大数据的安全现状与应对策略研究

桑运昌
(南京陆军指挥学院作战实验中心 南京 210045)

摘 要 大数据蕴藏着有价值的信息,但数据安全面临严峻挑战。在分析大数据基本特征的基础上,提出了当前大数据面临的安全挑战,并从监测机制、防范检测、响应水平和处理能力 4 个方面阐述了大数据安全的应对策略。

关键词 大数据,数据安全,数据挖掘
中图法分类号 TP309 **文献标识码** A

Study on Safety Status and Coping Strategies for Big Data

SANG Yun-chang
(Combat Experimental Center of Nanjing Army Command College, Nanjing 210045, China)

Abstract The data contain valuable information, but the data security is facing serious challenges. Based on the analysis of the basic characteristics of the data, we proposed the current security challenges large data faced, and described the data safety and countermeasures from four aspects, such as monitoring mechanisms, prevention and detection, response level and processing capability.

Keywords Large data, Data security, Data mining

随着互联网、物联网、云计算等技术的快速发展,以及智能终端、网络社会、数字地球等信息体的建设和普及,全球数据量出现爆炸式增长,仅在 2011 年就达到 1.8 万亿 GB。IDC 预计,到 2020 年全球数据量将增加 50 倍[1]。无疑,大数据时代已经到来。一方面,云计算为这些海量的、多样化的数据提供存储和运算平台,同时数据挖掘和人工智能从大数据中发现知识、规律和趋势,为决策提供信息参考。但是,大数据的发展将进一步扩大信息的开放程度,随之而来的隐私数据或敏感信息的泄露事件时有发生。面对大数据发展的新特点、新挑战,如何保障数据安全是我们需要研究的课题。

1 大数据的特征

大数据通常被认为是一种数据量很大、数据形式多样化的非结构化数据。随着对大数据研究的进一步深入,大数据不仅指数据本身的规模,也包括数据采集工具、数据存储平台、数据分析系统和数据衍生价值等要素。其主要有以下几方面特点。

1.1 数据量大

大数据时代,各种传感器、移动设备、智能终端和网络社会等都无时不刻地在产生数据,数量级别已经突破 TB,发展至 PB 乃至 ZB,统计数据量呈千倍级别上升。据统计,2012 年全球产生的数据量已达到 2.7ZB,2015 年将超过 8ZB。

1.2 类型多样

当前大数据不仅仅是数据量的井喷性增长,而且还包含着数据类型的多样化发展。以往数据大都以二维结构呈现,但随着互联网、多媒体等技术的快速发展和普及,视频、音频、图片、邮件、HTML、RFID、GPS 和传感器等产生的非结构化数据,每年都以 60%的速度增长。预计非结构化数据将占数据总量的 80%以上[2]。

1.3 运算高效

基于云计算的 Hadoop 大数据框架,利用集群的威力高速运算和存储,实现了一个分布式运行系统,以流的形式提供高传输率来访问数据,适应了大数据的应用程序;而且,数据挖掘、语义引擎、可视化分析等技术的发展,使得可以从海量的数据中深度解析和提取信息,掌控数据增值的"加速器"。

1.4 产生价值

价值是大数据的终极目的。大数据本身是一个"金矿产",可以从大数据的融合中获得意想不到的有价值的信息。特别是激烈竞争的商业领域,数据正成为企业的新型资产,企业部追求数据最大价值化。同时,大数据价值也存在密度低的特性,需要对海量的数据进行挖掘分析才能得到真正有用的信息,形成用户价值。以监控视频为例,连续的播放画面中可以产生价值信息的数据可能是仅仅其中的一两秒。

2 大数据时代的信息安全"隐患"

作为新兴产物,大数据仍面临一些亟待解决的安全问题。从核心价值角度来看,大数据关键在于数据分析和利用,但数据分析技术的发展,对用户隐私产生了极大的威胁。在大数据时代,想屏蔽外部数据商挖掘个人信息是不现实的[2]。目前,各社交网站均不同程度地开放其用户所产生的实时数据,这些实时数据被一些数据提供商收集,还出现了一些监测数据的市场分析机构。通过人们在社交网站中写入的信息、智

桑运昌(1986—),男,硕士,讲师,CCF 会员,主要研究领域为数据工程。

能手机显示的位置信息等多种数据组合，已经可以以非常高的精度锁定个人，进而挖掘出个人信息体系，用户隐私安全问题堪忧。面对现有的大数据的种种问题，笔者认为有以下几个方面需要改进。

2.1 非结构化数据对大数据存储提出新要求

在大数据之前，我们通常将数据存储分为关系型数据库和文件服务器两种。而当前大数据汹涌而来，数据类型的千姿百态也使我们措手不及。对于将占数据总量 80%以上的非结构化数据，虽然 NoSQL 数据存储具有可扩展性和可用性等优点，利于趋势分析，为大数据存储提供了初步解决方案，但是 NoSQL 数据存储仍存在以下问题：1）相对于严格访问控制和隐私管理的 SQL 技术，目前 NoSQL 还无法沿用 SQL 的模式，而且适应 NoSQL 的存储模式并不成熟；2）虽然 NoSQL 软件从传统数据存储中取得经验，但 NoSQL 仍然存在各种漏洞，毕竟它使用的是新代码；3）由于 NoSQL 服务器软件没有内置足够安全的验证方式，因此客户端应用程序需要内建安全因素，这又反过来产生了诸如身份验证、授权过程和输入验证等大量安全问题。

2.2 网络层的安全策略是端点数据安全的重点加固对象

常规的数据安全模式往往喜欢分层构建，这也是数据安全的常规做法[3]。现有的端点安全方式对于网络层的安全防护并不完美。一方面是大数据时代的信息爆炸，导致网端的非法入侵次数急剧增长，这对于网络层的考验十分严峻；另一方面，由于云计算的大趋势，现在的网络数据威胁方式和方法越来越难以预测和辨识，这给现有的端点数据安全模式造成了巨大的压力。在未来，网络层安全应当作为重点发展的一个层面。在加强网络层数据辨识智能化、结构化的基础上加上本地系统的相互监控协调，同时杜绝非常态数据的运行，这样就能够在网络层构筑属于大数据时代的全面的安全堡垒，弥补自身的缺陷。

2.3 本地策略的升级

由于大数据时代的数据财富化导致了大量的信息泄露事件，而这些泄露事件中来自内部的威胁更大，因此在本地策略的构建上需要加入对于内部管理的监控和监管手段。用纯数据的模式来避免由人为原因造成的数据流失和信息泄露。从这一点出发我们可以预想到在未来的数据安全模式中，管理者的角色权重逐渐分化，数据本身的自我监控和智能管理将代替一大部分人为的操作[4]。在本地安全策略的构建过程中，还要加强与各个环节的协调。由于现在的数据处理方式往往会依托于网络，因此在数据的处理过程中会出现大量的数据调用，在调用过程中就容易出现很大的安全威胁。

2.4 数据存储的问题

在传统端点的数据安全中，数据存储作为非法入侵的最后一站，被业界人士高度重视，他们为数据存储建立了全面完善的防护措施，这些非常值得借鉴，但是还需进一步的完善。这里的完善主要是数据存储隔离与调用之间的数据逻辑关系策划。这同样是为了适应现在的数据模式。

3 大数据安全的应对策略

3.1 建立信息系统安全事件监测机制，及时发现信息系统安全问题

在运维阶段，诸如如何及时发现异常行为、如何判断该用户是否被控制或穿了马甲、如何处理服务器出现的大量外连上传行为等问题很频繁。因此，政企用户需要建立一套有效的安全事件监控和预警措施，以能够在信息系统即将遭到攻击或已经遭到攻击时，快速、准确地发现攻击行为，并迅速启动处置和应急机制[1]；同时可以对信息系统的安全事件进行综合分析，了解当前整体系统的安全态势，为整体网络与信息安全规划提供有效的数据支持。

3.2 预先防范，提前做好安全性检查，全面提升主动检测能力

Web 应用的安全性成为越来越需要关注的问题，有近 40%的入侵是由于 Web 应用的问题造成的[5]。Applied Research 发表的一份调查报告表明，企业反馈超过一半的最频繁的攻击是针对 Web 应用的。这些攻击中有一半都出现在著名的“OWASP 十大威胁”名单中。面对这些持续而频繁的攻击，政企用户需要进行定期的安全检查，及时主动发现信息系统中存在的安全漏洞及潜在威胁。

3.3 提高安全事件的响应和处理能力

监控中发现的问题，以及在安全检查中对自身脆弱性的了解，为应急响应的处理提供了依据，同时依据自身及行业特点，建立安全知识库[6]。鉴于目前多数政企单位并不具备独立处理安全事件的技术实力，政府单位需要专业安全服务厂商提供安全事件的预警、响应和必要的技术支持，以提高政企单位信息部门的安全事件响应与处理能力。

3.4 通过强大的综合分析能力，为信息部门提供数据参考和决策支持

应随时了解信息系统的运行情况和安全状况、安全态势，在海量数据的基础上，对安全事件和安全态势进行综合分析，得出宏观的规律和各类不同事件相互联系的规律，为信息部门提供强有力的数据参考和决策支持。

参 考 文 献

[1] 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

[2] 陈明奇，姜禾，张娟，等. 大数据时代的美国信息网络安全新战略分析[J]. 信息网络安全，2012(8)：32-35

[3] 赛迪智库软件与信息服务研究所. 美国将发展大数据提升到战略层面[N]. 中国电子报(第 003 版)，2012-7-17

[4] 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

[5] 肖新斌，史召臣. 云计算引发的安全风险[J]. 信息安全与技术，2011(6)：13-14

[6] 胡光永. 基于云计算的数据安全存储策略研究[J]. 计算机测量与控制，2011，19(10)：2539-2541