基于大数据分析的网络安全态势感知

展 娜 杨志军
(北京交通大学信息中心 北京 100044)

摘 要 伴随着网络技术的飞速发展,网络安全问题也日益严峻。为了满足网络安全的需求,文中提出了基于大数据分析的网络安全态势感知体系架构。借助基于大数据分析技术的机器学习和数据挖掘算法,结合安全威胁情报分析,更加智能地洞察网络安全的态势,从而更加积极、灵活地去应对层出不穷的威胁和变化多端的风险。

关键词 态势感知,大数据,关联分析
中图法分类号 TP309 **文献标识码** A

Network Security Situational Awareness Based on Big Data Analysis

ZHAN Na YANG Zhi-jun
(Information Center, Beijing Jiaotong University, Beijing 100044, China)

Abstract With the rapid development of network technology, network security issues are becoming more and more serious. In order to meet the requirements of network security, a network security situational awareness architecture based on big data analysis was proposed. With the help of big data analysis technology-based machine learning and data mining algorithms, combined with security threat intelligence analysis, it can more intelligently understand the security of information network. In addition, it actively and flexibly responds to endless threats and unknown and variable risks.

Keywords Situational awareness, Big data, Correlation analysis

1 引言

在网络普及的当下,利用网络存在的漏洞和安全缺陷对网络系统进行的黑客攻击层出不穷,病毒木马植入、DDoS攻击、网络钓鱼、漏洞式攻击等网络攻击方式威胁着网络安全。为了满足网络安全需求,各内部网络部署了防火墙、入侵防御系统、防病毒系统、Web应用防火墙、网络行为审计、防病毒网关等相关网络安全设备。这些安全设备基本涵盖了边界网络安全、内部威胁检测、漏洞管理、安全审计的网络安全运行各层面的基本要求,它们在运行过程中会产生大量的日志文件。不同的安全设备会产生不同格式的日志文件,如果只是单独分析每个安全设备上的日志信息,就会割裂多源日志中的信息关联,从中得到的网络安全状态描述可能会片面或不准确,甚至产生漏报、误报现象,更不能有效评估网络的整体安全态势。

为了及时应对当前大数据时代下的网络安全威胁,研究基于大数据分析的网络安全态势感知技术迫在眉睫,以便能够实现多源不同格式数据高效采集、海量数据快速融合、分析模型智能有效、态势展示类型丰富的目标要求。利用大数据分析,对网络安全设备日志进行全方位的处理与分析,识别各式各样的网络攻击,实现对整体网络安全态势的实时感知。

2 网络安全态势感知

态势感知是一种基于环境的,动态、整体地洞察网络攻击的能力,是以安全大数据为基础,从全局的角度提升对网络攻击的辨别、处理、分析的一种方法。网络安全态势感知利用关联分析和可视化等技术对影响网络安全的诸多要素进行获取、理解,实现对网络安全状况的准确评估以及对未来网络安全趋

展 娜(1995—),女,硕士生,主要研究方向为移动与互联网络

杨志军(1967—),男,高级工程师,主要研究方向为移动与互联网络

势的预测，并以可视化的方式展现给用户。网络安全态势感知是对网络安全性进行定量分析的一种手段，是对网络安全性的精细度量，对保障网络安全起着非常重要的作用。

目前对于网络安全态势感知的研究有很多，文献[1]认为神经网络具有自主推理和感知的能力，并且可以模拟认知功能，包括学习、记忆、推理和感知；文献[2]在此基础上将小波神经网络融入到安全态势感知研究中，提出了一种网络安全态势感知模型；文献[3]提出了一种面向多步攻击的网络安全态势评估方法。本文提出的基于大数据分析的网络安全态势感知通过机器学习、威胁情报分析等方法实现对安全态势的分析和预测。

3　网络安全态势感知体系架构

由于安全设备太多，所提供的报警信息量太大，信息中真实报警与虚假报警混杂，这些海量的日志信息必须经过数据融合和关联分析等处理才能作为态势感知的信息来源。与此同时，大数据具备的数据量巨大、数据类型庞杂、价值密度低和处理速度快等特点能够满足网络安全态势感知对海量数据实时性、准确性、高效率的要求。因此，针对大规模网络空间中数据的海量、多模式、多粒度的特点，满足并行性、实时性数据处理的要求，将大数据技术引进网络态势感知领域，提出基于大数据的网络态势感知体系架构，包括数据采集、数据预处理、数据存储、大数据分析和态势展示 5 层，如图 1 所示。

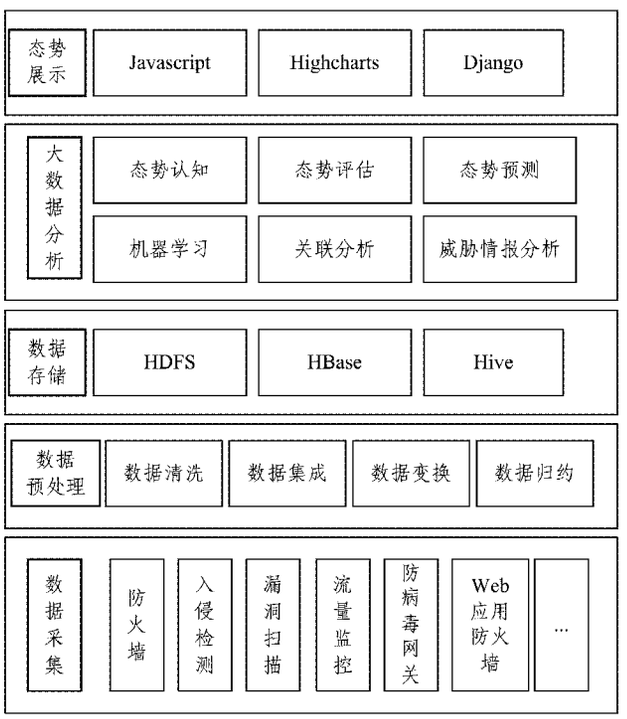

图 1　网络安全态势感知体系架构

3.1　数据采集

多源异构安全数据的主要来源为网络安全设备，包括防火墙、入侵防御系统、防病毒系统、Web 应用防火墙、流量控制等。从各种安全设备上收集格式多样的数据，对其进行初步整理后，将数据送入大数据平台，同时将数据存储在 Hive 数据仓库或 HBase 数据库内。

3.2　数据预处理

由于网络安全态势感知的数据来自不同的网络安全设备，因此这些报警信息和日志信息具有截然不同的数据内容和数据格式。我们需要对采集来的信息进行预处理，包括安全事件的格式统一化、对无用安全事件的过滤、对安全事件的时间排序、对重复安全事件的归并等，以便为网络安全态势感知提供更为确切的数据源，进而得到相对的网络安全态势。数据预处理包括数据清洗、数据集成、数据变换和数据归约。

(1)数据清洗：数据库中的数据有一些是不完整的、有噪声的和不一致的，数据清理试图填充缺失的值，光滑噪声并识别离群点，纠正数据中的不一致，将完整、正确、一致的数据存入数据仓库中。

(2)数据集成：把异构数据源集成到一个数据仓库中。

(3)数据变换：通过光滑聚集、数据范化、规范化、属性构造等方式将数据转换成适用于数据挖掘的形式。

(4)数据归约：数据归约技术可以用来得到数据集的归约表示，它小得多，但仍然接近于保持原数据的完整性，对归约后的数据集进行挖掘更有效，并产生相同或几乎相同的分析结果。

数据预处理应用大数据所提供的 Hadoop 基础平台和 MapReduce 分布式并行计算技术。

3.3　大数据分析

各种日志数据、网络流量数据经过归一化预处理后，存储于分布式 HBase 和 HDFS 中。下一步就是要对这些大数据进行关联分析，挖掘深层次、复杂的攻击行为，达到识别有计划的攻击，从而增加攻击检测率，主动感知网络中的异样以及整体的网络安全态势。

3.3.1　态势认知

态势认知是了解当前的状态，包括状态识别与确认(攻击发现)，以及对态势认知所需信息来源和素材的质量评价。简单来说，在网络安全领域，状态识别就是意识到有攻击发生(攻击发现)。状态确认则是确认攻击的类型、来源、属性和攻击目标等(攻

击确认)。首先,合理运用大数据分析工具构建攻击发现和分析模型,并可以借助机器学习等技术逐步实现对相关算法模型的持续改进。其次,充分发挥威胁情报的作用,用其他人观测到的外部攻击信息来帮助自己进行观测,这样除了能够提高自己的分析精度和速度外,还能从中发现自己在数据来源、分析模型等方面的不足,查缺补漏。对态势认知信息素材的质量评价(或者说是质量度量)不但关乎到信息本身的质量和可信度,还关乎由此形成的知识-情报-决策的质量和可信度,如真实性、完整性和时效性等。因此,可以看到态势认知的质量评价实际是以结果为导向的一套评价体系,信息质量评价不但支撑着态势认知,还会延伸到态势评估和态势预测,贯穿于整个态势感知体系的始终。

3.3.2 关联分析

关联分析即挖掘关联现象,从大量数据中发现事物、特征或者数据之间的、频繁出现的相互依赖关系和关联关系。通过对实时产生的海量多源异构报警信息和日志信息进行关联分析,可以快速发现并识别网络攻击,预测特定的安全事件的发展趋势,并及时对危害级别大的攻击进行预警。

现有的关联分析算法有基于攻击序列模板的事件关联方法、基于 Bayesian 分类器的关联方法、基于属性相近度的方法以及基于因果关系的关联方法等。每个关联算法都有自己的优点,为了提高关联分析的效率和准确性,我们可以在基于因果关系的关联方法的基础上引入属性相近度算法思想,即在判定事件的因果关系前,首先精简告警信息,接着判定事件之间的因果关系,最后生成攻击场景图。这样可以消除重复事件或相似事件,把事件误报概率降到最低。

大部分的安全事件并不是孤立产生的,它们之间存在着一定的联系,甚至很多是属于同一个报警,这样,这类的报警之间就存在着一定的相近信息。因此,我们需要将这属于同一类报警的信息合并为一条报警记录信息。有的警报信息只有一条,根据常识我们知道一次攻击就能成功的概率几乎为零,因此只有一条警报信息的很大可能是虚报警,这类信息我们应该去除。通过以上方法,我们就可以对警报信息进行关联和合并,最大程度地缩减警报的数量。其中,对于警报的相近度,我们是采用报警记录的特征属性的相近度来衡量的。

每个报警记录都是由一系列属性构成的,其中比较重要的属性就是源 IP、目的 IP、源端口、目的端口、协议类型、时间属性等。我们通过计算这些属性各自的相近度来比较两个报警之间的相似度。相近度的值区间为[0,1],相近度的值越大,代表对应的两个属性越相近,1 代表完全相同,0 代表完全不同。通过计算各属性相近度来聚合相似告警,消除冗余信息,接着针对消除冗余告警后的低阶告警,利用基于因果关系的关联分析算法找出告警之间的内在联系,给出完整的事件描述,及时地发现攻击者的入侵行为。

3.3.3 态势评估

进行态势评估时首先要建立一套态势感知量化评估指标体系,将指标体系作为量化评估的基准。评价指标体系主要由 4 部分组成:网络脆弱性、网络容灾性、网络所受威胁性、网络稳定性。

网络安全态势评估主要是判断潜伏在网络中的风险,通过对其各类因素进行分析,让我们对当前网络整体的安全状况及网络安全防御措施更加清晰明了[4]。在经过合理与全方位的评判后,清楚地掌握整体的网络安全态势。网络态势评估的目的是提高整个网络和系统的安全性,其着眼点在于整体的状况,与网络结构和网络业务紧密相关。文献[5]提出了基于时空关联分析的网络态势评估方法,通过对多角度安全信息的融合,挖掘出攻击间的因果关联;文献[6]提出了基于挖掘攻击场景的态势评估方法,通过还原攻击路径实现网络安全态势的评估。

3.3.4 态势预测

网络安全态势预测是指在过去和当前的态势评估的基础上,通过合理的预测方法,对网络整体或局部的安全态势在未来一段时间的发展趋势进行预测,它是网络安全预警的前提,只有进行准确的预测才能进行高效的预警。网络攻击具有的多样性和隐蔽性等特点使得安全态势变化是一个非常复杂的过程,而大数据技术具备的特点使其在网络安全态势预测方面的应用十分广泛。目前,随着对相关算法探索的不断深入,主要产生了基于灰色理论预测法、基于人工神经网络预测法、基于时间序列预测法[7]和支持向量机预测法等,它们有各自的特点和适用范围。基于灰色理论预测法只能反映安全态势的趋势,对网络安全态势预测的实时性精度有待提高;基于人工神经网络预测法可能会产生预测结果不稳定等问题;基于时间序列预测法是指通过时间序列的历史数据揭示安全态势随时间变化的规律,根据过去的变化趋势预测未来的发展;支持向量机预测法预测绝对误差小,保证了预测的正确趋势率,能准确

(下转第 69 页)

究[J]. 中国铁道科学，1998(3)：1-9.
[16] 刘炯. 基于专家系统的城轨 CBTC 列车运行调整研究[D]. 北京：北京交通大学，2007.
[17] SALIM V，CAI X. A genetic algorithm for railway scheduling with environmental considerations[J]. Environmental Modelling & Software，1997，12(4)：301-309.
[18] 姚丽君. 轨道交通运行调度系统的研究分析[D]. 南京：东南大学，2011.
[19] 牛惠民，陈明明，张明辉. 城市轨道交通列车开行方案的优化理论及方法[J]. 中国铁道科学，2011，32(4)：128-133.
[20] 孔维珍. 基于微粒群算法的城市轨道交通列车运行调整研究[D]. 兰州：兰州交通大学，2013.
[21] 肖枫. 基于模糊神经网络的城市轨道交通行车间隔时间优化研究[D]. 重庆：重庆交通大学，2013.
[22] 李锦，王联国. 基于细菌觅食优化算法的城市轨道交通调度优化[J]. 计算机工程与科学，2017，39(3)：586-592.
[23] 王智慧. 高速列车运行冲突检测与实时调度的模型算法与仿真平台研究[D]. 北京：北京交通大学，2017.
[24] KENNEDY J，EBERHART R. Particle swarm optimization[C]// IEEE International Conference on Neural Networks. 1995：1942-1948.

(上接第 52 页)

预测网络态势的发展趋势。

3.4　态势展示

通过将大数据平台中存储的数据和中间分析结果按照需求进行可视化呈现，以多视图、多维度的方式与用户进行交互，可以帮助我们全面掌握网络安全态势，从而应对日益复杂的网络安全形势。

多源异构的海量数据让我们很难掌握网络的安全状况，通过态势展示我们可以很清晰地了解网络的整体情况。网络态势可视化技术作为一项新技术，是网络安全态势感知与可视化技术的结合，将安全分析的结果和实时安全状态进行可视呈现，发现网络安全攻击和威胁，把握网络安全事件的发展趋势，全方位感知网络安全态势。这种方式提高了对数据的综合分析能力，能够有效降低误报和漏报概率，并且能够对某些网络安全威胁做出预测。

结束语　为了应对日益频繁和严重的网络安全攻击，将基于大数据分析的网络安全态势感知技术应用于网络安全中，不仅能够从整体上动态反映当前网络的安全状况，还能够预测未来的网络安全趋势。本文提出基于大数据的网络态势感知体系架构，包括数据采集、数据预处理、数据存储、大数据分析和态势展示 5 层。从态势认知、关联分析、态势评估和态势预测等多方面重点阐述了如何利用大数据分析进行多源日志的分析处理。在今后的研究中，需要进一步提高关联分析算法的精度，并且在网络安全态势可视化方面做进一步的探索。

参 考 文 献

[1] OGIELA M R，YOU I. Cognitive and secure computing in Information Management[J]. International Journal of Information Management，2013，33(2)：243-244.
[2] CONGH，CHAOW. NetworkSecuritySituationAwareness Based on the Optimized Dynamic Wavelet Neural Network[J]. IJ Network Security，2018，20(3)：593-600.
[3] 杨豪璞，邱辉，王坤. 面向多步攻击的网络安全态势评估方法[J]. 通信学报，2017，38(1)：187-198.
[4] 何骞. 网络安全态势评估若干关键技术研究[J]. 中国新通信，2016，18(14)：42.
[5] LIN Y C，WEI S T，FU L C. Grasping unknown objects using depth gradient feature with eye-in-hand RGB-D sensor[C]// IEEE International Conference on Automation Science and Engineering. IEEE，2014：1258-1263.
[6] WANG N，GONG X，LIU J. A new depth descriptor for pedestrian detection in RGB-D images[C]// International Conference on Pattern Recognition. IEEE，2012：3688-3691.
[7] 席荣荣，云晓春，金舒原，等. 网络安全态势感知研究综述[J]. 计算机应用，2012，32(1)：1-4.