form Newman & Newman (2009)
Middle childhood	6 – 12 years old	Friendship; concrete operations; skill learning; self-evaluating; team play.	• *Children use of Internet to deepen intimacy with friend's special relevance for children who are lonely or who suffer from social anxiety.* • *Bullying including identification of bully – victims, and findings that bullies are often admired by peer even though they are avoided. Clearer link between link between types of rejected children and literature on bullies and victims.* • *Expanded discussion of in-group favoritism and children's increasing ability to modify in-group and out-group attitudes based on social reality and multiple group membership.*
Early adolescence	12 – 18 years old	Physical maturation; Formal operations; Emotional development; Membership in the peer group; romantic & sexual relationship.	• *Research on crowd's affiliation in a Dutch sample highlights four underlying factors associated with crowd membership and their relationship to depression, delinquency & aggression.* • *New discussion of changes in frequency and intensity of parent-adolescence conflict, comparison to White & Black teens, Chinese youth & their parents.*
Later adolescence	18 – 24 years old	Autonomy from parents; gender identity; internalized morality; career choice.	• *Expansions of the idea that new levels of awareness of social injustice and societal inequality influence moral judgments.*

Figure 2. Typical engagement in online activities (adapted from Pieters & Krupin, 2010, pg 24)



A similar finding was found in UK in 2007, whereby a quantitative and qualitative research project was conducted on 1, 1511 children and young people's aged 9 to 19 years old (pre-adolescents and adolescents) on the use of the internet (Livingston & Helsper, 2007). The research output showed that girls do communicate more on the Internet than boys, but it concluded that it is the age that plays an important factor more than gender when it comes to involvement into risky e-behaviour.

Malaysian pre-adolescent and adolescent shows similar patterns of e-Behaviour such as reported by Pieters and Krupin (2010). In Malaysia, based on the survey made in 2009 and presented in 2010 Conference in India by Koay (2010) also show a similar pattern of e-Behaviour such as in Figure 2, among Malaysian pre-adolescents and adolescents. The details presented has been summarised in Table 6.

Some of the e-behaviour listed in Table 6 is associated to risky behaviours identified from Pieters and Krupin (2010) research. One that leads to a risky online behaviour is the downloading of some kind of media online especially downloading online game. Downloading online games are found popular among pre-adolescents compared to adolescents.

Accordingly, a research conducted by Grusser *et al.,* (2007) found that "*gaming has an addictive potential that is also mirrored by addiction-related cognitive components like significantly stronger positive outcome expectancies*" (p. 291). Positive outcome experiences are those experiences that add a reward to a particular activity. These positive outcome experiences may lead to expectations in other areas and lead the gamer to more destructive or addictive behaviour for example playing violent games (ex. Kingpin: Life of Crime, Soldier of Fortune, Unreal Tournament, Mortal Combat 4, and Wild 9) could encourage pre-adolescents and adolescents to become cruel to pets or other animals, get into fights with other children in school or college, react in disappointment, criticism or teasing with extreme and intense anger or use vulgar words, blame others for something or even take revenge. According to psychologist playing violent computer games for just a few minutes is much more harmful than watching violence on television or films (Braid & Craig, 2011). Braid and Craig further added other studies which found that kids who spend more time playing violent video games are more hostile and more likely to argue than other teens, kids who play violent games for less than 10 minutes tend to act aggressively shortly after playing.

In addition another study was conducted on 205 adolescents aged 10 – 14 years revealed that aggression or the external behaviour problems were identified among early adolescents due to Internet communication, the amount spent on online gaming and by playing first-person shooters (Holtz & Appel, 2011). Furthermore, Chen et al (2005) study related to online gaming-related crimes and other social influences in Taiwan discovered that out of 613 cases reported, young offenders aged between 15 and 20 years old makes up the highest statistic (63.3%) and the number of young victims are more which is 209 (34.1%). Thus, future studies should further explore on these young group empirically examining the hour used in playing online games and their external behaviour. As spending longer hour playing online games may have a possibility for pre-adolescents and adolescents to be involved in cyber-related crimes and being a victim.

As similar studies are done in other countries, it is significance to have a study in Malaysia on this area for "*online gaming market in Malaysia has grown constantly with the increase in the number of games, service providers, and gamers. It is a fact that Malaysia's younger generation is more involved in the interactive indoor games rather than playing outdoors*", (Hussein, Wahid, & Saad, 2009).

Besides online communication, downloading and gaming, engaging into other risky e-behaviours are also obvious among today's pre-adolescent and adolescent. The list of risky behaviours is as listed by Pieters and Krupin (2010) and summarized in Table 7.

Table 6. Four main uses of internet by Malaysian (summarised from Malaysian Communications & Multimedia Commission, 2009)

Purpose for Use of the Internet		
Percentage Base on 2663 Samples of Internet Users		
Getting Information		76.9%
Web Surfing	77.2%	
Goods or services	58.9%	
Related to health	34.9%	
From government/public	30.1%	
Real estate	9.7%	
Others	13.8%	
Communication by text		74.8%
Email	97.6%	
Chatting/messenger	66.5%	
Others	0.6%	
Leisure		50.1%
Music	78.9%	
Video or computer games	60.3%	
E-book, magazine, reading newspaper	53.9%	
Downloading a movie, images, etc	52.1%	
Listening to radio/Watching TV	45.4%	
Others	0.9%	
Social Networking/Online Community		46.8%
Facebook	76.1%	
Friendster	60.9%	
Myspace	26.7%	
Tagged	8.7%	
Hi5	6.8%	
Interactive online games	6.2%	
Twitter	4.9%	
Flickr	2.2%	
Linked In	1.5%	
Others	1.4%	

Besides that, Hussain (2011) indicated some of the other common e-Behaviour that seems to be prominent lately is cyber bullying (sending and posting harmful material or engaging in any form of social cruelty using the Internet). Pieters and Krupin (2010) did highlight in their research on the increase involvement of pre-adolescence and adolescence into cyber

Table 7. List of risky e-behaviours

No.	Risky Behaviour as Listed by Pieters & Krupin (2010)	Percentage Base on 1357 Sample Size
1	Allowing their home computer to become infected with a virus or other software	23%
2	Sharing password with friends	13%
3	Downloading programs without their parents knowledge	25%
4	Chatting with people they do not know in offline world	22%
5	Viewing or downloading X-rated content among boys especially 16 – 17 years old.	11% 35%
6	Boys are more likely than girls to have ever downloaded programs without parental knowledge or those with X-rated material.	45%
7	Give out personal information such as first name, age and e-mail address but 1/10 gives out photo of themselves, their school name, last name, cell phone number or the description of what they look like, parents name, home address and school address.	-
8	Girls (25%) especially 16 – 17 year old girls (43%) are more likely than boys to chat with people online they do not know in offline world)	-

bullying. Additionally their findings revealed that 1/10 youth (10%) admit engaging in some form of cyber bullying, kids are more likely to admit involvement into a "cyber-prank" (6%) than sending anonymous e-mails (3%), spreading rumours online (3%) forwarding private information without someone's permission (2%) or posting mean and hurtful information about someone online (2%). It was also recorded that 1/10 (9%) said that they have been approached online by someone they do not know, received a message of a bullying nature or had their password hacked. In fact the practice of computer hacking became prevalent in the 1990s and prosecutions for hacking related crimes increased (Nykodym, Ariss, & Kurtz, 2008). In another report, Melatdoust (2010) in her blog reported that "*cyber-bullying situation in Malaysia is rampant. For instance, 60 cases of cyber bullying were reported to Cyber Security Malaysia as of October 2007*". Cyber bullying is not a new phenomenon. Shukor (2006) noted back in 2006 that "*cyber bullying is common in schools in Klang Valley (Malaysia), and has in fact been around for at least five years*" (pg. 4)

Hussain (2011) added that Cyber stalkers (decent anonymous Internet users who steal identity to lure victim for evil intentions), cyber trafficking (illegitimate lucrative business through recruiting, transporting, transfer or harbouring of recipient of person with or without victims consent or knowledge), pornography (produce and disseminate images of child abuse), use cyber café which has no access restriction over the content of the web pages, chat rooms and other offensive material and online gambling which happens while in search for online games are the many other online behaviours that seem to be popular among today's teenagers. Moreover aggressive behaviours as illustrated above have been reported to be more significant among early adolescents due to their unique psychological and biological character (Ko et al., 2009). Cyber stalking is not new for Malaysians. CyberSecurity Malaysia in their report collected indicates that in 2009 the cases were 174 compared to 70 only in 2008 (The Star, 2010). Thereafter in 2011 the case of cyber stalking increased to 459 cases (Malaysian Computer Emergency, 2011). To

date there is no statistic indication of the age group record in the data presented especially focusing on pre-adolescent and adolescent in school environment. It is therefore an area to consider.

Other findings of e-behaviour that was also recorded through Pieters and Krupin's (2010) research are as pre-adolescence and adolescence get older they tend to hide what they do online (27% are from age 10 – 12, 54% age 13 – 15 and 56% age 16 – 17), clear the browser history (21%) and hide or delete text messages (20% with girls 23% and boys 17%).

PROPOSED STUDIES

In line to the facts and figures on the increase of Internet usage among Malaysian pre-adolescents and adolescents, future studies on the pre-adolescents and adolescents should be off great interest. As no specific quantitative research was identified to collect data from Malaysian pre-adolescents and adolescents on the risky online behaviour has been done especially among the age group between 10 – 19 years, further research will be significant.

As most of the research is related to Internet usage among Malaysian youths (Patrick, 2011; Qin, 2011; Munusamy & Ismail, 2009; Pawanteh & Rahim, 2000), there should be a quantitative profiling of pre-adolescents and adolescents risky e-behaviour, in considering the seriousness of the behaviour before being an adult. Considering Malaysian pre-adolescent and adolescent culture and parenting style are different from the Western culture (Keshavarz & Baharudin, 2009), this contributes further for an in-depth research. Areas such as gender and race may also be significance for this research, as Malaysia consists of various ethnics such as Malays, Chinese, Indians and others.

The output of the quantitative research would be useful for Malaysian government and authorities in understanding the e-Behaviour setting among pre-adolescence and adolescence in general and the urgency of the behaviours and thus making it easier to strategize action plan.

CONCLUSION

This paper has deliberated the current trends and patterns of activities among pre-adolescents and adolescents on Internet usage in general, and specifically focusing on Malaysia. The preliminary results from various studies reveal that the great concern and strategies have to be put into place as the age group in using the Internet has reduced. Children as young as pre-school are now exposed to the Internet. Routine activity of pre-adolescents and adolescents are changing, highlighting Internet as the basis, such as online chatting, online games, online searching and sending and receiving of mails. Some of these e-behaviours are linked closely to cyber-related crimes that may affect the use of e-Commerce or any dependent activities of Internet by organization in the future. Furthermore, the population growth of pre-adolescents and adolescents are increasing and the Internet may pose a very high risk at the later stage of life to this age group as the technology is advancing and more easily accessible globally. The relatively lenient sanctions associated with the complexity of the offences may further worsen the situation if no measures are taken now to curb the problems mentioned in this paper.

As the age of 10 to 19 years classified as both pre-adolescents and adolescents are the heavy user of e-Commerce and other Internet base facilities and services, strategies and plans related directly towards them should be designed. The association of Malaysian pre-adolescence and adolescence being cyber-crime victims are not clear and further work is expected to find out on the seriousness of the risky behaviour by profiling and conducting a quantitative research with appropriate sample size. Accordingly, further research is required to explore on gender's role and age group among Malaysia pre-adolescents and adolescents in encountering communicative online risks and factors that associated to it. Parents as well as other authoritative bodies should play a great role from the beginning of the pre-adolescents and adolescents Internet use. Knowing what is

the actual circumstances faced by Malaysian pre-adolescents and adolescents and the factors leading them can be a great assistance in plotting appropriate action plan by parents and relevant authorities.

ACKNOWLEDGMENT

We would like to acknowledge the considerable contribution of Mr. Ahmad Izham Bin Khairuddin from Cyber Security Malaysia for his valuable insight on the current cyber happening among Malaysian adolescents. Thanks are also due to Dr. Kirandeep Kaur for proof reading this paper.

REFERENCES

Australian Communications and Media Authority. (2010). *Trends in media use by children and young people*. Retrieved from http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=7&ved=0CGQQFjAG&url=http%3A%2F%2Fwww.acma.gov.au%2Fwebwr%2F_assets%2Fmain%2Flib310665%2Ftrends_in_media_use_by_children_and_young_people.doc&ei=buW-T6-uGMnNrQfNkZGwCQ&usg=AFQjCNG2pbmIODWx44hgSuphXqyHseC69g

Braid, J. B., & Craig, A. A. (2011). Violent video games and hostile expectations: A test of the general aggression model. *Personality and Social Psychology Bulletin*, *2011*, 1679–1686.

Chapman, A. (n.d.). Erikson's psychosocial developmental theory. Retrieved from http://www.businessballs.com/erik_erikson_psychosocial_theory.htm

Chen, Y. C., Chen, P. S., Hwang, J. J., Korba, L., Song, R., & Yee, G. (2005). An analysis of online gaming crime characteristics. *Internet Research*, *15*(3), 246–261. doi:10.1108/10662240510602672.

Cherry, K. (2012). *Erikson's theory of psychological development.* Retrieved from http://psychology.about.com/od/psychosocialtheories/a/psychosocial.htm

Christodoulaki, M., & Fragopoulou, P. (2010). SafeLine: Reporting illegal internet content. *Information Management & Computer Security*, *18*(1), 54–65. doi:10.1108/09685221011035269.

Digital Media Across Asia. (2010). *Malaysia internet penetration.* Retrieved from http://comm215.wetpaint.com/page/Malaysia+Internet+Penetration European Travel Commission. (2013). *Usage patterns and demographic.* Retrieved from http://www.newmediatrendwatch.com/markets-by-country/18-uk/148-usage-patterns-and-demographics

Grusser, S., Thalemann, R., & Griffiths, M. (2007). Excessive computer game playing: Evidence for addiction and aggression? *Cyberpsychology & Behavior*, *10*(2), 290–292. doi:10.1089/cpb.2006.9956 PMID:17474848.

Hall, G. S. (2006). Youths: Its education, regimen, and hygiene. *The Echo Library: Middlesex*, 7.

Holtz, P., & Appel, M. (2011). Internet use and video gaming predict problem behavior in early adolescence. *Journal of Adolescence*, *2011*, 34–58. PMID:20303580.

Household Use of the Internet Survey. (2005). *Household use*. Retrieved from http://www.skmm.gov.my/link_file/facts_figures/stats/pdf/Household_use_internet_survey2005.pdf

Hussain, R. (2011). Cyberspace task force for children protection. *International Journal of Academic Research*, *3*(2), 1001–1007.

Hussein, Z., Wahid, N. A., & Saad, N. (2009). Behavioral study on Malaysian game player experiences: How the embedded information inside a computer game affect players' behavior. In *Proceedings of the 9th Global Conference on Business & Economics,* Cambridge University, Cambridge, UK.

Internet World Statistic. (2010). *Malaysia: Internet usage stats and marketing report.* Retrieved from http://www.internetworldstats.com/asia/my.htm.

Internet world statistic, usage and population statistics. (2012). Retrieved from http://www.internetworldstats.com/stats.htm

Keshavarz, S., & Baharudin, R. (2009). Parenting style in a collectivist culture in Malaysia. *European Journal of Soil Science*, *10*(1), 66–73.

Kim, K., Ryn, E., Chon, M. Y., Yuen, E. J., Choi, S. Y., Seo, J. S., & Nam, B. W. (2005). Internet addiction in Korean adolescents and its relation to depression and suicidal ideation: A questionnaire survey. *International Journal of Nursing Studies, 43*(2006), 185–192.

KO. (2009). C. H., Yen, T. Y., Liu, S. C., Huang, C. F., & Yen, C. F. (2009). The Associations between aggressive behaviors and internet addiction and online activities in adolescents. *The Journal of Adolescent Health*, *44*, 598–605. doi:10.1016/j.jadohealth.2008.11.011 PMID:19465325.

Koay, H. C. (2010). Household use of internet survey 2009 Malaysia. In *Proceedings of the 7th International Conference Asia-Pacific Internet Research Alliance*, New Delhi, India.

Kshetri, N. (2009). Positive externality, increasing returns, and the rise in cybercrimes. *Communications of the ACM*, *2009*, 141–144. doi:10.1145/1610252.1610288.

Lee, W. (2000). Internet took the crown as top IT growth in Malaysia. *MalaysiaCnet*. Retrieved from http//www.malaysia.cnet.com/news/2000/04/21/200000421j.html

Livingstone, S., & Helsper, E. (2007). Taking risks when communicating on the internet: The role of offline social-psychological factors in young people's vulnerability to online risks. *Information Communication and Society*, *10*(5), 619–643. doi:10.1080/13691180701657998.

Malaysian Computer Emergency. (2011). *Reported incidents based on general incident classification statistics 2011.* Retrieved from http://www.mycert.org.my/en/services/statistic/mycert/2011/main/detail/795/index.html

Melatdoust, H. (2010). *Cyber-bullying in Malaysia.* Retrieved from http://cyberbullyinglaws.wordpress.com/2010/09/09/cyber-bullying-in-malaysia

Miniwatts Marketing Group. (2011). *Internet usage statistics: World internet users and population stats.* Retrieved from http://www.internetworldstats.com/stats.htm

Miniwatts Marketing Group. (2013). *Internet usage statistics: World internet users and population stats.* Retrieved from http://www.internetworldstats.com/stats.htm

Munusamy, K., & Ismail, M. (2009). Influence of gender role on internet usage pattern at home among academicians. *The Journal of International Social Research, 2/9*. Fall 2009. Retrieved from http://www.sosyalarastirmalar.com/cilt2/sayi9pdf/munusamy_ismail.pdf

NAS Recruitment Communications. (2009). *Internet usage in the United States.* Retrieved from http://www.nasrecruitment.com/docs/white_papers/Internet_Usage_United_States.pdf

Newman, B. M., & Newman, P. R. (2009). *Development through life: A psychological approach* (10th ed.). Balmont.

Nielsenwire. (2011). *Malaysian internet usage takes off in 2010*. Retrieved from http://blog.nielsen.com/nielsenwire/global/malaysian-internet-usage-takes-off-in-2010

Nykodym, N., Ariss, S., & Kurtz, K. (2008). Computer addiction and cyber crime. *Journal of Leadership. Accountability and Ethics*, *2008*, 78–85.

Patrick, C.-H., Wai, S. K., & Arumugam, C. C., Veeri, & Ang, P. H. (2011). Ethnic-based digital divide and internet use amongst Malaysian students. *Akademika*, *81*(1), 93–100.

Pawanteh, L., & Rahim, S. A. (2000). Who me? A cyberteen: Implications of Internet usage on realities and identities of Malaysian adolescents. *Asia Pacific Media Educator*, *9*, 43–58.

Paynter, J., & Lim, J. (2001). Drivers and impediments to e-commerce in Malaysia. *Malaysian Journal of Library and Information Science*, *6*(2), 1–19.

Philips, L. E. (2010). *Younger and younger, more kids are online.* Retrieved from http://www.emarketer.com/Article/Younger-Younger-More-Kids-Online/1008085

Pieters, A., & Krupin, C. (2010). *Youth online behaviour.* Retrieved from http://safekids.com/mcafee_harris.pdf

Qin, Y. S. (2011). *A study of Internet addiction among students of Sekolah Menegah Jenis Kebangsaan Pei Yuan, Kampar*. Retrieved from http://eprints.utar.edu.my/274/1/PY-2011-0802518.pdf

Sreejit, P. (2001). M'sia oldest search engine upbeat in trying times. *ZDNet.* Retrieved from http://www.zdnetasia.com/msia-oldest-search-engine-upbeat-in-trying-times-39001129.htm

The Star. (2010). *Cyberstalking a serious threat.* Retrieved from http://www.cybersecurity.my/en/knowledge_bank/news/2010/main/detail/1853/index.html

World Bank Group. (2012). *Internet users as percentage of population*. Retrieved from http://www.google.com.my/publicdata/explore?ds=d5bncppjof8f9_&met_y=it_net_user_p2&idim=country:MYS&dl=en&hl=en&q=internet%20usage#!ctype=l&strail=false&bcs=d&nselm=h&met_y=it_net_user_p2&scale_y=lin&ind_y=false&rdim=region&idim=country:MYS&ifdim=region&hl=en_US&dl=en&ind=false

Ybarra, M. L., & Mitchell, K. L. (2004). Youth engaging in online harassment: Associations with caregiver-child relationships, Internet use, and personal characteristics. *Journal of Adolescence*, *27*(3), 319–336. doi:10.1016/j.adolescence.2004.03.007 PMID:15159091.

Selvi Salome is currently a lecturer, School of Business Infrastructure. Infrastructure University Kuala Lumpur (IUKL). She received her BBA (Hons) degree from the Universiti Kebangsaan Malaysia, Malaysia in 1996 and Masters in Information Technology from University Kebangsaan Malaysia in 2003. She is currently a registered PhD candidate with University Tenaga Malaysia (UNITEN). Her research interests include Information Technology and psychological impact and causes towards the cyber citizen.

Manjit Singh Sidhu is currently the Head of Graphics and Multimedia, College of Information Technology (COIT), University Tenaga Nasional. He received his BSc (Hons) degree in Computer Science from the University of Wolverhampton, UK in 1997 and Masters in Information Technology from University Putra Malaysia in 2000. He completed his PhD in Computer Science from Universiti Malaya in 2007. He is a Chartered IT Professional Fellow, UK and a member of the British Computer Society. He is a senior member of Institute of Electrical and Electronics Engineers (IEEE), Computer and Communications Society, Malaysian National Computer Confederation (MNCC), and Associate Fellow of the Malaysian Scientific Association (MSA). His research interests include patterns of interactions in multimedia and virtual reality applications, 2-D & 3-D visualization, computer simulations and animation. Anuar Muslimen was a research assistant and now a Tutor attached to the college of foundation studies UNITEN.

BOOK REVIEW

Computer Forensics:
Cybercriminals, Laws, and Evidence

Reviewed by Szde Yu, Department of Criminal Justice, Wichita State University, Wichita, KS, USA

Computer Forensics: Cybercriminals, Laws, and Evidence

Marie-Helen Maras

372 pp.

$93.95

ISBN 978-144-9600-72-3

In her own words, Dr. Marie-Helen Maras wrote this book in an attempt to appeal to the individual who does not have a comprehensive legal or technical background in computer forensics. Indeed, most law enforcement officers and students who are considering a career in computer forensics do not or have not had sufficient training in both computer science and criminal justice. Even those criminologists who claimed specialty in cybercrime usually do not have enough technical knowledge on the application of technology. On the other hand, those who possess computer skills often are not familiar with the legal requirement and the sophistication of evidence integrity. A book that can narrow the gap is much needed in the field of computer forensics.

In totally 13 chapters, the author thoroughly introduces the basics of computer forensics and the cyber environment in which forensic works would be conducted. In chapter 1, the definition and typology of cybercrime are discussed. This is utmost important in that different types of cybercrime may require different skill sets to extract digital evidence, and they also usually generate different types of digital evidence, which entails different legal consideration in the presentation of evidence. In the following chapter, the concept of electronic evidence is introduced and the procedure of computer forensics is also described. The differences between public and private investigations are emphasized, and the rules of evidence within different areas of law are also addressed. In chapter 3, it first explains telecommunications and electronic communications data. Then it briefly but adequately introduces the laws that govern the privacy of personal data. In chapter 4, the author focuses on the legal protection for privacy and how it applies to computers. It discusses searches and seizures of computers

and electronic evidence. Chapter 5 talks about cybercrime statutes and the crimes they cover. The author makes a nice distinction among different types of cybercrime as the law that applies may differ. The crimes covered in the chapter include hacking, website defacement, writing and distributing malicious code, computer intrusions and attacks, cyberterrorism, different types of fraud, intellectual property theft, electronic espionage, cyberharassment, and cyberstalking. In chapter 6, the cyber environment which breeds multiple types of cybercrime (e.g. cyberbullying, identity theft, and online scam) is discussed. Cyberspace can be and should be seen as a social setting that creates or facilitates certain crimes. This is an important aspect that is usually missing in a book that is too technical to address the social environment of crime. Nonetheless, in this book the content also covers technical knowledge. In chapter 7, the tools that can be used to collect evidence are introduced. It also addresses the problems investigators may run into when extracting electronic evidence. Chapter 8 discusses what an investigator should do at a crime scene when digital evidence is involved. The procedure is detailed and the concept of evidentiary integrity is reinforced. Chapter 9 is about how to conduct a corporate investigation. This is very useful because computer forensics indeed is often used in a private domain rather than in criminal justice. This makes the book more appealing to a wider audience. Chapter 10 is about email forensics. Email forensics should be emphasized because email can potentially be used to commit crime or contain crucial evidence. The author craftily uses illustrations to explain the technical concepts involved in email forensics. Chapter 11 is about network forensics. It discusses the purpose of network forensics and the tools that can be useful for such purposes. In this chapter the author avoids bombarding the reader with too many technical terms. Instead, she adeptly talks about the network structure in a way easier to understand for people who lack extensive computer literacy. In chapter 12, mobile phone and PDA investigations are addressed. Considering how prevalent nowadays mobile phones are, it is important to know how computer forensics is applied to portable devices. However the discussion in this chapter is a bit limited as it fails to include some of more advanced and popular devices, such as iPad. The discussion of smartphones is also lacking in how the applications may facilitate crime and hinder forensic works. In chapter 13, what a computer forensics investigator should be prepared to face in a courtroom is also discussed. Evidence is never all about techniques. How it is interpreted and presented is as crucial as how it was collected. This chapter is particularly valuable for practitioners who may possess the skill but lack sufficient knowledge on the legal proceedings.

All in all, as mentioned, this book should be recommended to people who first started studying computer forensics. It is very helpful in establishing a conceptual framework for computer forensics, and it paints a fairly solid idea of what to expect in both the technical and legal aspects. It is finally a book about computer forensics that is suitable for readers who are primarily trained in social sciences but it will also suit technicians well. I particularly appreciate the fact that the author stresses the differences in different types of cybercrime separately in terms of their respective cyber environments, techniques required, evidence generated, and the legislation applied. As in all other fields, there are many subdivisions in computer forensics. You cannot expect to tackle all cybercrimes by knowing just one or two techniques. Understanding the broadness of cybercrime is crucial. In addition to the technical aspects of cybercrime, investigators would be better prepared if they familiarize themselves with the human aspects of cybercrime as well. After all, crime is committed by people. *Computer Forensics: Cybercriminals, Laws, and Evidence* does touch on the "criminal" part, although the main focus is still on the "crime".

Despite my overall appreciation for Dr. Maras' book, I offer some critique. In the book Dr. Maras seems to imply computer forensics is for the investigation of cybercrime. I must point out the fact that in today's society where

technology has been embedded in our everyday life, digital evidence can be applicable to any crime, because electronic devices could be used in any crime that is not normally construed as cybercrime. For example, a murderer could videotape his killing using his cell phone. It is not a cybercrime but the crucial evidence can be stored in the form of digital evidence. Email can be used for communication between criminals. Even though they did not commit crime through email, the email conversation could still serve as evidence. GPS devices also could store important evidence related to a criminal's whereabouts in a case that is not necessarily cybercrime. To think of computer forensics as cybercrime investigation is incomprehensive. Computer forensics ought to be seen as part of forensic science that could be applicable in any crime investigations. Moreover, computer forensics not only generates evidence in a legal sense, but also provides useful leads for investigation. This is to say even if the information extracted from electronic devices is not forensic evidence per se, computer forensics might still be able to offer clues for investigators to look for evidence somewhere else or to profile the suspect. For example, the websites a suspect frequently visits may not prove any crime but such information can help understand the suspect's hobby and interest. If the full utility of computer forensics can be clearly identified, it would be beneficial for the reader to understand the practicality of computer forensics and thus circumvents the misconception that suggests computer forensics is only relevant to cybercrime.

Moreover, in comparison with other books about computer forensics, some weaknesses of this book are noteworthy. Perhaps due to the intent to avoid overwhelming the readers who are not familiar with the operating system of a computer, in this book the discussion on how a computer system would affect the operation of computer forensics is limited. For example, the forensic tools working on a Windows system may not work properly on a mackintosh machine, not to mention other mobile devices. Usually textbooks on computer forensics would stress this but the author only shallowly addresses this in the book. This aspect is deemed important because in my opinion a digital system as to digital investigation is analogous to a physical crime scene in a street crime investigation. The structural environment affects how forensic works can or should proceed effectively. The book may need to emphasize and illustrate this more because it is the fundamentals of computer forensics. In addition, I would recommend more consideration on the "criminal" aspect of computer forensics. As mentioned, in this book the main focus is still on "crime". This is a norm in most books related to computer forensics. However, since Dr. Maras is a criminologist and a former investigator, it is a shame she does not integrate the human element into the discussion of digital evidence more extensively. For instance, how have modern technologies changed a criminal's modus operandi? What are the criminological theories that are most applicable to cybercriminals? Discussions in these aspects would engender a more comprehensive perspective. These aspects are generally lacking in books about computer forensics, but I think they would be especially suitable for this book, given the fact that the book title does emphasize cybercriminals. Finally, another weakness of this book is the failure to take into consideration the transnational nature of digital evidence when discussing the legislation governing computer forensics. With the increasing popularity of cloud storage, much digital evidence is not locally stored. Especially when it involves a foreign nationality, what are the legal and technical implications?

Granted, it is impossible to cover everything in a book and the speed of book writing can never catch up with the advancement of new technologies. Nonetheless, these critiques merely serve as suggestions. It seems the author intends to distinguish her book from other books on computer forensics by simplifying the computer science domain in the discussion. Although this might have been accomplished for good reason, I do believe introducing more of a criminological angle in supplement to the

deficiency of computer science would better distinguish this book, considering the author's backgrounds.

In conclusion, I would recommend this book to anyone, including practitioners, academics, and students, who is interested in computer forensics. It addresses not only the technique, but also the law and the criminal background. Dr. Maras' writing style renders great readability and the organization of the book is superb. It can serve as a good textbook for it provides practical exercise, critical thinking questions, and review questions at the end of each chapter, although they are not consistently offered for every chapter. Nonetheless, students would have sufficient knowledge to learn about and some insightful questions for research and brainstorming. For those who are seeking more advanced knowledge and detailed techniques on computer forensics, however, there might be more suitable books available.

REFERENCES

Maras, M. (2012). *Computer forensics: Cybercriminals, laws, and evidence*. Sudbury, MA: Jones & Bartlett Learning.

Szde Yu is a former military investigator from Taiwan in charge of computer security and criminal investigation. He is currently an Assistant Professor of Criminal Justice in the School of Community Affairs at Wichita State University in USA. He used to serve in the College Information Security Oversight Committee at the state university of New York. With educational background in both computer science and criminology, Dr. Yu's research interests involve cybercrime, computer forensics, cyber-psychology, and criminal profiling. He has published journal articles on subjects such as email forensics, digital piracy, and cyber-profiling. In research, he particularly appreciates interdisciplinary collaboration. In teaching, Dr. Yu has taught terrorism, research methods, criminological theory, police problems, and criminal investigations.

International Journal of Digital Crime and Forensics

An official publication of the Information Resources Management Association

Mission

The mission of the **International Journal of Digital Crime and Forensics (IJDCF)** is to provide and foster a forum for advancing research and development of the theory and practice of digital crime prevention and forensics. IJDCF addresses a broad range of digital crimes and forensic disciplines that use electronic devices and software for crime prevention and investigation. This journal informs a broad cross-sectional and multi-disciplinary readership ranging from the academic and professional research communities, to industry consultants and practitioners. IJDCF publishes a balanced mix of high quality theoretical and empirical research articles, case studies, book reviews, tutorials, and editorials.

Subscription Information

IJDCF is published Quarterly: January-March; April-June; July-September; October-December by IGI Global. Full subscription information may be found at www.igi-global.com/IJDCF The journal is available in print and electronic formats.

Institutions may also purchase a site license providing access to the full IGI Global journal collection featuring more than 100 topical journals in information/computer science and technology applied to business & public administration, engineering, education, medical & healthcare, and social science. For information visit www.igi-global.com/isj or contact IGI at eresources@igi-global.com.

Copyright

Correspondence and questions:

Editorial: Chang-Tsun Li
Editor-in-Chief
IJCDF
ijdcf@dcs.warwick.ac.uk

Subscriber Info: IGI Global
Customer Service
701 E. Chocolate Avenue
Hershey PA 17033-1240, USA
Tel: 717/533-8845 x100
E-Mail: cust@igi-global.com

The *International Journal of Digital Crime and Forensics* is currently listed or indexed in: Applied Social Sciences Index & Abstracts (ASSIA); Bacon's Media Directory; Cabell's Directories; Compendex (Elsevier Engineering Index); DBLP; GetCited; Google Scholar; INSPEC; JournalTOCs; Library & Information Science Abstracts (LISA); MediaFinder; Norwegian Social Science Data Services (NSD); SCOPUS; The Index of Information Systems Journals; The Standard Periodical Directory; Ulrich's Periodicals Directory