•科学前沿•

DOI:10.15961/j.jsuese.201600352

基于大数据的网络安全与情报分析

陈兴蜀[1], 曾雪梅[1*], 王文贤[1], 邵国林[2]

(1. 四川大学 网络空间安全研究院, 四川 成都 610065; 2. 四川大学 计算机学院, 四川 成都 610065)

摘　要:随着IT技术和通信技术的发展,网络环境日趋复杂,云计算和虚拟化等技术的应用,使得主机边界、网络边界也变得动态和模糊。同时,网络攻击频繁,隐蔽性、持续性、趋利性等高级网络威胁增多。而传统网络安全与情报分析技术受数据来源单一、处理能力有限、部署依赖于物理环境等因素的限制,导致对威胁情报的获取、分析、利用能力不足,且对网络安全态势的感知与预测能力有限,不能有效解决当前和未来所面临的网络安全挑战。作者以大数据技术给网络安全与情报分析研究带来的挑战与机遇为线索,回顾大数据的内涵,分析当前网络安全与情报分析面临的困境,梳理大数据和网络安全与情报分析的关系,阐述大数据技术对传统安全分析方法的改变。大数据技术在安全领域应用形成大数据安全分析这一新型安全应对方法,通过紧扣安全数据自身的特点和安全分析的目标,应用大数据分析的方法和技术,解决网络安全与情报分析中的实际问题。一方面,批量数据处理技术、流式数据处理技术、交互式数据查询技术等大数据处理技术解决了高性能网络流量的实时还原与分析、海量历史日志数据分析与快速检索、海量文本数据的实时处理与检索等网络安全与情报分析中的数据处理问题;另一方面,大数据技术应用到安全可视分析、安全事件关联、用户行为分析中,形成大数据交互式可视分析、多源事件关联分析、用户实体行为分析、网络行为分析等一系列大数据安全分析研究分支,以应对当前的网络安全挑战。大数据安全分析技术在APT攻击检测、网络异常检测、网络安全态势感知、网络威胁情报分析等方面已经得到应用,但是,当前的网络安全形势仍不容乐观:高级网络威胁与攻击的有效检测方法缺乏;未知复杂网络攻击与威胁预测能力不足;缺乏度量网络安全态势评估结果的评价体系,关键资产与网络整体的态势评估指标体系不完善,网络安全态势感知评估方法缺少针对性;网络威胁情报信息分析的新型数据源数据获取难度大,缺乏威胁情报共享标准,尚未建成规模化、一体化的现代威胁情报中心和开放的威胁情报综合服务平台。围绕这些问题,需要研究高级网络威胁发现方法、复杂网络攻击预测方法、大规模网络安全态势感知技术、威胁情报数据采集与共享技术,并在高级网络威胁早期检测、隐蔽性和持续性网络通信行为检测、基于大数据分析的网络特征提取技术、综合威胁情报的高级网络威胁预测、非公开网络情报采集等关键技术上实现突破,以提升大数据对网络信息安全的支撑能力,增强网络信息安全风险感知、预警和处置能力。

关键词:大数据;网络安全;情报分析

中图分类号:TP391.4　　**文献标志码**:A　　**文章编号**:2096-3246(2017)03-0001-12

Big Data Analytics for Network Security and Intelligence

CHEN Xingshu[1], *ZENG Xuemei*[1*], *WANG Wenxian*[1], *SHAO Guolin*[2]

(1. Cybersecurity Research Inst., Sichuan Univ., Chengdu 610065; 2. College of Computer Sci., Sichuan Univ., Chengdu 610065, China)

Abstract: With the development of IT and communication technology,the network environment is becoming more and more complicated, and the perimeters of host and network become dynamic and fuzzy due to the application of cloud computing and visualization technology.At the same time,network attacks become more frequent and advanced network threats with evasive and persistent behavior and profit-chasing are also increasing.However,due to the limit of data source and process ability and device deployment relied on physical environment,traditional network security and intelligence techniques are inefficient on the acquisition capability,analytical abilities and utilize capacity of threat intelligence,and the

收稿日期:2017 – 04 – 23

基金项目:国家自然科学基金资助项目(61272447)

作者简介:陈兴蜀(1968—),女,博士,教授,博士生导师. 研究方向:云计算及大数据安全;可信计算.E-mail: chenxsh@scu.edu.cn

* 通信联系人 E-mail:zengxm@scu.edu.cn

网络出版时间:2017 – 05 – 22 09 : 54 : 54　**网络出版地址**:http://kns.cnki.net/kcms/detail/51.1773.TB.20170522.0954.001.html

perception and prediction ability of the network security situation is limited,so it cannot solve the current and future network security challenges efficiently.The chances and challenges caused by big data for network security and intelligence analysis are took as this paper clue.first of all,the connotation of big data is reviewed,and the current dilemmas in network security and intelligence analysis are analyzed,and then the relationship between big data with network security and intelligence analysis is explored,and the changes of traditional security analysis brought by big data technologies are parsed.Big data security analysis,a new security method is formed after the big data technologies was applied in cyber security field.The value of big data security analysis embodies in solving practical problems in network security and intelligence analysis through the methods and technologies of big data analysis under sticking to the purpose of security analysis and the character of security data itself.On the one hand,big data processing technology,such as bulk data processing technology,streaming data processing technology,interactive data query technology,can solve the issues of data processing in the high-performance network traffic real-time restore and analysis,massive historical log data analysis and rapid retrieval,massive text data real-time processing.On the other hand,big data technologies are applied in security visual analysis,security event association and network user behavior analysis,a series research branch of big data security analysis are formed,such as big data interactive visual analysis,multi-source event correlation analysis,user entity behavior analysis,network behavior analysis and so on.Big data security analysis technologies have applied in APT attack detection,network anomaly detection,network security situation perception,network threat intelligence analysis,etc.Even though some achievements have been made in big data based network security and intelligence analysis,the current network security situation is still not optimistic.The effective detection method of advanced network threats and attacks is lacking.The detection and prediction result to unknown complex network attacks is undesirable.The measurement system for evaluation methods of network security situation awareness is needed,and large-scale network security situation awareness indicator system for key assets and entire network is incomplete yet,and the evaluation methods are no pertinence.It is difficult to acquire data from new type data sources of threat information and the standards of threat intelligence sharing are needed to be researched further.Large-scale and integrate threat intelligence center and open service platform aren't yet build.Around the above problems,it needs to be studied that the methods of advanced network threat discovery,complex network attack prediction,large-scale network security situational awareness and threat information collection and sharing technology and needs some key technical breakthroughs such as early detection of advanced network threats,concealment of continuous network communication behavior detection,big data analytics based network feature extraction technology,intelligent-based advanced network threat forecast,non-public network intelligence collection,so as to improve the big data supporting capability for network information security and enhance network information security risk perception and disposal capacity.

Key words: big data;network security;intelligence analysis

随着IT技术与通信技术的发展，计算能力和网络带宽迅速提升，云计算、物联网、社交网络等新兴服务兴起，数据正以前所未有的速度增长和累积，大数据时代已经到来。大数据技术具有从数量巨大、结构复杂、类型众多的数据中快速获得有价值信息的能力，能够揭示传统手段所看不到的内容和变化趋势，是当前学术界、产业界甚至各国政府关注的热点。大数据技术为信息安全产业发展带来新的机遇。

然而随着网络环境的日趋复杂，网络攻击频繁出现，具有广泛性、隐蔽性、持续性、趋利性的网络攻击与信息窃取已经从个人蔓延到金融、通信、能源、航空、交通等许多领域，对公民、企业及国家信息安全构成了严重威胁，应对新的网络安全问题需要基于长时间的历史数据与多源信息开展网络安全分析；随着互联网应用、交互方式的激增，网民在网络空间的活动具有多样性、灵活性、持续性等特征，挖掘有效的情报需要关联更多的信息。2012年3月，Gartner指出："信息安全问题正在变成一个大数据分析问题，大规模的安全数据需要被有效地关联、分析和挖掘"[1]。应用大数据分析技术能够对当前和历史的各种类型数据进行关联分析与检索，帮助安全管理者实时洞悉安全情报和安全态势，快速做出判断和响应。

作者以大数据技术给网络安全与情报分析研究带来的机遇为线索，首先回顾大数据的内涵，梳理大数据和网络安全与情报分析的关系；然后，探讨大规模安全数据的处理与分析关键技术，列举大数据技术在网络安全与情报分析中的典型应用；最后，分析大数据环境下网络安全与情报分析的研究趋势。

1 大数据安全分析

1.1 大数据的内涵

大数据是具有数量巨大（volume）、来源多样（variety）、生成极快（velocity）、多变（variability）等特征，且难以用传统数据体系架构有效处理的包含大量数据集的数据[2]。美国国家标准技术研究所（NIST）大数据公共工作组定义和分类小组认为大数据的上述4V特性是驱动数据密集型应用向新架构转变的动力[3]。业界普遍认为从大量的数据中抽取信息，处理的数据越多获得的价值越大。中国信息通信研究院发布的《大数据白皮书（2016年）》指出大数据是新资源、新技术、新理念的混合体[4]。从资源的视角来看，大数据是新资源，体现一种全新的资源观；从技术的视角看，大数据代表了新一代数据管理与分析技术；从理念的视角看，大数据定义了一种全新的思维角度，即带来了"实事求是"的新内涵——数据驱动与数据闭环。

大数据技术是使大数据中所蕴含的价值得以挖掘

和展现的一系列技术与方法，包括数据采集、预处理、存储、分析挖掘、可视化等[2]。从大数据技术研究领域的角度，目前大数据主要涉及2个不同的技术领域：一是，致力于研发可以扩展至PB、EB、ZB级别的大数据存储平台；二是，大数据分析（big data analysis/analytics），关注在短时间内处理大量不同类型的数据集，分析出高价值信息，是体现大数据核心价值的关键。

1.2 传统网络安全与情报分析的困境

1.2.1 网络安全分析的困境

随着IT架构日益复杂，各种应用不断涌现，数据和业务更加集中，网络和应用的边界越来越模糊，体现出随业务、资源等变化的动态性，基于单一边界、控制点的传统网络安全设备难以有效掌握整个网络或系统的安全状态。

日志、网络流量等数据用于安全分析已经很成熟，但是，由于保留和分析大量数据所消耗的成本较高，系统日志与主机活动等数据一般保留一段时间后被删除。为实现对网络或云计算平台等系统的全面安全分析，需要从全局的角度获取安全分析所需数据，包括：网络数据包、日志、资产状态、业务信息、漏洞信息、身份认证与访问信息、用户行为信息、配置信息等，可能还需要来自互联网的外部情报信息等数据。这些数据产生的速度越来越快，且数据类型涵盖结构化、半结构化和非结构化，呈现出大数据的特点。传统网络安全检测方法受数据源保留时间、数据分散和数据处理能力的限制，无法有效应对。

网络攻击手段不断更新，技术复杂性增加，僵尸网络、特种木马与蠕虫、高级持续威胁（advanced persistent threat，APT）等网络攻击目标性和趋利性增强，显示出长期性、多路径性、复合性、隐蔽性等攻击特征，传统网络攻击检测技术难以有效检测具有长期性、隐蔽性的新型网络攻击。

1.2.2 情报分析的困境

随着云计算、移动互联等信息技术的迅猛发展，互联网上威胁情报信息的信息源越来越多，传统的情报分析工具因数据源单一、大规模数据关联效能低，无法满足新常态下的情报挖掘分析需求，包括：建立高效智能的外部信息源搜索、信息采集方法；对内部和外部采集的大量非结构化数据进行快速处理和存储；实施多源数据复杂关联、快速检索与情报跟踪等。

1.3 大数据安全分析

大数据分析技术应用在安全领域，与传统安全技术相结合，诞生了新型安全应对方法——大数据安全分析（big data analysis for security）。大数据安全分析技术以前所未有的规模和速度实现对海量多源异构数据集的存储、处理与分析，解决传统网络安全与情报分析的问题：

1）解决内部数据源（网络流量、安全设备日志、系统日志、用户行为信息等）和外部数据源（漏洞信息、威胁情报信息等）的大规模数据的采集、预处理与存储问题；

2）解决流式数据的实时分析和大规模历史数据的离线分析问题，实现信息与网络安全态势智能洞悉，主动、弹性地应对新型复杂的威胁和未知多变的风险；

3）解决日志、网络流量、威胁情报、用户行为等多源异构数据的快速复杂关联分析与检索问题，实现多尺度、多维度、细粒度的安全事件深入分析与跟踪。

大数据安全分析技术帮助安全分析者及决策者获得全面掌握IT活动的新视角和基于数据驱动的决策支持。基于大数据的网络安全与情报分析是紧扣安全分析数据自身的特点和安全分析的目标，应用大数据分析的方法和技术，解决实际网络安全问题的技术。思科公司提出的OpenSOC将大数据安全分析技术应用到安全管理平台（security operations center，SOC）中，构建针对网络包和流的大数据安全分析框架，实现网络异常的实时检测[5]。云安全联盟（CSA）在《安全智能中的大数据分析》中以案例的方式阐述了大数据安全分析技术对网络安全分析的改变[6]：Zions Ban公司在Hadoop系统上使用Hive查询大幅提高传统SIEM工具海量数据检索时间，从原来的20 min～1 h提高到1 min；HP实验室应用大规模图推理方法，基于从大型企业收集的20亿HTTP请求数据集、从ISP收集的10亿DNS请求数据集和从世界范围超过900家企业收集的350亿网络入侵检测系统告警数据集等，识别企业网络中被恶意软件感染的主机、访问的恶意域名，对在ISP收集的数十亿的DNS请求和响应数据构成的TB级DNS事件进行分析，识别僵尸网络及网络中的恶意活动；François，J开展的BotCloud研究项目利用MapReduce分析涉及1 600万主机的7.2亿Netflow数据，建立主机关联关系，跟踪僵尸网络中的命令–控制（C&C）通道，识别僵尸网络中的感染主机；美国RSA实验室应用大数据分析技术，基于对攻击者行为模式的分析，实现APT攻击检测。

2 网络安全大数据分析关键技术

2.1 大数据处理技术

大数据的计算模式可以分为批量计算（batch computing）和流式计算（stream computing）两种形态[7]。伯克利大学AMP实验室提出了数据分析的软件栈（Berkeley Data Analytics Stack，BDAS），从大数据计算模式的角度，将大数据处理技术分为3种类型[8]：批量数据处理技术、流式数据处理技术、交互式（interaction）数据查询技术。

2.1.1 批量数据处理技术

批量计算先进行数据的存储，再对存储的静态数据进行集中计算，如图1所示。批量数据处理技术实现了大规模静态数据的高吞吐处理，并以其吞吐量大为显著特征。复杂的批量数据处理通常的时间跨度为数十分钟到数小时。由于批量数据处理技术在应对大量持久数据时表现极为出色，因此在网络安全领域常被用于网络全流量分析、日志分析等历史数据分析中，也用于欺诈检测、APT检测等。

图 1 批量数据处理示意图

Fig.1 Batch data processing

Hadoop和Apache Spark是典型的可用于大数据批量处理的架构。Hadoop由HDFS分布式文件系统负责静态数据的存储，并通过MapReduce将计算逻辑分配到各数据节点进行数据计算和价值发现。赛门铁克数据共享计划——世界情报网络环境（WINE）项目平台使用MapReduce对收集的550万恶意软件样本数据、30 TB基于声誉的安全数据集、10万垃圾邮件样本数据、以及来自世界7 500万个传感器的安全威胁遥测数据集等大规模安全数据集进行高效处理，实现基于严格实验方法的安全基线过程[9]。也有许多学者采用MapReduce对网络流量数据进行处理，实现对僵尸网络的检测[10−12]。Apache Spark是运行在分布式计算集群上的大规模数据处理的快速和通用引擎，是一个新兴的大数据处理引擎。Spark系统提供一个集群的分布式内存抽象，实现内存计算机制，在批量数据处理上表现相当出色；在Spark官方网站上给出logistic回归运算实验证明，基于内存运算的Spark速度比Hadoop MapReduce快100倍，有取代MapReduce的趋势[13]。

2.1.2 流式数据处理技术

流式计算是实时产生、实时计算，其结果反馈往往也具有及时性的一种数据处理方法，如图2所示。流式数据处理技术可将到来的流式数据在内存中直接进行实时计算，数据处理延迟短、实时性强。流式数据处理技术很适用于处理必须对变动或峰值做出及时响应并且关注一段时间内变化趋势的数据分析场景，其数据处理的时间跨度通常为数百毫秒到数秒。

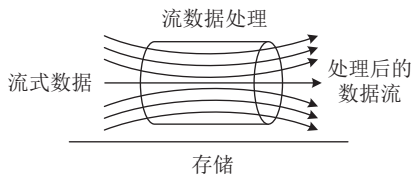

图 2 流式数据处理示意图

Fig.2 Streaming data processing

典型的流式数据计算架构有Twitter的Storm、Apache Spark中的Spark Streaming。Storm是一种侧重于极低延迟的流处理开源框架。数据处理的单位是单条，默认情况下提供“至少一次（at least once）”的处理保证，但这意味着某些情况下如果遇到失败可能会处理多次，且Storm无法确保按照特定顺序处理消息。Storm很适用于实时处理高性能网络流量，如网络会话流还原或流汇聚[14]。Spark Streaming是大规模流式数据处理的新贵，它是Spark核心API的一个扩展，先将流式计算分解为一系列短小的批处理作业，其核心机制是接收实时流的数据，并根据一定的时间间隔拆分成一批批的数据，然后通过Spark Engine处理这些批数据，最终得到处理后的一批批结果数据[15]。与Storm相比，Spark Streaming有一定的延时，适用于对实时性要求稍低的数据分析。

流式数据处理的实时性优势与批量数据处理应对大量持久数据优势具有明显的互补特征，可以满足多种应用场景下不同阶段的数据计算要求。

2.1.3 交互式数据查询技术

网络安全与情报分析强调以人作为安全分析主体和需求主体，将人的认知能力应用到安全分析过程中[15]。例如，网络安全人员在分析已发生的网络安全问题时，通常先分析网络整体运行状态定位异常时刻，再根据异常时刻聚焦网络主体，观察主体特定特征向量的时序特征。针对复杂的网络安全事件，往往还需要根据发现的异常主体，寻找和跟踪时序变化相似的主体等[16]。大数据交互式查询技术基于HBase、Hive、MongoDB等NoSQL类型的数据存储，构建相应的数据索引，使得基于历史数据的交互式查询通常的时间跨度为数十秒到数分钟，能够支持PB级日志数据的秒级检索，是实现网络安全与情报交互式分析的关键技术之一。与非交互式数据处理相比，交互式数据处理灵活、直观、便于控制。

交互式查询系统的典型代表系统有Apache Spark系统和Google的Dremel系统。Spark的内存计算机制使其天生具有对数据的快速交互式查询处理能力。Spark提供强大的交互式分析引擎Spark Shell和交互查询引擎Spark SQL，可以直接对其弹性分布式数据集进行操作，并快速返回结果。目前，已有许多学者利用Spark构建其大数据安全分析框架[17−18]。Dremel[19]是一个可扩展的交互式即时查询系统，通过结合多树状执行过程和列状数据结构，能做到几秒内完成对万亿张表的聚合查询，它是对MapReduce交互查询能力不足的补充。

2.2 大数据安全分析技术

大数据安全分析是大数据环境下网络安全与情

报分析的核心。其中，安全可视分析、安全事件关联分析、用户行为分析等是大数据安全分析技术的热点。

2.2.1　安全可视分析技术

数据可视化可以帮助分析者一眼洞悉数据背后隐藏的信息并转化为知识及智慧[20]，因此，许多学者提出将网络安全数据以可交互的图形图像的方式表现出来，借助人的视觉处理能力观察网络安全数据中隐含的信息以帮助网络安全分析人员感知和理解网络安全问题，并逐渐形成了网络安全可视化这一新兴的交叉研究领域[18,21]。早在1995年Bechker等[22]就提出对网络流量状况进行可视化。之后，网络安全可视化领域的学者提出了许多新颖的可视化设计，实现了交互式可视分析工具NVisionIP[23]、Visflow-Connect-IP[24]、NVisionCC[25]，用于网络流量的异常监测和入侵检测等。

网络安全可视化的一般步骤是[26]：首先，确定网络安全分析人员关心的问题，也就是有什么数据，需要从数据中获得什么信息；然后，设计可视化结构来表示数据，建立数据到可视化结构的映射；最后，设计缩放、聚焦、回放、关联和更新等人机交互功能，实现人与可视化工具的交流。

网络安全可视化采用的主要图形有节点连接图、网格图、矩阵图、点阵图、柱状图、队列图、平行坐标图（parallel coordinate plot）、散点图、热力图、力导向图（force-directed graph）、地图等[26]，分析过程中常用到统计分析方法、多视图协同分析技术、大规模网络处理与大规模图形数据处理等。与传统的网络安全可视化不同，大数据安全可视化面临2个问题：一方面是，网络安全数据的规模，即如何提出新的可视方法以帮助安全分析人员分析大规模、高纬度、多来源、动态演化的网络安全数据并实时做出决策。另一方面是，创造符合网络安全分析人员心理映像的大数据可视化表征，能够让安全分析人员一眼发现大数据中隐含的安全问题。近几年，已有学者开始逐步深入地探讨如何快速处理大规模流量时序数据，以及如何可视化大规模网络监控对象的流量变化。例如，Fischer等[27]设计了一个基于Web的视觉分析应用程序NVisAware，对网络数据流进行监视和可视分析，并设计了基于Spark的网络安全态势可视化工具NStreamAware。

2.2.2　安全事件关联分析技术

随着网络规模的日益扩大，网络安全事件呈指数级增长，安全事件相互之间存在错综复杂的关系。例如，有些安全事件由同一个攻击行为产生，有些存在因果关系，还有的安全事件是由一系列的攻击行为组成的复杂攻击。网络安全事件关联分析技术需要将各种复杂的网络安全事件进行充分关联，找出它们之间的关系，去掉冗余后给出完整的事件描述，以及时发现网络攻击者的入侵行为。大数据分析技术在海量数据关联分析上具有明显优势，被广泛应用于海量网络安全数据的深度关联分析与基于历史数据的宽时间周期内多类型安全事件智能关联分析和复杂事件处理（complex event processing，CEP）[28]。

按照关联对象的不同，可将基于大数据的安全事件关联分析方法分为4类[29]：

1）安全设备报警关联分析。该类方法针对海量且不断产生的主机日志、防火墙日志、入侵告警等安全告警数据，应用大数据处理技术，过滤与系统无关的虚假安全事件和冗余安全事件，通过事件之间存在的相似关系、因果关系等对事件进行聚合处理，获得更精简准确的安全报警。例如，通过报警记录之间的属性（源IP、目的IP、源端口、目的端口、协议类型、时间等）相似性度量，对安全事件进行分类合并，实现报警信息的精简[30]。

2）网络和主机关联分析。该类方法提取表征网络流量和主机异常的特征，通过共同属性特征的综合关联，实现对网络安全的监测。

3）不同领域安全事件关联分析。该类方法综合利用来自不同领域的各类安全事件间的内在联系，对安全事件进行关联分析，实现网络攻击检测。例如，利用网络拓扑结构与不同设备报文TTL之间的关联关系过滤未达到攻击目标的虚假告警[31]；利用主动扫描工具获得的主机脆弱信息、主机配置信息与安全告警之间漏洞相关性关系以过滤与目标主机系统无关的告警；利用外部威胁情报、网络主机IP信息、告警信息之间IP相关性关系识别超级告警事件。

4）攻击步骤关联分析。该类方法根据多步攻击等先验知识，使用攻击图、攻击树或攻击序列的方式描述已知攻击事件的因果关系、时序关系等，将事件的关联分析转化为图模式匹配、子树匹配或字符串序列匹配等，实现网络攻击检测、网络态势评估与预测。例如，根据先验知识构建包含攻击事件、攻击事件发生的前提条件、攻击事件造成的影响的三元组，通过匹配攻击事件的前因和后果，分析2个攻击事件之间是否存在因果关系，并实现关联操作；利用描述网络主机的连接关系、脆弱信息，以及攻击规则库、攻击者属性等之间的关联关系，生成以主机为节点的攻击图，用于网络安全分析[32]。

2.2.3　用户行为分析技术

在企业内外网无法完全分开的情况下，企业的IT管理人员发现，即便用最先进的安全产品防止了黑客的攻击，却无法根治“内鬼”。用户行为分析（user behavior analytics，UBA）成为IT安全行业解决

该问题的新技术，用于发掘“不知道的未知情况”。用户在使用网络应用与服务时，会在系统中留下痕迹，其行为出现在网络流量、日志记录、审计跟踪记录等处。UBA技术通过对用户上述信息的收集，并根据信息中用户留下的数字痕迹，建立一条用户行为基准线（例如，用户活跃时间、使用服务类型、使用服务的频率等），描述用户的“正常行为”。UBA技术可用于反数据窃取和反诈骗中，以帮助组织检测内部威胁、有针对的攻击和金融诈骗。为了更准确地识别威胁，终端、应用、网络和外部威胁等除用户外的其他实体也被关注。应用大数据关联分析技术，将这些实体的行为与用户行为进行关联分析，用于保护组织免受来自内部与外部的威胁。Gartner将用户与实体行为分析（UBEA）技术作为2016年十大信息安全技术之一[33]、2017年十大战略技术趋势之一[34]。

网络行为是用户行为在网络流量上的体现。网络行为分析（network behavior analysis，NBA）是一种通过监测网络流量，关注网络流量异常和偏离正常操作的行为，以增强网络安全性的方法。NBA是未知网络攻击检测的一把“利剑”，通常基于NetFlow/IP-FIX中的源IP地址、目的IP地址、源端口、目的端口、包数量、流字节数等属性构成的特征向量刻画网络用户行为，实现对网络的分析和持续自动评估，检测网络攻击、网络异常、高级威胁和不良行为[35–42]。

从整体流程上来看，用户行为分析包括确定需求，数据采集，数据预处理（集成、清洗、转换），应用相关方法进行模式挖掘，挖掘结果评估，挖掘结果分析与应用等步骤，其中，数据预处理、方法的选取、方法的评估是用户行为分析最重要的3个步骤。用户行为分析过程中通常使用统计分析、聚类分析、关联规则分析、时序数据挖掘分析等大数据分析技术。

3 大数据技术在网络安全与情报分析中的应用

大数据技术为网络安全与情报分析注入新的技术源动力，提高网络信息安全攻击检测、风险感知、情报分析能力，形成“数据驱动安全”的网络安全与情报分析新思维。

3.1 APT攻击检测

在全球网络信息化程度高速发展的大背景下，具有隐蔽性、渗透性和针对性的APT攻击日益增多，使国家、企业的网络信息系统和数据安全面临严峻挑战。例如，2010年发现的震网病毒（Stuxnet）以关键工业基础设施为目标，延缓了伊朗核项目建设长达2年；2012年5月发现的超级恶意软件火焰病毒（Flame）利用Windows操作系统的漏洞入侵个人电脑，获取中东多国的大量机密信息；2016年黑暗力量（Black Energy）不仅入侵了乌克兰的电力系统，还攻击了其矿业和铁路系统。APT攻击隐蔽性强，其攻击空间路径和攻击渠道不确定，大多数传统的安全解决方案无法抵御这种新型攻击[43]。扩大时间和空间范围进行数据关联分析是检测APT攻击的最有效途径之一[31]。由于基于网络大数据分析的安全检测技术可以实现海量网络安全数据的深度关联分析，也可对宽时间周期内的多类型安全数据智能关联，因此在检测APT攻击方面具有明显优势。

美国RSA实验室提出的Beehive系统[44]通过大数据分析技术在短时间内处理大量日志信息，检测组织机构中的资源使用模式，发现以往会被忽视的策略违背与恶意软件感染，并根据APT攻击多个阶段的行为与正常通信存在细小的行为差异，关联检测到的看似孤立的事件，发现攻击者APT入侵的证据。2012年，Giura等在《Science》发表研究成果提出了一种可成功检测APT攻击的方法[45]。该方法基于攻击树的概念，建立了一个概念攻击模型——攻击金字塔；攻击金字塔在顶层包含可能的攻击目标（例如敏感数据、高层职员、数据服务器），采用横向平面表示与攻击相关联的事件环境（例如用户平面、网络平面、应用平面或物理平面）；提出的检测框架先将所有记录在组织中的可能与安全相关的事件分组为多个场景（contexts），再在每个场景及跨场景中使用MapReduce并行处理，应用不同的算法检测可能的恶意活动。

近几年，有许多学者研究基于大数据分析方法的APT攻击检测[46–48]。作者团队开展了基于大数据的网络行为分析方法研究，关注对隐蔽性、持续性可疑通信行为的检测，探索APT等高级网络攻击早期检测方法，目前已在实际校园网上检测到一些可疑的隐蔽性持续性加密通信。

3.2 网络异常检测

网络异常检测一直是网络安全领域内最为活跃的研究分支之一，包括对流量突变、设备失效、越权资源访问、可疑主机等的检测，其本质原理是探寻表征目标对象属性、状态与变化的特征，然后构建检测模型，对违背策略或偏离正常行为模式的行为进行判定。近几年，大数据技术越来越多的应用到网络异常检测中，尤其是基于大数据的网络用户行为分析技术的应用，极大提高了当前网络异常检测的准确率[37–38,41,49]。

基于行为特征和机器学习的方法实现了网络异常分析建模和异常检测过程的自动化，应用基于深度学习的大数据分析技术能够实现数据特征的快速自动抽取，解决建模过程中对专家知识的依赖。2015年，黑帽（Black Hat）大会上360公司的王占一做了题为《深度学习在流量识别中的应用》的主题演

讲，将深度学习应用到网络流量协议分类和未知协议检测中，对常见应用协议识别的准确率达到90%以上，对未知协议的识别达到80%，且识别过程在不区分协议是否加密时对整个网络流的可识别率达到54.94%[50]。商业化网络空间安全解决方案提供公司Deep Instinct于2015年11月在美国旧金山成立，宣称其应用深度学习技术的安全解决方案能够抵御未知攻击，能够及时检测0day漏洞的威胁和APT攻击[51]。

作者团队开展了基于四川大学校园网流量、日志、上网认证等信息的网络安全大数据分析研究，应用大数据可视分析、关联分析、交互式分析技术，实现对校园网网络的异常检测。利用平行坐标图实现对网络流量应用分类的交互式分析（如图3所示），及时、全面地掌握不同的区域（学生区、行政区、教学区等）在给定的时间段内的流量分类情况，定位各类网络流量的最大贡献区域，及时发现异常或非法流量（例如赌博、色情等违规流量）。利用关联分析技术和基于力导向图的可视化呈现，实现对邮件系统的协同攻击检测（如图4所示），根据对邮箱账号登录信息的统计与过滤，在被探测的邮箱账号、新出现的攻击源、持续出现的攻击源、攻击源网段之间绘制连接线，帮助网络管理者快速获得针对邮箱账户的协同攻击情况。为解决数据量大导致可视化速度慢的问题，对数据做变换与降维处理，并通过调整图形的节点斥力和引力参数，提高图形绘制算法收敛速度，以满足具有大规模数据的邮箱账户协同探测攻击状态的可视化要求。

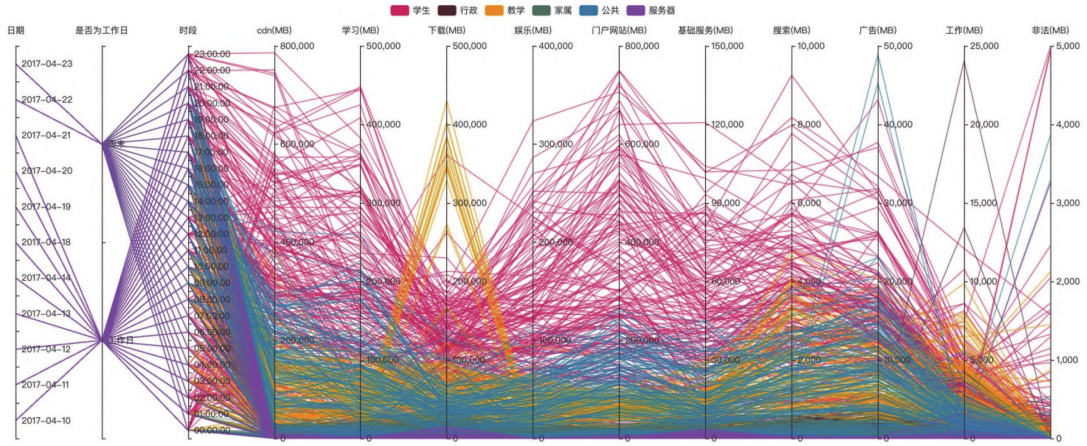

图 3　区域流量应用分类交互式平行坐标图

Fig.3　Regional flow classification interactive parallel coordinates graph

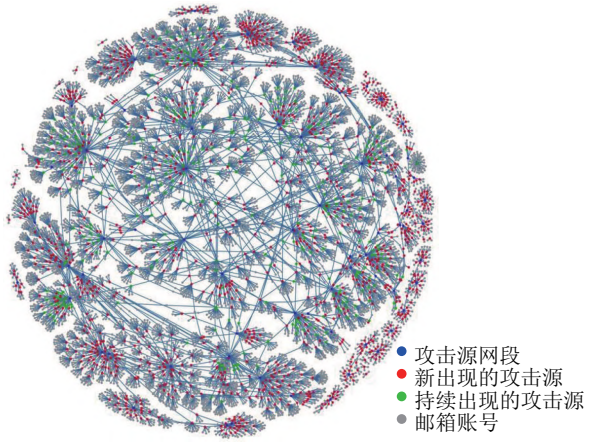

图 4　邮箱账户协同探测攻击检测力导向图

Fig.4　Email accounts collaborative attack detection force-directed graph

3.3　网络安全态势感知

面对不断增加的多层面网络安全威胁和安全风险，企业和组织需要及时发现网络中的异常事件，实时掌握网络安全状态，由过去的“亡羊补牢”转向事前自动评估，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（network security situational awareness，NSSA）技术能够综合各方面的安全因素，通过安全要素的获取、理解、评估与可视，从总体上反映网络安全状况，并对网络安全的发展趋势进行预测和预警[52]。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。NSSA已经成为当前网络安全领域的研究热点。

网络安全领域许多企业纷纷提出安全大数据态势感知方案或构建安全大数据态势感知预警平台。阿里云云盾基于其云平台，以软件即服务（SAAS）的方式提供网络态势感知服务，旨在为客户开启“上帝视角”，让用户“看见”威胁；360公司推出的态势感知及安全运营平台NGSOC，利用大数据技术，对本地全量数据进行采集和存储，以情报为驱动，实现对网络的持续监控、安全分析与威胁溯源。学术界在基于大数据的网络安全态势感知、态势评估、态势预测等方面开展了大量研究[53–56]。作者团队研发了校园网络业务及安全态势分析平台NTCI.NUBA（如图5所

示），对四川大学校园网出口网络流量、数据中心流量、安全设备日志、身份认证数据等多源数据实时采集，基于Hadoop和Spark构建数据的存储与分析架构，实现校园网整体网络安全态势感知、网络流量业务安全态势感知、服务器业务安全态势感知和网络安全事件实时监测等。

图 5 网络安全态势实时感知

Fig.5 NTCI.NUBA real-time network security situation awareness

3.4 网络威胁情报分析

威胁情报是通过大数据、分布式系统或其他特定收集方式获取，包括漏洞、威胁、特征、行为等一系列证据的知识集合和可操作性建议。威胁情报为传统防御方式带来了有效补充，立足于攻击者的视角，依靠其广泛的可见性以及对整个互联网风险及威胁的全面理解，帮助用户更好地了解威胁，使用户能够准确、高效的采取行动，避免或减少网络攻击带来的损失[57]。完整的安全威胁分析体系由情报源、融合与分析、事件响应3个环节组成[58]。

当前国内外主要的网络安全威胁情报提供商有FireEye、iSightParnters、CrowdStrike、Symantec、微步在线等20多家，其威胁情报产品与服务从反恐怖主义、防范网络犯罪、漏洞检测、恶意软件清理等多个角度满足不同用户的特定需求。作者团队研发了网络空间信息安全数据采集与综合分析平台。平台依据不同的主题需求，从网站、论坛、博客、微博、微信、自媒体APP、贴吧等采集特定主题数据，并对采集的数据进行热点话题发现与跟踪、情感分析、自动摘要、溯源和扩散等分析，实现情报的采集与分析综合服务。

4 网络安全与情报分析的研究趋势

当前的网络安全形势仍不容乐观，缺乏高级网络威胁与攻击的有效检测方法，在复杂网络攻击与威胁预测、大规模网络安全态势感知，以及威胁情报信息的收集、共享与分析处理方面还存在大量尚待解决的问题，如何解决这些问题将是未来几年网络安全与情报分析的研究趋势。

4.1 高级网络威胁发现方法研究

4.1.1 高级网络威胁的早期检测方法研究

新型木马、僵尸网络、APT等攻击是经过精心策划的高级隐蔽性攻击，且由多个阶段的攻击组成，具有较高的检测难度。尽早发现网络中的高级网络威胁，减少攻击损失，具有重要的实际意义。在高级网络威胁的早期，攻击者常采用C&C服务方式[59]或是DNS隐蔽信道方式进行通信[47]，大部分C&C服务器常采用各种DGA、Fast Flux、Double Flux等方法隐蔽服务器地址。文献[57,59]指出，通过对DGA域名的检测，可以发现C&C服务器，从而实现对APT、僵尸网络等高级网络攻击的早期检测。目前，DGA域名检测方法主要有基于DNS流量特征[57,60]和基于域名文本特征分析的检测[61]等，但检测准确率还有待提高。另外，如何利用已发现的DGA域名，并分析其与请求域名IP集、域名解析IP集的关系，进一步发现控制网络是尚待研究解决的问题。

4.1.2 隐蔽性、持续性异常通信行为的检测方法研究

隐蔽性、持续性的高级网络攻击在潜伏期间通常采用对抗性隐蔽技术，将网络通信数据隐匿于大量正常流量、通用协议以及合法服务中。因此，如何针对隐蔽性、持续性网络威胁特点，设计出能够区分隐蔽性、持续性网络通信行为与正常通信行为的特征抽象与行为描述方法，在大量网络数据流中有效识别出隐蔽性、持续性异常通信行为，是当前需要解决的问题。

4.1.3　多维度高级网络威胁协同检测方法研究

高级网络威胁固有的复杂、低频特性，导致在检测模型构建阶段无法获得足够的分析和训练样本，综合威胁情报信息等多维度信息有助于快速判断检测结果集的准确性，提高攻击检测效率和检测精度。然而每个维度的信息结构、含义都不一样，需要解决多源异构数据关联、融合问题，并在此基础上研究基于多维度信息的高级网络威胁协同检测方法。

4.2　复杂网络攻击预测研究

复杂网络攻击是指由多个单独攻击组合而成的攻击过程，它具有依赖关系、选择关系、并列关系。预测未来的攻击对网络管理者来说是个巨大的挑战。目前，主要基于攻击图、攻击树等方式对网络安全风险进行评估及网络攻击的预测[49,62]。但是，这类方法基于已经暴露的脆弱性等静态信息，而且攻击图等方式无法考虑各种不确定性因素，对未来攻击的预测可靠性和准确性不高。面对当前日益复杂化的网络攻击，需要研究可靠性和准确性更高的方法对复杂网络攻击进行预测。在大数据环境下，来自不同领域的大量网络安全信息的获取与存储成为可能，研究基于入侵检测告警、活动响应、行为模式、威胁情报等多源数据的关联分析方法、用户行为分析方法等，实现对复杂网络攻击的预测是网络安全分析的一个研究热点。

4.3　网络安全威胁态势感知技术研究

2016年4月19日，习近平总书记在中国网络安全和信息化工作座谈会上发表重要讲话，在论述安全和发展的关系时，特别强调了NSSA的重要性、工作任务和工作目标[63]。并指出“维护网络安全，首先要知道风险在哪里，是什么样的风险，什么时候发生风险”，“感知网络安全态势是最基本最基础的工作”。

4.3.1　NSSA评价体系研究

态势指状态和形势，是一个比较抽象的概念。而什么样的态势为好，好到什么程度，往往只是一种感觉。无论打分还是分级都缺少科学依据，客观性、说服力不强。其根本原因在于没有明确、形式化的态势定义，且缺乏度量评估结果优劣的指标与方法，无法形成对态势评估的共识。因此，开展NSSA评价体系研究，使NSSA评估研究具有明确的目标与方向是当前具有重要意义的研究内容。

4.3.2　NSSA指标体系研究

在大数据环境下，需要关联更多的数据源对网络系统的安全状态进行表示。已有的NSSA指标体系基于使用层次结构表示网络系统，无法展现网络元素之间错综复杂的关系，不利于挖掘多源多属性数据内部潜在的态势信息，需要选择并且扩展用于态势感知的特征测度，建立合理完善的面向关键资产与面向网络整体的NSSA指标体系。

4.3.2　NSSA评估方法研究

目前，NSSA评估的方法多种多样，数据融合领域几乎所有的理论方法都被应用到网络安全态势评估中，存在大量重复研究现象，甚至为了增加理论深度而使用各种数学方法。需要针对NSSA评估对象，选择合适的评估方法，并结合具体问题，有针对性地进行改进和优化，提高评估的准确性和效率。

4.4　威胁情报获取、感知与共享问题研究

威胁情报驱动的信息安全防御已成为业界公认的信息安全发展方向，备受学术界和产业界关注[64]。如何快速准确地获取和分析威胁信息，从中挖掘有价值的情报，并实现面向领域、面向用户的情报感知是当前威胁情报分析领域面临的重要问题与关键任务。

4.4.1　威胁情报的获取

在大数据时代，威胁情报的获取来源、媒体形态、内容形式等得到极大的丰富，需要从多个维度进行信息的搜集和整合，从而在全景视角的高度上实现有效的威胁情报感知。云计算、虚拟化技术的出现极大地扩展了网络的规模，以软件定义网络（SDN）为代表的新型通信网络架构增加了网络部署与监控的灵活性和可扩展性。新型网络结构下威胁情报的来源与采集方法、情报数据的采集频率、传递方法与传输路径、情报数据的汇集方法等都是当前需要解决的问题。

除了基于相关基础设施和网络流量等获取威胁情报之外，QQ、微博、微信、论坛、Facebook、Twitter等网络社交媒体的兴起，使威胁情报的搜集边界朝着面向用户和服务的信息内容分析方向快速发展。从网络社交媒体中搜寻攻击者或某些核心人物的资料与言论等为情报分析工作提供了更广阔的数据源。这些可用的威胁情报源中，网络社交媒体等大部分为受限网络，无法使用常规网络爬虫技术快速获得有价值的情报信息，需要研究新型情报获取工具，突破受限网络对资源访问的限制。

4.4.2　威胁情报的感知

在威胁情报感知时，一是，需要对来自各个渠道的海量情报进行融合，不可避免地需要解决多源异构数据的关联、推理、融合等分析问题；二是，如何从用户需求出发，从海量情报信息中快速提取出与用户实际业务息息相关的威胁情报，是威胁情报分析时需要解决的问题。

4.4.3　威胁情报的共享

在整个防御过程中，单个或几个用户的威胁情报信息往往是不够的，只有实现“共享”，最大限度地将海量的安全威胁情报信息汇集起来，才能使威胁情报信息不断被收集、丰富、分析、再收集，从而形成一个有效的闭环，以最大化情报的内在价值。美国早

在2003年的《网络空间安全国家战略》中就提出了建立信息共享与分析中心，确保能够接收实时的网络威胁和漏洞数据。中国尚未建成规模化、技术完备的集数据截获、分析、转化、共享与利用为一体的现代化威胁情报中心。在威胁情报共享的标准和协议上，国外学者已抢先制定了OpenIOC和STIX等一系列标准与协议，而中国学者仍处于摸索和尝试阶段，无论在技术手段、标准制定，还是分析应用方面，都需要继续努力。

5 结 语

大数据技术已逐渐深入到许多网络空间安全问题的处理和解决方案中，改变了网络空间安全与情报分析的研究格局，提高了高级网络攻击检测、信息安全风险感知与威胁情报分析处理等网络安全防御技术水平。但是，在网络空间安全新形势下，在进攻和防御的激烈对抗中，还需要继续利用大数据技术，综合运用多源数据，探索复杂网络攻击遏制、感知网络信息安全风险感知、预警和处置、情报共享和研判的新技术，提升大数据对网络安全与情报分析的支撑能力。

参考文献：

[1] Gartner.Information security is becoming a big data analytics Problem[EB/OL].[2012-03-23].https://www.gartner.com/doc/1960615/information-security-big-data-analytics.

[2] 大数据标准化白皮书(2016年)[R].北京:全国信息技术标准化技术委员会大数据标准工作组中国电子技术标准化研究院,2016.

[3] NIST Big Data Public Working Group .Draft NIST big data interoperability framework:Volume 1,definitions[EB/OL].[2017-03-20].http://bigdatawg.nist.gov/home.php.

[4] 大数据白皮书(2016年)[R].北京:中国信息通信研究院(工业和信息化部电信研究院),2016.

[5] Cisco.OpenSOC: Big data security analytics framework [EB/OL].[2017-03-20].http://opensoc.github.io/.

[6] Cloud Security Alliance.Big data analytics for security intelligence[EB/OL].[2017-03-20].https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Big_Data_Analytics_for_Security_Intelligence.pdf.2013.

[7] Sun Dawei,Zhang Guangyan,Zheng Weimin.Big data stream computing: Technologies and instance[J].Journal of Software,2014,25(4):839–862.[孙大为,张广艳,郑纬民.大数据流式计算:关键技术及系统实例[J].软件学报,2014,25(4):839–862.]

[8] Franklin M.The berkeley data analytics stack:Present and future[C]// Proceedings of the 2013 IEEE International Conference on Big Data.Santa Clara:IEEE,2013:2–3.

[9] Dumitras T,Shou D.Toward a standard benchmark for computer security research:The worldwide intelligence network environment (WINE)[C]//Proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security.New York:ACM,2011:89–96.

[10] Francois J,Wang S,Bronzi W,et al.Botcloud:Detecting botnets using mapreduce[C]//Proceedings of the 2011 IEEE International Workshop on Information Forensics and Security (WIFS).New York:IEEE,2011:1–6.

[11] Singh K,Guntuku S C,Thakur A,et al.Big data analytics framework for peer-to-peer botnet detection using random forests[J].Information Sciences,2014,278:488–497.

[12] Kaushik G,Patil S,Chawla T.Botnet detection techniques with data mining using MapReduce[J].Journal of Basic and Applied Engineering Research,2014,2(10):869–873.

[13] Apache Spark™ -lightning-fast cluster computing[EB/OL].[2017-03-20].http://spark.apache.org/.

[14] Ma Ke,Li Lingjuan.Distributed real time stream data clustering algorithm and its implementation based on storm[J].Journal of Nanjing University of Posts and Telecommunications(Natural Science Edition),2016(2):104–110.[马可,李玲娟.分布式实时流数据聚类算法及其基于Storm的实现[J].南京邮电大学学报(自然科学版),2016(2):104–110.]

[15] Zhao Ying,Wang Quan,Huang Yezi,et al.Collaborative visual analytics for network traffic time-series data with multiple views[J].Journal of Software,2016,27(5):1188–1198.[赵颖,王权,黄叶子,等.多视图合作的网络流量时序数据可视分析[J].软件学报,2016,27(5):1188–1198.]

[16] Cheng Xueqi,Jin Xiaolong,Wang Yuanzhuo,et al.Survey on big data system and analytic technology[J].Journal of Software,2014,25(9):1889–1908.[程学旗,靳小龙,王元卓,等.大数据系统和分析技术综述[J].软件学报,2014,25(9):1889–1908.]

[17] Zhao Kejun,Ge Liansheng,Liu Yang,et al.Sclable security analysis platform based on Hadoop and Spark[J].Journal of Huazhong University of Science & Technology (Natural Science Edition),2016(Sup.1):25–28.[赵科军,葛连升,刘洋,等.基于Hadoop和Spark构建可扩展的网络安全分析平台[J].华中科技大学学报(自然科学版),2016 (增刊1):25–28.]

[18] Marchal S,Jiang X,State R,et al.A big data architecture for large scale security monitoring[C]//Proceedings of the 2014 IEEE International Conference on Big Data.Anchorage:IEEE,2014:56 – 63.

[19] Melnik S,Gubarev A,Long J J,Romer G,Shivakumar S,Tolton M,Vassilakis T.Dremel: Interactive analysis of web-scale datasets[J].Proceedings of the VLDB Endowment,2010,3(1/2):330–339.

[20] Ren Lei,Du Yi,Ma Shuai,et al.Visual analytics to-wards big data[J].Journal of Software,2014,25(9):1909–1936.[任磊,杜一,马帅,等.大数据可视分析综述[J].软件学报,2014,25(9):1909–1936.]

[21] 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.

[22] Becker R A,Eick S G,Wilks A R.Visualizing network data[J].IEEE Transactions on Visualization and Computer Graphics,1995,1(1):16–28.

[23] Lakkaraju K,Yurcik W,Lee A J.NVisionIP:Netflow visualizations of system state for security situational awareness[C]//

Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security.New York:ACM,2004: 65 – 72.

[24] Yin X,Yurcik W,Slagell A.VisFlowConnect-IP:An animated link analysis tool for visualizing netflows[C]//Proceedings of the 2005 FLOCON-Network Flow Analysis Workshop (Network Flow Analysis for Security Situational Awareness).Pittsburgh PA:Carnegie Mellon University, 2005:1–4.

[25] Yurcik W,Meng X,Kiyanclar N.NVisionCC:A visualization framework for high performance cluster security[C]// Proceedings of the 2004 ACM workshop on Visualization and Data Mining for Computer Security.NewYork:ACM,2004: 133 – 137.

[26] Zhao Ying,Fan Xiaoping,Zhou Fangfang,et al.A Survey on Network Security Data Visualization[J].Journal of Computer-Aided Design&Computer Graphics,2014,26(5): 687–697.[赵颖,樊晓平,周芳芳,等.网络安全数据可视化综述[J].计算机辅助设计与图形学学报,2014,26(5):687–697.]

[27] Fischer F,Keim D A.NStreamAware:Real-time visual analytics for data streams to enhance situational awareness[C]// Proceedings of the Eleventh Workshop on Visualization for Cyber Security.New York:ACM,2014:65 – 72.

[28] Cerullo G,Coppolino L,D'Antonio S,et al.Enabling convergence of physical and logical security through intelligent event correlation[M]//Intelligent Distributed Computing Ⅸ.Berlin:Springer,2016:427 – 437.

[29] Fu Yu,Li Hongcheng,Wu Xiaoping,et al.Detecting APT attacks: A survey from the perspective of big data analysis[J]. Journal on Communications,2015,36(11):1–14.[付钰,李洪成,吴晓平,等.基于大数据分析的APT攻击检测研究综述[J].通信学报,2015,36(11):1–14.]

[30] Zhang Shuying.Network security event correlation analysis and situation assessment prediction technology research[D]. Changchun:Jilin University,2012.[张淑英.网络安全事件关联分析与态势评测技术研究[D].长春:吉林大学,2012.]

[31] Tian Zhihong,Wang Bailing,Zhang Weizhe,et al.Network intrusion detection model based on context verfication[J]. Journal of Computer Research and Development,2013,50(3): 498–508.[田志宏,王佰玲,张伟哲,等.基于上下文验证的网络入侵检测模型[J].计算机研究与发展,2013,50(3):498–508.]

[32] Zhong Shangqing.Research on network security based on host-based attack graph[D].Beijing:Journal of Beijing University of Posts and Telecommunications,2012.[钟尚勤.基于主机攻击图的网络安全性研究[D].北京:北京邮电大学,2012.]

[33] Gartner.Gartner identifies the Top 10 technologies for information security in 2016[EB/OL].[2017-03-20].http://www.gartner.com/newsroom/id/3347717.2016.

[34] Gartner.Gartner's Top 10 strategic technology trends for 2017[EB/OL].[2017-03-20].http://www.gartner.com/smarterwithgartner/gartners-top-10-technology-trends-2017/.

[35] Zheng Liming.Key technologies research on traffic anomaly detection and optimization for large-scale networks[D]. Heifei:National University of Defense Technology,2012.[郑黎明.大规模通信网络流量异常检测与优化关键技术研究[D].合肥:国防科学技术大学,2012.]

[36] Zhou Tao.Abnormal network behavior detection technology based on statistical learning[J].Big Data Research,2015, 1(4):38–47.[周涛.基于统计学习的网络异常行为检测技术[J].大数据,2015,1(4):38–47.]

[37] Bhuyan M H,Bhattacharyya D K,Kalita J K.Network anomaly detection:Methods,systems and tools[J].IEEE Communications Surveys & Tutorials,2014,16(1):303–336.

[38] Hu Yangrui,Chen Xingshu,Wang Junfeng,et al.Anomalous traffic detection based on traffic behavior characteristics[J]. Netinfo Security,2016(11):45–51.[胡洋瑞,陈兴蜀,王俊峰,等.基于流量行为特征的异常流量检测[J].信息网络安全,2016(11):45–51.]

[39] Shao Guolin,Chen Xingshu,Yin Xueyuan,et al.Profiling structure-stability-based server traffic: Behavior models and system[J].Journal of University of Electronic Science and Technology of China,2017,46(1):102–108.[邵国林,陈兴蜀,尹学渊,叶晓鸣.基于流量结构稳定性的服务器网络行为描述:建模与系统[J].电子科技大学学报,2017,46(1): 102–108.]

[40] Yan Hao.Network user behavior analysis base on traffic monitoring and measurement[D].Beijing:Beijing University of Post and Telecommunications,2011.[延皓.基于流量监测的网络用户行为分析[D].北京:北京邮电大学,2011.]

[41] Li Qiao,He Hui,Fang Binxin,et al.Awareness of the network group anomalous behaviors based on network trust[J]. Chinese Journal of Computers,2014,37(1):1–14.[李乔,何慧,方滨兴,等.基于信任的网络群体异常行为发现[J].计算机学报,2014,37(1):1–14.]

[42] Liu J,Liu F,Ansari N.Monitoring and analyzing big traffic data of a large-scale cellular network with Hadoop[J].IEEE Network,2014,28(4):32–39.

[43] 周涛.大数据与APT攻击检测[J].信息安全与通信保密,2012(7):23.

[44] Yen T F,Oprea A,Onarlioglu K,et al.Beehive:Large-scale log analysis for detecting suspicious activity in enterprise networks[C]//Proceedings of the 29th Annual Computer Security Applications Conference.New York:ACM,2013:199 – 208.

[45] Giura P,Wang W.Using large scale distributed computing to unveil advanced persistent threats[J].Science,2012,1(3): 93–105.

[46] Marchetti M,Pierazzi F,Colajanni M,et al.Analysis of high volumes of network traffic for advanced persistent threat detection[J].Computer Networks,2016,109:127–141.

[47] Zhao G,Xu K,Xu L,et al.Detecting APT malware infections based on malicious DNS and traffic analysis[J].IEEE Access,2015,3:1132–1142.

[48] Zhang Xiaosong,Niu Weina,Yang Guowu,et al.Method for APT prediction based on tree structure[J].Journal of Uni-

versity of Electronic Science and Technology of China,2016,45(4):582–588.[张小松,牛伟纳,杨国武,等.基于树型结构的APT攻击预测方法[J].电子科技大学学报,2016,45(4):582–588.]
[49] Bhuyan M H,Bhattacharyya D K,Kalita J K.Network anomaly detection:Methods, systems and tools[J].IEEE Communications Surveys & Tutorials,2014,16(1):303–336.
[50] Wang Z.The applications of deep learning on traffic identication[EB/OL].[2017-03-20].https://www.blackhat.com/docs/us-15/materials/us-15-Wang-The-Applications-Of-Deep-Learning-On-Traffic-Identification-wp.pdf.
[51] Musthaler L.How to use deep learning AI to detect and prevent malware and APTs in real-time[EB/OL].[2017-03-20].http://www.networkworld.com/article/3043202/security/how-to-use-deep-learning-ai-to-detect-and-prevent-malware-and-apts-in-real-time.html.
[52] Zhang Jianfeng.Research on key technologies of net work security assessment[D].Heifei:School of National University of Defense Technology,2013.[张建锋.网络安全态势评估若干关键技术研究[D].合肥:国防科学技术大学,2013.]
[53] Han Weihong,Sui Pinbo,Jia Yan.Security situation analysis and prediction system for large-scale network SSAP[J].Netinfo Security,2012(8):11–14.[韩伟红,隋品波,贾焰.大规模网络安全态势分析与预测系统YHSAS[J].信息网络安全,2012(8):11–14.]
[54] Guan Lei,Hu Guangjun,Wang Zhuan.Research on network security situational awareness technology based on big data[J].Netinfo Security,2016(9):45–50.[管磊,胡光俊,王专.基于大数据的网络安全态势感知技术研究[J].信息网络安全,2016(9):45–50.]
[55] Zhao Meng.Network security situation awareness based on big data[J].Netinfo Security,2016(9):90–93.[赵梦.基于大数据环境的网络安全态势感知[J].信息网络安全,2016(9):90–93.]
[56] Wu J,Ota K,Dong M X,et al.Big data analysis based security situational awareness for smart grid[J].IEEE Transactions on Big Data,2016,PP(99):1.
[57] Li Juntao,Shi Yong,Xue Zhi.APT detection based OH DNS traffic and threat indigence[J].Information Security and Communications Privacy,2016(7):84–88.[李骏韬,施勇,薛质.基于DNS流量和威胁情报的APT检测[J].信息安全与通信保密,2016(7):84–88.]
[58] MacDonald N.Prevention is futile in 2020:Protect information via pervasive monitoring and collective intelligence[EB/OL].[2017-03-20].https://www.gartner.com/doc/2500416/prevention-futile-protect-information-pervasive.
[59] García S,Uhlíř V,Rehak M.Identifying and modeling botnet C&C behaviors[C]//Proceedings of the 1st International Workshop on Agents and CyberSecurity.New York:ACM,2014:1.
[60] Grill M,Nikolaev I,Valeros V,et al.Detecting DGA malware using NetFlow[C]//Proceedings of the 2015 IFIP/IEEE Symposium on Integrated Network and Service Management,Ottawa:IEEE,2015:1304 – 1309.
[61] Zhang Weiwei,Gong Jian,Liu Qian,et al.Light weight domain name detection algorithm based on morpheme features[J].Journal of Software,2016,27(9):2348–2364.[张维维,龚俭,刘茜,等.基于词素特征的轻量级域名检测算法[J].软件学报,2016,27(9):2348–2364.]
[62] Chen Xiao Jun,Fang Binxing,et al.Inferring attack intent of malicious insider based on probabilistic attack graph model[J].Chinese Journal of Computers,2014,37(1):62–72.[陈小军,方滨兴,谭庆丰,等.基于概率攻击图的内部攻击意图推断算法研究[J].计算机学报,2014,37(1):62–72.]
[63] 习近平.在网络安全和信息化工作座谈会上的讲话(2016年4月19日)[N].人民日报,2016–04–26(2).
[64] 中国信息通信研究院.网络与信息安全产业白皮书(2015年)[EB/OL].2015.12.http://www.cac.gov.cn/files/pdf/baipishu/informationsecurity.pdf.

陈兴蜀,女,博士,教授,博士生导师。四川大学网络空间安全研究院常务副院长,网络与可信计算研究所所长。全国信息安全标准化技术委员会委员,教育部信息安全教学指导委员会委员,中央网信办云计算服务安全专家组副组长,全国信息安全标准化技术委员会大数据安全工作组副组长,国际标准化组织ISO/IEC专家,四川省学术与技术带头人,四川省有突出贡献的优秀专家。

陈兴蜀主持了国家科技支撑项目、国家自然科学基金项目等20余项国家级、省部级科研课题。发表高水平论文100余篇,其中SCI/EI检索论文60余篇。主持编写GB/T 31167—2014《信息安全技术 云计算服务安全指南》国家标准,主编GB/T 31168—2014《信息安全技术 云计算服务安全能力要求》。获四川省科技进步二等奖(排名第1);荣获中国互联网发展基金会网络安全专项基金2016网络安全优秀教师奖(全国共评选8名)。

(编辑 赵 婧)

引用格式:Chen Xingshu,Zeng Xuemei,Wang Wenxian,et al.Big data analytics for network security and intelligence[J].Advanced Engineering Sciences,2017,49(3):1–12.[陈兴蜀,曾雪梅,王文贤,等. 基于大数据的网络安全与情报分析[J].工程科学与技术,2017,49(3):1–12.]