大数据驱动的安全协同生态建设

鲍旭华[1,2]，曲晓东[1,2]，郑新华[1,2]
1. 360企业安全集团，北京 100015；2. 大数据协同安全技术国家工程实验室，北京 100015

摘要
大数据技术发展给网络安全领域带来了挑战和机遇。新技术和新模式伴随着数据泄露、个人隐私风险、数据跨境流动、数据滥用等一系列安全风险，系统地介绍了大数据安全保障思路以应对这些风险。同时，大数据技术的发展为安全产业能力提升带来了巨大的机会，将会在大数据技术、智能安全模式和安全产业协同层面分别发挥作用。

关键词
大数据安全；安全协同；安全产业生态；国家工程实验室

中图分类号：TP391　　文献标识码：A　　doi: 10.11959/j.issn.2096-0271.2018033

Big data driven security collaborative ecological construction

BAO Xuhua[1,2], QU Xiaodong[1,2], ZHENG Xinhua[1,2]
1. 360 Enterprise Security Group, Beijing 100015, China
2. National Engineering Laboratory for Big Data Collaborative Security Technology, Beijing 100015, China

Abstract
Big data technology in network security field has brought challenges and opportunities at the same time. New technology and new model spring up with data breaches of privacy risks, cross-border data flows, data misuse and a series of security risks. The idea of big data security to deal with these risks was introduced. At the same time, the development of big data technology has brought great opportunities to the improvement of security industry capacity. Big data technology, intelligent security mode and security industry synergy will be developed.

Key words
big data security, security collaborative, security industry ecology, National Engineering Laboratory

1 引言

大数据是一场革命，它将改变人们的生活、工作和思维方式。我国高度重视大数据的战略意义和社会意义，国务院于2015年出台《促进大数据发展行动纲要》，并将“实施国家大数据战略，推进数据资源开放共享”纳入“十三五”规划。《促进大数据发展行动纲要》提出“加大大数据关键技术研发、产业发展和人才培养力度，着力推进数据汇集和发掘，深化大数据在各行业创新应用，促进大数据产业健康发展”，强调“推进基础研究和核心技术攻关。围绕数据科学理论体系、大数据计算系统与分析理论、大数据驱动的颠覆性应用模型探索等重大基础研究进行前瞻布局，开展数据科学研究，引导和鼓励在大数据理论、方法及关键应用技术等方面展开探索”。《中华人民共和国国民经济和社会发展第十三个五年规划纲要》提出：实施国家大数据战略，加快海量数据采集、存储、清洗、分析发掘、可视化、安全与隐私保护等领域关键技术攻关；完善大数据产业公共服务支撑体系和生态体系，加强标准体系和质量技术基础建设。

大数据意味着巨大的挑战。海量数据聚集在带来巨大价值的同时，也带来数据泄露、黑客入侵等安全风险，一旦发生危害将导致重大的损失。要化解安全风险，就必须突破大数据安全的核心技术，而且必须依靠我国的自主力量。习近平总书记在2016年的网络安全和信息化工作座谈会上指出“核心技术受制于人是我们最大的隐患。”突破数据安全核心技术，要从基础技术、通用技术、非对称技术、“杀手锏”技术、前沿技术、颠覆性技术等方面入手，一方面立足自主创新、自立自强；另一方面坚持开放交叉和协同创新，形成大数据安全技术协同创新平台。

大数据也意味着巨大的机遇。大数据为安全技术的发展提供了新的、强大的驱动力，海量多源异构的数据为更深入的安全分析提供了可能。大数据为智能化的安全运营提供了基础，无论是人工协作还是机器智能，都以此建立了新的模式。大数据为安全产业生态协同创造了条件，安全供应商的合作和协同带来的收益将远大于竞争。

2 大数据安全隐患和威胁

目前，我国各行业的大数据应用风起云涌，大数据在国民经济发展中发挥越来越大的作用，但是，大数据的安全问题也日益凸显，形势不容乐观。当前，我国在大数据安全领域存在以下几方面的问题。

一是数据泄露事件层出不穷。数据是重要资产，大数据意味着高价值的财富，容易遭受敌对分子的攻击和窃取。近年来，我国发生了多次重大数据泄漏事件，近日（2018年4月），饿了么、百度外卖、美团等外卖平台的个人信息发生泄露；在2017年，我国就发生过58同城全部简历数据泄露、优酷1亿用户账号泄露和浙江省松阳县警方侦破7亿条个人信息案件等重大事件。如果不采取更多措施，可以预计今后还将发生更多的数据泄漏事件。

二是个人信息安全保护形势严峻。随着“互联网+”战略的推进，越来越多的业务运营会采集个人信息，并对个人信息进行存储、处理，甚至共享。个人信息的非法收集、泄露、滥用等已成为社会关注的焦点问题。2018年1月，我国发生了“支付宝年度账单事件”，诱导用户同意收集个人信息并向第三方提供；2017年9月，在10款应用

隐私条款评审结果发布会上，全国信息安全标准化技术委员会秘书处有关负责人指出，长期以来，我国普遍存在隐私条款笼统不清、未给用户足够的选择权、大量收集与所提供服务无直接关联的个人信息、私自共享和转让个人信息等问题。

三是数据跨境流动问题影响信息产业“走出去”步伐。随着我国“一带一路”、企业“走出去”等战略的实施，国内电商、社交、游戏、移动互联网等优势领域企业正向境外进行大规模扩展，由此带来的个人信息和重要数据出境行为也日益频繁。2018年5月25日，欧盟将正式实施《通用数据保护条例（GDPR）》，其目的就是保障欧盟的数据安全，对在欧盟开展业务的境外企业有很严格的管理规定。美国、澳大利亚、新加坡等国家也颁布了类似的法规条例。如果数据跨境流动的问题处理不好，不仅影响对外业务的开展，也将影响我国境内数据的安全。

四是数据滥用成为重大安全隐患。通过大数据分析，可以发掘许多有价值的信息，甚至可能影响国家安全。例如，近期爆发的Facebook数据泄露事件表明，对社交数据的分析利用可以操作舆情，从而最终影响国家的选举结果，这就是数据滥用的结果。目前，我国平台型互联网公司掌握着大量数据，基于这些数据进行挖掘和利用，可以对经营管理乃至社会稳定产生重要影响；近期暴露的“大数据杀熟”事件就体现了这些公司有滥用数据的冲动。

五是大数据安全核心技术薄弱。美国禁止对中兴通讯股份有限公司供应芯片一事，暴露了我国核心信息技术受制于人的危险情况，在大数据领域也是如此。虽然我国许多企业提出了“去IOE”的口号，但是，大数据系统的芯片、操作系统、数据库等核心产品仍是国外垄断，我国国产系统仍无法替代西方的产品，距离形成完整的生态链则更遥远。一旦在大数据领域发生类似的事件，我国各大数据系统也将被“掐脖子”。

3 保障大数据安全

与传统的保障数据安全相比，保障大数据安全面临新的形势和新的风险。这种形势的变化来自大数据集带来的更大的安全威胁以及大数据生态复杂化引入的新风险。具体而言，与传统数据安全相比，大数据安全需要从两个维度来考虑：第一是安全防护的对象，包括系统和信息（数据）两类；二是安全需求的来源，包括入侵者和参与者两类。这两个维度共同构成了大数据安全内涵的4个象限，如**图1**所示。

- 信息隐私：在参与者使用合法途径获取授权数据的前提下，防止其结合外部知识，分析得到隐私信息。
- 信任机制：提供安全机制，使得参与者可以控制自己的哪些数据以什么形式被其他参与者获取。
- 数据安全：防止入侵者使用非法途径，获取非授权数据。
- 系统防护：防止入侵者使用非法途径获取系统控制权限或损害系统的正常运行。

大数据安全的保障需要4个方面的协作，任何短板都会带来隐患。

4 大数据技术驱动

长期以来，在信息安全的攻防对抗中，防守一方总处于被动局面，其根本原因是信息的不对称。攻击者可以自由选择入侵的对象、时机和方法，而防守方却只能时刻保持警惕，随时迎战任何可能的威胁。

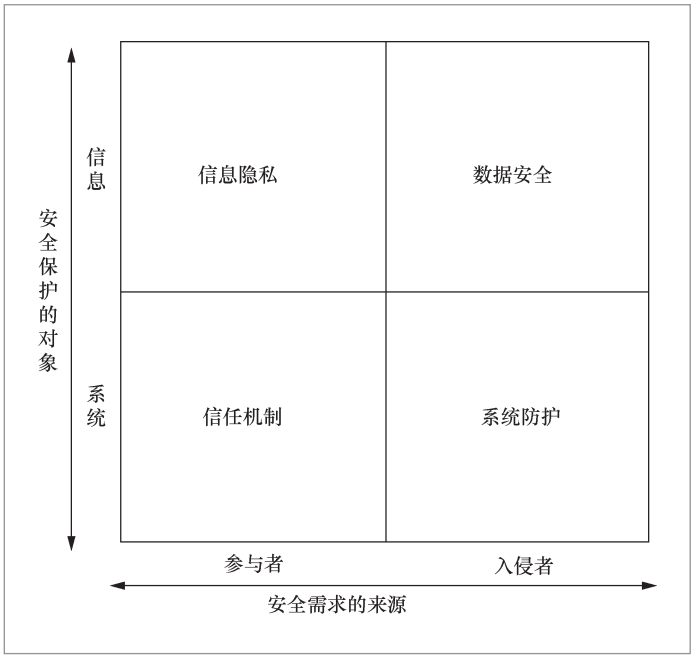

图1 大数据安全内涵

大数据带来的核心驱动力就是描绘出攻击者的行为模式，为防守方提供情报支持，减少这种信息不对称的情况。

数据量级不同时，处理的方法也不同。传统的安全检测以特征和规则为基础，需要一个特征提取和规则编制过程。但随着攻击方法的丰富和复杂，这个过程的成本越来越高。大数据技术的发展，使汇集海量数据进行协同分析、省略人工提取过程成了一种新思路，即数据量级协同。例如，360公司的QVM杀毒引擎采用了有穷向量机的机器学习方法，在超过100亿个样本库的基础上，不断进行迭代学习。新出现的恶意样本会被引擎自动识别，并成为下次迭代学习的基础。

传统安全信息和事件管理（security information and event management，SIEM）与安全操作中心（security operation center，SOC）产品通过提供关联算法和规则来检测高级威胁，同样需要人工定制的先验知识。但是以高级持续性威胁（advanced persistent threat，APT）为代表的外部威胁越来越复杂，隐蔽性也越来越强。对企业的安全团队来说，新型攻击出现太快，使得先验知识难以获取。异构数据协同的思路是将多个安全检测设备同时作为数据来源，进行多源数据协同分析，利用部分先验知识将微小的线索联系起来，由点及面，发现攻击行为。例如，安全业界普遍认为，传统的边界防御很难彻底抵御入侵者。对边界和内网中的终端、应用、网络等各种行为建立画像和基线，以用户和实体为核心，使用用户实体行为分析（user and entity behavior analytics，UEBA），综合利用统计模型和机器学习等方法，发现异常行为，将是更为有效的方式。

云地数据协同是指本地安全设备与云端威胁情报进行协同，以获取最新的先验知识。攻击者也需要考虑成本和收益的问题，一种新的攻击方法不会只出现一次，而是会被反复使用。对于特定企业第一次遭受的攻击，在网络中可能已经反复出现并被发现过多次了。因此，对于本地的安全防护系统来说，从云端获取最新的威胁情报，将成为基本的安全能力之一。

5 安全产业协同

数据协同和智能协同可以带来安全能力的提升，但最为重要的是，真正的革命将来自产业协同。协同带来的利益是双向的，一旦实现这种协同，安全供应商可以更专注独有的功能或服务，而用户可得到更强大的安全能力。产业协同可能有以下多种方式。

第一是自发式协同，各个厂商提供应用程序编程接口（application programming interface，API），供其他

厂商和客户直接调用，目前多数国际安全企业的协同采用这种方式。这种方式的优点是形式灵活，缺点是接口和服务质量的不统一容易造成混乱。

第二是联盟式协同，几个厂商组成对等的联盟，协商彼此交换的内容，例如PaloAlto Networks等厂商共享威胁情报的网络威胁联盟（Cyber Threat Alliance，CTA）。这种方式的优点是接口统一，缺点是同质化和封闭性。

第三是生态式协同，不同类型的厂商有组织地形成一个生态系统，采用开放透明的平台提供服务。这种方式的优点是稳定性和包容性强，缺点是需要足够开放稳定的平台。

Gartner的一份2015年的研究报告[①]认为，“到2019年，全球2 000强企业50%的对外服务和解决方案花费将通过不到10家组织生态系统的战略供应商提供。”采用生态系统供应商有诸多优点。首先，每类安全产品或服务都会有多家供应商在生态系统内，彼此良性竞争，可以为用户提供更多样化的选择。其次，安全产品和服务的种类在不断增加，彼此的联动也越来越复杂，生态供应商负责组织彼此间的协同，可以大幅提高管理效率。最后，对于不断涌现的新兴厂商提供的先进技术，企业客户进行尝试会面临较大的风险和代价，由生态供应商逐步尝试就可以减少这种风险。

安全技术的复杂性给客户带来了诸多怀疑，而生态供应商想要赢得企业客户，就必须提供一个开放、公正的平台。同一篇Gartner的报告提出：“到2019年，对供应商价值网络（包括分包商）中的运营和安全活动的透明需要，将使得对供应商安全和风险管理解决方案和服务的需求增加30%。”这样才可以形成良性循环，就像消费者市场中的苹果生态系统，最终为消费者提供了高质量的服务，同时使应用开发商和平台方提供方获利。

未来的安全产业生态协同（如**图2**所示）至少会包含3种角色：一是安全产品和服务的供应商，与传统供应商相比，其主要区别在于提供丰富的接口，能够与其他产

① Predicts 2016: IT vendor ecosystems must be re-evaluated based on agility, collaboration and risk

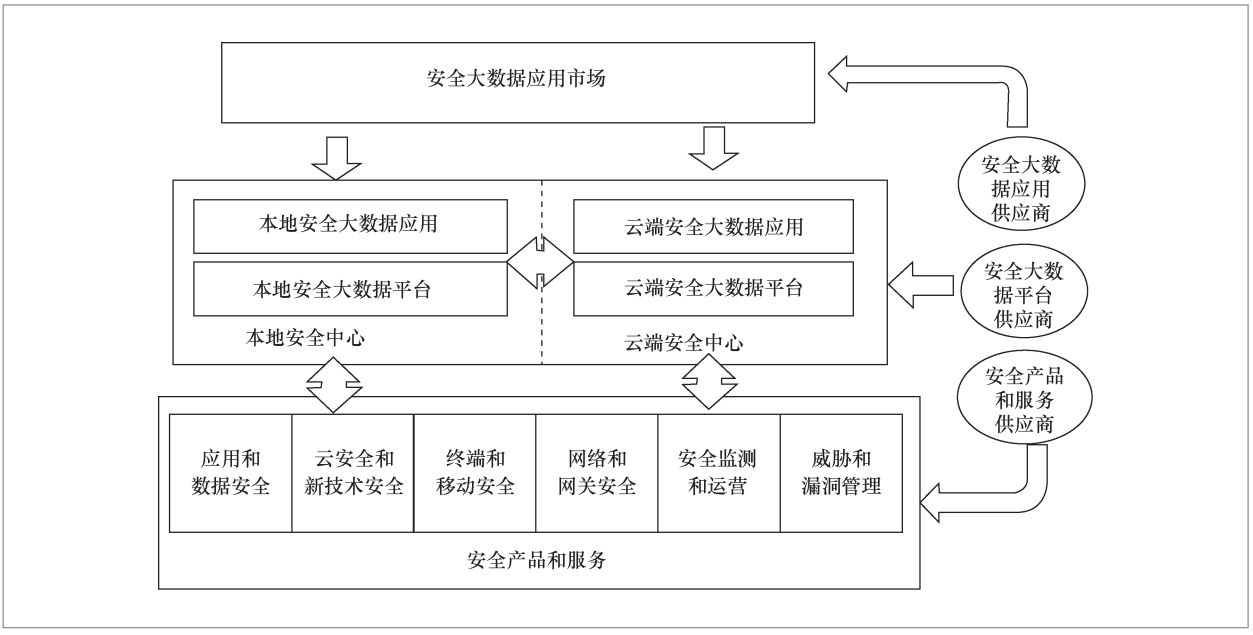

图 2　安全产业生态协同

品和服务以及大数据平台连接；二是安全大数据平台的供应商，包括本地平台和云端平台，这些平台可以收集产品和服务的信息，对接外部情报，进行综合分析，并对产品和服务提供协同指令，是未来安全生态的大脑；三是安全大数据应用供应商，他们在大数据平台的数据和功能基础上提供专项分析能力，以适应复杂多变的外部威胁和内部需求。三者各司其职，利益分享，为安全能力的提升共创未来。

6 生态协同机制

大数据协同安全技术国家工程实验室建成后，在技术支撑、产品服务、试点应用、资金数据4个方面形成“跨部门、跨区域、跨行业的具有全国性示范效应平台”。以企业为主体，以市场为导向，推进“产学研用测”的合作（如图3所示），实现体制机制、合作模式、创新人才培养三大突破，提高自主创新能力，加速产业结构的转型升级，整合产学研资源而形成的创新力，搭建大数据协同安全创新服务平台，提供大数据协同安全共性支持平台，吸引众多创新安全企业进入这个领域并支持它们发展。

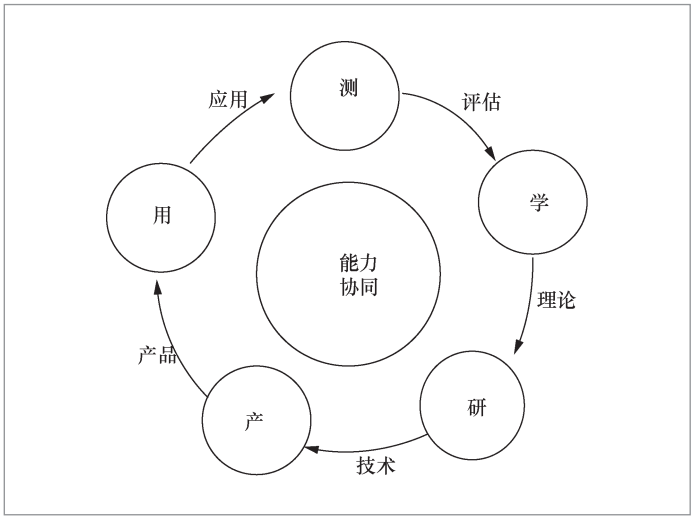

图 3 “产学研用测”合作循环

该实验室集中突破23项核心关键，形成图4中覆盖数据、系统、网络、终端的7类协同技术（情报协同、病毒协同、漏洞协同、入侵协同、众测协同、应急协同、溯源协同）。例如采用终端安全技术发现泄露的涉密文件，就可以通过溯源协同关联网络安全、系统安全和数据安全，在网络和系统中追溯涉密文件泄露的途径和过程，进而对数据安全的防护进行针对性的强化，从而达到监测与处置的协同效果。

该实验室基于云计算和虚拟化技术构建大数据协同安全共性支持平台，为生态合作伙伴提供基础研发支撑环境。大数据协同安全共性支持平台采用云计算和虚拟化技术，搭建标准统一、功能完善、系统稳定、安全可靠、集中统一的开放式、可扩展的基础支撑平台，以复杂数据关联分析、大数据搜索、大规模机器学习等技术支撑大数据安全分析，为大数据协同安全的各项工作提供流式和批量数据接入，非结构化、半结构化和结构化数据存储，批处理计算、迭代计算、流式计算和高性能计算等基础支持。大数据协同安全共性支持平台已经对公众发布了一系列公开项目，具体如下。

- XLearning人工智能开源平台：大数据与深度学习相融合，基于Hadoop Yarn完成了TensorFlow、MXNet、Caffe、Theano、PyTorch、Keras、XGBoost等常用深度学习框架的集成，是典型的“AI on Hadoop”的实现。XLearning则可以帮助人工智能开发者实现调度的统一和服务器资源的复用。随着平台算法库的不断增容和优化，开发者工作难度将大大降低。
- 分布式拒绝服务（distributed denial of service，DDoS）攻击威胁信息共享系统：是最大的面向整个互联网提供DDoS监测和告警服务的系统之一。目前，

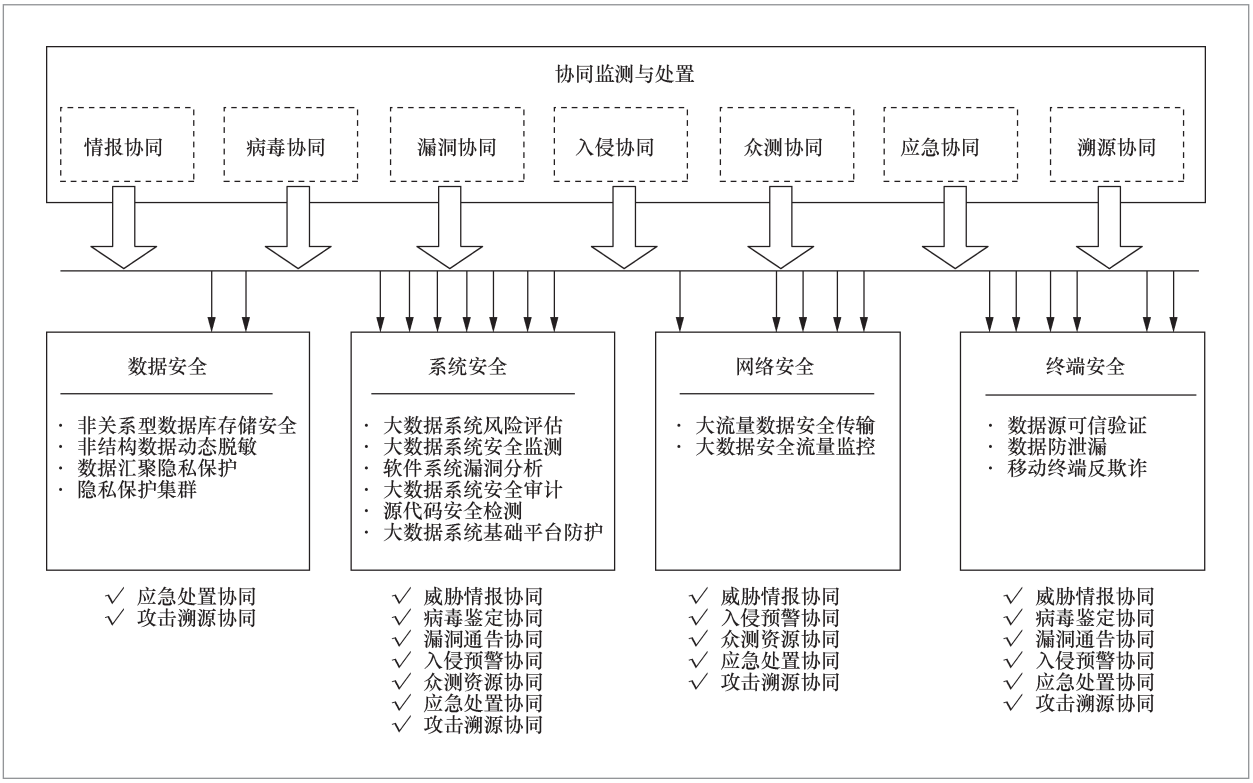

图 4　7 类协同技术

平均每天发生超过3万个DDoS攻击。对于全球大规模的DDoS事件，该系统能够在第一时间成功探测出来，在准确度和及时性方面都有非常好的表现。来自不同公司和组织的工程师都在基于该系统进行日常安全维护和研究工作。

● 网络扫描信息共享系统：是专门记录网联网扫描活动的系统。针对互联网中被攻击者广泛应用的网络扫描进行被动监测，本系统记录扫描来源、目的、频度等信息，平均每天记录超过1万个扫描IP地址的活动，供分析使用。

● 威胁情报共享系统：是面向公众开放的威胁情报数据查询服务系统，于2015年开始对外开放。安全厂商、政企用户等经线上注册并审核通过后，都可以在系统对域名、IP地址、样本信息等进行查询，系统将反馈云端数据查询结果。技术人员结合这些信息进行分析，可以对定位安全威胁发挥关键作用。系统具有关联分析和海量数据两大特色，可以将用户提交的查询信息关联起来，协助用户进行线索拓展，挖掘出在企业或组织内部分析中无法发现的更多线索。

7 结束语

本文介绍了大数据技术发展为网络安全领域带来的挑战和机遇，阐述了数据泄露、个人隐私风险、数据跨境流动、数据滥用等一系列安全风险以及应对这些风险的大数据安全保障体系。同时，大数据技术的发展也为安全产业能力提升带来了机遇，大数据协同安全技术国家工程实验室会在不同层面开展工作，推动整个安全产

业对这种机遇的把握，包括大数据驱动的安全技术发展、人机智能安全协作模式、安全产业生态协同等。

参考文献:

[1] 冯登国, 张敏, 李昊. 大数据安全与隐私保护[J]. 计算机学报, 2014, 37(1): 246-258.
FENG D G, ZHANG M, LI H. Big data security and privacy protection[J]. Chinese Journal of Computers, 2014, 37 (1): 246-258.

[2] 王珊, 王会举, 覃雄派, 等. 架构大数据: 挑战、现状与展望[J]. 计算机学报, 2011, 34(10): 1741-1752.
WANG S, WANG H J, QIN X P, et al. Architecting big data: challenges, studies and forecasts[J]. Chinese Journal of Computers, 2011, 34(10): 1741-1752.

[3] NIST. NIST big data security and privacy subgroup: NIST SP 1500-4[S]. 2015

作者简介

鲍旭华(1977-)，男，博士，360企业安全集团战略高级工程师、研究主任。毕业于中国科学院信息安全国家重点实验室。在信息安全领域从业十余年，主要研究方向为安全态势感知和DDoS防范，曾在中国信息安全测评中心、信息安全共性技术国家工程研究中心、绿盟科技、华为等单位担任管理和研究职务。申请发明专利5项，发表学术论文10余篇，出版专著《破坏之王：DDoS攻击与防范深度剖析》。目前兼任网络安全监测预警关键技术北京市工程实验室副主任、国家信息中心安全大数据开发与治理中心副秘书长、工业互联网标准委员会安全组组长等。

曲晓东(1971-)，男，360企业安全集团高级副总裁，主要负责销售、供应链、政府关系及品牌运营与管理。拥有20年媒体行业从业经历，是IT行业的知名专家，中国计算机学会理事，中国计算机学会青年科技工作者论坛(YOCSEF)荣誉委员，中国计算机学会信息安全专家委员会常务委员。2012年加入360公司，先后担任公关、市场业务部副总裁，2014年开始负责360公司的企业安全业务，参与投资和管理网神、网康等多家子公司，在管理融合、企业文化融合与品牌融合等方面均具有丰富经验。支持团队获得中国通信学会科学技术奖一等奖、北京市科学技术奖二等奖等。指导完成工业和信息化部电子信息产业发展基金项目、国家发展和改革委员会信息安全专项、工业和信息化部重大专项、国家高技术研究发展计划等多项国家级项目。

郑新华(1980-)，男，360企业安全集团战略顾问，中国系统工程学会监事，北京信息科技大学硕士校外导师。长期从事计算机、软件工程和系统工程领域的研究和应用，主持或参与制定了我国多部大数据、威胁情报相关的标准，发表论文20余篇，参与编制《大数据领导干部读本(第2版)》《大数据标准化白皮书(2018)》《中国网络安全与信息化产业桔皮书(2017年)》等。

收稿日期: 2018-04-24