DOI: 10.3724/SP.J.1224.2014.00307

“大数据处理中的基础理论与关键技术”专刊

大数据金融背景下商业银行客户信息保护研究

董纪昌，焦丹晓，张 欣，宋子健，李秀婷

(中国科学院大学管理学院，北京 100190)

摘 要: 在大数据金融背景下，伴随新技术、新业务形态不断出现，客户信息保护成为商业银行面临的具有挑战性的新命题。本文结合国内的案例，首先分析了大数据金融对商业银行客户信息保护的影响机制，认为大数据技术背景下，客户信息被赋予了三大新的特点：数据量大、数据价值高、数据泄露破坏性强。客户信息数据流主要经过数据搜索、数据传输、数据存储三个环节，信息泄露主要通过信息脱敏处理、数据存储管理及内部违规三条渠道，存在黑客攻击、内部违规、管理疏漏三个主要原因。然后，结合信息泄露的环节、原因、结果进行了具体的案例分析。最后，从法律、管理和技术角度提出大数据金融背景下商业银行客户信息保护措施建议。

关键词: 大数据金融; 商业银行; 客户信息保护

中图分类号: F83 **文献标识码**: A **文章编号**: 1674-4969(2014)03-0307-12

引言

2012 年 10 月，IDC（Internet Data Center）发布的关于中国大数据技术和服务市场的首份报告中指出，中国大数据市场规模将会从 2011 年的 7 760 万美元增长到 2016 年的 6.17 亿美元，复合增长率达 51.4%，市场规模增长近 7 倍。数据规模不断增大给金融行业带来新的机遇与挑战，我国金融行业正逐渐步入大数据时代。商业银行等传统金融机构也积极布局大数据金融，依托自身的资金优势打造自身网络平台体系，或与成熟电商、支付平台联手寻求布局大数据金融的先机；借助互联网信息技术完善数据体系，探索新的大数据金融业态[1]。

与传统金融行业相比，大数据金融一方面表现出透明度更强、参与度更高、协作性更好、中间成本更低等一系列优势，另一方面也面临着诸多新问题，包括软件和数据的处理能力、资源共享和数据管理等。这些挑战广泛地分布在大数据的存储、分析、管理和数据安全等各个方面[2]。大数据金融的互联网属性也带来了网络身份确认、假冒网站、交易欺诈等一系列问题，造成了层出不穷的数据泄密事件，给商业银行客户信息保护提出了挑战。

此外，长期以来我国对个人信息的保护缺乏足够的重视，无论是在制度设计还是操作层面上对个人信息保护都相当薄弱[3]。2014 年 4 月中国银行业监督管理委员会和中央银行在联合下发的《关于加强商业银行与第三方支付机构合作业务管理的通知》中，特别强调了商业银行要做好客户信息安全与保密工作。虽然我国在这方面取得了一定的进展，但与国外成熟体系相比还有较大

收稿日期: 2014–08–01; 修回日期: 2014–08–15

基金项目: 国家自然基金项目(71173213)；国家自然基金青年基金项目（71203217）；中国博士后科学基金项目（2013M540129）

作者简介: 董纪昌（1974–），男，教授，博士生导师，研究方向为房地产金融、互联网金融。

焦丹晓（1987–），女，博士研究生，研究方向为互联网金融、房地产金融。E-mail: jdxfantasy@163.com

张 欣（1989–），女，博士研究生，研究方向为房地产金融、互联网金融。E-mail: zhangxin411@mails.ucas.ac.cn

差距。目前我国颁布的法律法规尚不能覆盖大数据金融下产生的客户信息安全问题，而相关规范性文件与法律文件不能相互呼应，无法形成一个立体化监管体系[4]。

事实上，相关法律法规的缺位、隐私权保护法制理念的淡薄、客观的经济利润，催生了环环相扣的信息交易利益链条。低成本、高收益的暴利诱使一些人铤而走险，从组织内部非法获取信息。相关调查显示，黑客盗取个人信息后，可以直接侵入别人账号、邮箱而获取有用的信息，也可以打包销售；此类交易非常简单，通过专门的网络地下黑客论坛或腾讯即时聊天工具在线交易，多次卖给不同客户来谋取巨额利润。

2013 年数据泄密的方式和影响范围不断升级。一方面信息泄密的途径和方式进一步复杂化及多样化，既有利用系统及应用漏洞攻击的方式，也有基于商业软件植入病毒的方式[5]；另一方面，信息泄密的领域进一步扩大，既涵盖了衣食住行等个人隐私，也包括了金融安全、商业秘密及国家秘密等领域[6]。同时，移动互联网及云计算的普及大大提升了信息泄密的风险，由这些创新信息化模式导致的信息泄密事件呈现快速增多的态势。

尽管商业银行经过数年的发展，积累了丰富的风险控制和客户信息管理经验，对于客户的身份认证、资金监管、账户安全有强大的市场优势，但是，大数据金融发展快、覆盖广、管理弱、技术水平要求高等一系列特点，也在一定程度上为某些机构或者个人的非法操作带来了可能空间，增大了客户信息泄露和被违规利用的风险。根据国家互联网应急中心数据显示，2014 年 2 月，境内感染网络病毒的终端数为 220 万余个，境内被篡改网站数量为 12 428 个，信息系统安全漏洞为 699 个。信息安全事件的频现将大数据金融的信息安全问题推向了风口浪尖，有效防范信息安全风险成为银行业广泛探讨的焦点问题。伴随新技术、新业务形态不断出现，客户信息保护成为商业银行面临的具有挑战性的新命题。在大数据金融快速发展的背景下，关于客户信息保护方面的研究具有巨大的理论价值和现实意义。

1　文献回顾

国外学者关于大数据处理技术应用于金融行业的研究起步较早，W. Breymann、A. Dias 等在原有金融数据分析方法的基础上，提出了检验高频金融数据之间的极值依赖关系的模型[7]。J. Bughin、M. Chui、J. Manyika 针对大数据处理与云存储等技术在商业银行金融数据管理方面的应用，提出了若干建议[8]82。

在客户信息保护方面，国外学者也进行了一些研究，K. Kim 和 B. Prabhakar 指出，由于互联网的开放性，网络交易的安全性将是其未来发展的重点[9]。M. Sathye 的研究表明，阻碍个人用户使用网络银行系统的主要原因就是对于交易安全的担忧[10]。伴随金融领域大数据处理技术的普及，一些学者研究了客户信息保护所面临的一些新问题。用户选择网络银行的主要考虑因素是网络银行的安全性，包括网络操作系统的稳定程度、信息的质量和隐私保护等方面。Z. Liao 和 M. T. Cheung 指出，大多数网银用户不满意网络银行的安全性并对金融交易过程中存在的风险感到担忧[11]。P. Selvapriyavadhana 研究发现，新兴移动银行服务业中暴露出来的客户信息保护问题越来越严重[12]。

国内关于大数据金融对于商业银行影响的研究为数不多。邱峰指出，鉴于大数据金融自身存在诸多亟待解决的缺陷及商业银行所处的特殊地位，目前大数据金融尚无法撼动、取代商业银行，二者之间应是互相合作、优势互补，商业银行应采取多种有效措施，迎接大数据金融的崛起[13]。冯娟娟基于商业银行视角，重点分析网络支付、网络借贷、金融搜索等新型大数据金融运作模式的特征，揭示商业银行在大数据金融领域的优势与短板，并指出商业银行要通过寻求合作共赢、重视客户体验、发掘培养人才、提升科技水平等

提升核心竞争力[14]。梁璋、沈凡提出了新金融模式下银行如何将资本、客户和政策等方面的资源优势转化成为其与新金融势力博弈或融合过程中的筹码，从而推动我国金融业和商业的全面升级[15]。

国内关于大数据金融背景下客户信息保护的相关学术研究几乎空白，仅有的一些探讨主要集中在金融机构从业人员的职业伦理等方面。快钱支付清算信息有限公司副总裁顾卿华认为，大数据金融的快速发展对数据安全和系统稳定性要求极高，企业要想为客户提供创新、高效的服务，必须要注重网络安全[16]。中国建设银行电子银行部副总经理于潇在《主题对话：大数据时代的金融创新与挑战》中表示，在大数据金融的时代，应该用好现有的静态数据，并管理好客户的动态行为数据，加强合作，尤其是 Web 数据的合作，更多地谋求共赢，从而为客户提供全方位的服务，加强客户信息的安全性，提升银行风险防范能力[17]。爱投资网创始人赵春霞表示，大数据金融行业在保护客户资料上应该向比较成熟的金融机构学习，做到客户资料信息的碎片化处理，不同部门的每个员工都只能掌握客户的一部分信息，这样才能大大降低信息泄露的风险，保障客户资料的安全[18]。

对于关于客户信息数据的管理的研究由来已久，总体来看客户数据管理系统发展分为三个阶段。第一代客户数据管理系统以 PC、办公软件的运用为主要标志。它实质上是一个客户管理信息系统（management information system，MIS），在 20 世纪 80 年代中期至 20 世纪 90 年代中期得到迅速发展。该系统以结构化数据作为主要处理和存储对象，实现了文档写作和数据统计电子化，主要面向办公室管理人员和事务处理人员，将客户信息载体从传统的纸介质方式转向了比特方式，一般仅限于银行部门内部的数据计算和统计[19]。1995 年，Gerstuer 提出“以网络作为中心的计算”模式，产生了第二代客户信息管理系统[20]。第二代客户数据管理系统以实现工作流程自动化为目的，以网络和协同工作技术为主要特征。它对银行客户数据管理的发展趋势产生极大影响，从此，银行客户数据管理系统以网络作为中心，以信息和工作流为主要处理内容。第三代客户数据管理系统以充分利用银行局域网为主要标志，以知识的管理信息化为核心，建立在银行的数据网络平台上，可以提高和丰富企业员工的知识共享体系和学习功能，利用银行客户数据管理的使用提高了系统运行效率[21]。

2 大数据金融对商业银行客户信息保护的影响机制分析

大数据金融是指利用大数据开展的金融服务，即针对海量数据，经过互联网、云计算等信息化处理方式，结合传统金融服务，开展资金融通、创新金融服务，是传统金融行业与互联网精神相结合的新兴领域。它的特色在于依托云计算的分布式处理、分布式数据库、云存储和虚拟化技术，对海量数据进行挖掘[22]。大数据金融的发展为商业银行客户信息保护带来了技术可行性，同时由于大数据时代数据价值的提升，也对客户信息保护形成了严峻的考验。

表 1 梳理了近年来国内外数据泄露事件的涉及客户及涉密方式。虽然客户信息安全保护上出现的问题和事故在起因上各不相同，但从结果上来看，都造成了巨大的经济损失。在大数据技术背景下，客户信息被赋予了新的三大特点：数据量大、数据价值高、数据泄露破坏性强。毫无疑问，互联网时代的云存储技术为海量数据提供了存储空间，数据库建设、数据计量单位的数倍提升彰显了“ Big Data ”时代的新特点；数据挖掘、机器学习等智能挖掘技术的深入发展使得对数据处理、提取价值的能力增强，数据价值提高，社会继而进入“ 数据为王 ”的时代；正因为数据量增大且价值提高，数据泄露的破坏性才越发增强，大数据技术对数据泄露的后果产生了杠杆效应，小的技术漏洞会导致极大的损失。结合案例分析，目前商业银行客户信息泄露主要存在三种情形：黑

表 1 国内外数据泄露案例总结

序号	企业	曝光时间	涉及客户	泄密方式
1	美国万事达信用卡公司	2005–06–17	1 390 万份信用卡客户资料被盗	黑客入侵了“信用卡第三方付款处理器”的网络系统
2	Visa 信用卡公司	2005–06–17	2 200 万份信用卡客户资料被盗；其中 Visa 中国分公司的 3 100 个账户、近 500 张牡丹国际卡遭泄密	黑客入侵了“信用卡第三方付款处理器”的网络系统
3	汇丰银行（香港观塘裕民坊分行）	2008–05	近 16 万客户数据遗失，包括账号号码、姓名、交易金额	一部载有这些数据的计算机服务器遗失
4	汇丰银行（瑞士私人银行部）	2010–03–11	涉 1.5 万名客户，占该行私人银行全球客户总数的 15%	IT 部门内部员工盗窃数据
5	花旗银行	2011–06–08	近 20 万用户资料被盗，包括账户信息。3 400 名信用卡客户损失约 270 万美元	网站遭到黑客攻击，客户信息被盗取
6	民生银行	2011–07	涉及 2 名钻石级客户，共计 200 余万元；其中，一名歌手的信用卡被盗用 120 万	信用卡中心的客服人员私自为已销卡但未销户的钻石级客户补办信用卡，大肆透支
7	韩国农协银行	201–17	3 500 万名用户信息遭泄漏，包括用户姓名、电话号码、身份证号码等	网络遭黑客攻击，用户信息泄露，在线交易系统曾一度瘫痪
8	招商银行信用卡中心	2012–03	数千份个人信息遭泄露	内部员工为谋取个人利益向他人出售客户信息
9	中国工商银行	2012–03	数千份个人信息遭泄露	内部员工为谋取个人利益向他人出售客户信息
10	中国人寿	2013–02	泄露了 80 万份投保人的个人信息	合作网站升级操作失误
11	英国离岸金融业	2013–04	200 多万份邮件等文件泄密	范围涉及 170 个国家的 13 万富豪，是具有重大影响的金融安全事件
12	博思艾伦咨询公司	2013–06	数十万份客户资料	内部雇员利用 U 盘及账户凭证获得了对关键性系统的访问权，将取得的资料提供给媒体记者
13	Corporate Car Online	2013–09	个人信息泄露的客户数量超过 85 万名，导致成千上万份信用卡信息被曝光	第三方供应商合作处理敏感数据带来潜在风险，黑客通过其安全漏洞获取了对数据的访问权
14	Adobe 公司	2013–01	近 300 万客户的敏感信息及个人数据遭到泄密，包括客户名称、加密信用卡或借记卡账号等	黑客利用其产品源代码上存在的安全漏洞入侵
15	MongoHQ 公司	2013–01	给数百位云用户带来直接影响，受到间接影响的用户数量则可能成千上万	内部支持应用程序相关的安全控制机制失效
16	花旗银行韩国分行	2013–12	打印有 34 000 余名贷款客户信息的 A4 纸质资料外泄，包含客户的名字、手机号、贷款额、利息率和工作单位等	内部人员为了提高自己的贷款业务成绩，将公司电算网存储的客户信息泄露给非法放贷业主
17	美国大型零售商 Target	2013–12	4 000 万份支付卡数据泄露，包含姓名、电子邮箱地址及其他信息；影响到 1.1 亿用户，损失成本高达 4.2 亿美元	合作厂商遭恶意邮件侵入，进而透过收款机窃取信用卡信息
18	渣打银行	2013–12	103 287 条客户信息被泄露，包括客户的名字、身份证号、电话号码和工作单位等	银行 IT 中心外包企业的职员通过 USB 盗取客户信息
19	个人信用评估公司 Korea Credit Bureau (KCB)	2014–01	泄露 KB 国民卡、乐天卡及 NH 农协卡多家公司 1.04 亿条用户个人信息，包括姓名、电话号码、住所、身份证号码及贷款交易内容、信用卡认可免税书等敏感的信用信息	内部职员在受信用卡公司委托开发电脑程序的过程中，非法收集和泄露信用卡公司的客户信息
20	支付宝	2014–02	支付宝用户 20 GB 海量信息被泄露，包括客户的实名、手机、电子邮箱、家庭住址、消费记录等	支付宝的前技术员工利用工作之便，在职期间多次在公司后台下载支付宝用户资料
21	众筹平台 Kickstarter	2014–02	部分用户信息被盗取	黑客攻击
22	Ebay 公司	2014–02	1.45 亿客户的电邮地址、加密后的密码、出生日期及住址等数据被盗取	员工凭入证外泄

客攻击网站，盗取用户信息；内部员工利用用户信息谋取个人利益；管理问题导致客户信息保护不力。

图 1 从客户信息泄露内外因的角度进行了数据统计，图 2 分析了客户信息泄露的动机。结果显示，外部因素是客户信息泄露的主要因素，Verizon 还进一步指出，在调查采样的 10 万次数据泄露安全事件中，92%的攻击手段属于以下范畴：犯罪软件（各种以控制系统为目的的恶意软件）、WEB 应用攻击、DOS 拒绝服务攻击、网络间谍、POS 入侵、支付卡信息窃取[23]9。谋取经济利益依然是窃取客户信息的主要动机。

本文在以上分析的基础上总结了大数据金融对商业银行客户信息保护的影响机制，如图 3 所示。

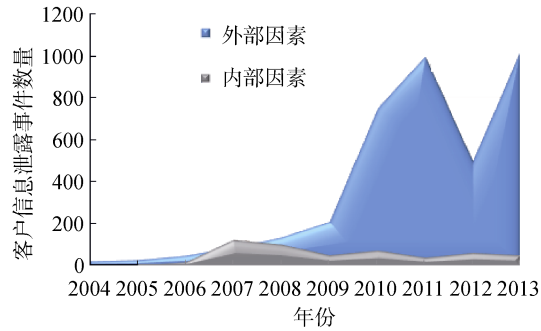

图 1　客户信息泄露内外部原因

数据来源：文献[23]8。

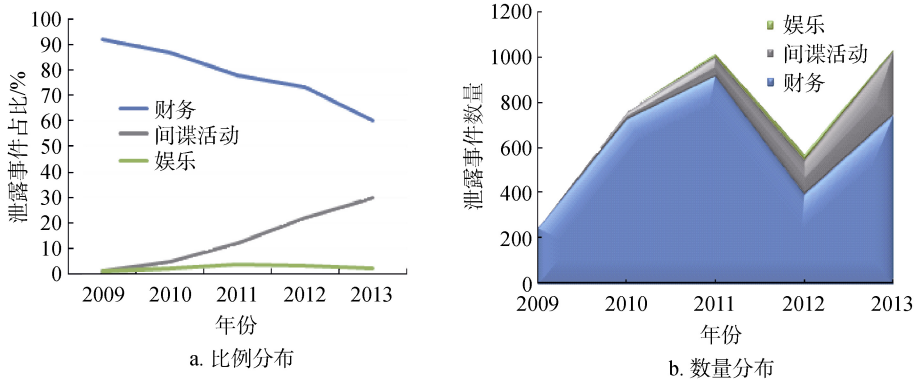

图 2　客户信息泄露动机

数据来源：文献[23]9。

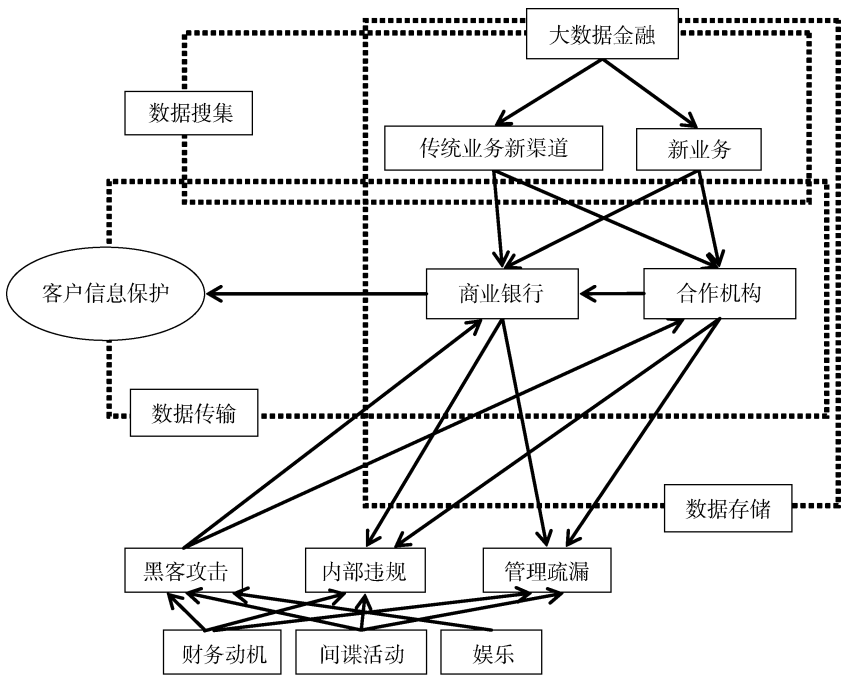

图 3　大数据金融对商业银行客户信息保护的影响机制图

大数据背景下，商业银行客户信息数据流主要经过数据搜索、数据传输、数据存储三个环节[24]，信息泄露主要通过信息脱敏处理、数据存储管理及内部违规三条渠道，存在黑客攻击、内部违规、管理疏漏三个主要原因[25]。互联网技术快速发展的背景下，信息处理水平大幅增强使得金融信息价值数倍提升，数据的背后掩藏着的是惊人的财富交易。巨额的经济利益必然导致基于财务动机、间谍活动的黑客攻击的频率提升、内部违规风险增强，同时给银行客户信息保护相关的管理工作提出了严峻的挑战。在大数据技术的背景下，客户信息泄露具有其必然性因素，而其最终有效的处理手段则必然依赖于法律法规的细化、技术水平的改进和管理水平的提升。下面根据大数据金融背景下信息泄露的渠道分析其产生的必然性。

1）客户信息脱敏处理。

由于自身技术水平短板，商业银行往往将大数据平台业务外包给一些合作公司，这种外包业务发展速度较快，而与之匹配的管理模式发展明显滞后，使客户信息面临外泄风险。银行在与软件公司开展系统开发外包合作过程中，未能做好客户信息保护的工作，提供给外包商的工作设备和数据均未进行彻底的数据清理等工作，并且外包开发人员可以通过其系统便捷地访问银行前置业务系统，由此导致外包开发方可以随意地进入银行客户信息系统，获取银行客户信息。

2）客户数据存储管理。

客户信息进入到银行的管理系统之后，管理存储环节也是信息泄露的一个主要渠道。

首先，在数据存储管理环节，客户信息存在泄露风险的首要原因是数据存储硬件设备管理不严。其主要体现在部分银行未对其内部办公计算机的 USB 等外部设备接口使用加以监管和控制，使得不法人员有机会通过外部存储设备，将储存在互联网云存储或者计算机内存中的客户信息等相关文件数据进行非法拷贝，导致客户信息外泄。

其次，客户信息管理系统维护不审慎。例如，2011 年中国银行在进行系统升级过程中，有关人员出现误操作，直接导致众多客户信息出现混乱，导致客户信息外泄，造成了重大损失。

再次，重要数据加密处理级别不够。部分银行受其自身技术水平的限制，对其互联网交易平台系统使用的加密算法安全级别不够，大多采用将密钥明文固化于加密程序中。这种方式使得外部不法分子可以轻易获取密钥、算法，并在此基础上进行暴力破解客户信息。

最后，客户信息系统用户权限管理混乱。部分银行的客户信息管理系统存在诸多弊端，如对于系统的部分重要用户给予过高权限、未定期提醒客户修改登陆口令、用户权限岗位调整不同步导致权限混乱等。

3）部门内部违规。

除管理客户信息流程上的渠道漏洞之外，还有不能被忽视的一点，就是银行自身的内部监管漏洞。前文中提到，目前法律上对于客户信息保护的边界仍未有一个明确的界定，因此，各银行普遍将客户信息保护视作保密管理、信息安全管理或征信管理的一部分，没有明确定义客户信息保护的范畴，这就使得各部门在履行各自义务的过程中，不仅未能形成互补，甚至存在相互掣肘的情况。

除了部门职责界定不清之外，部分商业银行还存在内部问责机制缺失的现象，即部分商业银行在客户信息保护方面的相关规章制度基本停留在强调自身工作人员提高保密意识的层面上，而在避免客户信息泄露的发生、泄露事件应急处理和责任追究方面缺乏明确规定。

此外，大数据金融的发展也会造成无孔不入的信息收集行为，严重危害到个人信息主体的合法权益；客户隐私的界限难以划清、如何在客户隐私与商业营销之间进行取舍，这是商业银行不得不面对的问题，也是立法、监管等部门必须从顶层加以研究和关注的问题。

建立在传统金融上的客户信息保护面临着全

面颠覆，需从制度机制、内控机制、保密机制到防控机制进行重新定位，如客户信息的采集、使用、保密、保密例外，纸质和电子信息的集中统一管理，第三方合作及外包单位信息保护力度的监督、审查，信息技术科技含量、防火墙、身份识别与认证、数据加密、数字签名、第三方认证及网络安全监控等技术的跟进和及时更新等。

3 案例分析

案例一：江苏银行“泄密门”事件。

2012 年 2—4 月，江苏银行上海金桥支行单方面凭借宜信普惠信息咨询（北京）有限公司提供的授权查询书，在没有与客户发生任何业务关系的情况下，擅自查询客户个人信用报告，涉及客户 3.2 万名。并在获取客户信息之后，违规将部分查询结果提供给宜信普惠信息咨询（北京）有限公司。宜信普惠信息咨询（北京）有限公司利用非法获得的大量客户信息，进行大数据分析，并向部分客户推销保险产品，给客户造成严重困扰。

案例分析：

1）信息泄露环节：数据存储环节。

2）信息泄露原因：银行部门内部违规。

3）信息泄露结果：客户个人隐私遭泄露，江苏银行上海金桥支行被通报并责令整改。

案例二：携程旅行网“泄密门”事件。

2014 年，在线旅游行业巨头携程旅行网（简称携程）用户信用卡信息遭泄露事件引发了用户的恐慌。携程与包括中国银行、招商银行等十余家国内银行合作进行网上业务，在用户不知情的情况下，擅自存储用户信用卡信息。由于系统存在漏洞，导致大量用户银行卡信息存在泄露风险，该信息的泄露可能直接引发信用卡盗刷风险。

案例分析：

1）信息泄露环节：数据搜索环节、数据传输环节。

2）信息泄露原因：黑客攻击、管理疏漏。

3）信息泄露结果：引发合作银行客户恐慌，携程删除非法存储的客户信用卡数据。

上述案例的分析结果表明，银行泄露客户个人信息的事件频繁发生，商业银行中的客户信息保护工作确实存在重大问题。特别是在当前的大数据金融背景下，客户信息安全的重要性更加明显。当前，不论是商业银行还是大数据金融企业，其管理制度、内控能力和信息保密的相关规定、执行力度都是不够的；对于客户信息保护所采取的技术和手段远远达不到应对大数据金融给银行业带来的冲击所需要的客观要求；对于接触客户信息的人员监管和审核未能达标，容易出现内部员工受利益驱使泄露客户信息的情况；此外，客户对于自身信息的自我保护意识还有待加强。

大数据金融对于客户信息保护提出更高层次的要求和基准[26]，这是商业银行应当首先考虑和应对的主要问题。客户信息保护不仅需要商业银行自身加强内部管理、及时更新和改进客户信息保护技术和手段、完善相关部门人员的考核机制与培训管理，也需要合理的法规政策督促与客户自我保护意识的提升。

4 建议

大数据金融企业和商业银行进行大数据金融创新时，往往因为追求便捷性与时效性，忽视客户信息保护。我国商业银行目前正处于发展大数据金融业务的关键时期，更应当注意加强客户信息安全建设。

1）落实相关法律法规，完善内控机制。

（1）在法律法规指导下制定相应规章制度，完善自身监管。

银行内部有关部门应当认真学习借鉴国外商业银行客户信息保护的，并结合我国实际情况，制定相应规章制度，严格遵守“最小授权、最小知悉范围、最小信息量”的客户信息使用原则。此外，还应建立行业涉信人员从业资格制度，明确银行对其内部工作人员、具有合作关系的第三方工作人员或机构的泄露、丢失客户信息的行为

负有同等责任。

（2）完善客户信息泄露的应急处理制度。

除了应当从预防的角度制定规章制度之外，对于客户信息外泄的应急处理环节也要有相应的明文规定，应急处理制度应使遭信息泄露的客户得到合理的赔偿，使泄露者、买卖者的违法成本高于收益，从根本上威慑、遏制盗卖个人信息的行为。可以考虑如下规制手段：①设定最低的损害赔偿标准；②在举证责任分配及举证责任能力赋予上向处于弱势地位的受害者倾斜；③对于造成客户重大损失的相关部门或人员，依法追究其民事责任。

（3）银行内部系统权限分级，减少越权操作行为。

在移动互联网银行快速发展的背景下，越来越多的金融业务是依靠网络平台完成的，客户、柜台操作人员及技术管理人员等不同职能的人群需要有不同级别的系统访问权限（图 4）。

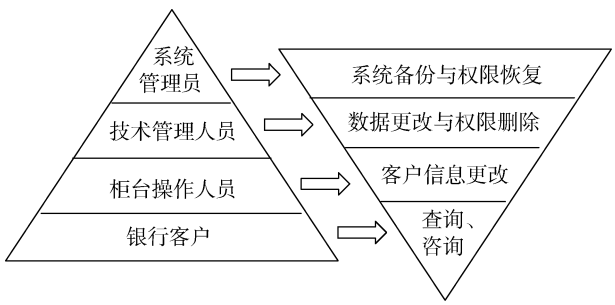

图 4　银行内部系统权限分级

对于拥有基础权限的客户与柜台操作人员，要实行单人单口令认证机制；对于拥有高权限的技术管理人员，在进行系统维护时要实行双人多认证机制；对于拥有最高权限的系统管理员，需要实行多重认证机制和多步验证机制，并在最终执行前，将信息传递给该信息管理渠道上的每一位管理员，尽最大可能将误操作与个人私自操作的可能性降到最低。

2）提高客户自我信息保护意识。

在大数据金融时代，各种新型支付网站的建立增大了客户操作风险，并提高了信息被盗的可能性。为此，商业银行应当通过印制专门的宣传彩页或用户使用手册等方式加强对线上用户信息保护的宣传力度，利用广播、新闻、网络等宣传媒体及时报道有关网银客户信息被盗的案例来提高客户自我信息保护意识，不定期通过手机短信或电子邮件的方式为客户发送网银安全使用措施及升级提示灯信息来培养客户良好的操作习惯和风险识别能力。

3）大数据客户信息认证和传输安全的技术实现。

在大数据时代，客户信息保护的关键在于技术的创新[8]76。针对大数据金融背景下客户信息认证和传输过程，本文提出以下可操作性措施。

（1）采取多重信息验证，确保客户信息认证过程安全。

如图 5 所示，在大数据金融迅猛发展的背景下，相应的网络窃取信息技术更是获得了极大的发展，因此单纯地通过认证密码和 Pin 码或者防火墙等网络认证结合的方式已经不能满足网络银行系统的身份认证需求。大型银行的网上银行多采用软硬件结合的双因子认证方式作为身份认证的辅助解决方案。

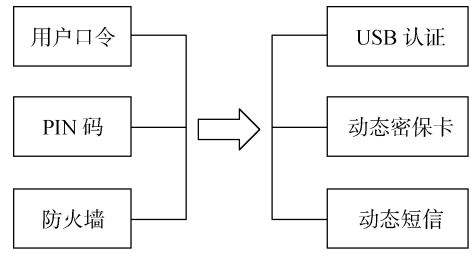

图 5　大数据金融背景下客户信息认证技术转变

（2）完善数据加密算法，确保数据传输安全。

大数据金融使得银行内部的网络与外部通信网络之间的联系更加紧密，因此，客户信息数据在生成之后传送到银行系统的过程中就难免会与外界网络进行更大范围的接触与传递，也就更容易受到外界主动和被动的攻击[27]。

为保证数据传输的安全性，银行应采用加密方式传输数据，并且应当将密钥与加密文件进行

分离存储和传送。中央银行和四大国有商业银行在密文传输上都有一套自己的先进方法和技术，其他商业银行应在借鉴他们的经验的基础上，结合自身的具体情况加以改进和完善，开发出属于自己的加密技术。

4）采用云计算技术和 Hadoop 分布式系统重构数据库。

当今主流电商均采用了云计算技术手段和 Hadoop 分布式系统来进行数据库构建，这点可供商业银行借鉴。

（1）商业银行信息存储模式的局限性。

在大数据金融到来之前，商业银行客户信息存储系统大多采用的是数据库集群模式。但是，随着大数据在金融行业的应用越来越广，数据库集群模式在技术上所具有的局限性也就越来越明显：硬件复杂度过高，实施架设难度较大，构建系统成本、系统运营维护成本较高；数据库安全性和数据集可扩展性提升空间极小，在数据传输过程中，容易受到外部攻击，导致客户信息泄露；随着设备量的增加和应用的复杂化，系统兼容性问题凸显，存在重大隐患[28]。

（2）云计算技术手段和 Hadoop 分布式系统。

①云计算新型分布式网络计算架构的实现。

云计算是一种新型分布式的网络计算架构，特别适用于为各种网络应用提供计算、存储、网络、软件等在线服务（图 6）。云计算架构是由大规模低端服务器所组成的服务器集群，其作用是提供海量存储空间和提供大规模数据的处理能力，该架构具有可靠性、扩展性及高可用性等特点。目前主流电子商务平台首选云计算平台 Hadoop 架构来进行网页搜索、大数据分析等，如淘宝、百度和网易等。

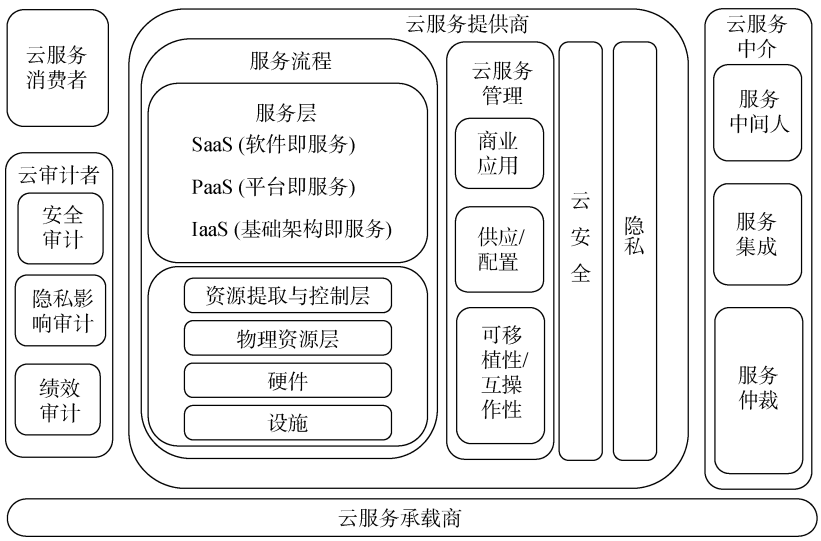

图 6　云计算分布式网络计算架构

②Hadoop 分布式系统的技术实现。

MapReduce 由 Google 公司设计，是非关系型数据管理和分析技术中最为典型的代表，主要用于对集群上的大数据集进行并行计算处理。Hadoop 分布式系统是在 Google 公司提出 MapReduce 技术之后，于 2004 年由一个开源组织 Apache 发布的一个分布式计算框架，用于模仿和实现 Google 云计算的主要技术[29]。

图 7 为 MapReduce 的数据处理过程。它的计算流程和基本原理可以概括为两步：第一步，将大数据集分解为成百上千个小数据集，每个（或若干个）数据集分别由集群中的一个节点进行处理并生成中间结果；第二步，将处理生成的中间结果通过大量的节点进行合并，并形成最终结果。通过数据的分解与合并过程，解决了在系统层面数据库集群所难以解决的扩展性、容错性等问题[30]。

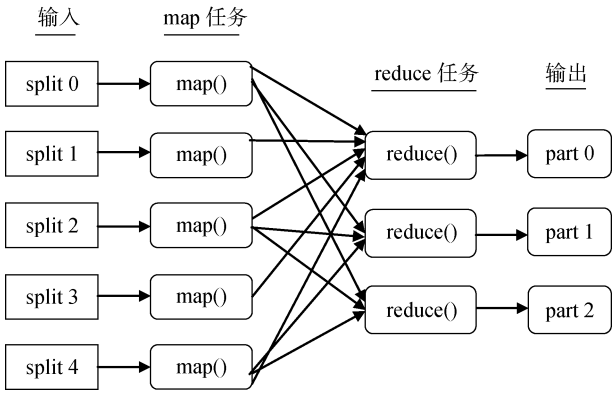

图 7 **MapReduce** 数据处理过程

MapReduce 免费开源，其构建于异构廉价服务器搭建可弹性伸缩的大规模集群的基础之上，采用并行的、分布式的处理和分析大规模数据的方式，构建成本远远低于数据库集群所采用的并行式数据库[31]。数据库集群与 MapReduce 两种大型数据的处理模式比较分析见表 2。

通过表 2 不难看出，MapReduce 大数据处理模式具有相对显著的优势。传统的数据库系统实时响应能力较强，但对于 TB 级或 PB 级的大数据集，数据挖掘的检索速度则急剧下降。引入云计算与 MapReduce 的分布式系统架构可以充分利用两者的优势，实现高效率的数据挖掘与决策支持，使得现有的数据处理模式效率大大增强，同时也使得客户信息保护力度得以加强。

表 **2** 大规模数据处理模式的比较分析

模式	扩展性	完整性与安全性	灵活性	数据处理能力	处理效率	异构存储	成本
数据库集群	有限	低	低	主要支持结构化数据	低	欠缺	高
MapReduce	可扩展	高	高	支持复杂的数据处理	高	高支持	低

5）明确第三方合作机构选择标准。

商业银行由于受限于自身技术水平短板的制约，系统开发等业务大多采用外包模式，在此过程中难免会将部分客户信息发送至第三方机构，因而银行对于第三方机构的技术考察应该进一步加强[32]。

目前主流的支付平台大多采用牺牲空间维度来满足客户对于时间维度的需求。例如，以支付宝为代表的第三方自营支付平台和以拉卡拉为代表的第三方合作支付平台均为了保证客户使用的便捷性而跳过了客户支付时银行验证的环节，这样就会使得数据的空间维度需求相对升高，并且同时会使得客户信息泄露危险级别提升。

因此，商业银行应当选择数据处理能力优秀的外包公司，保证在客户支付环节尽可能缩短与银行验证通信的时间，这样就可以避免客户信息因为驻留第三方支付平台而导致外泄的情况；同时，高技术标准的第三方合作机构选择必然带来商业银行成本的上升，在寻求可信赖合作伙伴的同时，应力求合作模式、营销渠道的创新以避免过高的成本压力。

总之，商业银行应该加大对于技术层面的人力和物力投入，加快开发自主互联网业务平台，将现有的系统模型进行进一步改进和扩展，最终达到每个客户所对应的密钥算法都是独立存在的，这样即使黑客破解了一个客户的密文也无法破解其他客户的密文，从最大程度上保护客户资产安全。

参考文献

[1] 谢平, 尹龙. 网络经济下的金融理论与金融治理[J]. 经济研究, 2001, 4(6): 25.

[2] Russom P. Big Data Analytics [J]. TDWI Best Practices Report, 2011(1): 25.

[3] 吕德宏, 汝璇卿, 叶建洋. 借鉴国外经验拓展国有商业银行个人理财业务[J]. 浙江金融, 2007(10): 18-19.

[4] 唐友伟. 个人金融信息保护工作亟待完善[J]. 金融会计, 2012(4): 64-66.

[5] Lehmann E, Neuberger D. Do Lending Relationships Matter?: Evidence from Bank Survey Data in Germany [J]. Journal of Economic Behavior & Organization, 2001, 45(4): 339-359.

[6] 鲁晓明. 论网络银行客户隐私权的保护[J]. 广东商学院

学报, 2005(2): 89-92.

[7] Breymann W, Dias A, Embrechts P. Dependence Structures for Multivariate High-Frequency Data in Finance [J]. Quantitative Finance, 2003, 3(1): 1-14.

[8] Bughin J, Chui M, Manyika J. Clouds, Big Data, and Smart Assets: Ten Tech-Enabled Business Trends to Watch [J]. McKinsey Quarterly, 2010, 56(1).

[9] Kim K, Prabhakar B. Initial Trust, Perceived Risk, and the Adoption of Internet Banking [C]// Proceedings of the Twenty First International Conference on Information Systems. Association for Information Systems, 2001: 537-543.

[10] Sathye M. Efficiency of Banks in a Developing Economy: the Case of India [J]. European Journal of Operational Research, 2003, 148(3): 662-671.

[11] Liao Z, Cheung M T. A Multi-dimensional Decision Framework to Support Corporate Bond Investment [J]. Journal of Decision Systems, 2012, 21(2):161-170.

[12] Selvapriyavadhana P. Mobile Banking Services on Data Protection Analysis In Networking [J]. IJRCCT, 2014, 3(4): 526-529.

[13] 邱峰. 大数据金融对商业银行的冲击和挑战分析[J]. 吉林金融研究, 2013(8): 10.

[14] 冯娟娟. 大数据金融背景下商业银行竞争策略研究[J]. 现代金融, 2013(4): 14-16.

[15] 梁璋, 沈凡. 国有商业银行如何应对大数据金融模式带来的挑战[J]. 新金融, 2013 (7): 47-51.

[16] 顾卿华. 快钱：互联网金融注重网络安全是先决条件[EB/OL]. (2014-03-18)[2014-08-01]. http://news.xinhuanet.com/fortune/2014-03/18/c_119827631.htm.

[17] 于潇. 于潇：互联网金融时代应该管理好客户动态行为数据[EB/OL]. (2014-01-15)[2014-08-01]. http://finance.sina.com.cn/hy/20140115/112017960365.shtml.

[18] 赵春霞. 爱投资赵春霞：理性面对互联网金融[EB/OL]. (2014-04-25)[2014-08-01]. http://news.xinhuanet.com/info/2014-04/25/c_133289622.htm.

[19] 石少功, 刘向晖. 论电子商务服务业产业集群的形成[J]. 未来与发展, 2008(11): 25-28.

[20] 井润田, 左齐. 信贷风险管理系统的开发实践[J]. 中国管理科学, 2002, 10(5): 30-34.

[21] 汤代禄. 互联网的变革[M]. 北京：电子工业出版社, 2007: 35-37.

[22] Dias A, Embrechts P. Dynamic Copula Models for Multivariate High-Frequency Data in Finance [J]. Manuscript, ETH Zurich, 2004: 9.

[23] Verizon. 2014 Data Breach Investigations Report [EB/OL]. [2014-08-01]. http://www.verizonenterprise.com/DBIR/2014/.

[24] Golany B, Storbeck J E. A Data Envelopment Analysis of the Operational Efficiency of Bank Branches [J]. Interfaces, 1999, 29(3): 14-26.

[25] Hamilton A. The Financial Revolution: the Big Bang Worldwide [M]. Viking, 1986: 45-55.

[26] 赵乐峰, 杜凯. 规范发展我国 P2P 网络借贷平台的思考[J]. 金融教学与研究, 2012(3): 33-36.

[27] LaValle S, Lesser E, Shockley R, et al. Big Data, Analytics and the Path from Insights to Value [J]. MIT Sloan Management Review, 2013, 21(1): 79.

[28] Aronova E, Baker K S, Oreskes N. Big Science and Big Data in Biology: From the International Geophysical Year through the International Biological Program to the Long Term Ecological Research (LTER) Network, 1957–present [J]. Historical Studies in the Natural Sciences, 2010(1): 183-224.

[29] 朱珠. 基于 Hadoop 的海量数据处理模型研究和应用[D]. 北京：北京邮电大学, 2008: 7-20.

[30] 王润华. 基于 Hadoop 集群的分布式日志分析系统研究[J]. 科技信息, 2009(15): 60.

[31] 多雪松, 张晶, 高强. 基于 Hadoop 的海量数据管理系统[J]. 微计算机信息, 2010(13): 202-204.

[32] Tusnady G E, Dosztanyi Z, Simon I. PDB_TM: Selection and Membrane Localization of Transmembrane Proteins in the Protein Data Bank [J]. Nucleic Acids Research, 2005, 33(S1): 275-278.

Customer Information Protection of Commercial Banks under the Background of Big Data Finance

Dong Jichang, Jiao Danxiao, Zhang Xin, Song Zijian, Li Xiuting

(*School of Management*, *University of Chinese Academy of Sciences*, *Beijing* 100190, *China*)

Abstract: Under the background of big data finance and with the appearance of new technology and new commercial activities, customers' information protection has become a new challenge to commercial banks. This paper analyses the influence mechanism of big data finance on customer information protection of commercial banks through case study domestically and internationally. We think that, under the background of big data technology, customers' information has three new characteristics: the volume of the information is big; the information is more valuable; and the divulging of information is destructive. Data processing includes three steps, which are data searching, data transmission and data storage. Information divulging happens through information desensitization treatment, data storage management and internal regulation breaking. The main reasons include hacker attacks, internal regulation breaking and management oversight. Then combined with the information divulging channels, reasons and results, we made specific case analysis. At last, from the perspective of law, management and technique, we put forward some advices/suggestions on the customer information protection under the background of big data finance.

Keywords: big data finance; commercial bank; customer information protection