文章编号:1671-4598(2019)10-0229-05 DOI:10.16526/j.cnki.11—4762/tp.2019.10.047 中图分类号:TN915.08 文献标识码:A

基于大数据的网络安全态势感知及主动防御技术研究与应用

刘冬兰，刘 新，张 昊，于 淼，马 雷，赵晓红
（国网山东省电力公司电力科学研究院，济南 250003）

摘要：针对电力信息系统网络安全态势感知及主动防御问题，介绍了网络安全态势感知相关概念及技术；为了监控网络安全态势，研究了利用大数据分析技术开展基于多源日志的网络安全态势感知，提出了态势感知平台部署架构及主动防御模型思想，并将其技术应用于某电力公司网络信息系统环境；通过在公司内外网网络出口部署全流量数据采集分析器，对原始网络流量进行实时采集和存储，并借助大数据可视化分析工具与丰富的数据展示组件，实现对分析结果的多维度图形化直观展现；经实验测试实现了攻击事件及安全态势的实时监控预警，保障了公司信息系统的安全稳定运行。

关键词：大数据；网络安全；态势感知；主动防御

Research and Application of Network Security Situation Awareness and Active Defense Based on Big Data Technology

Liu Donglan, Liu Xin, Zhang Hao, Yu Hao, Ma Lei, Zhao Xiaohong
(State Grid Shandong Electric Power Research Institute, Jinan 250003, China)

Abstract: In view of the problem of network security situation awareness and active defense of power information system, this paper introduces the related concepts and technologies of network security situational awareness. In order to monitor network security problems, a network security situation awareness technology based on multi-source logging methods by utilizing big data analysis is proposed. The deployment architecture of situation awareness platform and the idea of active defense model are proposed and applied to the information system environment of a certain electric power company. We deployed network traffic security analyzer in the export of company's internal and external network. It can acquire and storage the original network traffic in real time. By using the big data visualization analysis tool and rich data display component, the realization of the multidimensional graphical visualization of the analysis results is presented. Through the experimental test, it realizes the real-time monitoring and early warning of the attack event and security situation, and guarantees the safe and stable operation of the company's information system.

Keywords: big data; network security; situation awareness; active defense

0 引言

近年来，针对能源电力行业的网络攻击越来越多，而国家电网公司是全国能源供给核心，电网安全关乎全国经济安全。电力行业的信息化发展水平和速度居各行业前列，并且安全防御体系的建设一直是电力行业信息化工作的重点。目前电力信息系统网络内已部署了丰富的安全设备，包括防火墙、WEB应用防火墙、入侵检测系统（IDS）、入侵防御系统（IPS）、数据库审计设备等，安全监测的内容不断细化，但由于传统安全设备的相互孤立性，产生的安全情报数据呈现出分散和孤立的共性，传统的分析方法和模型已不能满足目前的安全现状需求，究其原因在于传统安全设备相互孤立，产生的安全日志数据的种类和数量增长迅速，呈现出数据量大、异构的特性，而目前缺失数据统一收集、处理和分析的系统化技术手段。规范提升对这部分数据的分析、挖掘和有效利用，是改进当前安全威胁防御方式亟待解决的问题。

国家电网公司正在大力推进坚强智能电网和泛在电力物联网建设，电网数字化和智能化程度不断提高，伴随着各个信息系统的交互性要求也越来越高，网络信息安全给电力信息系统及泛在电力物联网建设提出了更高的要求。比如相关安全数据的采集和存储能力、信息系统安全威胁的发现感知能力、立体化纵深防御能力等方面，都面临着相比过去传统信息系统的安全防护体系更高的技术和管理规范化要求。随着相关安全威胁情报数据的数量及数据种类不断增多，数据以指数级快速增长，大量数据的采集、存储、融合、分析变得越来越困难。当前，网络攻击行为呈现出复杂化、分布化等特点，防火墙、入侵检测等网络安全设备已经不能满足网络空间变化的需求。因此，需要研究新技术来感知预警网络中的攻击等异常事件，提高网络安全防护能力。

网络安全态势感知技术能够全面感知网络安全威胁态

收稿日期:2019-08-03； **修回日期**:2019-08-17。
基金项目:国网山东省电力公司科技项目(52062617002V)。
作者简介:刘冬兰(1987-),女,云南宣威人,硕士,高级工程师,主要从事网络安全,大数据,区块链等技术方向的研究。

势、洞悉网络及应用运行健康状态、通过全流量分析技术实现完整的网络攻击溯源取证，帮助安全人员采取针对性响应处置措施。网络安全态势感知技术研究涉及到大数据分析、数据融合处理等相关技术[1]。通过对采集的网络安全日志数据进行全流量分析处理，可挖掘网络中的攻击事件，进而感知公司的网络安全态势。

针对电力信息系统大数据环境下的安全威胁和各类网络攻击，研究了基于大数据的电力信息系统网络安全态势感知及主动防御技术。通过在公司网络出口处部署全流量数据采集器，对安全日志、网络流量、APT 等多维度异构数据源进行全要素的信息采集，并采用高性能的处理器对数据进行处理存储，采用大数据相关技术对攻击数据包进行回溯分析，对攻击态势进行图形化直观展示，可以提供从攻击预警、识别和分析取证的全面分析能力。

1 网络安全态势感知相关概念及技术

1.1 网络安全态势感知相关概念

态势感知是一种基于环境的、动态、整体地洞悉安全风险的能力，是以安全大数据为基础，通过在一定时间和空间内对公司网络信息系统的全部数据流量进行采集获取，进而匹配攻击关联规则，并对未来可能发生的攻击事件进行预测。整个态势感知过程包括态势要素获取、态势理解和态势预测[2-5]。网络态势是指对各类网络设备的运行数据、网络中发生的用户访问及攻击等行为构成的网络状态。网络态势感知是在大规模网络环境中，对网络运行中的安全要素进行获取、理解、分析、展示并预测最近的发展趋势[6]。

网络安全态势感知通过对网络中的全部数据流量进行实时采集，再运用数据融合、数据挖掘技术对流量进行分析处理，再采用大数据可视化技术进行直观展示，实时展示网络的运行安全状况，为网络安全提供保障[7-9]。通过实时感知网络中的安全态势，可以及时发现网络中的攻击事件，并及时对公司信息系统存在的漏洞进行加固处理，避免和减少公司网络被恶意人员攻击的风险。

1.2 网络安全态势感知相关技术

目前，网络空间数据呈爆发式增长，各单位都加大了网络安全设备的投入，网络拓扑也越来越复杂，随着业务的增多，业务系统平台也逐渐增多。但是，网络攻击的手段和方法却越来越先进，涌现了很多新型的零日漏洞，同时，随之出现了很多自动化和智能化的攻击工具，导致网络威胁逐渐增多，造成的损失也逐渐增多。为了实时、准确地显示整个网络的安全态势，需要对网络数据流量进行全流量采集、数据预处理、融合分析等环节，实时监控网络数据流量。网络安全态势感知涉及的技术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[10-11]。

数据融合技术是指利用计算机对按时序获得的若干观测信息，在一定准则下加以自动分析、综合，以完成所需的决策和评估任务而进行的信息处理技术。这个处理过程主要是对具有相似特征的数据进行整合，按照数据的关联性进行合并等融合处理，从而得到更为准确、可靠的结论[11-12]。

数据挖掘是指从大量的数据中通过算法搜索隐藏于其中信息的过程，通过统计、在线分析处理、情报检索、机器学习、专家系统和模式识别等诸多方法来挖掘出有用的信息，找出能被大家理解的信息和知识的过程[10,13]。

网络安全态势特征提取技术是通过对网络数据流量进行融合处理后，与攻击特征库规则进行对比分析，找出能体现网络实时运行状况的数据特征以及攻击特征，从而能够判断出网络是否安全或者网络被黑客攻击面临的威胁情况[14]。网络安全态势特征提取是评估预测网络态势的前提，目前网络安全态势特征提取方法主要有模糊层次分析法、层次分析法、德尔菲法和综合分析法[11,15]。网络安全态势预测是网络安全态势感知的重要环节，主要通过对网络数据流量分析出网络的运行状况，再运用科学的理论方法推测网络在未来可能发现的变化。由于网络态势在不同时间段彼此相关，可以根据安全态势的变化预测未来可能受到的网络攻击，从而可以有针对性地对网络设备配置合理的安全策略，采取主动防御的思想，预防大规模网络安全事件的发生[11-12]。

可视化技术是指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[16]。目前已有很多研究将可视化技术和可视化工具应用于网络安全态势感知领域，通过可视化方式实时展现国内外对公司网络信息系统的攻击情况，可以准确定位出攻击源及攻击目标，从而更方便帮助用户从安全态势图上快速找出安全威胁，更直观显示出网络安全状态。

2 态势感知平台部署架构

基于大数据技术的网络安全态势感知预警平台部署架构如图 1 所示，部署架构主要如下：前端服务器主要分为 TSA 服务器、IDS 服务器、防火墙等服务器，每种类型的前端服务器都为大数据分析技术的网络安全态势感知预警平台提供数据来源，供态势感知预警监测系统进行数据分析与检索。

采集服务器负责对 TSA、IDS、APT、IPS 等前端安全服务器数据进行集中收集，并对数据进行过滤，缓存，简单范式化等处理操作。

预处理服务器汇总所有采集服务器上报的数据，并对上报数据进行统一的范式化处理，对采集的数据进行数据归并、数据清洗等操作，并根据不同业务把数据存储到不同的存储系统上。

预处理完成的数据都存放在 hadoop 服务器上，并利用 hadoop 的存储与分析能力，对数据进行关联统计与数据挖掘，形成结果数据并导入检索引擎，供 web 服务器查询数据。

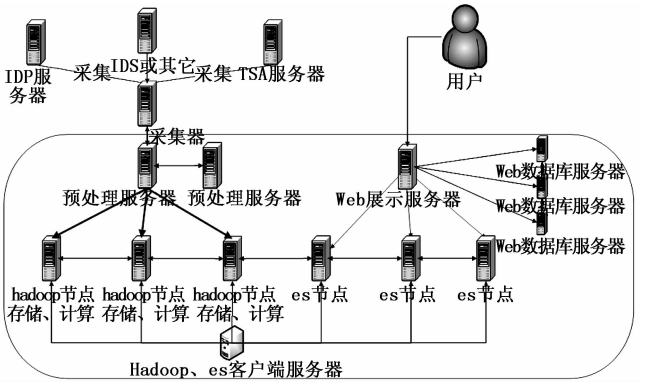

图 1　基于大数据的态势感知平台部署架构

Es 节点（Elasticsearch）服务器对 hadoop 服务器形成的结果数据，进行海量数据的存储与简单的二次统计，并提供接口给 web 服务器检索数据。

客户端服务器针对整个态势感知预警平台，提供自动化运维与监控服务，运维人员通过客户端服务器提供的接口去配置与管理系统平台的任务调度与运维监控。

Web 服务器主要分为 Web 数据库服务器和 Web 展示服务器。Web 数据库服务器主要是存放态势感知预警平台系统的业务功能数据；Web 展示服务器利用业务服务器的基础数据和态势感知预警平台系统的数据按业务功能管理与威胁数据分析两大功能进行数据可视化展现。

3　网络安全态势感知及主动防御模型研究

随着网络规模的扩大以及网络攻击复杂度的增加，防火墙、防病毒、IDS、IPS、安全审计等众多的安全设备应用广泛，各省电力公司加大了网络安全建设的投入力度，先后建立并部署了各种类型的安全设备或系统，如 APT 系统、IDS、IPS、防火墙、杀毒软件等。但基于特征规则的传统安全设备只能检测已知网络攻击，存在较高漏报和误报。而且以纯粹攻防理念为主导的 APT 类产品仅对恶意代码样本进行分析，与业务结合性差，无法做到攻击溯源与取证，仍需人工分析在海量数据中寻找线索，对信息安全专业人员的分析帮助有限。同时，虽然安全运营中心（SOC）和安全信息和事件管理（SIEM）对安全系统的大量日志进行了整合，但数据源单一，而且缺乏提供精准分析的能力与手段，安全分析人员从这些海量数据中分析出有效线索无异于大海捞针。事实上，无论是传统安全设备与系统，还是 SOC 等，都是保证网络安全的重要组成部分。但是，这些部分彼此间相对隔离，信息不能及时共享，针对发生的网络安全事件，依靠人工效率极低且时间滞后，无法发挥各部分最大的安全性能。

为了加强网络安全管理，我们提出了基于多源日志的网络安全态势感知预警技术，将多维度、多层次的安全数据源进行整合，构建基于大数据分析技术的态势感知预警监测平台，充分利用数据之间的内在关系进行深度分析和挖掘，发展全面的态势感知和高效精准的分析技术，可以大大地提升网络空间态势感知与预警监测能力。

研究开发的网络安全态势感知预警平台通过内建以机器学习和智能分析算法为基础的多种网络安全分析模型，采用 TEZ，SPARK 并行计算框架，并利用数据挖掘，文本分析，流量分析，全文搜索引擎，实时处理等方式来对数据进行深度的分析与挖掘，结合模型库内的入侵检测模型、网络异常行为模型、设备异常行为模型，实时甄别未知的安全威胁。通过对公司信息系统进行实时监控，构建了“事前发现、事中控制、事后追踪”的主动防控的信息安全管理体系，变被动防御为主动防控，完善已发生的信息事件应急处置机制，通过主动防御技术措施的实施和管理体系化的创新，提高对未知安全威胁的发现和甄别能力。网络安全态势感知平台主要涉及以下 3 种分析模型的深度研究。

3.1　关联分析模型

网络安全设备产生的日志描述了安全事件的详细过程，针对网络中疑似攻击事件会产生报警，可对相应的报警日志信息进行关联分析，对攻击数据包进行回溯分析，查看相应的 TCP 会话等过程，最终还原出攻击事件的整个过程。通过采用基于相似度的报警关联分析，可以降低关联分析的难度，把报警日志的数量精确到最小，提高了关联分析的准确性。基于相似度的报警关联分析过程如下：

1）提取报警日志中的主要属性，主要包括：时间、攻击源位置、攻击源 IP、攻击类型、攻击目标 IP、攻击目标、所属单位，并将攻击报警属性还原成相应的攻击事件，形成原始报警；

2）通过对重复报警进行合并，对报警进行分类匹配，聚合有较高相似度的报警，生成聚合报警；

3）对聚合后的报警的各个属性定义报警属性相似度的计算方法，并对每个属性分配权重；

4）计算两个聚合报警的相似度，通过比较相似度阀值，判断需不需要对聚合报警进行再次报警；

5）采用基于时间窗口的报警选择方法来选择适量的报警进行关联比较，并构建关联知识库，挖掘原始报警之间的关联关系，绘制报警事件关联图；

6）根据聚合报警事件输出属于同一类的报警信息，生成相应的安全事件。安全事件按照时间节点记录了发生的攻击类型、攻击源 IP、攻击目标 IP、攻击类型、攻击次数等详细信息。

3.2　融合分析模型

由于安全设备采集的日志可能存在冗余性，所以需要对多个设备采集的日志信息进行融合处理，才能生成更准确的安全态势。通过对单源日志报警信息进行相似度的报警关联分析，可以还原每个攻击事件的原始过程。针对多源安全事件，采用 Dempster/Shafer 证据理论（D－S 证据理论）方法进行日志融合判别[17-20]，对安全事件的可信度进行评估，进一步提高准确率，减少安全设备的误报率。采用 D－S 证据理论对网络安全攻击事件进行融合的过

程为：

1）针对安全事件的多维度属性，包括：时间、攻击源位置、攻击源 IP、攻击类型、攻击目标 IP、攻击目标、所属单位，采用多维信息的分域、层次融合方式，利用初始信息确定融合所需的概率分布的近似算法对安全事件数据进行融合处理；

2）对安全设备原始报警日志，根据初始信任分配方法，分配信息度相似函数；

3）通过采用 D－S 的组合规则，计算安全设备报警日志融合之后的安全事件的可信度。

3.3 攻击要素分析模型

通过在网络出口处部署全流量采集设备，可以对全部数据流量进行实时采集，但是采集的全流量数据信息巨大，需要进一步分析出真实的攻击事件。攻击要素分析主要包括如下过程：

1）通过对大量网络攻击实例进行研究，建立攻击特征库；

2）把攻击特征库加入大数据态势感知平台，自动匹配攻击事件；

3）根据攻击事件类型分析被攻击服务器主机上开放的服务端口可能发生的漏洞；

4）建立平台当前网络环境的漏洞知识库；

5）发现攻击异常流量，生成攻击事件；

6）通过与漏洞知识库进行比较，确认攻击事件。

4 实验结果分析

目前公司部署的基于大数据的网络安全态势感知预警平台已接入信息外网出口流量及 18 个重要资产服务器流量、内网出口流量及 24 个重要资产服务器流量、2 台 IPS、2 台防火墙、信息外网 APT 系统，共部署高性能硬件服务器 7 台，其中 ES 节点 3 台，hadoop 节点 3 台，前端展示服务器 1 台。截至目前，已获取威胁事件数据 8 TG，记录威胁数据条数 6 180 兆条，现已加入互联网威胁情报 32 万余条。通过态势感知预警平台监控到前端设备网络流量如图 2 所示。

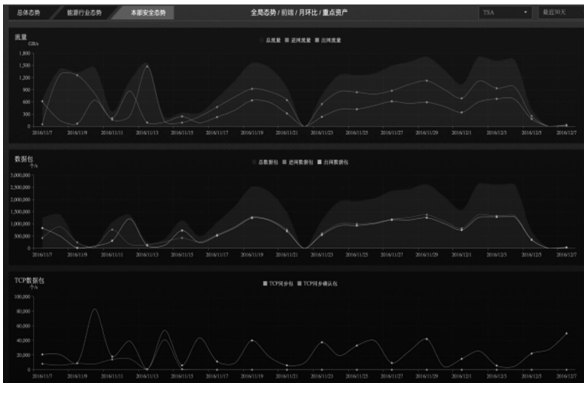

图 2 监控前端设备网络流量

以 2018 年 6 月数据为例，共发现外网威胁事件 674 311 条，重点资产威胁事件 1 870 条，其中，触发 Web 攻击类 2650 次，DDoS 攻击 2300 次，浏览器扫描类 275 500 次；从威胁情报命中占比来看，威胁 ip 占比 53%，威胁域名占比 32%，威胁 URL 占比 12%；从攻击方式看，主动攻击占比 63%，被动攻击占比 37%；从攻击来源看，国内占比 66.84%，美国占比 33.03%。公司外网重点资产被攻击态势展示结果如图 3 所示，威胁事件及占比分析结果如图 4 所示，威胁源国家占比数如图 5 所示。2018 年 10 月、11 月威胁环比数据如图 6 所示，各类攻击威胁次数差异 Top5 如图 7 所示，攻击源 ip 和目的 ip 的行为画像如图 8 所示。从态势感知预警的分析结果中，我们可以清晰地看到公司信息系统面临的安全威胁，从而及时制定相应的安全加固措施，保障了公司信息系统的安全稳定运行。

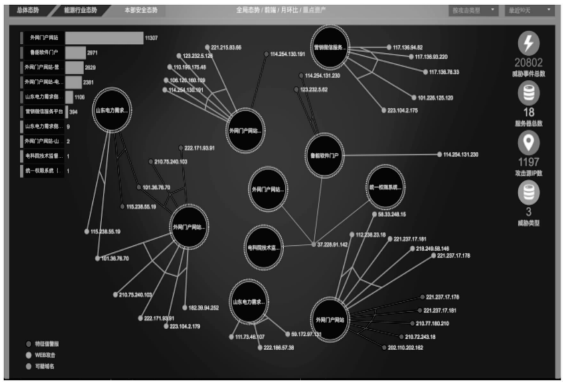

图 3 重点资产被攻击情况

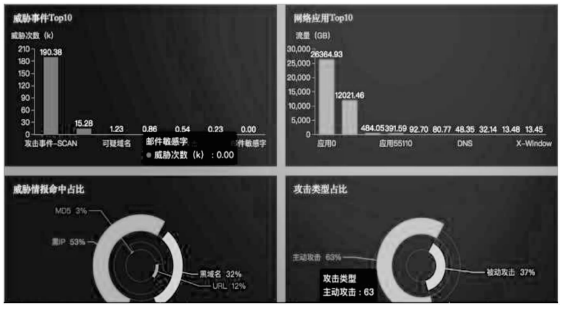

图 4 威胁事件及占比情况

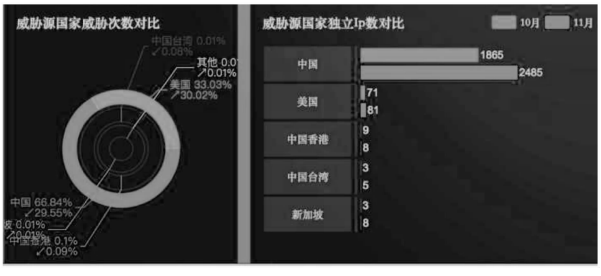

图 5 威胁源国家占比数

5 结语

为了解决日益严重的网络安全威胁，通过采用态势感知技术实时感知预测网络安全威胁。通过在介绍网络安全

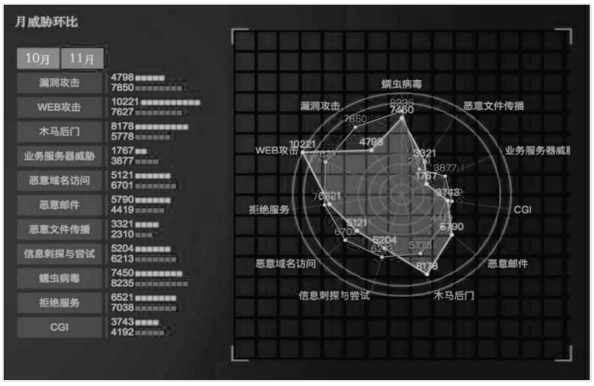

图 6　月威胁环比情况

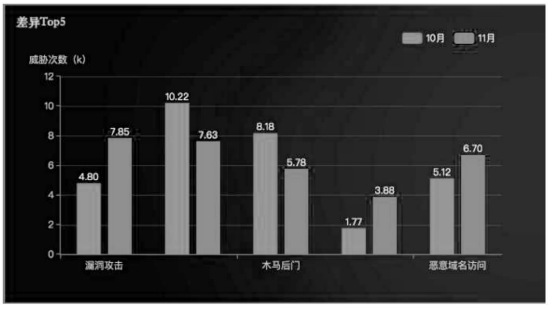

图 7　各类攻击威胁次数差异 Top5

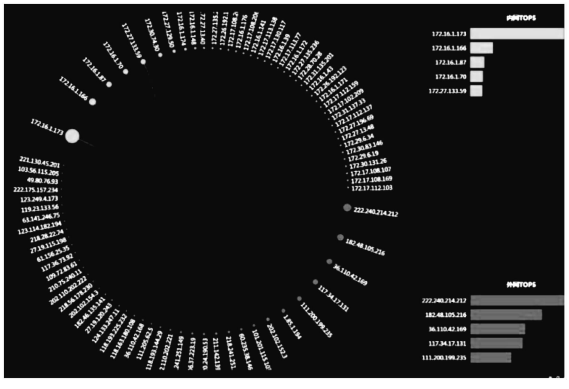

图 8　攻击 IP 行为画像

态势相关概念和技术的基础上，对网络安全态势感知及主动防御技术进行了深入研究，重点研究了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技术，并将其技术应用于公司网络信息系统环境。通过在公司内外网网络出口部署全流量数据采集分析器，对原始网络流量进行实时采集和存储，采用大数据分析技术对安全威胁进行实时智能分析，对网络攻击过程进行回溯分析，并借助大数据展示组件，实现对分析结果的多维度图形化直观展现。通过构建基于大数据的网络安全态势感知预警平台，实时监控公司电力信息系统面临的安全威胁，及时制定相应的安全加固措施，保障了公司信息系统的安全稳定运行。

参考文献：

[1] Lenders V，Tanner A，Blarer A. Gaining an edge in cyberspace with advanced situational awareness [J]. Security & Privacy IEEE，2015，13 (2)：65－74.

[2] 龚　俭，藏小东，苏　琪，等. 网络安全态势感知综述 [J]，软件学报，2017，28 (4)：1010－1026.

[3] Bass T. Intrusion detection systems and multisensor data fusion：Creating cyberspace situational awareness [J]. Communications of the ACM，2000，43 (4)：99－105.

[4] Endsley M R. Toward a theory of situation awareness in dynamic system [J]. Human Factors，1995，37 (1)：32－64.

[5] Franke U，Brynielsson J. Cyber situational awareness a systematic review of the literature [J]. Computers & Security，2014，46：18－31.

[6] 傅里辉，刘俊丽，陈双喜. 基于大数据的系统安全态势感知研究 [J]. 时代农机，2016，43 (11)：49－50.

[7] 韦　勇，连一峰，冯登国. 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.

[8] 王慧强，赖积保，朱　亮，等. 网络态势感知系统研究综述 [J]. 计算机科学，2006，33 (10)：5－10.

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

[10] 曹蓉蓉. 大数据环境下网络安全态势感知研究 [J]. 数字图书馆期刊，2014，117 (02)：11－16.

[11] 宋一非. 基于大数据的信息安全威胁感知处置平台 [D]. 天津：天津大学，2016.

[12] 毛军礼，汲锡林. 基于大数据的网络态势感知体系架构 [J]. 无线电通信技术，2018，44 (3)：217－223.

[13] 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.

[14] 方　圆. 基于大数据的网络安全态势感知 [J]，科技风，2016，12 (1)：96－98.

[15] Zhang M J，Li S P，Li X. Research on technologies of underwater feature extraction and target location based on binocular vision [A]. 2015 27th Chinese Control and Decision Conference (CCDC) [C]. 2015：5798－5804.

[16] Shiravi H，Shiravi A，Ghorbani A A. A survey of visualization systems for network security [J]. IEEE Transactions on Visualization and Computer Graphics，2012，18 (8)：1313－1329.

[17] 高湛军，李思远，彭正良，等. 基于网络树状图和改进 D－S 证据理论的配电网故障定位方法 [J]. 电力自动化设备，2018，38 (6)：65－71.

[18] 张翠玲，王大志，江雪晨，等. 基于选择判据和贴近度的 D－S 证据融合方法 [J]. 计算机测量与控制，2015，23 (10)：81－85.

[19] 李　月，徐余法，陈国初，等. D－S 证据理论在多传感器故障诊断中的改进及应用 [J]. 东南大学学报：自然科学版，2011，41 (S1)：102－106.

[20] Guan X，Yi X，He Y. An improved Dempster－Shafer algorithm for resolving the conflicting evidences [J]. Information Fusion，2005，11 (12)：68－75.