doi:10.3772/j.issn.2095-915x.2018.02.010

基于文献分析的信息安全领域发展与趋势研究

管洋洋[1] 康翠翠[1] 李镇[1] 曹自刚[1] 高剑波[1] 牛温佳[2] 谭建龙[1] 郭莉[1]

1. 中国科学院信息工程研究所 北京 100093;
2. 北京交通大学 北京 100044

摘 要 本文以信息安全领域相关的国际知名期刊和会议等数据为基础，采用大数据集下的数据挖掘技术，实现顶级科研论文的大数据分析原型系统。本文研究了国际信息安全领域的主要研究单位、研究进展以及研究趋势。同时，研究为中国科学院和创新研究院在信息安全学科方向和研究力量布局方面提供分析结果和决策依据。

关键词： 信息安全；研究现状；发展趋势；文献计量
中图分类号： G35
开放科学（资源服务）标识码（OSID）

The Development And Trend Analysis of Information Security Field Based On Academic Publications

GUAN Yangyang[1] KANG Cuicui[1] LI Zhen[1] CAO Zigang[1] GAO Jianbo[1] NIU Wenjia[2]
TAN Jianlong[1] GUO Li[1]

1. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China;
2. Beijing Jiaotong University, Beijing 100044, China

Abstract Based on the data related to internationally renowned journals and conferences in the field of information security, this study used the data mining technology under large data sets to construct the big data analysis prototype system for top research papers. This paper studied the major research institution,

基金项目： 中国科学院信息学工程研究所重点项目资助（Y520051101）。

作者简介： 管洋洋（1988–），硕士，研究方向：信息安全，数据处理，Email：guanyangyang@iie.ac.cn；康翠翠（1987–），研究方向：机器学习，数据挖掘；李镇（1987–），硕士，研究方向：信息安全，大数据处理；曹自刚（1987–），博士，研究方向：信息安全，大数据处理；高剑波（1966–），博士，研究方向：复杂性科学与大数据研究；牛温佳（1982–），博士，研究方向：人工智能、机器学习和数据挖掘；谭建龙（1974–），博士，研究方向：算法设计、数据流管理技术研究和信息安全；郭莉（1969–），博士，研究方向：信息安全，网络安全，数据流处理。

research progress, and research trends in the field of international information security. Moreover, this research also provided the academic results and decision-making basis for the Chinese Academy of Sciences and the Institute of Innovation in the direction of information security disciplines and the layout of research power.

Keywords：Information security; current status; future development; bibliometric

1 引言

近年来，随着互联网的发展和移动智能终端的普及，中国的互联网技术已经扩展到了千家万户。据中国互联网络信息中心发布的《中国互联网络发展状况统计报告》指出，截至2015年6月份，我国网民规模已达6.68亿，互联网普及率为48.8%。从城市到农村，从儿童到老人，中国逐步进入了全民联网的时代[1]。与此同时，人们在互联网的影响下，生活方式在无形中逐步发生改变。我们不仅从互联网上可以得到信息获取和休闲娱乐的个性化需求，还可以获得一系列的基本生活服务，如交水电费、购买火车票和飞机票，甚至教育和医疗等民生服务。互联网技术的发展，智能终端应用水平的不断提高，成为了推动社会进步和国家经济发展的巨大力量。然而，伴随着互联网在人类生活中的进一步深化和融合，个人隐私信息泄露等网络安全问题层出不穷[2-3]。

针对互联网安全问题，我国目前已全面加强针对网络安全的依法管理和科学管理，加速构建网络安全保障体系[4]，以确保我国人民的切身利益和安全、推动互联网以及相关产业的快速健康发展。然而，确保网络与信息安全需要大批具有高素质、高技术水平的复合型人才。虽然我国信息安全人才队伍建设发展迅速，且成绩显著，但是仍然不够系统化、规模化、体系化以及前瞻化。

为了促进我国在信息安全领域更好的发展，我们爬取了计算机学会推荐国际学术刊物上关于信息安全领域相关的论文数据，采用数据挖掘技术，研究国际信息安全领域的主要研究单位、研究重点以及研究进展[5]，从群体研究趋势中把握未来信息安全趋势，了解国际科研单位的研究现状、技术水平和前瞻研究方向，为我们的信息安全研究力量布局方面提供分析结果和决策依据[6-10]。

2 数据采集

为了让分析结果更具可靠性和全面性，我们通过于2012年公布的《中国计算机学会推荐国际学术刊物》，结合人工筛选找出信息安全方面的顶级国际期刊和顶级国际会议，筛选出相关的国际期刊和会议的信息。事实上，大多数这些刊物的出版商都是IEEE、ACM、Elsevier以及Springer中的一家。因此，我们省略了一下其他数据来源，选择科研论文数据量相应比较大的这四个国际主流论文出版社IEEE、ACM、Elsevier和Springer，在信息安全领域进行近十年来科研论文数据的采集工作。利用基于网络爬虫的定点采集方法，针对数据

doi:10.3772/j.issn.2095-915x.2018.02.010

源对采集频度的限制问题，采用代理 IP 技术实现数据的高效采集。论文主要分布在四个出版社中 52 个主要期刊上，其中国际会议有 the Annual International Conference on Information Security and Cryptology（ICISC）、Conference on Selected Areas in Cryptography（SAC）、Theory of Cryptography、International Conference on Practice and Theory in Public Key Cryptography（PKC）、The IEEE International Conference on Trust, Security and Privacy in Computing and Communications（TrustCom）等，国际杂志有 Computer Fraud & Security、IEEE Transactions on Information Forensics and Security（TIFS）等。

最后，整个数据库可以分成两个结构表。主表 paper 以及论文被引用表 cited_by。其中，科研论文主表的字段设计如表 1 所示。

表 1 信息安全领域的科研论文数据主表 paper 字段

论文题目（title）、
摘要（abstract）、
关键词（keywords）、
出版商（publisher）、
版页（pages）、
发表位置（published_In）、
出版日期（publication_date）、
发行日期（issue_date）、
赞助者（sponsorer）、
第一作者（first_author）、
第一作者单位（first_author_unit）、
合作者（co_author）、
链接（url）、
主键（p_id）

科研论文被引用表的字段设计如表2所示。

表 2 信息安全领域的科研论文数据被引用表 cited_by 字段

主键（c_id）
被引用完整信息（c_info）
论文 ID（p_id）
主键（p_id）

3 数据挖掘与分析

3.1 数据库的组成

通过上述描述可知，信息安全领域科研论文数据集主要从 Springer、ACM、IEEE 以及 Elsevier 四个国际主流出版商上进行数据采集。从 2005 年开始截至到 2015 年 7 月份，共采集到了信息安全领域的 28,748 篇国际会议论文、国际期刊论文和书籍的相关信息。通过初步整理，我们获得了该数据库国际科研论文出版的一些基本情况，如图 1 所示。

图 1 展示了 2005 年至 2015 年 7 月十年间信息安全领域科研论文逐年的发表情况。由于进行数据采集时，2015 年尚未结束，因此，2015 年的数据统计并不完整。可见在信息安全领域，Springer 出版社是信息安全领域论文发表的巨头。从该图中可以看出，早在 2005 年与 2006 年，信息安全领域的国际研究就已经相当活跃。其次，2011 年是国际信息安全领域科研论文发表最多的年份，信息安全问题在这一年受到了极大的重视。随后的 2012 年至 2014 年，信息安全领域的研究热度依然不减。2011 年，美国中央情报局网站遭黑客联盟 LulzSec 攻击，韩国电信运营商 SK Communications 的信息外泄事件造成韩国用户账户信息大量遭到曝光，美移动运营商指定厂商安装窥探用户隐私软件，装载 Android 系统的智能手机和其他智能工具受到了大量病毒程序的侵扰，甚至国内知名网站 CSDN 证实 600 余万用户资料被泄露，新浪微博突然出现大范围“中毒”，多名用户账号疑似被黑且自动发送垃圾信息。这一年的互联网安全受到了极大的威胁，互联网及其相关产业

的蓬勃发展受到了一定的影响，互联网的信息安全问题再一次得到了广泛的重视[5]。

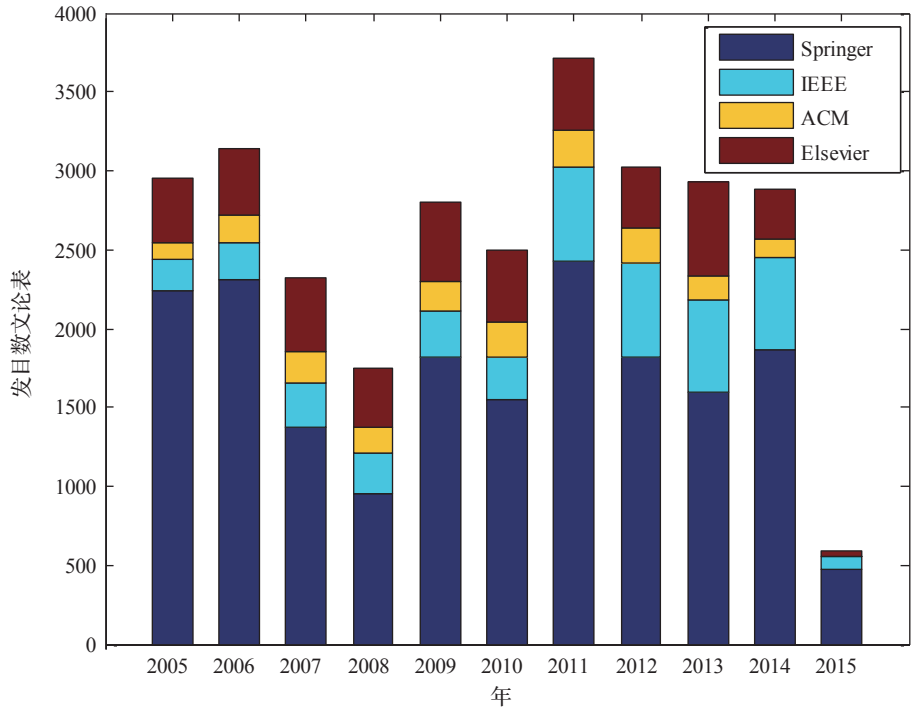

图 1 2005–2015 年信息安全领域科研论文数据集逐年论文发表情况

此后，2012 年全球最大天然气生产商遭到电脑病毒攻击，继 2013 年的“棱镜门”时间之后，2014 年，美国中情局前特工爱德华·斯诺登持续不断地向世人再次揭露美国国家安全局、英国国家通信总局 (GCHQ) 以及其他政府的监听计划，表明被关注监听的不仅仅是那些大企业。现在，互联网信息安全不仅关系到网民用户的个人利益，还与国家安全密不可分，涉及到国家政治、军事和经济等各个方面，影响到国家的安全和主权。显然，互联网信息安全已经成为二十一世纪世界十大热门课题之一。

3.2 各国科研论文发表情况

图 2、3 分别给出了 2005 年至 2015 年各国在信息安全领域的研究情况，包括主要研究单位数目、科研论文发表数目和论文作者数。在该科研论文数据库中，存在 10000 多个不同的国际研究单位和更多的科研人员，遍布 77 个不同的国家和地区。通过图 2 可知，科研论文的成果产出数量与研究单位数量有一定的相关关系，但并非绝对。图 3 分别列出了根据科研论文产出成果排名统计得到前十五个主要研究国家在信息安全领域的研究情况。这十五个国家分别是美国、法国、英国、中国、德国、日本、以色列、印度、澳大利亚、比利时、加拿大、瑞士、新加坡、韩国以及荷兰。其中，中国排名第四。可见，尽管中国在信息安全领域的相关研究起步较晚，但是成绩显著。然而，美国在信息安全领域的科研论文成果产出遥遥领先。

doi:10.3772/j.issn.2095-915x.2018.02.010

其科研单位数量，科研人员和科研论文发表数量都是我国相关研究的两倍之多，我国与其差距依然显著。从图 3.4 可以看出，我国的科研人员人均论文产出处于中等偏上水平，排名第七。人均论文发表数目超越我国的国家有：日本、法国、以色列、美国、德国、印度。期间，2005 到 2015 年中国信息安全科研作者人均论文产出排名分别是 12，6，2，3，13，5，11，4，3，12，6，在 2007 年和 2012 年左右成绩显著。

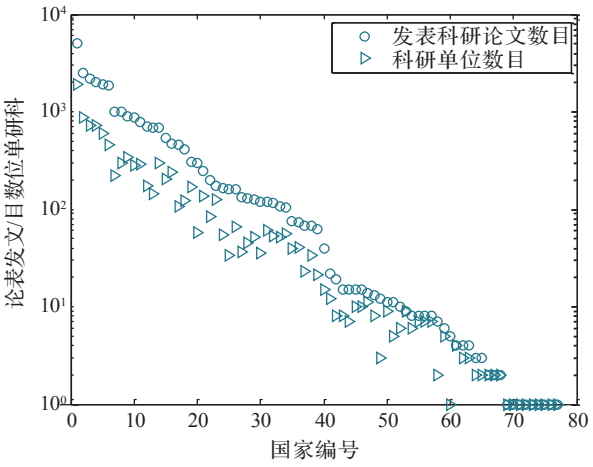

图 2 2005–2015 年信息安全领域各国研究情况统计

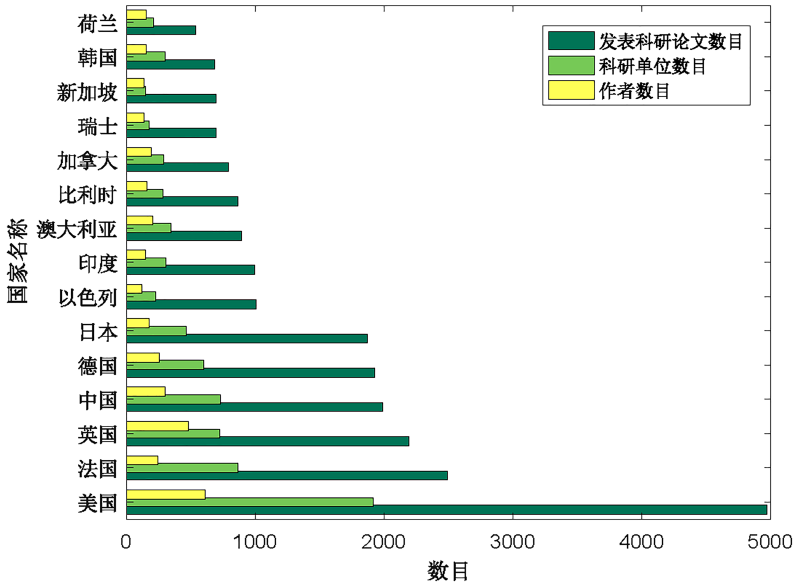

图 3 2005–2015 年前十五个主要国家科研论文发表数量、科研单位数及科研作者数统计

3.3 论文分类

为了研究信息安全领域世界各国的关注点以及各方向的发展状况，我们首先人工阅读引用率高的经典论文熟悉各个方向，查阅相关文献 [1]，总结出信息安全领域的五大研究方向，分别是密码学、网络安全、信息对抗、信息系统安全、信息内容安全。本文中，我们利用机器学习算法，将每篇论文归到其中最适合的一类中。因为每篇论文可能不仅仅只涉及其中的一个方向，所以采用监督的方式去分类会丢失掉该论文还属于其他类别的性质，于是我们采用了无监督方式进行聚类分析 [11-13]。主题模型是文本挖掘领域非常流行的方法，在文档分类、聚类中都有大量的应用。因此，为了进一步分析该科研论文数据集的研究内容和研究方向，我们对该科研论文数据集的所有标题和摘

要进行了进一步的处理，包含分词、去除停用词、词干提取等步骤。最后采用无监督主题模型（Topic Model）方法，对科研论文进行了文档分类 / 聚类分析。在这里，我们采用了经典的 Latent Dirichlet Allocation（LDA）主题模型算法 [14]，提取出每篇论文的特征，然后通过 K-means 聚类算法 [15]，以及人工分析，将所有学术论文分类到信息安全领域的五大一级学科中。该数据库在五个一级学科的分布情况如图 5 所示。

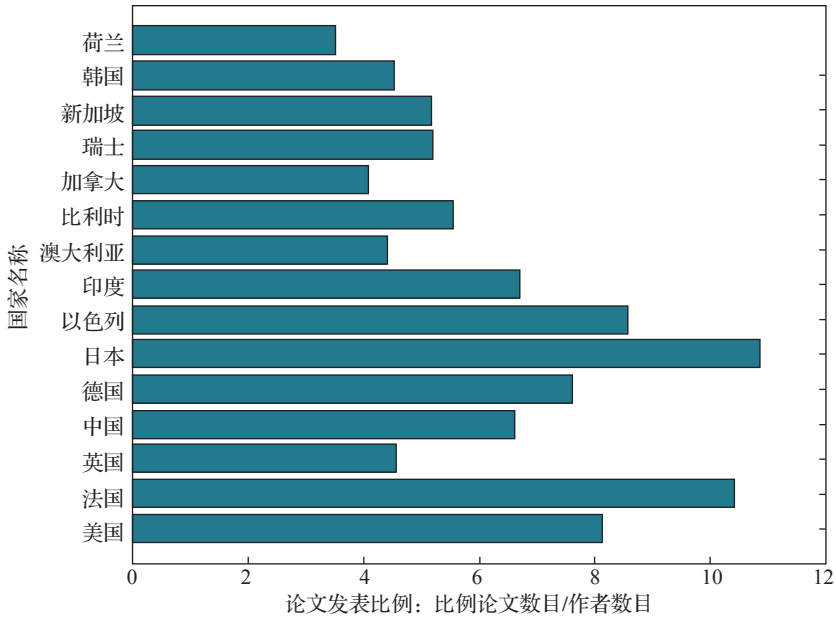

图 4 2005–2015 年前十五个主要国家人均论文发表数

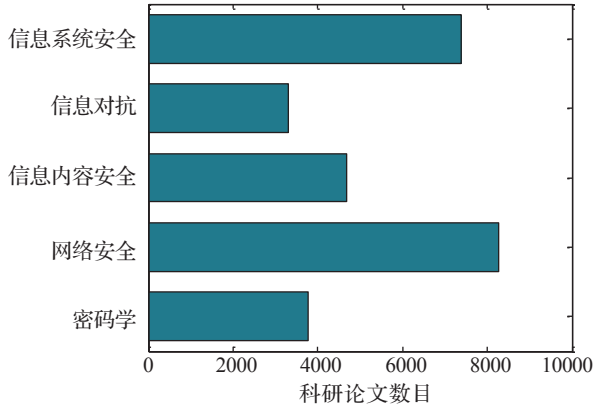

图 5 2005–2015 年信息安全领域 5 大学科论文发表情况

从图 5 中可以看出，随着互联网的不断发展，网络安全一直是信息安全领域比较关注的方向，这从今年国务院学位委员会、教育部新设立一级学科“网络空间安全”可以想象到网络安全将会继续得到较大的重视和发展。

图 6 和图 7 展示了信息安全领域各一级学科在 2005 年至 2015 年的逐年国际研究进展。由于 2015 年的数据不全，我们可以暂且忽略该年。从图 6 中，我们可以看到各方向的发展状况。随着网络技术的普及，网络安全的研究热度一直不减，每年均有大量的研究成果。其次信息系统安全的研究也比较重视。从图 6 中可以发现，在 2011 年和 2012 年，各个方向分别达到了研究的热潮，此后一直保持着较好的研究热度。从各年比例图 7 中也可以看出，网络安全在 5 大学科中占据主要地位，信息系统安全也占较大比重，而其他三门学科均处于较稳定的状态。

doi:10.3772/j.issn.2095-915x.2018.02.010

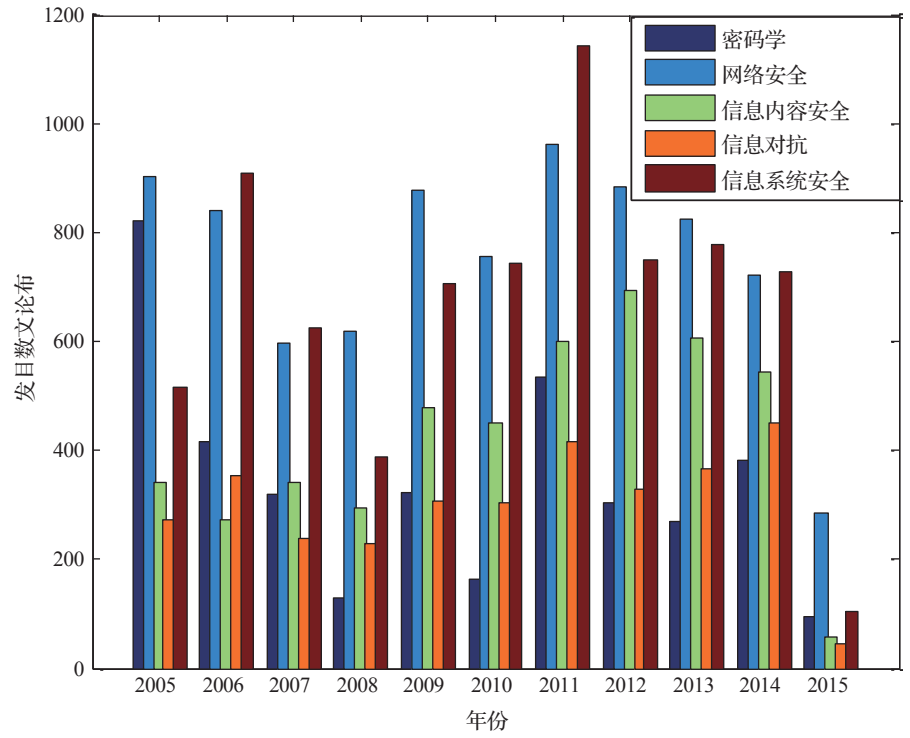

图 6 2005–2015 年信息安全领域 5 大学科科研情况

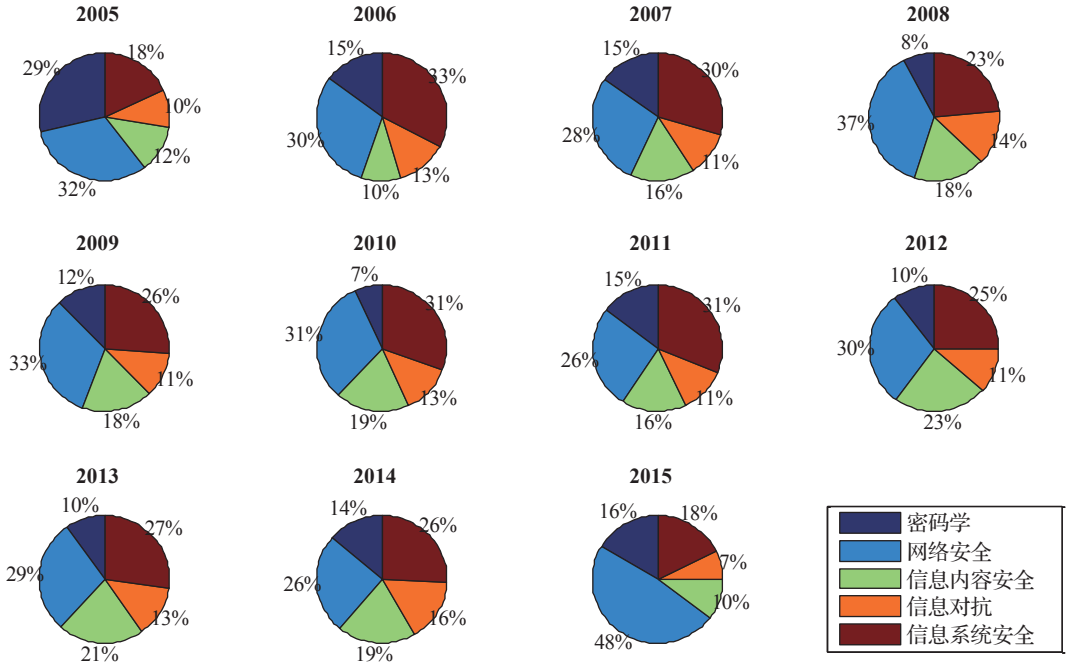

图 7 2005–2015 年信息安全领域 5 大学科科研比例图

了解了国际上总体的情况，图 8 和图 9 给出了中国在这五个学科上的研究发展情况。从柱形图和各年比例图中可以看出，中国于 2012 年加大了信息安全领域的研究力度，信息安全学科的建

设得到更高的重视，和国际相比起步较晚。并在网络安全、密码学方向较为重视，其中网络安全在 2012 年取得较好的研究成果，信息内容安全和信息系统安全则刚刚发展起来，处于起步阶段。

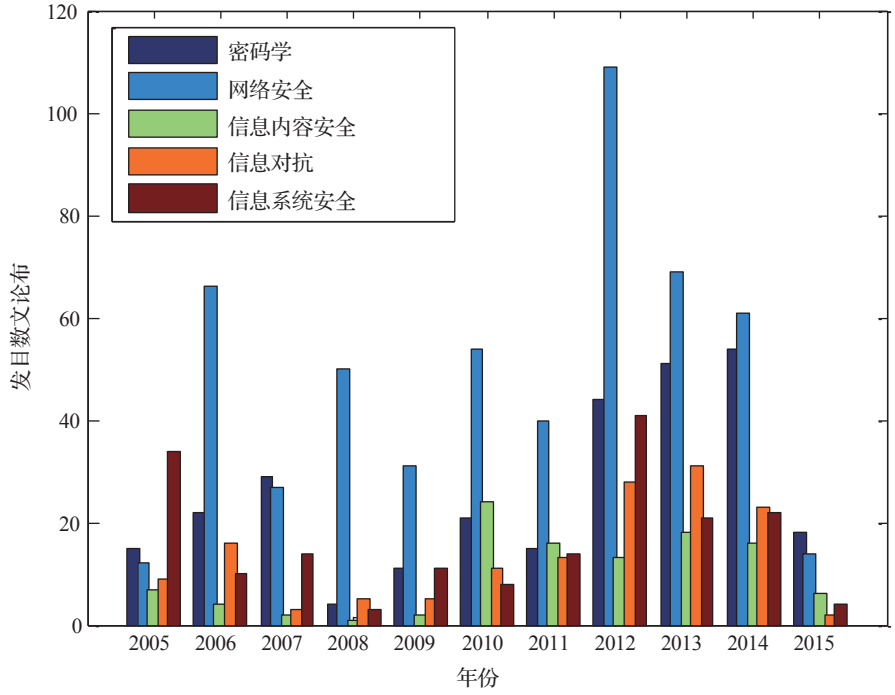

图 8 2005–2015 年信息安全领域 5 大学科科研情况 – 中国

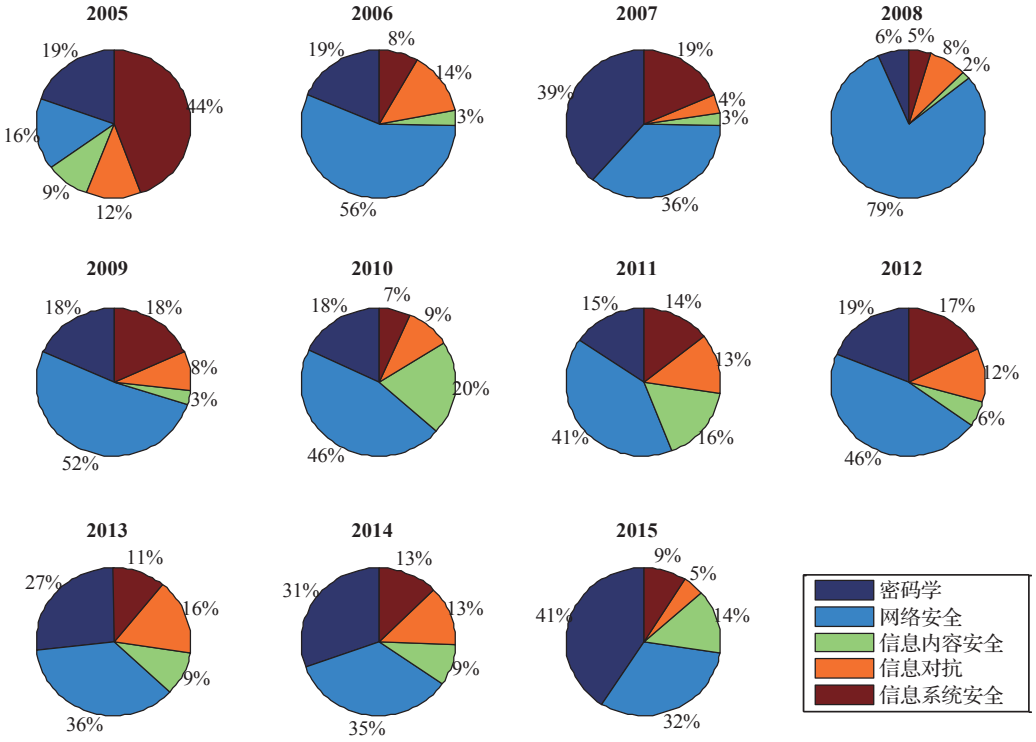

图 9 2005–2015 年信息安全领域 5 大学科科研比例图 – 中国

doi:10.3772/j.issn.2095-915x.2018.02.010

图 10 是中国在信息安全领域的论文发表情况与国际的比较图。可以看出，中国在信息安全领域的研究起步较晚，但是已取得不错的研究进展。具体地说，中国主要是在密码学和网络安全方面取得了一定的研究成果，而且从 2012 年起，每年都取得了一定的成绩。而在信息内容安全、信息对抗、信息系统安全三个方向虽然比之前有了很小的发展，但是和国际总体相比还是相差很多，科研成果不到国际科研成果的 10%，发展较为缓慢。

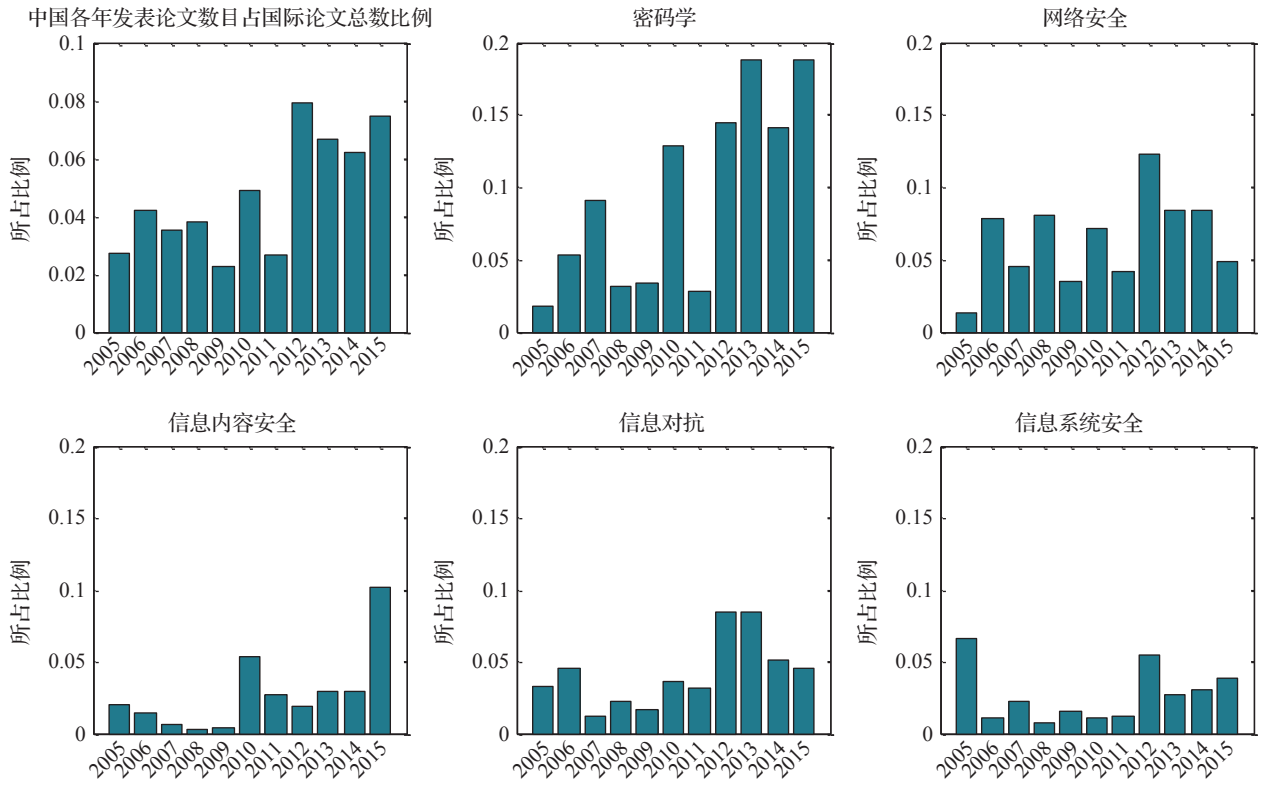

图 10 2005–2015 年信息安全领域中国在世界上的发展情况

3.4 科研论文被引用情况

此外，该科研论文数据集还采集了论文被引用信息数据库子集，可以统计得到科研论文数据集中论文的被引用情况，如图 11 所示。由于后发表的文章都是引用先前发表的文章，所以图 11 显示的下降趋势是合理的。

图 12 拟合出了论文与该论文被引用次数的关系曲线。在该数据集 2005 年至 2015 年的科研论文中，有将近 8000 条论文存在被引用记录，部分论文被引用次数小于 5 次。从图 12 中可以看出，500 多篇科学论文被引用次数超过 20 次，个别科研论文的被引用次数高达近 300 次。根据原始数据拟合出被引用次数关于论文 ID 的拟合曲线 。从图 12 可以直观地看到少部分论文的被引用率极高，而大部分论文的引用率都很低。结合这个幂律反比的拟合曲线公式，我们将论文之间的被引用关系建立一个关系网络，可以发现这个网络的度服从 power

law，即，形成了一个无尺度网络。所以我们的论文之间的被引用关系已经和现实中的网络包括 WWW，社交网络，PPI 网络一样自成为一个体系 [16-17]。

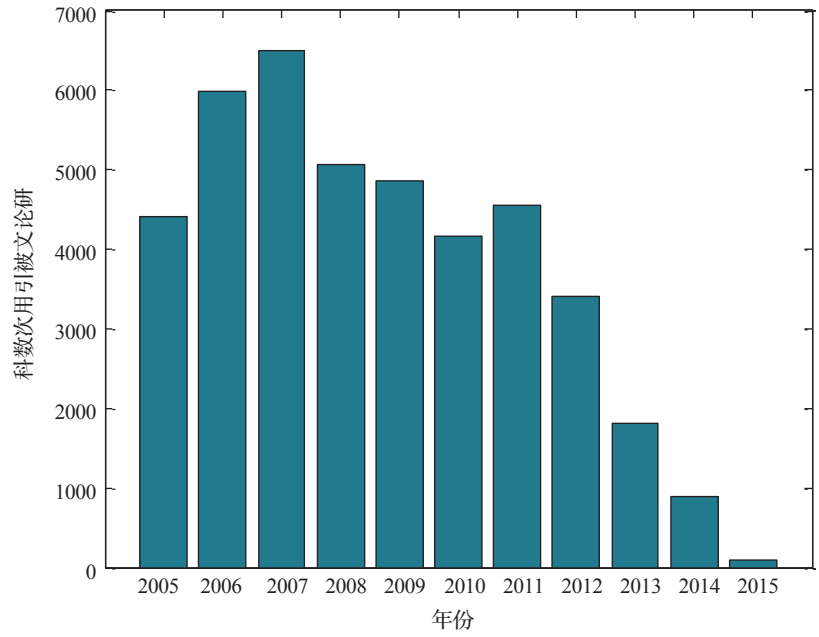

图 11 2005–2015 年信息安全领域科研论文各年引用情况

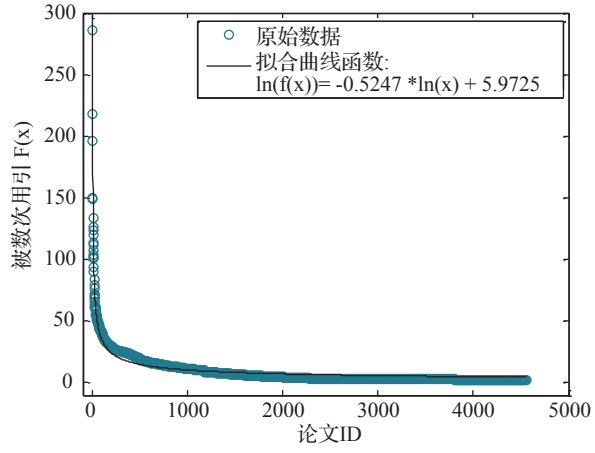

图 12 2005–2015 年信息安全领域所发表论文 / 书籍的被引用情况

综合以上的分析可以知道，要想在信息安全领域得到更好的发展，科研团队和科研资金的投入是必不可少的。从图 10 的分析中我们可以大致了解国内在信息安全领域的研究现状，以及与走在国际最前面的科研队伍的差距。首先，我国要始终牢牢抓住网络安全这个国际主流，加大研究力度，争取取得更好的研究成果。其次，在密码学方面，中国从 2012 年起取得了较大的进步，科研论文占国际密码学论文的比例相比于其他四个方向都要高，所以中国可以试着再加强密码学的研究，追上国际的步伐，争取成为国际上密码学研究水平较高的国家。此外，相对于当今比较重视的信息系统安全，我国的科研论文所占比例比较低，发展也比较平缓，必须加以重视。最后，在信息内容安全和信息对抗领域，我国的科研依然处于起步阶段，科研成果占国际科研成果的比例依然很低。因此，总体而言，我国需要加大在信息安全领域的投入和建设，以确保在这个全民联网的大数据时代，维护我国人民的切身利益以及捍卫国家的主权和安全。

doi:10.3772/j.issn.2095-915x.2018.02.010

4 总结

为了给信息安全领域中高质量研究成果的碰撞、沟通与协作提供良好的基础认识，我们采集了来自 Springer、ACM、IEEE 以及 Elsevier 四个国际主流出版商，几乎覆盖近十年来国内外主要研究机构在信息安全领域上全部的科研论文成果产出。并对该科研论文数据集进行了一系列的数据分析，使研究人员对国内外信息安全领域的研究情况、研究进展、研究趋势以及主要的研究单位和国家都有了更为细致的了解，为科学院和创新研究院在信息安全学科方向和研究力量布局方面提供了具有一定可信度的分析结果和决策依据。有利于建设我国信息安全保障体系，发展信息安全技术与产业，培养各层次创新型信息安全人才。

参考文献

[1] 佚名 . CNNIC 发布第 36 次《中国互联网络发展状况统计报告》[J]. 中国信息安全 , 2015(8):19-19.

[2] 袁永波 , 胡元蓉 . 探析大数据时代下的网络安全问题 [J]. 网络安全技术与应用 , 2015(2):165-165.

[3] 温洲 . 信息安全领域的研究 [J]. 科技与企业 , 2015(8):76-76.

[4] 张焕国 , 王丽娜 , 杜瑞颖 , 等 . 信息安全学科体系结构研究 [J]. 武汉大学学报（理学版）, 2010, 56(5):614-620.

[5] 温亮明 , 王军 , 余波 . 基于论文产出视角的高校图书馆科研实力研究——以“985 工程”高校为例 [J]. 情报工程 , 2015, 1(5):107-118.

[6] 刘京旋 , 杜永萍 , 杜晓燕 , 等 . 学术网络中科研人员影响力分析方法研究 [J]. 情报工程 , 2015, 1(6):83-89.

[7] 黄存东 . 关于计算机网络信息安全问题的技术研究 [J]. 软件 , 2013(34):140-141.

[8] 徐丞 . 探讨计算机信息网络安全技术及未来发展方向 [J]. 信息系统工程 , 2014(3):82.

[9] Feng D G, Wang X Y. Progress and Prospect on Information Security Research in China[J]. Journal of Computer Science & Technology, 2006, 21(5):740-755.

[10] 崔蓉 . 计算机信息网络安全技术及发展方向 [J]. 信息技术与电脑 , 2010(10):12.

[11] Ott R L, Longnecker M. An Introduction to Statistical Methods and Data Analysis[M]. Massachusetts: Duxbury Press, 2008.

[12] Tang J, Wu S, Sun J, et al. Cross-domain collaboration recommendation[C]. Proceedings of the Proceedings of the 18th ACM SIGKDD international conference on Knoledege discovery and data mining, Beijing, China, ACM, 2012:1285-93.

[13] Abbas A, Faiz A. Usefulness of digital and traditional libraries in higher education[J]. International Journal of Services Technology and Management, 2013, 19(1-3):149-161.

[14] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. J Machine Learning Research Archive, 2003(3):993-1022.

[15] Duda R O, Hart P E, Stork D G. Pattern classification[M]. Wiley, 2001.

[16] Pengbin G, Weiwei W, Bo Y. Co-authorship network analysis in improvisation theory research [C]. Proceedings of the 2012 International Conference on Information Management, Innovation Management and Industrial Engineering, IEEE, 2012:244-8.

[17] Lee D H, Brusilovsky P, Schleyer T. Recommending collaborators using social features and MeSH terms[J]. Proceedings of the American Society for Information Science and Technology, 2011, 48(1):1-10.