 DOI: https://doi.org/10.12345/bdai.v6i2.24583

Application of Big Data Technology in Computer Information Security

Xin Zhang

Information Department of the 981st Hospital of the Joint Logistics Support Force, Beijing, 100000, China

Abstract

Computer as an information integration equipment, needs to carry out a lot of information collection and processing, in the context of the Internet, information is easy to appear information leakage and information distortion and other conditions, affecting the security of information. In view of this situation, it is necessary for relevant personnel to strengthen the attention to computer information security, combined with big data technology, to implement information security. This paper starts from the computer information security, talk about the security problems and causes of computer information, and then combined with big data technology, with the help of the integrated characteristics of big data set, the security problems are integrated, on this basis to study the information security guarantee strategy, in order to promote the development and progress of the computer industry.

Keywords

big data technology; computer information security; data security

大数据技术在计算机信息安全中的应用研究

张鑫

联勤保障部队第 981 医院信息科，中国·北京 100000

摘　要

计算机作为信息的整合设备，需要进行大量的信息收集与处理，在互联网背景下，信息就很容易出现信息泄露以及信息失真等状况，影响信息的安全性。针对这一状况，就需要相关人员加强对计算机信息安全的重视，结合大数据技术，对信息安全进行落实。本文就从计算机信息安全入手，浅谈计算机信息存在的安全问题以及成因，然后结合大数据技术，借助大数据集的集成性特征，将安全问题进行整合，在此基础上研究信息安全的保证策略，以推动计算机行业的发展与进步。

关键词

大数据技术；计算机信息安全；数据安全

1 引言

大数据技术是指大数据的应用技术，指引行业发展的风向标，涵盖各类大数据平台、大数据指数体系等大数据应用技术。在计算机中，计算机需要大量的信息集成与收集计算，作业环节就很容易出现信息缺失、信息泄露、信息失真以及信息损坏等状况，影响计算机功能的发挥。此背景下，计算机信息作业就需要加强对信息安全的重视，要求相关人员结合大数据技术，分析计算机作业环节的信息安全隐患，并且结合需要，制定大数据在计算机中的应用策略，保证信息的安全性，推动计算机行业的发展。

【作者简介】张鑫（1992–），男，回族，中国河北承德人，本科，助理工程师，从事医疗大数据应用、人工智能系统使用中的泄密风险研究。

2 大数据技术与计算机信息安全概述

大数据技术是指通过收集、存储、处理和分析海量、多样化的数据，以提取有价值的信息和洞察的技术。其核心特征通常包含大规模数据处理、多样化数据类型、实时分析、分布式存储和计算以及数据挖掘与机器学习等[1]。如今，大数据技术广泛已经应用于金融、医疗、电商、交通、社交媒体等领域，用于实现预测分析、个性化推荐、风险控制、客户行为分析等功能。常见的信息安全问题如图 1 所示。

计算机信息安全（Information Security, 简称 InfoSec）是指通过采取技术、管理和法律等手段，确保计算机系统和网络中的信息在存储、传输和处理过程中的机密性、完整性和可用性的作业。其目标是保护信息不受未经授权的访问、破坏、篡改或丢失[2]。随着互联网的发展和数字化转型的推进，信息安全成为每个组织和个人需要重视的问题。

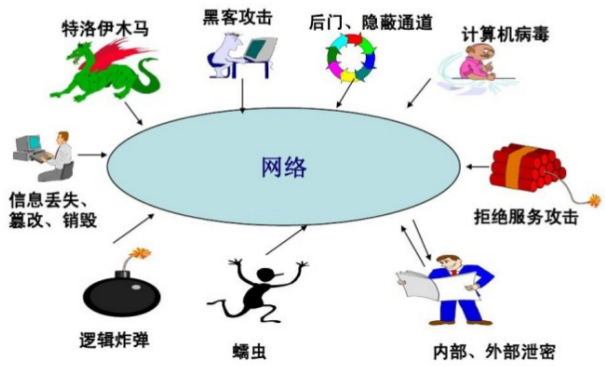

图 1 常见的信息安全问题

3 常见的计算机信息安全问题

首先是恶意软件影响，一些病毒、木马、蠕虫、勒索病毒以及间谍软件等会监视用户活动并窃取敏感信息，如密码和银行账户信息；其次是网络攻击，攻击者通过大量无效流量使目标服务器或网络瘫痪，导致合法用户无法访问，还会拦截并篡改用户与目标系统之间的通信，以窃取或修改数据；然后是数据泄露，如果敏感数据（如密码、身份证号、信用卡信息等）没有加密存储，可能被盗取。如果数据传输过程中未使用加密（如未使用 HTTPS），可能会遭到窃听或篡改。此外还存在漏洞与补丁管理不当、跨站脚本（XSS）、SQL 注入（SQL Injection）、权限提升（Privilege Escalation）、跨站请求伪造（CSRF）、数据篡改、供应链攻击、不安全的 API 以及未授权访问等。随着计算机信息安全问题不断演变，新的攻击方式和威胁层出不穷。图 2 为计算机信息安全防护体系。

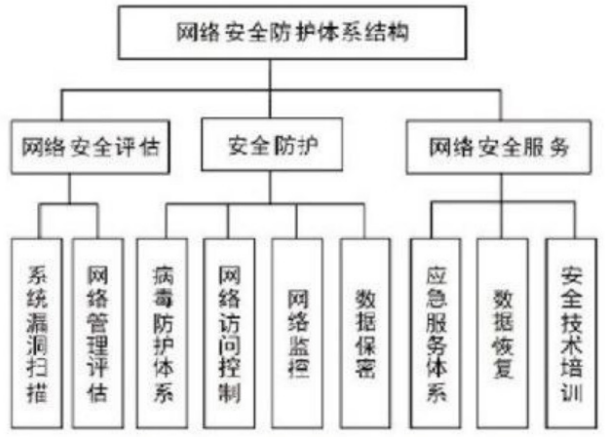

图 2 计算机信息安全防护体系

4 大数据技术在计算机信息安全中的应用

4.1 应用在威胁检测与预警中

在计算机信息安全中，利用大数据技术进行实时威胁检测与预警是应对复杂安全威胁的重要手段。大数据技术可以通过收集、分析和实时处理海量数据，帮助识别潜在的安全威胁，并迅速发出预警，提升防御能力。

首先，大数据技术可以通过流式处理技术（如 Apache Kafka、Apache Flink 等）实时收集和处理网络流量、事件日志等信息，这样可以在威胁发生的瞬间进行捕获与分析。

其次，使用流处理框架（如 Apache Flink、Apache Storm 等）对海量实时数据进行实时分析，检测网络流量中的异常模式、非法访问、恶意代码等。通过事件关联技术将不同来源的安全事件进行关联，识别潜在的复杂攻击行为。

然后，大数据平台能够通过历史数据和攻击情报构建攻击行为模型。通过分析攻击者的行为、攻击工具、攻击链条等，预测并识别未来可能发生的攻击。还能够结合大数据和人工智能技术，通过模拟不同安全场景，进行风险评估和预测。例如，可以模拟 DoS 攻击对网络的影响，预测各种攻击模式对企业资源的危害。

此外，在获取足够信息之后，某些大数据处理任务可以在靠近数据源的边缘设备上进行处理，减少数据传输的延迟。尤其是在物联网（IoT）环境中，边缘计算可以实时分析设备的行为并检测潜在的安全问题。并且利用分布式计算和存储平台，将安全事件的检测分布到不同的节点上。通过分布式处理能力，系统能够对大规模数据进行并行分析，快速识别跨地域、跨网络的复杂攻击。

综上，大数据技术在实时威胁检测和预警中的应用通过实时数据收集、流数据分析、机器学习、自动化响应等技术手段，显著提高了网络安全系统对复杂攻击的响应能力和防护效果。

4.2 应用在大规模日志分析中

日志数据是网络和系统活动的重要记录，包含了大量关于操作、事件、错误和警告的信息。通过对大规模日志数据的分析，安全团队可以有效发现潜在的安全威胁、识别攻击模式、并及时采取相应的防护措施。实际作业环节，就可以将大数据技术应用到该环节。

首先，为了处理和存储海量日志数据，可以使用分布式存储系统（如 Hadoop HDFS、Amazon S3、Elasticsearch 等）存储日志。这些分布式存储系统具备扩展性，能够应对日志量的增长，并支持高效地查询和检索。

其次，对于实时日志流的处理，可以使用流处理框架（如 Apache Kafka、Apache Flink、Apache Storm 等）对日志数据进行实时分析。这些框架可以帮助在日志生成的瞬间检测到异常或攻击行为（例如：流量暴增、重复登录失败等）。

然后，日志数据通常是非结构化或半结构化的，因此需要进行清洗和预处理。通过数据清洗技术，去除冗余信息、标准化日志格式，并将关键信息（如 IP 地址、用户行为、访问路径等）提取出来，以便后续分析。

此外，需要借助大数据平台，通过定义规则或过滤器来检测异常事件。例如，如果某个 IP 在短时间内发起过多的登录请求，系统会识别为暴力破解攻击。通过这种规则分

析，可以快速捕获常见的攻击模式。

综上，大数据技术在大规模日志分析中的应用使得安全团队能够在面对大量和多样化的日志数据时，高效地识别、分析和响应安全威胁，为安全防御策略的制定提供支持。

4.3 应用在数据加密与隐私保护中

随着大数据技术的广泛应用，数据量的激增和数据来源的多样化，如何有效地保护敏感数据成为一个重要问题。在计算机信息安全中，就可以将大数据技术应用在该环节。

数据加密环节，可以通过多种方法进行数据的加密，比如加密算法，加密算法又可以分为对称加密与非对称加密，其中，对称加密使用相同的密钥进行数据加密和解密，如 AES（Advanced Encryption Standard）。对称加密适合于大数据环境下处理大量数据时，速度较快且计算开销较小，但密钥管理和分发是其挑战之一。非对称加密使用一对密钥（公钥和私钥），如 RSA 算法。非对称加密一般用于加密小规模的数据，如传输密钥、身份验证信息等，适合保护密钥交换和敏感数据传输，但在处理大规模数据时速度较慢。

隐私保护环节，常见技术包括以下几种。首先是差分隐私，差分隐私是一种通过添加噪声来确保分析结果不会泄露个体信息的技术。其目标是确保即使攻击者知道大部分数据，也无法通过查询结果推断出关于某个特定个体的信息。在大数据分析中，差分隐私常常用于保护用户数据，在进行统计分析或机器学习模型训练时，确保数据不会泄露个人隐私信息；其次是数据脱敏，数据脱敏（Data Masking）是指通过修改数据的某些敏感部分（如替换为匿名或随机化的值），使得即便数据被外部人员访问，也无法识别出敏感信息。在大数据环境中，数据脱敏可广泛应用于测试、开发以及数据分析等场景，确保敏感数据的隐私不被泄露；然后是数据访问控制与审计，在大数据环境中，可以通过定义用户角色，并为每个角色分配访问权限，可以有效管理数据访问。并且开展数据访问审计，记录所有的用户访问行为，帮助企业监控不正常的数据访问活动，并及时发现潜在的安全威胁。

综上，通过加密技术、隐私保护，企业可以有效保障大数据环境中的敏感数据和用户隐私，降低数据泄露和滥用的风险。

4.4 应用在自动化安全响应与修复中

因为数据量巨大、数据源繁多，且攻击者的技术手段也不断进步，所以计算机信息安全保障环节，自动化安全响应与修复显得尤为重要，需要相关人员通过以下手段进行设计。

自动化安全响应环节，首先需要通过大数据技术（如流数据处理、实时日志分析）对网络流量、系统日志、应用日志进行实时监控，自动检测异常行为或攻击迹象。例如，利用SIEM(Security Information and Event Management)系统，通过自动化分析和关联安全日志，可以识别出潜在的安全威胁；其次要设计攻击检测与自动化响应措施，可以直接入侵检测和防御（IDS/IPS），IDS/IPS 系统通过自动分析网络流量和数据包内容，能够实时识别异常流量和攻击模式，触发自动响应机制进行防御或隔离，减轻攻击对系统的影响[3]；然后要重视自动化事件分类与优先级设定，可以通过大数据分析技术，自动将安全事件分类，并根据事件的严重程度和影响范围为事件分配优先级。这有助于安全团队集中精力处理高优先级的事件，而低优先级的事件则可以通过自动化方式处理。

自动化安全修复环节，首先，大数据环境中，可以通过自动化的漏洞扫描工具（如 Nessus、OpenVAS 等）对系统、应用、数据库等进行定期扫描，识别出潜在的漏洞。对于已知漏洞，系统能够自动推送修复补丁或配置变更；其次，数据环境中的许多安全问题源于不当的配置。自动化工具能够持续审计系统、数据库、网络设备等的配置，发现不安全的配置（如默认密码、过于宽松的权限设置等）并进行修复；然后，在遭受攻击或数据损坏的情况下，自动化灾难恢复（DR, Disaster Recovery）系统能够在大数据环境中恢复受影响的数据和服务[4]。通过备份系统和自动化恢复流程，攻击后系统可以迅速恢复正常运行。综上，在大数据环境中，自动化安全响应与修复是确保信息安全的关键组成部分。

5 结语

大数据技术在应用过程中对海量的数据进行分析和整理，应用工具软件将数据的分析、收集、管理等进行集合，实现了数据的灵活性提升，但在日常应用过程中，还存在着信息安全问题。根据大数据技术的科学性和系统性，从根本上避免破坏计算机信息安全的行为，使用相关检测技术时也可以全面地保护计算机的信息安全。

参考文献

[1] 彭烈华. 大数据技术在计算机网络信息安全管控中的应用 [J]. 智能物联技术, 2024, 56 (05): 124-127.

[2] 李旭. 大数据技术在计算机信息安全中的应用分析 [J]. 中国信息界, 2024, (01): 93-96.

[3] 张敏. 大数据技术在计算机信息系统中的应用 [J]. 数字技术与应用, 2024, 42 (02): 115-117.

[4] 张梅. 大数据技术在计算机网络信息安全管理中的应用 [J]. 哈尔滨职业技术学院学报, 2023, (04): 115-117.