科学通报 2015 年 第 60 卷 第 5-6 期：427 ~ 432
专辑：大数据研究与应用 观 点 www.scichina.com csb.scichina.com
《中国科学》杂志社 SCIENCE CHINA PRESS

# 大数据安全与自主可控

陈左宁*, 王广益, 胡苏太, 韦海亮

国家并行计算机工程技术研究中心, 北京 100080
* 联系人, E-mail: husutai@163.com

2014-08-15 收稿, 2014-10-17 接受, 2015-01-22 网络版发表
国家自然科学基金重点项目(51436007)和国家重点基础研究发展计划(2013CB228304)资助

**摘要** 继互联网、物联网、云计算之后, 大数据已成为当今信息技术领域的发展热点. 大数据在带来“大”价值的同时, 也存在“大”安全问题. 大数据的基本特征对计算设施、存储、网络、信息资源等提出了更高的安全要求, 传统的信息安全手段和管理机制已经跟不上大数据时代的信息安全形势发展. 本文在研究大数据安全新特点的基础上, 分析了我国大数据发展面临的信息基础设施自主可控程度低、安全防护技术和手段不足等问题; 阐述了自主可控对大数据安全的重要性和意义, 明确了解决大数据安全的根本之道在于实现我国主要信息产品、设备和技术的自主设计制造, 并总结了我国在大数据安全领域自主可控产品的发展现状. 大数据安全事关国家安全, 本文最后从加强大数据战略规划和安全体系建设、构建中国特色自主可控的发展路线、强化大数据技术在信息安全领域的创新应用等3个方面, 探讨提出了解决我国大数据安全的策略和办法, 以确保我国大数据时代的信息安全逐步朝着体系化、规范化和技术自主可控的方向发展.

**关键词**
大数据安全
自主可控
信息基础设施
网络攻击
技术创新

大数据安全包括2个层面的含义: 保障大数据安全和大数据技术用于安全. 前者是指保障大数据计算过程、数据形态、应用价值的处理技术, 涉及大数据自身安全问题; 后者是利用大数据技术提升信息系统安全效能和能力的方法, 涉及如何解决信息系统安全问题[1]. 大数据安全与自主可控相互促进、相互影响, 美国“棱镜门”事件让我们警醒, 只有自主可控才是解决大数据安全的根本出路. 我国信息技术自主可控程度和安全管理处于较低水平, 各行各业数据信息平台存在大量安全隐患. 因此, 我们必须加强大数据战略规划和安全体系建设, 构建中国特色自主可控的技术路线, 打造具有自主知识产权的软硬件产业链, 强化大数据技术在信息安全领域的应用, 以筑起大数据时代国家安全的铜墙铁壁.

## 1 我国大数据安全形势与存在问题

### 1.1 大数据安全的新特点

当前, 随着数据的进一步集中和数据量的增大, 传统的信息安全手段已经不能满足大数据时代的信息安全要求, 对大数据进行安全防护变得更加困难, 数据的分布式处理也加大了数据泄露的风险.

(1) 云计算设施为数据窃密创造条件, 安全威胁将持续加大. 随着大数据、云计算技术的发展和应用, 越来越多的大数据出现在云端, 而大数据在云端的集中存储处理, 使得安全保密风险也向云端集中, 一旦云端服务器违规外联或被攻击, 海量信息可在瞬间被集中窃取.

(2) 大数据成为网络攻击的重点目标, 加大了信息泄露风险. 大数据的“大”, 体现在数据被不断地处

**引用格式**: 陈左宁, 王广益, 胡苏太, 等. 大数据安全与自主可控. 科学通报, 2015, 60: 427–432
Chen Z N, Wang G Y, Hu S T, et al. Independence and controllability of big data security (in Chinese). Chin Sci Bull, 2015, 60: 427–432, doi: 10.1360/N972014-00812

理和利用后, 其价值会越来越大, 正因为如此, 大数据更易成为攻击者重点关注的大目标, 从而意味着大风险. 美国“棱镜门”事件显示, 美国通过云计算和大数据技术, 利用收集的公开数据并进行分析所获得的开源情报占其情报总量的80%~90%, 凸显了大数据时代信息泄露风险不断加剧.

(3) 大数据成为高级可持续威胁(APT)攻击载体, 应用于网络攻击手段. 数据挖掘和数据分析等大数据技术可以被攻击者用来发起高级可持续威胁攻击. 攻击者将APT攻击代码隐藏在大数据中, 利用大数据发起僵尸网络攻击, 能够同时控制大量傀儡机并发起攻击, 使得攻击更加精准, 从而严重威胁网络安全[2].

大数据的新特征对信息基础设施、存储、网络、信息资源等提出了更高的安全要求. 在大数据应用的整个过程, 需要关注传输数据的机密性保护, 采用大数据存储的隐私保护、备份技术, 研究关系型/非关系型数据库的大数据挖掘安全机制以及关注大数据发布审计技术; 此外, 针对APT攻击的防御技术也是需要研究的重点. 当前, 有关大数据安全的研究和实践已经逐步展开, 包括科研机构、政府组织、企事业单位、安全厂商等在内的各方力量, 正在积极推动与大数据安全相关的标准制定和产品研发, 以便为大数据的大规模应用奠定更加安全和坚实的基础.

### 1.2 我国大数据安全隐患与问题

大数据的发展使我国信息安全面临新的挑战, 现有的信息基础设施和安全手段已经不能满足大数据时代的信息安全要求.

(1) 国家信息基础设施自主可控程度低, 数据安全面临严重威胁. 众所周知, 无论是我国政府部门、企业的信息系统, 还是个人的PC、平板电脑、智能电话, 很多都采用了国外公司的产品. 全球IT巨头的思科、IBM、谷歌、高通、英特尔、苹果、甲骨文、微软等公司, 在过去几十年的中国信息化进程中, 一直扮演着重要角色. 我国通用处理器市场被英特尔和AMD等跨国公司垄断, 整机市场被IBM, HP和Sun等少数国际厂商瓜分, 操作系统、大型数据库等基础软件大部分来自国外企业, 思科等厂商所生产的网络设备在我国也占据市场主导地位. 这种现状严重影响着我国在大数据时代信息安全的基础.

(2) 国家大数据安全防护技术和手段不足, 难以提供有效安全保障. 目前我国的行业云承载着事关国家国计民生、经济运行的业务系统和数据, 云计算的发展必将导致信息在收集、传输、储存、处理等各个环节上进一步集中, 将使得信息安全问题成为中国云建设的焦点问题. 云计算环境采用虚拟化技术和多租户服务模式, 硬件资源的高度整合及网络架构的统一, 使得传统安全中的物理边界消失; 大数据在云端的集中存储, 使得原本分散在用户终端的安全保密风险向云计算中心集中, 一旦云端服务器遭到入侵, 或者云端系统提供方在系统中留有后门, 信息安全的堡垒将大门洞开; 数据拥有权与物理控制权的分离, 以及云端数据存储位置的不确定性, 使得数据所有者难以监管数据的安全性; 而作为云计算核心技术的虚拟化技术, 其安全体系的不明确和安全机制的不完善, 也带来了不容忽视的安全风险. 同时, 攻击隐藏在云中, 给安全事件的追踪分析增加了困难. 此外, 云计算服务软件分发和移动互联网接入的开放性, 也为网络攻击提供了更多的路径, 成为新的安全威胁. 而信息安全防护技术的演进和手段的创新远没有跟上大数据非线性增长的步伐, 对大数据进行安全防护变得日益困难, 大数据安全隐患更加凸显[2].

## 2 大数据安全呼唤自主可控

### 2.1 自主可控是实现大数据安全的根本出路

解决大数据安全的根本之道在于实现国家主要信息产品、设备和技术的自主设计制造[3]. 美国“棱镜门”事件曝出美国窃取全球多个国家的数据信息. 棱镜计划的实施方式为: 让美国各大技术公司配合, 由美国国家安全局通过各大技术公司的产品“后门”进入他们的信息系统, 并由此获取相关信息[4]. 这意味着如果我们的信息基础设备和行业云使用未经安全认证的外国、外资背景的厂商的设备或云服务, 将无法保障业务系统及其数据的安全性, 给国家信息安全形成潜在的威胁. 因此, 要想摆脱以棱镜计划为代表的美国政府各种计划的监听, 根本之道就是在于建立安全可信的信息系统, 而其基础是对关键设备的自主可控, 实现国家主要信息产品、设备和技术的可控管理使用. 从当前信息技术发展水平来看, “西强我弱”的情势会长期存在, 我国的信息产业还需要

更加开放，购买使用国外产品、设备和技术不可避免. 但前提是安全必须确实保障，这是发展与安全的辩证关系. 因此，对新技术新应用应做到“先审后用、能控则放、用中管控、安全审计”，不断提高对信息技术漏洞隐患的分析与发现能力、对技术产品和系统运用的风险评估能力，才能趋利避害，实现信息安全可控.

自主是安全基础，可控是安全目的. 实现自主可控虽有一定难度，但它是一个必然选择，是一个底线，是信息安全保障工作的根本出路.

### 2.2 我国大数据安全领域自主可控产品的发展现状

大数据安全涉及大数据应用过程的采集安全、存储安全、挖掘安全、发布安全等多个环节. 国家经过近几年的投入和产业发展，面向大数据安全的自主可控产品已经从注重研发到注重应用，从自我发展到产业整合，在市场化和产业化的道路上迈出了坚实步伐[5].

在网络传输设备这一领域，国内厂商生产的设备，无论是技术、质量还是价格，基本可以替代国外的产品，甚至是替代国外的高端产品. 比如锐捷网络公司全球第一台同时支持云数据中心和云园区网的核心交换机Newton 18000，是目前全球最高配置的核心交换机之一. 这表明国产设备已经从产品技术性能上、满足个性化需求等方面胜任或超过国外品牌.

在计算机软硬件方面，国产化自主可控产品已经初步具备体系化发展的基础. 在芯片方面，已经形成自主RISC指令架构的申威体系、基于MIPS架构的龙芯体系、基于SPARC架构的飞腾体系等，初步实现了CPU的自主可控；同时也在引进X86，ARM等主流架构，期望走出引进消化、吸收再创新之路. 在操作系统层面，借鉴开源的LINUX发展模式，国内已发展了中标、麒麟等操作系统，并可以适配国产的申威、龙芯、飞腾等处理器. 国内数据库产品已经有南大通用、达梦、人大金仓等公司研发的产品，虽然通用数据库技术积累不足还处于追赶者的位置，但在面向大数据处理的新型数据库方面，国内产品和国外产品已处于同一水平[1)].

在云计算系统方面，云计算已成为国家“十二五”规划最重要的战略部署之一，标志着自主云计算研发进入加速发展期. 2013年7月，阿里巴巴成为首个成功“去IOE”的中国公司. 阿里云“飞天”云计算平台已在金融服务、政府管理、医疗健康、气象、电子商务等多个领域应用. 曙光公司也凭借自主可控的创新研发实力积累的强大技术优势，已经掌握了包括云基础设施、云管理平台、云安全、云存储、云服务等一系列云计算核心技术与产品，可以为用户提供“端到端”云计算自主可控的整体解决方案.

在大数据分析处理设备方面，华为、浪潮、曙光等公司推出的大数据一体化解决方案，将为我国通信数据统计，互联网/移动互联网的日志和用户行为分析，物联网/传感器网络的数据监控和追踪分析，以及金融交易数据的离线统计和挖掘等众多行业的“大数据运营战略”提供“落地”工具.

可以说，我国大数据安全领域的自主软硬件产品发展势头良好，已经能满足一定范围的业务应用需求，为构建我国自主可控的大数据安全奠定了一定基础.

但是，从数据应用的角度看，我国仍然处于大数据时代的发展初期. 国内各行业、企业对大数据的安全防护与安全应用仍处于研究与摸索阶段. 虽然像腾讯、阿里巴巴等互联网公司的数据应用效果已较为明显，但对用户数据的保护方面还有很多漏洞. 目前，国内行业、企业的大数据分析技术与平台还存在信息容易泄露、安全技术落后、防控能力不足问题，亟待在加强自主可控技术产品使用的同时，从信息安全体系建设、大数据安全技术应用等方面加以解决[6,7].

## 3 构建我国自主可控的大数据安全之路

### 3.1 加强大数据战略规划和安全体系建设

保护大数据安全不单是技术问题，更是关系国家根本利益的战略性问题. 我们必须从事关国家安全的战略高度出发，加强国家信息安全体系的顶层设计，对大数据安全发展做出正确的战略判断和战略预见，确定国家大数据安全建设的总体思路、战略目标和优先秩序，理清大数据共享与保密、效益与安全的关系，制定大数据安全发展综合规划和行动计

1) 从“棱镜门”看国外软硬件产品的可替代性研究，中国软件行业协会财务及企业管理软件分会海比研究，2013

划, 整合健全信息安全组织管理机构, 强化统一领导和协调, 促使信息安全逐步朝着体系化、规范化、协调化和融合化的方向发展[8].

加快建设健全自主可控的大数据安全体系, 已成为当前我国信息化发展进程中亟待解决的问题. 大数据安全体系建设是一项系统工程, 由管理者、使用者、建设者, 管理对象、管理工具等要素组成, 要从策略、技术、管理、人员等各个方面综合考虑, 抓住重点, 协同发展. 我国亟须在国家统一的信息安全政策指导下, 针对大数据安全的新特点, 加快出台有关大数据挖掘收集、存储传输与处理利用的法律法规、条令条例、管理办法和标准规范; 建立严格的大数据服务、技术与产品准入制度, 限定大数据服务提供商必须在国家法律法规允许的范围内开展运营; 完善制度化的防范机制, 包括设备采购机制、信息定密机制、安全预警机制、应急响应机制、人员管理和专业队伍培养机制等, 通过强制性的内在安全建设, 提升大数据安全保障能力.

### 3.2 构建中国特色自主可控的发展路线

目前, 国外厂商还占据着我国信息产业的技术优势和市场垄断地位, 我国计算机用户也已形成固定的使用习惯, 导致自主软硬件产业要想完全通过所谓市场"公平"竞争得到发展非常困难. 因此, 必须立足我国信息环境的基本现状, 选准突破口, 依靠国家强有力的政策措施, 坚持产业化推进, 构建具有中国特色的自主可控技术发展路线.

(1) 自主可控技术的发展要以国家各类重要信息基础设施的保密信息平台建设为突破口, 推进国产芯片、国产网络设备、国产操作系统、国产数据库和国产云平台、云存储、云安全等关键软硬件产品的大规模应用, 分期分批实现关键软硬件的自主化. 重要信息基础设施作为培育自主软硬件产品的重要阵地, 通过内部使用, 可为市场树立样板, 进而带动普通用户使用自主软硬件产品. 对于保密信息平台的自主化要遵循"顶层规划, 强制推行; 统一部署, 分步实施; 加强测评, 充分验证"的建设思路, 以应用为牵引, 构建支持自主可控大数据安全的产业链条, 形成长效发展的机制; 完善面向保密应用领域的科研与采购机制, 使自主软硬件产品有条件、有渠道进入采购市场; 支持企业加强科研配套能力建设, 支持实力企业牵头实现自主一体化平台发展, 并对优势单位给予持续稳定的政策、市场与科研支持, 以保持其又好又快发展.

(2) 自主可控大数据安全技术的发展要以产业化为基础, 坚持以适用、可靠为基本原则, 按照核心环节可控的产业链要素配置. 一是通过产业化基地建设, 在产品研发、规模应用、服务保障体系等方面加强互动, 推动产业快速发展. 要做好产业化技术支撑, 全国产自主研制的整机厂商、应用系统开发商以及现已成立的相关联盟要密切配合, 提供配套的技术支持和服务, 互补共赢. 二是通过组建大数据产业联盟, 实现大数据产业链上下游的有效结合. 积极探索产业化运作模式, 探索生态圈内的市场开拓和商业模式, 使各"链环"厂商得到更好的发展, 共同建立起健康的"链条". 通过推进产业化成熟发展, 广泛开展合作, 认真分析用户需求, 提供优质服务, 真正使自主软硬件产品与技术在我国大数据领域得到广泛应用.

(3) 自主可控安全技术的发展要在应用环节上下功夫. 应用自主可控技术除了从政策导向上予以大力支持外, 相关部门应大力宣传和推广自主可控技术, 提升用户对于自主可控软硬件产品的市场认知, 引导用户最终选择自主可控产品. 同时, 要加大对盗版软件和技术专利侵权的执法力度, 加强大数据知识产权保护, 综合运用著作权、专利权、商标权、商业秘密权、反不正当竞争等多种保护手段, 全方位综合保护自主可控技术在大数据领域应用的持续健康发展.

### 3.3 强化大数据技术在信息安全领域的创新应用

大数据在面临自身安全问题的同时, 也给信息安全的发展带来新机遇[4]. 普遍认为, 大数据将会是整个安全行业发生重大转变的启动因素. 大数据分析将给信息安全领域包括信息安全事件管理、网络监控、用户身份认证和授权、身份管理、欺诈检测治理等在内的大多数产品类别带来足以改变市场的变化. 大数据技术将为安全分析提供新的可能性, 为我们提供一个更宽广的新视角, 帮助我们更前瞻地发现安全威胁, 提升我们信息安全防护系统的安全能力和安全效果. 因此, 我们要紧紧围绕云计算和大数据时代的安全保密要求, 加大对基于大数据的关键安全技术研发的资金投入, 抢占基于大数据的安全技术先机. 一方面针对云计算和大数据特点, 把可信计

算作为应对大数据安全威胁需要重点突破的技术方向，将可信计算纳入国家关键信息基础设施的安全体系，基于国产安全平台构建可信、可控的安全运行环境，加强信息安全防护技术的攻研，完善云计算和大数据安全技术体系；另一方面利用云计算和大数据技术，推进新型信息安全技术的创新，重点攻研高级可持续威胁攻击预测建模、基于大数据的网络攻击追踪技术、大数据支持的智能驱动安全技术、云安全服务技术、浏览器虚拟化技术、多级安全虚拟化桌面技术[9]，推动新型信息安全技术发展，形成自主核心技术优势，提高我国大数据安全技术水平.

## 4 结束语

大数据安全已经引发国际新一轮的技术竞赛，是信息安全领域新的技术增长极[2)]. 国际上，美国等发达国家已率先启动大数据安全技术研究和应用，我们在大数据安全领域面临新的挑战. 技术上的自主可控是大数据安全工作的根本出路，要组织和动员各方面力量，加强大数据安全战略规划和体系建设，提高大数据技术自主创新能力，力争在较短的时间内摆脱关键设备和技术受制于人的局面，逐步形成大数据传输、存储、挖掘、发布以自主可控技术和安全设备为主的格局.

## 参考文献


1 Zhang N, Zhang Y Y, Hu K, et al. Big Data Security Technology and Application (in Chinese). Beijing: Posts & Telecom Press, 2014 [张尼, 张云勇, 胡坤, 等. 大数据安全技术与应用. 北京: 人民邮电出版社, 2014]

2 Zhou D S. Risks and Measures of Information Security in the Age of Cloud Computing and Big Data (in Chinese). Research Report, Computer Science and Technology, Jiangnan Institute of Computing Technology. 2014 [周东升. 云计算和大数据时代的信息安全风险与对策. 研究报告, 计算机科学与技术, 江南计算技术研究所. 2014]

3 Li G J, Cheng X Q. Big data research: Significant strategic areas of future science & technology and economic social development (in Chinese). Bull Chin Acad Sci, 2012, 27: 647–657 [李国杰, 程学旗. 大数据研究: 未来科技及经济社会发展的重大战略领域. 中国科学院院刊. 2012, 27: 647–657]

4 Zhou L H. Big data security panic under prism (in Chinese). New Econ Wkly, 2013, 9: 81–85 [周路菡. 棱镜下的大数据安全恐慌. 新经济导刊. 2013, 9: 81–85]

5 Wang Y Z, Jin X L, Cheng X Q. Network big data: Present and future (in Chinese). Chin J Comput, 2013, 36: 1–15 [王元卓, 靳小龙, 程学旗. 网络大数据: 现状与展望. 计算机学报, 2013, 36: 1–15]

6 Bai J. When big data meets information security (in Chinese). Inform Secur Commun Priv, 2013, 5: 12–14 [白洁. 当大数据遇上信息安全. 信息安全与通信保密, 2013, 5: 12–14]

7 Hu Y, Hao Y Z. Using data to govern and strengthen the country—brief discussion of our country's big data national strategy (in Chinese). J Lover, 2013, 7: 4–8 [胡泳, 郝亚洲. 数据治国与数据强国——简论我国的大数据国家战略. 新闻爱好者, 2013, 7: 4–8]

8 Hu K, Liu D, Liu M H. Research on security connotation and response strategies for big data (in Chinese). Telecommun Sci, 2014, 2: 112–117 [胡坤, 刘镝, 刘明辉. 大数据的安全理解及应对策略研究. 电信科学, 2014, 2: 112–117]

9 Dou W C, Jiang C. Big data: Technical ecosystem and problem discovery (in Chinese). ZTE Technol J, 2013, 19: 8–16 [窦万春, 江澄. 大数据应用的技术体系及潜在问题. 中兴通讯技术, 2013, 19: 8–16]


2) BigData Across the Federal Government, http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_fact_sheet_final_1.pdf.

# Independence and controllability of big data security

CHEN ZuoNing, WANG GuangYi, HU SuTai & WEI HaiLiang

*National Research Center of Parallel Computer Engineering & Technology, Beijing 100080, China*

As a result of the Internet, the Internet of Things, and cloud computing, big data is becoming a focal area in current IT research. Besides providing significant value, big data also brings about large security problems. The key characteristics of big data require higher standards for computing infrastructure, storage, networks, information resources, and so on. Traditional information security measures and management mechanisms can no longer keep up with the requirements of information security in the big data era. Based on research on new security features of big data, this article analyses the problems faced by big data development in our country, such as the low degree of independence and controllability of the information infrastructure, and the lack of secure protection technologies and measures. The article also describes the importance and meaning of independence and controllability of the security of big data, suggests that the fundamental way of solving big data security is the independent design and manufacture of our main information products, devices, and technologies, and summarizes the current development of independent and controllable products in the big data security area. Big data security is related to the security of our country, and this article finally proposes various policies and methods to solve big data security in our country from three aspects, namely, enhancing strategic planning and construction of a security system for big data, building independent and controllable development guidelines with Chinese characteristics, and strengthening the innovative application of big data technology in the information security area. These steps would ensure that in the big data era, the development of information security in our country follows the direction of systematization and standardization, as well as independent and controllable technologies.

**big data security, independent and controllable, information infrastructure, network attack, technical innovation**

doi: 10.1360/N972014-00812