文章编号：1005-8451（2015）02-0076-03

大数据环境下信息安全分析

张德栋，祝咏升，司　群
（中国铁道科学研究院　电子计算技术研究所，北京　100081）

摘　要：本文对大数据涉及的关键技术进行概括，分析了大数据时代信息安全所面临的挑战和机遇，最后给出了我国大数据环境下的信息安全建设的建议。

关键词：大数据；互联网；信息安全
中图分类号：U29-39　　**文献标识码：**A

Analysis of information security in big data environment

ZHANG Dedong, ZHU Yongsheng, SI Qun
(Institute of Computing Technologies, China Academy of Railway Sciences, Beijing　100081, China)

Abstract: Based on the analysis of key technologies related to the big data, the challenges and opportunities of information security in big data environment were analyzed, the constructive suggestions of information security in big data environment were proposed.

Key words: big data; Internet; information security

大数据中包括个人隐私、上网轨迹等有价值的信息，如果对电子邮件、搜索记录、交谈记录、文件传输记录、社交网站行为等海量数据进行分析，并关联现实中的一些个人行为（如信用卡、电话录音等），个人的生活状况几乎可以得到还原。随着对大数据信息的深入挖掘，信息的开放程度将进一步扩大。大数据在给信息安全行业提出新的挑战的同时，也将推动信息安全技术的发展。本文对大数据涉及的关键技术进行概括，分析大数据时代信息安全所面临的挑战和机遇，最后给出我国大数据环境下的信息安全建设的建议。

1　大数据特征

1.1　数据量大

各种传感器、终端设备、网络设备等会产生大量数据。据估计，2013 年互联网上的数据量达到 667 EB。

1.2　数据类型多

与传统的结构化数据相比，非结构化数据增长迅速，数据类型涉及音视频、图片、GPS 信息、网络日志等。

1.3　价值密度低

在大数据中，信息的价值需要从海量数据中挖掘分析得到。相对数据总量，有用数据显得微乎其微。以监控视频为例，在 24 h 的视频监控中，真正满足用户需要的信息可能只有几秒钟。

1.4　处理速度快

随着高速运算和存储、数据挖掘、语义引擎等技术的发展，对大数据的处理速度、解析深度将达到前所未有的程度，预计到 2020 年全球数据使用量将会达到 35.2 ZB。

2　大数据的关键技术

2.1　采集技术

大数据采集是指将分布的、异构的数据抽取集中，数据采集是后期数据清洗、转换、集成以及联机分析处理、数据挖掘的基础。比较典型的有系统日志采集和网络数据采集。大数据采集技术需要突破的技术有：分布式高速数据爬取、高速数据全映像、高速数据解析和转换等技术。

2.2　预处理技术

主要实现对已采集数据的抽取和清洗等操作。抽取：将不同结构和类型的数据转化为单一的或便于处理的构型，数据抽取是数据快速分析处理的基

收稿日期：2014-10-08
作者简介：张德栋，助理研究员；祝咏升，助理研究员。

础。清洗：过滤掉大数据中没有价值或不关心的内容，提取出有效数据。

2.3 多数据融合

数据融合是一个多级、多层面的数据处理过程，主要完成对来自多个信息源的数据进行自动检测、关联、估计和组合等处理。

2.4 数据挖掘分析

大数据挖掘分析主要解决复杂数据结构、多种类型、海量数据的有效处理问题。面向结构化数据的统计分析、特征提取、挖掘等技术相对较成熟；而对非结构化数据的处理方法主要以模式识别、机器学习等人工智能技术为主。

2.5 存储与处理技术

大数据存储和处理技术需重点解决复杂结构化、半结构化和非结构化大数据管理与处理技术，主要解决大数据的可存储、可表示、可处理、可靠性及有效传输等几个关键问题。开发可靠的分布式文件系统；突破分布式非关系型大数据管理与处理技术，异构数据的数据融合技术，数据组织技术，研究大数据建模技术；突破大数据索引技术；突破大数据移动、备份、复制等技术。

3 大数据环境下信息安全的挑战与机遇

3.1 大数据环境下信息安全的挑战

（1）大数据加大隐私泄露风险

数据集中存储使得数据泄露的风险逐渐增大，而大数据的来源范围非常广阔，涉及社交网站、交易信息、位置信息、行为轨迹、电子邮件等，对这些数据进行关联分析基本能够还原一个人的个人行为及生活轨迹，势必对用户隐私产生威胁。这些个人隐私信息被泄露后，其人身安全可能受到影响；同时，由于互联网管理制度的落后，没有对互联网中隐私数据的所有权和使用权进行界定和制定合理的标准，使得用户隐私泄露后，用户权限不能得到维护。

（2）大数据成为网络攻击的显著目标

网络技术的发展为不同领域、不同行业之间实现数据资源共享提供条件。对于大数据的数据整合和分析可以获得一些敏感和有价值的数据，这些数据会吸引更多的潜在攻击者，使得大数据成为更具吸引力的攻击目标。同时，数据的大量聚集，使得黑客在将数据攻破之后以此为突破口获取更多有价值的信息，降低了黑客的攻击成本。

（3）大数据成为高级可持续攻击的载体

高级可持续攻击（Advanced Persistent Threat，APT 攻击）的特点是攻击时间长、攻击空间广、单点隐藏能力强。首先，在大数据环境下，黑客可以利用大数据扩大 APT 攻击的效果。利用大数据黑客可以发起僵尸网络攻击，控制大量傀儡机，并对目标机发起攻击。其次，在数据环境下，大数据价值密度低的特点，安全分析很难聚焦于某个价值点上，APT 攻击代码更容易隐藏，增大检测工具发现 APT 攻击的难度。

（4）大数据安全存储新要求

大数据环境下，数据类型多样，非结构化数据占绝大多数。虽然 NoSQL 技术具有可扩充性特点，但其本身存在安全漏洞。同时，面对海量的数据，常规的安全扫描手段需要耗费过多的时间，已经无法满足安全需求。同时，安全防护手段的更新升级速度无法跟上数据量非线性增长的步伐。大数据对安全存储技术提出了新的要求。

3.2 大数据环境下信息安全的机遇

大数据给信息安全提供新挑战的同时也给信息安全提供了新的机遇。大数据提供了一个弹性架构，为安全分析提供了新的可能性，对海量数据进行分析能够在以往无法达到的广度和深度上，判断网络异常行为、异常流量，找出数据中的风险点。对实时安全和商务数据结合在一起的数据进行预防性分析，可识别钓鱼攻击，防止诈骗和阻止黑客入侵。网络攻击行为总会留下蛛丝马迹，这些痕迹都以数据的形式隐藏在大数据中，利用大数据技术整合计算和处理资源有助于更有针对性地应对信息安全威胁，找到攻击的源头。将大数据和云计算相结合，可以为信息安全提供更加智能的工具。

利用大数据技术，通过自动化分析处理与深度挖掘相结合，可以将很多事中、事后处理，转向事前自动评估预测、应急处理，由被动防御转变为主动防护。结合大数据分析技术，安全厂商可以做到从数据收集分析到安全管理策略下发，再到效果评

估的一整套安全解决方案。

4 保障我国大数据信息安全的建议

4.1 加大大数据信息安全体系建设

大数据技术领域的竞争，关系国家安全和未来，大数据安全技术的落后意味着国家安全在数字空间出现漏洞。“斯诺登事件”使得各国政府认识到大数据对国家信息安全的重要性，我国政府应加快大数据领域相关政策的出台，着重在大数据技术研发、人才培养、信息安全体系建设等方面给予政策支撑，加强顶层设计和政策支持，是大数据时代的客观要求。

4.2 加快大数据安全技术研发

大数据在数据的收集、处理以及存储等方面对信息安全技术提出了新的要求。以杀毒软件为例，在海量数据扫描一个恶意软件或代码可能需要几天甚至十几天的时间。因此，需要针对大数据环境，加大信息安全技术和信息安全产品的研发，提高我国大数据信息安全技术水平。

4.3 提高信息化产品的国产化水平

在我国信息化建设的进程中，普遍存在依赖国外技术装备的情况，这对我国尤其是关键行业的信息安全威胁巨大。鉴于此，国家应在政策层面给予引导，企业注重技术研发，努力提高我国信息化的国产化水平。

4.4 加强敏感数据监管力度

大数据加大了敏感信息泄露的风险，面对大数据蕴含的信息价值诱惑，黑客的攻击动机将更加强烈，手段层出不穷，相比于传统数据泄露或者黑客攻击事件，大数据时代一旦数据分析结果泄露，对整个企业甚至整个行业可以说是毁灭性打击，因此在大数据环境下要加强对敏感数据的监管，建立健全相应规章制度，规范大数据的使用流程和使用权限。

4.5 积极落实国家信息安全技术防范重点措施

加快发展技术先进、安全可靠、自主可控的网络核心技术替代产品，提高信息基础设施特别是重点部门的国产化程度。强化网络技术、信息安全技术、数据挖掘技术等方面的人才培养，努力克服关键技术受制于人的软肋，建立我国在国际网络控制的主导地位。

5 结束语

大数据将成为新一次技术变革的基石，面对挑战和机遇并存的大数据信息安全问题，需在加强传统信息安全建设的同时，加快面向大数据的信息安全技术研发，只有大数据技术和信息安全协同发展，大数据才可以真正成为新时代的驱动力。

参考文献：

[1] 周路菡．棱镜下的大数据安全恐慌 [J]. 新经济研究，2013（9）：81-85.

[2] 孟小峰，慈　祥．大数据管理：概念、技术与挑战 [J]. 计算机研究与发展，2013，50（1）：146-169.

[3] 潘柱廷．高端信息安全与大数据 [J]. 信息安全与通信保密，2012（12）：19-20.

[4] 方世敏．大数据面临的信息安全问题分析 [J]. 信息安全，2013（19）.

[5] 邬贺铨．大数据时代的机遇与挑战 [J]. 求是，2013（4）：47-49.

[6] 王　倩，朱宏峰，刘天华．大数据安全的现状与发展 [J]. 计算机与网络创新生活，2013（16）：66-69.

责任编辑　陈　蓉

高速铁路系统试验国家工程实验室——客运服务系统试验室

高速铁路系统试验国家工程实验室——行车安全监控和应急平台试验室