大数据环境下网络安全态势感知研究

□ 曹蓉蓉 / 南京政治学院上海校区军事信息管理系 上海 200433

摘要：随着网络规模和应用的迅速扩大，网络安全威胁不断增加，单一的网络安全防护技术已经不能满足需要。网络安全态势感知能够从整体上动态反映网络安全状况并对网络安全的发展趋势进行预测，大数据的特点为大规模网络安全态势感知研究的突破创造了机遇。文章在介绍网络安全态势相关概念和技术的基础上，对利用大数据开展基于多源日志的网络安全态势感知研究进行了探讨。

关键词：网络安全，态势感知，大数据，数据融合，态势预测
DOI：10.3772/j.issn.1673—2286.2014.02.003

1 引言

随着计算机和通信技术的迅速发展，计算机网络的应用越来越广泛，其规模越来越庞大，多层面的网络安全威胁和安全风险也在不断增加，网络病毒、Dos/DDos攻击等构成的威胁和损失越来越大，网络攻击行为向着分布化、规模化、复杂化等趋势发展，仅仅依靠防火墙、入侵检测、防病毒、访问控制等单一的网络安全防护技术，已不能满足网络安全的需求，迫切需要新的技术，及时发现网络中的异常事件，实时掌握网络安全状况，将之前很多时候亡羊补牢的事中、事后处理，转向事前自动评估预测，降低网络安全风险，提高网络安全防护能力。

网络安全态势感知技术能够综合各方面的安全因素，从整体上动态反映网络安全状况，并对网络安全的发展趋势进行预测和预警。大数据技术特有的海量存储、并行计算、高效查询等特点，为大规模网络安全态势感知技术的突破创造了机遇，借助大数据分析，对成千上万的网络日志等信息进行自动分析处理与深度挖掘，对网络的安全状态进行分析评价，感知网络中的异常事件与整体安全态势。

2 网络安全态势相关概念

2.1 网络态势感知

态势感知（Situation Awareness，SA）的概念是1988年Endsley提出的，态势感知是在一定时间和空间内对环境因素的获取，理解和对未来短期的预测。整个态势感知过程可由图1所示的三级模型直观地表示出来。

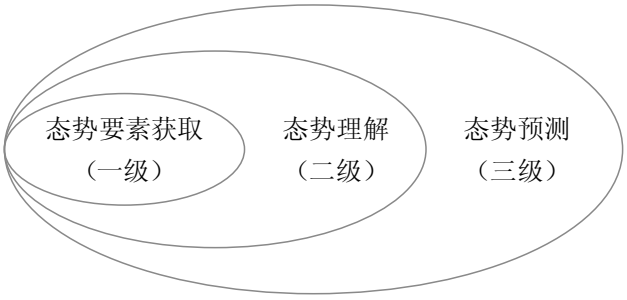

图1 态势感知的三级模型

所谓网络态势是指由各种网络设备运行状况、网络行为以及用户行为等因素所构成的整个网络当前状态和变化趋势。

网络态势感知（Cyberspace Situation Awareness，CSA）是1999年Tim Bass首次提出的，网络态势感知是在大规模网络环境中，对能够引起网络态势发生变化的安全要素进行获取、理解、显示以及预测最近的发展趋势。

态势是一种状态、一种趋势，是整体和全局的概念，任何单一的情况或状态都不能称之为态势。因此对态势的理解特别强调环境性、动态性和整体性，环境性是指态势感知的应用环境是在一个较大的范围内具有一定规模的网络；动态性是态势随时间不断变化，态势信息不仅包括过去和当前的状态，还要对未来的趋

势做出预测；整体性是态势各实体间相互关系的体现，某些网络实体状态发生变化，会影响到其他网络实体的状态，进而影响整个网络的态势。

2.2 网络安全态势感知

网络安全态势感知就是利用数据融合、数据挖掘、智能分析和可视化等技术，直观显示网络环境的实时安全状况，为网络安全提供保障。借助网络安全态势感知，网络监管人员可以及时了解网络的状态、受攻击情况、攻击来源以及哪些服务易受到攻击等情况，对发起攻击的网络采取措施；网络用户可以清楚地掌握所在网络的安全状态和趋势，做好相应的防范准备，避免和减少网络中病毒和恶意攻击带来的损失；应急响应组织也可以从网络安全态势中了解所服务网络的安全状况和发展趋势，为制定有预见性的应急预案提供基础。

网络安全态势感知的主要任务包括风险感知和事件感知两个方面。风险感知包括网络资产感知和网络脆弱性感知，网络资产感知是指自动、快速发现和收集大规模网络资产的分布情况、更新情况、属性等信息；网络脆弱性感知是分析、发现网络的脆弱性，对脆弱性进行统一标识和管理，网络脆弱性包括不可见脆弱性和可见脆弱性。事件感知主要包括安全事件感知和异常行为感知，安全事件感知是指能够确定安全事件发生的时间、地点、起因、经过和结果；异常行为感知是指通过异常行为判定风险，以弥补对不可见脆弱性、未知安全事件发现的不足，主要面向的是感知未知的攻击。

3 网络安全态势感知相关技术

对于大规模网络而言，一方面网络节点众多、分支复杂、数据流量大，存在多种异构网络环境和应用平台；另一方面网络攻击技术和手段呈平台化、集成化和自动化的发展趋势，网络攻击具有更强的隐蔽性和更长的潜伏时间，网络威胁不断增多且造成的损失不断增大。为了实时、准确地显示整个网络安全态势状况，检测出潜在、恶意的攻击行为，网络安全态势感知要在对网络资源进行要素采集的基础上，通过数据预处理、网络安全态势特征提取、态势评估、态势预测和态势展示等过程来完成，这其中涉及许多相关的技术问题，主要包括数据融合技术、数据挖掘技术、特征提取技术、态势预测技术和可视化技术等。

3.1 数据融合技术

由于网络空间态势感知的数据来自众多的网络设备，其数据格式、数据内容、数据质量千差万别，存储形式各异，表达的语义也不尽相同。如果能够将这些使用不同途径、来源于不同网络位置、具有不同格式的数据进行预处理，并在此基础上进行归一化融合操作，就可以为网络安全态势感知提供更为全面、精准的数据源，从而得到更为准确的网络态势。数据融合技术是一个多级、多层面的数据处理过程，主要完成对来自网络中具有相似或不同特征模式的多源信息进行互补集成，完成对数据的自动监测、关联、相关、估计及组合等处理，从而得到更为准确、可靠的结论。数据融合按信息抽象程度可分为从低到高的三个层次：数据级融合、特征级融合和决策级融合，其中特征级融合和决策级融合在态势感知中具有较为广泛的应用。

3.2 数据挖掘技术

网络安全态势感知将采集的大量网络设备的数据经过数据融合处理后，转化为格式统一的数据单元。这些数据单元数量庞大，携带的信息众多，有用信息与无用信息鱼龙混杂，难以辨识。要掌握相对准确、实时的网络安全态势，必须剔除干扰信息。数据挖掘就是指从大量的数据中挖掘出有用的信息，即从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中发现隐含的、规律的、事先未知的，但又有潜在用处的并且最终可理解的信息和知识的非平凡过程（Nontrivial Process）[1]。数据挖掘可分为描述性挖掘和预测性挖掘，描述性挖掘用于刻画数据库中数据的一般特性；预测性挖掘在当前数据上进行推断，并加以预测。数据挖掘方法主要有：关联分析法、序列模式分析法、分类分析法和聚类分析法。关联分析法用于挖掘数据之间的联系；序列模式分析法侧重于分析数据间的因果关系；分类分析法通过对预先定义好的类建立分析模型，对数据进行分类，常用的模型有决策树模型、贝叶斯分类模型、神经网络模型等；聚类分析不依赖预先定义好的类，它的划分是未知的，常用的方法有模糊聚类法、动态聚类法、基于密度的方法等。

3.3 特征提取技术

网络安全态势特征提取技术是通过一系列数学方法处理，将大规模网络安全信息归并融合成一组或者几组在一定值域范围内的数值，这些数值具有表现网络实时运行状况的一系列特征，用以反映网络安全状况和受威胁程度等情况。网络安全态势特征提取是网络安全态势评估和预测的基础，对整个态势评估和预测有着重要的影响，网络安全态势特征提取方法主要有层次分析法、模糊层次分析法、德尔菲法和综合分析法。

3.4 态势预测技术

网络安全态势预测就是根据网络运行状况发展变化的实际数据和历史资料，运用科学的理论、方法和各种经验、判断、知识去推测、估计、分析其在未来一定时期内可能的变化情况，是网络安全态势感知的一个重要组成部分。网络在不同时刻的安全态势彼此相关，安全态势的变化有一定的内部规律，这种规律可以预测网络在将来时刻的安全态势，从而可以有预见性地进行安全策略的配置，实现动态的网络安全管理，预防大规模网络安全事件的发生。网络安全态势预测方法主要有神经网络预测法、时间序列预测法、基于灰色理论预测法。

3.5 可视化技术

网络安全态势生成是依据大量数据的分析结果来显示当前状态和未来趋势，而通过传统的文本或简单图形表示，使得寻找有用、关键的信息非常困难。可视化技术是利用计算机图形学和图像处理技术，将数据转换成图形或图像在屏幕上显示出来，并进行交互处理的理论、方法和技术。它涉及计算机图形学、图像处理、计算机视觉、计算机辅助设计等多个领域。目前已有很多研究将可视化技术和可视化工具应用于态势感知领域，在网络安全态势感知的每一个阶段都充分利用可视化方法，将网络安全态势合并为连贯的网络安全态势图，快速发现网络安全威胁，直观把握网络安全状况。

4 基于多源日志的网络安全态势感知

随着网络规模的扩大以及网络攻击复杂度的增加，入侵检测、防火墙、防病毒、安全审计等众多的安全设备在网络中得到广泛的应用，虽然这些安全设备对网络安全发挥了一定的作用，但存在着很大的局限，主要表现在：一是各安全设备的海量报警和日志，语义级别低，冗余度高，占用存储空间大，且存在大量的误报，导致真实报警信息被淹没。二是各安全设备大多功能单一，产生的报警信息格式各不相同，难以进行综合分析整理，无法实现信息共享和数据交互，致使各安全设备的总体防护效能无法得以充分的发挥。三是各安全设备的处理结果仅能单一体现网络某方面的运行状况，难以提供全面直观的网络整体安全状况和趋势信息。为了有效克服这些网络安全管理的局限，我们提出了基于多源日志的网络安全态势感知。

4.1 基于多源日志的网络安全态势感知要素获取

基于多源日志的网络安全态势感知是对部署在网络中的多种安全设备提供的日志信息进行提取、分析和处理，实现对网络态势状况进行实时监控，对潜在的、恶意的网络攻击行为进行识别和预警，充分发挥各安全设备的整体效能，提高网络安全管理能力。

基于多源日志的网络安全态势感知主要采集网络入口处防火墙日志、入侵检测日志，网络中关键主机日志以及主机漏洞信息，通过融合分析这些来自不同设备的日志信息，全面深刻地挖掘出真实有效的网络安全态势相关信息，与仅基于单一日志源分析网络的安全态势相比，可以提高网络安全态势的全面性和准确性。

4.2 利用大数据进行多源日志分析处理

基于多源日志的网络安全态势感知采集了多种安全设备上以多样的检测方式和事件报告机制生成的海量数据，而这些原始的日志信息存在海量、冗余和错误等缺陷，不能作为态势感知的直接信息来源，必须进行关联分析和数据融合等处理。采用什么样的技术才能快速分析处理这些海量且格式多样的数据？大数据的出现，扩展了计算和存储资源，大数据自身拥有的Variety支持多类型数据格式、Volume大数据量存储、Velocity快速处理三大特征，恰巧是基于多源日志的网络安全态势感知分析处理所需要的。大数据的多类型数据格式，可以使网络安全态势感知获取更多类型的日

志数据，包括网络与安全设备的日志、网络运行情况信息、业务与应用的日志记录等；大数据的大数据量存储正是海量日志存储与处理所需要的；大数据的快速处理为高速网络流量的深度安全分析提供了技术支持，为高智能模型算法提供计算资源。因此，我们利用大数据所提供的基础平台和大数据量处理的技术支撑，进行网络安全态势的分析处理。

关联分析。网络中的防火墙日志和入侵检测日志都是对进入网络的安全事件的流量的刻画，针对某一个可能的攻击事件，会产生大量的日志和相关报警记录，这些记录存在着很多的冗余和关联，因此首先要对得到的原始日志进行单源上的关联分析，把海量的原始日志转换为直观的、能够为人所理解的、可能对网络造成危害的安全事件。基于多源日志的网络安全态势感知采用基于相似度的报警关联，可以较好地控制关联后的报警数量，有利于减少复杂度。其处理过程是：首先提取报警日志中的主要属性，形成原始报警；再通过重复报警聚合，生成聚合报警；对聚合报警的各个属性定义相似度的计算方法，并分配权重；计算两个聚合报警的相似度，通过与相似度阀值的比较，来决定是否对聚合报警进行超报警；最终输出属于同一类报警的地址范围和报警信息，生成安全事件。

融合分析。多源日志存在冗余性、互补性等特点，态势感知借助数据融合技术，能够使得多个数据源之间取长补短，从而为感知过程提供保障，以便更准确地生成安全态势。经过单源日志报警关联过程，分别得到各自的安全事件。而对于来自防火墙和入侵检测日志的的多源安全事件，采用D-S证据理论（由Dempster于1967年提出，后由Shafer于1976年加以推广和发展而得名）方法进行融合判别，对安全事件的可信度进行评估，进一步提高准确率，减少误报。D-S证据理论应用到安全事件融合的基本思路：首先研究一种切实可行的初始信任分配方法，对防火墙和入侵检测分配信息度函数；然后通过D-S的合成规则，得到融合之后的安全事件的可信度。

态势要素分析。通过对网络入口处安全设备日志的安全分析，得到的只是进入目标网络的可能的攻击信息，而真正对网络安全状况产生决定性影响的安全事件，则需要通过综合分析攻击知识库和具体的网络环境进行最终确认。主要分为三个步骤：一是通过对大量网络攻击实例的研究，得到可用的攻击知识库，主要包括各种网络攻击的原理、特点，以及它们的作用环境等；二是分析关键主机上存在的系统漏洞和承载的服务的可能漏洞，建立当前网络环境的漏洞知识库，分析当前网络环境的拓扑结构、性能指标等，得到网络环境知识库；三是通过漏洞知识库来确认安全事件的有效性，也即对当前网络产生影响的网络攻击事件。在网络安全事件生成和攻击事件确认的过程中，提取出用于对整个网络安全态势进行评估的态势要素，主要包括整个网络面临的安全威胁、分支网络面临的安全威胁、主机受到的安全威胁以及这些威胁的程度等。

5 结语

为了解决日益严重的网络安全威胁和挑战，将态势感知技术应用于网络安全中，不仅能够全面掌握当前网络安全状态，还可以预测未来网络安全趋势。本文在介绍网络安全态势相关概念和技术的基础上，对基于多源日志的网络安全态势感知进行了探讨，着重对基于多源日志的网络安全态势感知要素获取，以及利用大数据进行多源日志的关联分析、融合分析和态势要素分析等内容进行了研究，对于态势评估、态势预测和态势展示等相关内容，还有待于进一步探讨和研究。

参考文献

[1] 张云涛,龚玲.数据挖掘原理与技术[M].北京:电子工业出版社,2004.

[2] 席荣荣.网络安全态势感知综述[J].计算机应用,2012,32(1):1-4.

[3] 陈秀真,郑庆华,管晓宏,等.层次化网络安全威胁态势量化评估方法[J].软件学报,2006,17(4):885-897.

[4] 龚正虎,卓莹.网络态势感知研究[J].软件学报, 2010,21(7):1605-1619.

[5] 韦勇,连一峰,冯国登.基于信息融合的网络安全态势评估模型[J].计算机研究与发展,2009,46(3):353-362.

[6] 刘鹏,孟炎,吴艳艳.大规模网络安全态势感知及预测[J].计算机安全,2013(3):28-35.

[7] HALL D L. Mathematical Techniques in Multi-sensor Data Fusion [M]. Boston: Artech House, 2012: 125-137.

作者简介

曹蓉蓉（1963-），南京政治学院上海校区军事信息管理系副教授，研究方向：信息安全、信息系统。E-mail: crr1001@163.com

Research of Network Security Situation Awareness under Big Data Environment

Cao Rongrong / Department of Military Information Management, Shanghai Branch of Nanjing Political College, Shanghai, 200433

Abstract: With the rapid expansion of the network scale and its applications, network security threats continue to increase, a single network security technology could not meet the requirement. Network security situation awareness can dynamically reflect the overall network security and predict network security development trends. Characteristics of big data create opportunity for research breakthrough of large scale network situation awareness. Based on concept and technique introduction of network security situation awareness, this article discusses about research of network security situation awareness based on multi-source journal by utilizing big data analysis.

Keywords: Network security, Situation awareness, Big data, Data fusion, Situation prediction

（收稿日期：2014-01-08）