

国家高技术研究发展计划 (863计划)

课题编号: 2003AA1Z2330

计算机审计数据采集与处理技术 研究报告

国家863计划审计署课题组 著



清华大学出版社

计算机审计数据采集与处理技术 研究报告

国家863计划审计署课题组 著

江苏工业学院图书馆
藏书章

清华大学出版社
北京

内 容 简 介

我国改革开放以来的经济发展和科技进步,推进了国家审计的信息化步伐。金审工程的启动,使计算机审计从现场审计方式发展到联网审计方式。本研究报告是对联网审计中的数据采集、数据转换、数据存储、数据处理等若干关键技术创新研究成果的综合阐述,将为金审工程联网审计系统的建设和应用提供理论和技术支持。

本研究报告可供国家审计、内部审计、社会审计行业的领导干部,从事计算机联网审计的业务和技术人员参阅,也可供关心审计信息化的专家、学者和 IT 人士参阅。

版权所有,翻印必究。举报电话:010-62782989 13501256678 13801310933

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

本书防伪标签采用特殊防伪技术,用户可通过在图案表面涂抹清水,图案消失,水干后图案复现;或将表面膜揭下,放在白纸上用彩笔涂抹,图案在白纸上再现的方法识别真伪。

图书在版编目(CIP)数据

计算机审计数据采集与处理技术研究报告/国家 863 计划审计署课题组著.

—北京:清华大学出版社,2006.7

ISBN 7-302-13118-X

I. 计… II. 国… III. ①计算机应用—审计—数据采集—研究报告
②计算机应用—审计—数据处理—研究报告 IV. F239.1

中国版本图书馆 CIP 数据核字(2006)第 055605 号

出 版 者:清华大学出版社 地 址:北京清华大学学研大厦

<http://www.tup.com.cn> 邮 编:100084

社 总 机:010-62770175 客户服务:010-62776969

组稿编辑:王 青

文稿编辑:陆涸晨

印 刷 者:清华大学印刷厂

装 订 者:北京鑫海金澳胶印有限公司

发 行 者:新华书店总店北京发行所

开 本:185×260 印张:29 插页:3 字数:664 千字

版 次:2006 年 7 月第 1 版 2006 年 7 月第 1 次印刷

书 号:ISBN 7-302-13118-X/F·1543

印 数:1~3000

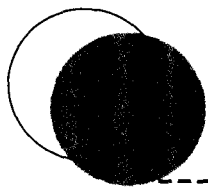
定 价:59.00 元

自主創新探索
計算機網路會計
的技術和方法。

丙戌年春

李宏義





前 言

PREFACE

20 世纪 90 年代以来,随着民主法制的快速推进,改革开放的不断深化,经济持续的快速发展,社会主义市场经济体制的初步建立,信息技术的广泛应用和逐步普及,国家管理社会的手段和方式发生着深刻的变革。这就使作为国家管理社会重要方面的审计机关,不仅要全面履行法定职责,而且必须适应信息化的严峻挑战。1998 年,审计署党组做出了建设审计信息化系统,即金审工程的战略决策。提出了审计信息化建设的总体目标,即经过若干年的努力,逐步建立起与国家信息化建设相配套的、在信息化环境下对国家财政财务收支的真实、合法和效益状况实行“预算跟踪+联网核查”的审计模式,逐步实现审计工作网络化、数字化、智能化,使审计工作从单一的事后审计转变为事后审计与事中审计相结合、从单一的静态审计转变为静态审计与动态审计相结合、从单一的现场审计转变为现场审计与联网审计相结合,全面履行法定审计职责。

为了实现上述目标,2002 年正式启动的金审工程一期建设项目,重点实施了以现场审计和审计管理为主的应用系统、网络系统、安全系统等内容的基础性建设,开展了联网审计的应用试点,并取得了显著成效。但实现审计工作网络化、数字化、智能化的关键是网络化,即网络环境下的联网审计。而实施联网审计必须对传统的审计方式、手段进行革命性变革,创新审计理念、模式、制度、方式与技术,例如,对反映财政财务收支真实、合法和效益状况的电子数据和信息系统的安全、可靠、经济的各项数据,如何采集、分析、处理、利用等,都需要从理论上搞清,并提出具有指导意义的操作规程、技术方法。为此,在科技部的大力支持下,2003 年审计署成立了“计算机审计数据采集与处理技术”课题组,并申请列入了国家“863 计划”。本课题的根本目的是总结金审工程一期建设尤其是联网审计方式的实践经验,从理论上提炼联网审计的若干理念、标准、模式等,为金审工程二期项目建设提供理论基础和技术支持。

课题研究采取了在总结以往实践经验的基础上,重点进行理论提升和技术创新并重的研究方法。一是对金审工程一期项目建设的经验教训、成果做法,特别是金融、海关等联网审计试点所取得的审计技术



和方法、审计组织和方式、审计管理和规范等,进行了认真的理论总结和技术提炼;二是利用国家“863计划”高层次的研究领域,依靠国家“863”专家的技术指导,充分利用国家“863计划”的有关成果,包括Linux操作系统、海量数据存储系统等,提升联网审计的技术含量;三是紧密结合金审工程二期项目重点建设联网审计的实际需要,实施各项科技成果和研究技术的应用,在中央部门预算执行审计中,对联网审计开展了较为系统和具有创新的示范性运用。

本课题对计算机联网审计必须解决和明确的下述关键问题进行了深入研究。

- 联网审计的定义。研究认为,联网审计是审计机关对国家财政和需要进行经常性审计监督且关系国计民生的重要行业或单位,在网络环境下实施的以“预算跟踪+联网核查”为主要特点的审计。
- 联网审计的组网模式。研究提出了“集中式”、“分布式”和“单点式”三种方式,针对被审计单位信息系统的数据和网络结构情况,实施不同的组网模式。
- 联网审计的数据采集。研究提出了哪些数据是审计必须采集的重要数据和关联数据;采用数据采集接口和数据采集模板、不同组网模式下的数据采集方法、数据采集点的选择、数据采集频度确定等采集策略,以及数据源的识别、增量数据的提取等数据采集技术。
- 联网审计的数据转换。研究提出了建立审计中间表、审计元数据和数据验证的概念、规则和技术方法。
- 联网审计的海量数据存储。研究提出了采用数据集中和分区管理、集中式和分散式等存储方式,并结合国家“863计划”成果“高可扩展海量数据存储技术”在本课题进行了SAN技术的应用示范。
- 联网审计的数据处理与分析。研究提出了在单点采集情况下设置审计前置服务器,在数据集中的情况下设置中心服务器;在审计前置服务器上设置根据审计经验进行实时自动审计的审计预警系统,在中心服务器上设置审计数据分区管理,基于审计分析模型、多维数据挖掘分析、科学算法和审计经验等的集中审计分析,对多个数据采集点审计前置服务机的各种监控等处理技术。
- 联网审计的系统安全。重点研究了数据采集、数据传输、数据存储和数据分析的安全防范和应对策略,设计研发了保护数据采集时联网审计双方系统安全的“单刀双掷网络开关”。
- 联网审计的制度规范。研究发现,实行联网审计,不仅要解决技术问题,更要解决思想理念、法律规定、标准规范等问题。因此,课题组对此进行了深入研究,并形成了相关研究意见。

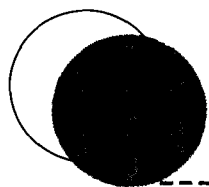
2005年9月,“计算机审计数据采集与处理技术”课题研究成果通过了科技部组织的国家“863”专家组鉴定验收。专家们通过实地考察、实际操作、理论论证等,认为课题紧密结合我国计算机审计实践,进行了关键技术研究;在联网审计组网模式、审计前置服务器设置、审计预警、中心服务器对前置机的流程控制等方面有技术创新;采用了“863计划”已有的成果;课题组形成14份研究报告、2篇论文、1本专著,申请了1项专利等,一致同意该课题通过验收。专家组还强烈希望课题组对研究所形成的技术报告整理出版,供有关专家、学者和广大审计人员参考,并在现有基础上,继续作为国家“863计划”课题,进行更加深入的研究。为此,我们将课题研究所形成的总体技术设计、组网模式技术、数据采集与转换技术、数据存储与处理技术、系统安全技术、应用环境技术、应用示范技术等11个部分的技术研究成果整理成专著予以出版。



本课题研究过程中,得到了科技部领导、高新技术司的大力支持,国家“863 计划”专家组自始至终给予了具体的专业指导。审计署领导给予了高度重视和大力支持,特别是李金华审计长不仅对课题研究极为重视和关怀,而且给予了许多重要指导,提出了许多关键性意见。审计系统的诸多领导和审计人员直接或间接地对课题研究给予了支持和帮助。清华大学、哈尔滨工业大学、南京审计学院、中国软件与技术服务股份有限公司、上海中标软件有限公司等给予了技术支持。对于各方的专业指导、努力合作和大力支持,我们在此表示衷心感谢。

联网审计作为一种崭新的计算机审计理论、技术和方式,正在冲击着长期以来传统的审计作业方式,既改变着人们的习惯意识和传统思维,也改变着审计系统现行的组织方式和管理方式。联网审计从提出、试点,到全面实施,正在和必将有力地提升我国审计监督的综合能力,更好地发挥审计机关在维护经济建设秩序、促进国民经济健康发展、推进政府廉洁高效建设方面的重要作用。联网审计作为一种计算机审计的理论、技术和方式,也正在和必将推进着我国具有中国特色现代审计方式的建设进程。然而,由于联网审计理论性强、技术要求高、业务和管理方式复杂,实践经验缺乏,很多理论、技术和标准等问题还有待深入研究,也有待金审工程建设联网审计系统的实践。本课题专著报告提出的联网审计理论和技术难免存在一些不足或错误,真诚希望各界学者和审计系统专家给予批评指正。同时,我们将根据专家们的意见,按照已审批的课题目标,进行更加深入的理论技术研究和应用实践,逐步形成较为完整的联网审计理论体系、技术方法、组织方式、法律制度、管理规范等理论技术构架,为金审工程实施联网审计提供更为坚实的理论基础和技术支持。随着我国计算机审计尤其是联网审计的全面实施,必将迎来具有中国特色现代审计的灿烂明天。

2005 年 10 月



目 录

CONTENTS

第一部分 总体技术设计篇	1
第 1 章 项目概况	2
1.1 课题名称	2
1.2 课题领域	2
1.3 课题责任人及依托单位	2
1.4 课题依托单位及技术与条件保障	2
1.5 课题组研究人员	4
1.6 课题研究周期	6
1.7 总体设计报告编制依据	6
1.8 课题验收情况	6
第 2 章 总体技术设计	7
2.1 课题研究目标和原则	7
2.2 研究内容和考核指标	8
2.3 课题研究总体技术方案	9
2.4 课题技术路线设计	16
第 3 章 联网审计组网模式技术设计	17
3.1 系统组网模式	17
3.2 核心技术研究	23
第 4 章 数据采集与转换技术设计	26
4.1 数据采集转换特征	26
4.2 审计数据采集策略	28
4.3 审计数据采集技术	29
4.4 审计数据转换技术	31
4.5 审计数据验证技术	34
第 5 章 数据存储与处理技术设计	36
5.1 审计数据存储技术	36
5.2 审计数据分析技术	39



5.3 数据存储核心技术研究	41
5.4 数据处理核心技术研究	50
第6章 应用示范环境技术设计	54
6.1 应用示范网络环境技术设计	54
6.2 数据采集转换环境技术设计	56
6.3 审计数据传输环境技术设计	57
6.4 海量数据存储环境技术设计	57
6.5 数据分析处理环境技术设计	57
6.6 实验环境安全系统技术设计	58
6.7 实验环境设施设备技术设计	58
第7章 应用示范软件技术设计	60
7.1 应用示范软件总体构架	60
7.2 对前置机控制系统设计	62
7.3 数据采集转换系统设计	62
7.4 审计预警系统技术设计	63
7.5 数据传输系统技术设计	63
7.6 数据存储系统技术设计	64
7.7 审计分析系统技术设计	64
7.8 应用平台系统技术设计	66
第8章 数据采集与处理的安全技术设计	68
8.1 联网审计的安全设计思路	68
8.2 关键技术实现方案	71
8.3 不同采集方式的安全保护	79
第二部分 组网模式研究篇	81
第9章 需求分析	82
9.1 系统需求	82
9.2 研究和建设目标	85
第10章 组网模式	87
10.1 组网模式设计原则	87
10.2 组网模式的选择	88
10.3 总体结构	89
10.4 集中式组网模式	90
10.5 分布式组网模式	99
10.6 单点式组网模式	108
第三部分 组网核心技术研究篇	121
第11章 需求分析	122



第 12 章 核心技术研究	123
12.1 研究原则	123
12.2 核心技术研究	124
12.3 安全技术研究	131
12.4 网络冗余研究	138
12.5 组网模型结构设计	140
12.6 网络开关结构设计	141
12.7 内网数据接收设计	142
12.8 组网模式应用示范	142
第四部分 采集与转换技术研究篇	143
第 13 章 审计数据采集与转换特征研究	144
13.1 被审计单位系统特点	144
13.2 数据采集与转换特征	145
第 14 章 数据采集策略	147
14.1 数据选择	147
14.2 数据接口使用	148
14.3 采集模板使用	149
14.4 联网审计数据采集策略	149
第 15 章 数据采集技术	155
15.1 数据源识别技术	156
15.2 数据交换处理技术	160
15.3 增量数据提取技术	166
第 16 章 审计数据转换技术	170
16.1 概念	170
16.2 数据转换必要性	172
16.3 数据转换系统设计	173
16.4 两种不同转换方式	177
16.5 数据转换规则	177
第 17 章 审计数据验证	181
17.1 数据验证的重要性	181
17.2 不同阶段的数据验证	182
17.3 数据验证的技术和方法	185
第五部分 存储与处理技术研究篇	189
第 18 章 需求分析	190
第 19 章 计算机审计数据存储的技术和方法	192
19.1 审计数据的存储模式	192



19.2 审计数据的存储技术	195
第20章 审计数据的组织和管理	198
20.1 审计数据的组织	198
20.2 审计数据的管理	198
20.3 审计数据管理工具	220
20.4 审计数据存储的实施	220
第21章 计算机审计数据分析技术和方法	223
21.1 通用性基础数据库的建立	223
21.2 审计中间表的建立	223
21.3 具体的数据分析	228
第六部分 存储核心技术研究篇	239
第22章 联网审计存储需求分析	240
第23章 审计署存储系统解决方案	242
23.1 设计原则	242
23.2 总体性能指标	242
23.3 方案拓扑图	243
23.4 解决方案介绍	243
第24章 硬件设备配置需求	249
第25章 在 SAN 上使用 DB2 分区数据库	250
第七部分 处理核心技术研究篇	253
第26章 课题背景及研究思路	254
26.1 国内外研究现状	254
26.2 主要研究内容	256
第27章 数据处理技术基础研究及应用分析	257
27.1 联机分析处理技术	257
27.2 数据挖掘技术	261
27.3 OLAP 及数据挖掘技术的应用分析	265
第28章 基于关联规则的审计特征智能提取研究	273
28.1 问题的提出	273
28.2 审计特征与体系架构研究	274
28.3 单机环境下的审计特征智能提取研究	278
28.4 审计关联规则分析	283
第29章 孤立点检测技术在审计中的应用	286
29.1 引言	286





29.2 基于孤立点检测的数据处理方法	286
第 30 章 基于 Benford 法则的数字分析法	293
30.1 概述	293
30.2 Benford 法则	295
30.3 Benford 法则的应用研究	298
30.4 数字分析法的软件实现	303
30.5 数字分析法在海关业务审计中的应用	305
第 31 章 异常数据的审计专业判断	309
第八部分 安全技术研究篇	311
第 32 章 联网审计对安全的要求及对策	312
第 33 章 联网审计的网络环境分析	313
33.1 安全建设总体方向与组网模式分析	313
33.2 系统整体结构及示意	314
第 34 章 系统网络安全模型	316
34.1 面临的威胁及防御关键技术	316
34.2 安全模型	317
第 35 章 各关键技术实施方案	322
35.1 数据采集前置机相关安全技术	322
35.2 网络隔离系统	326
35.3 防火墙及入侵检测系统	330
35.4 PKI 信任体系	334
35.5 通信平台安全技术	336
35.6 服务器、工作站及网络设备安全	339
35.7 安全管理及应急措施	344
35.8 安全漏洞检测	346
第 36 章 不同采集方式的具体方案	348
36.1 审计系统联网的整体部署	348
36.2 集中式采集方式	349
36.3 分布式采集方式	350
36.4 单点式采集方式	351
第 37 章 PKI 国内现状及标准	353
第九部分 安全核心技术研究篇	355
第 38 章 研究背景及内容	356
38.1 研究背景	356



38.2 研究内容	356
第 39 章 安全风险分析	358
39.1 外界威胁及风险分析	358
39.2 安全性分析技术	360
第 40 章 组网安全	368
40.1 组网模式	368
40.2 安全方案	369
第 41 章 数据采集安全	374
41.1 数据采集系统分析	374
41.2 安全网闸技术及其安全性问题	375
41.3 数据采集过程中的数据安全性问题	377
第 42 章 数据传输安全	378
42.1 联网审计系统的数据传输方式	378
42.2 数据传输安全策略	379
42.3 VPN 技术	379
42.4 传输过程中数据完整性的保证策略	380
第 43 章 数据存储安全	381
43.1 数据存储系统所面临的威胁	381
43.2 传统的安全方案	382
43.3 远程数据备份方案	383
43.4 数据存储安全策略	383
第 44 章 安全管理体系	385
44.1 安全管理资产	385
44.2 安全管理角色	385
44.3 安全管理制度	386
第 45 章 安全管理体系	387
45.1 基本定义	387
45.2 现有的灾备与恢复技术	387
45.3 基于数据流的应用级灾备方案	388
第十部分 应用示范环境建设篇	391
第 46 章 应用示范网络环境技术设计	392
46.1 应用示范网络构架选择	392
46.2 应用示范网络构架设计	395
第 47 章 数据采集转换环境技术设计	399
47.1 数据采集组网研究	399



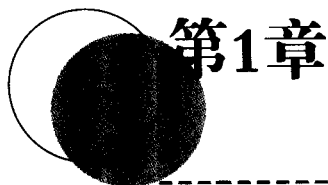
47.2 数据采集前置机环境	401
47.3 通信资源的选择	402
第 48 章 海量数据存储环境技术设计	403
48.1 审计数据传输环境的研究	403
48.2 数据存储环境的实现	403
第 49 章 数据分析处理环境技术设计	405
49.1 在前置机对数据进行预处理	405
49.2 在数据中心对数据进行分析处理	407
第 50 章 审计数据备份环境技术设计	408
50.1 备份技术	408
50.2 备份技术的选择与实现	409
第 51 章 应用环境安全系统技术设计	410
51.1 安全风险分析	410
51.2 系统组网安全	412
51.3 数据采集安全	412
51.4 数据传输安全	413
51.5 数据存储安全	413
第 52 章 应用示范环境设施或设备的选型与设计	416
52.1 设备选型原则	416
52.2 网络设施的设计	417
52.3 设备配置的设计	417
第十一部分 应用示范篇	419
第 53 章 应用示范软件总体构架	420
53.1 审计前置机	420
53.2 数据中心应用构架	424
53.3 “现场审计实施系统”	425
第 54 章 对前置机控制系统设计	426
54.1 对前置机设置的控制	426
54.2 对前置机功能的控制	427
54.3 对前置机安全的控制	427
第 55 章 数据采集转换系统设计	428
55.1 数据采集设计	428
55.2 数据转换设计	429
第 56 章 审计预警系统技术设计	430
56.1 系统级预警	430



56.2 审计经验级预警	430
第 57 章 数据传输系统技术设计	432
第 58 章 数据存储系统技术设计	433
58.1 海量数据存储	433
58.2 数据分区管理	433
第 59 章 审计分析系统技术设计	434
59.1 多维数据库及联机分析	434
59.2 审计分析模型和构建技术	435
59.3 审计中间表和创建技术	436
59.4 计算机审计方法技术	437
59.5 行业指标分析技术	437
59.6 通用查询和专用分析器	438
第 60 章 应用平台系统技术设计	444
60.1 应用平台功能结构	444
60.2 应用系统技术路线	445
第 61 章 应用示范的标准规范设计	446
61.1 数据结构标准规范设计	446
61.2 数据接口标准规范设计	446
61.3 应用规则标准规范设计	447
61.4 应用模板标准规范设计	447
后记	448

第一部分

总体技术设计篇



第1章

项目概况

CHAPTER 1

课题名称

- 本课题名称为“计算机审计数据采集与处理技术”。
- 本课题编号为 2003AA1Z2330。

课题领域

本课题领域为国家高技术研究发展计划(“863 计划”)——信息技术(863—100)。

- 主题(重大专项): 软件重大专项。
- 所属专题名称: 重大应用示范。

课题责任人及依托单位

- 课题责任人: 刘家义。
- 课题依托单位: 审计署计算机技术中心。

课题依托单位及技术与条件保障

1.4.1 课题依托单位

本课题的依托单位为审计署计算机技术中心。计算机技术中心现有的技术基础包括 3 个方面。

① 培养了一支既懂得计算机审计业务,又懂得计算机审计技术的计算机审计骨干和专家队伍。计算机技术中心现有技术人员 27 人,其中高级职称 18 人,有硕士和博士研究生学位的 5 人,来审计署工作 10