View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Xiamen University Institutional Repository

学校编码：10384　　　　　　　　　　　　　　分类号_____密级
学　　号：23120080150534　　　　　　　　　UDC

厦门大学

博 士 学 位 论 文

电子取证关键技术研究及在云计算平台上的应用

Key Technologies of Digital Forensic and its Application on Cloud Computing Platform

吴鸿伟

指导教师姓名：郭东辉　教授
专　业 名 称：
论文提交日期：2013 年　月
论文答辩时间：2013 年　月
学位授予日期：2013 年　月

答辩委员会主席：
评　　阅　　人：

2013 年　月

厦门大学博硕士论文摘要库

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（　　　　　　　　　　　　　　）课题（组）的研究成果，获得（　　　　　　　　　）课题（组）经费或实验室的资助，在（　　　　　　　　　）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

年　　月　　日

厦门大学博硕士论文摘要库

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（　　　）1.经厦门大学保密委员会审查核定的保密学位论文，于　　　年　　月　　日解密，解密后适用上述授权。

（　　　）2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年　　月　　日

厦门大学博硕士论文摘要库

厦门大学博硕士论文摘要库

摘　要

电子数据取证在各类违法犯罪的电子证据获取中得到广泛应用，并于 2012 年被全国人大列为新的法定证据类型，从法律层面确认了电子证据的效力。因此对电子数据取证方法的研究具有很好的实际意义和应用价值。

电子数据取证需对电子数据进行分析，以提取相关证据及线索，关键词匹配、加密数据分析和算法重构是其中的关键技术。通过关键词匹配能快速检索定位出所需要的电子数据；而重要的数据经常被加密保护，对该类型的数据进行解密分析，才能提取重要线索；电子数据的多样化，增大了取证数据的复杂度，对不同类型的数据取证进行重构支持，能有效提升取证的效率。为了全面提升电子数据取证的效率，本文进行了电子数据取证关键技术研究，本文的创新点主要体现在：

（一）提出基于 GPU 的关键词加速匹配算法，解决了电子数据分析匹配速度和准确性的平衡问题。给出一种基于多 GPU 的正则匹配引擎，并采用折半分组优化算法解决了有限状态机在大规模正则集合情况下的空间爆炸问题，提升匹配效率及准确性。

（二）提出基于 GPU 的加密数据分析算法，解决了现有加密数据分析的复杂度问题。给出了一种基于字符串的复杂密码快速遍历算法克服了特殊口令搜索的遍历问题，并设计出基于 GPU 的通用解密框架，降低了加密数据分析的复杂度。

（三）提出基于 GPU 的解密算法可重构方法，解决电子数据取证数据的多样化问题。针对基本解密模块的内部算法给出了模块可重构的设计方案，根据解密系统的可并行计算特点设计系统调用可重构算法，提升解密系统的工作效率。

最后，基于本论文的理论和方法开发了一套高速的电子数据分析系统，以处理海量的电子数据，并在厦门超算中心的取证云平台上集成应用。

关键词：　电子数据取证；关键词匹配；数据解密；可重构计算；云计算；

厦门大学博硕士论文摘要库

厦门大学博硕士论文摘要库

ABSTRACT

Recently, electronic evidence has been related with not only computer crimes, but also other kinds of crime. And it is the first time being regarded as a kind of evidence in law since the 11th National People's Congress in 2012.Therefore, it is necessary for us to do some research on digital forensic.

In order to obtain evidences from electronic data, it is necessary to make an analysis on these data. And there are three key technologies: keyword matching, data decryption and reconfigurable computing. Keyword matching makes it possible for us to get the right data that needed. Unfortunately, these data are always encrypted into ciphertext by password. We should try to decrypt these ciphertext before you get the plaintext. Besides, the variety for electronic data makes it more difficult to analysis. Thus, we have to do some research on reconfigurable computation. There are several contributions in this thesis:

(1) A regex approach based on GPUs is presented to meet with the request of both speed and accuracy during keyword matching. In this approach, we design a regex engine based on multi-gpus, and use the binary algorithm to overcome the space over-expanded problem with DFA in huge matching mode.

(2) An encryption data analysis approach based on GPU is presented to overcome traditional analysis methods of encrypted data. In this approach, we design a complicated password generating algorithm based on string token, to generate special passwords. Moreover,we design a common data decryption architecture based on GPU. This makes it easier to analyse encrypted data.

(3) An decryption algorithm reconfigurable approach based on GPU is presented to deal with the problem of data variety during digital forensic. In this approach, we design a reconfigurable architecture for decryption algorithm in module level. We also design the other reconfigurable architecture for the system Scheduling level. Through the two reconfigurable architectures, we improve the efficience of the whole decryption system dramatically.

Finally, we develop a high-speed electronic data analysis system based on the theories and approaches proposed on this thesis, in order to process huge of electronic data. And this system has been integrated into the cloud of digital forensic in Xiamen Supercomputing Center.

厦门大学博硕士论文摘要库

Key Words: Digital Forensic; Keyword Match; Data Decryption; Reconfigurable Computing; Cloud computing;

厦门大学博硕士论文摘要库

目 录

第 1 章.引言 …… 1

1.1.研究背景和意义 …… 1

1.2.关键技术及其研究进展 …… 2

1.2.1.关键词匹配技术 …… 2

1.2.2.数据解密分析技术 …… 4

1.2.3.算法可重构技术 …… 5

1.3.关键问题及主要研究内容 …… 7

1.4.论文的章节安排 …… 8

第 2 章.电子数据取证技术基础与算法原理 …… 10

2.1.电子数据取证技术相关概念 …… 10

2.1.1.电子数据取证相关定义 …… 10

2.1.2.电子数据取证的发展 …… 11

2.1.3.电子证据的数据类型及特性 …… 12

2.1.4.电子数据取证的流程 …… 13

2.2. GPU 架构及编程原理 …… 14

2.2.1. GPU 的基本架构与计算原理 …… 14

2.2.2.基于 CUDA 的 GPU 编程原理 …… 15

2.3.正则匹配算法 …… 17

2.3.1.正则表达式定义 …… 17

2.3.2.正则表达式匹配算法 …… 18

2.3.3.经典正则优化算法 …… 24

2.4.数据解密算法 …… 27

2.4.1.散列算法及其应用 …… 28

2.4.2.密钥加密算法及其应用 …… 30

2.4.3.解密分析方法 …… 32

2.5.可重构计算技术 …… 32

2.5.1.可重构计算简介 …… 32

2.5.2.可重构计算分类 …… 33

厦门大学博硕士论文摘要库

2.5.3.数据解密可重构技术······35
2.6.本章小结······35
第 3 章.关键词匹配技术及算法设计······36
3.1.正则匹配算法分析······36
3.1.1. NFA 和 DFA 对比分析······36
3.1.2.正则的 DFA 状态爆炸问题······37
3.1.3.正则表达式分组优化算法······38
3.2.基于多 GPU 的匹配引擎设计······41
3.2.1.正则匹配算法并行化分析······41
3.2.2.并行匹配引擎架构设计······41
3.2.3.任务均分算法设计······43
3.2.4. CPU 和 GPU 的并行优化······45
3.3.实验及结果对比分析······46
3.3.1. GPU 与 CPU 匹配性能对比分析······46
3.3.2. GPU 与其它硬件平台的匹配性能对比······48
3.3.3.折半分组优化算法对 DFA 状态数的影响分析······49
3.4.本章小结······49
第 4 章.数据解密分析技术及算法设计······51
4.1.解密算法框架设计······51
4.1.1.口令遍历模块设计······51
4.1.2.哈希类通用解密框架设计······52
4.1.3.文档类通用解密框架设计······54
4.2. MD5 解密算法分析······56
4.2.1.算法设计及优化······56
4.2.2.解密流程设计及实验分析······57
4.3. Office 加密文件解密算法优化分析······59
4.3.1. Office 加密原理分析······59
4.3.2. office 安全机制分析及推导······60
4.3.3.解密流程设计及实验对比分析······62

厦门大学博硕士论文摘要库

4.4. 7z 加密文件解密算法优化分析 ······ 65

4.4.1. 7z 加密原理分析 ······ 65

4.4.2. 7z 安全机制分析及推导 ······ 68

4.4.3.解密流程设计及实验对比分析 ······ 71

4.5.本章小结 ······ 73

第 5 章.算法可重构技术分析 ······ 75

5.1.设计方法 ······ 75

5.1.1.数据划分 ······ 75

5.1.2.粒度选择 ······ 76

5.1.3.平台设计 ······ 77

5.2.模块级可重构设计 ······ 78

5.2.1.基本单元划分 ······ 78

5.2.2.接口设计及配置功能 ······ 78

5.2.3.单元组合 ······ 79

5.2.4.案例设计及分析 ······ 80

5.3.系统级可重构设计 ······ 85

5.3.1.基本模块划分 ······ 85

5.3.2.接口设计及配置功能 ······ 85

5.3.3.系统调度设计 ······ 86

5.3.4.案例设计及分析 ······ 87

5.4.实验对比分析 ······ 92

5.5.本章小结 ······ 93

第 6 章.电子数据取证关键技术的运用 ······ 94

6.1.取证云技术简介 ······ 94

6.2.取证云平台的总体架构 ······ 97

6.3.关键技术的应用 ······ 99

6.3.1.正则匹配过滤器设计及实现 ······ 99

6.3.2.数据解密系统设计及实现 ······ 101

6.3.3.关键技术的集成应用 ······ 104

厦门大学博硕士论文摘要库

6.4.本章小结 ········ 112

第 7 章.总结与展望 ········ 113

7.1.工作总结 ········ 113

7.2.工作展望 ········ 114

参考文献 ········ 116

读博士期间发表的论文及专利 ········ 124

读博士期间参与的课题项目及获奖情况 ········ 126

致　　谢 ········ 128

厦门大学博硕士论文摘要库

CONTENTS

1. Introduction... 1

1.1 Movitation... 1

1.2. Key problems and their research & development.. 2

1.2.1. Keyword matching techonology.. 2

1.2.2. Data decryption and analysis techonology 4

1.2.3. Algorithm reconfigurable computing techonology.................................... 5

1.3.Key problems and our main works 7

1.4. Arrangement of the thesis 8

2. Base conceptions and algorithm principles of digital forensis 10

2.1.Related conceptions 10

2.1.1.Definitions of digital forensis 10

2.1.2. The development of digital forensis.. 11

2.1.3. Data types and characteristics.. 12

2.1.4. The procedure of digital forensis 13

2.2.GPU architecture and program 14

2.2.1. GPU architecture and how it works 14

2.2.2. GPU programming based on CUDA 15

2.3.Algorithms of regex match.. 17

2.3.1.Definitions of regex match... 17

2.3.2.The matching algorithms of regular expression...................................... 18

2.3.3.Classical regex match algorithms 24

2.4.Data decryption algorithms 27

2.4.1.Hash algorithms and its applications 28

2.4.2. Data encryptions and its applications 30

2.4.3. The analysis methods of data decryption.. 32

2.5.Reconfigurable computing technology... 32

2.5.1.Instruction 32

2.5.2. Classify of reconfigurable computing.. 33

2.5.3. Reconfigurable computing of data decryption.. 35

2.6. Sections conclusion.. 35

3. Keyword matching technology and algorithm design....................... 36

厦门大学博硕士论文摘要库

3.1. Regex matching algorithms..36

3.1.1. Comparison between NFA and DFA..36

3.1.2.DFA status explosion of multi-regexes...37

3.1.3.Binary algorithm of regex..38

3.2. The engine of keyword matching based on multi-gpus..................................41

3.2.1. How to parallelize for regex algorithm..41

3.2.2. The architecture design of keyword matching engine..................................43

3.2.3. Task division design..45

3.3. Experiment and result analysis..46

3.3.1. Matching performance comparison between GPU and CPU.........................46

3.3.2. Matching performance comparison between GPU and other platforms.........48

3.3.3. How the binary algorithm affects the status count for DFA.........................49

3.4. Section conclusion..49

4. Data decryption technology and algorithm design.........................51

4.1. Designs of decryption framework...51

4.1.1. Design of password exhaustion module..51

4.1.2. Design of common framework for hash decryption.....................................52

4.1.3. Design of common framework for document decryption..............................54

4.2. MD5 analysis and reverse..56

4.2.1. Algorithm design and optimizition...56

4.2.2. Reverse design and experiment analysis...57

4.3. Office document analysis and reverse...59

4.3.1.The analysis of encryption algorithms in Office documents..........................59

4.3.2.The analysis of office document security..60

4.3.3. Reverse design and experiment analysis...62

4.4.7z document analysis and reverse..65

4.4.1. The analysis of encryption algorithms in 7z documents...............................65

4.4.2. The analysis of office document security..68

4.4.3. Reverse design and experiment analysis...71

4.5.Section conclusion...73

5. Algorithm reconfigurable computing analysis................................75

5.1.Design analysis..75

5.1.1.Data division...75

厦门大学博硕士论文摘要库

5.1.2. Granularity selection 76

5.1.3.Platform design 77

5.2. Reconfigurable computing design in module level 78

5.2.1. Base units division 78

5.2.2. Interface design and how to configure 78

5.2.3. Units combination 79

5.2.4. Case design and analysis 80

5.3.Reconfigurable computing design in system level 85

5.3.1. Base module division 85

5.3.2. Interface design and how to configure 85

5.3.3. The design of system schedule 86

5.3.4. Case design and analysis 87

5.4. Experiment and result analysis 92

5.5. Section conclusion 93

6.An application of key technologies in digital forensis 94

6.1. Introduction of the cloud of digital forensis 94

6.2.The whole architecture 97

6.3. Applications of key technologies 99

6.3.1. The design and implementation of regex keyword matching 99

6.3.2. The design and implementation of data decryption system 101

6.3.3. Key technologies Integrated into the cloud of digital forensis 104

6.4. Section conclusion 112

7. Summary and future work 113

7.1. Summary 113

7.2. Future work 114

References 116

Published papers & patents list 124

Participated projects & awards list 126

Acknowledgement 128

厦门大学博硕士论文摘要库

厦门大学博硕士论文摘要库

Degree papers are in the “Xiamen University Electronic Theses and Dissertations Database”. Full texts are available in the following ways:
1. If your library is a CALIS member libraries, please log on http://etd.calis.edu.cn/ and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库