doi:10.11835/j.issn.1000-582X.2016.02.010

大数据与安全可视化

向 宏[a,b],张 瑜[b],胡海波[a,b]
(重庆大学 a.信息物理社会可信服务计算教育部重点实验室; b.软件学院,重庆 400044)

摘 要:IT 技术的飞速发展开启了大数据时代。海量的数据信息带来更多的数据价值的同时,各类安全问题也随之而来。通过数据可视化技术能够帮助充分全面并且及时地找出系统中可能存在的安全威胁,评估系统安全,保证基础设施的安全。研究从数据角度出发,结合大数据的特征对安全数据进行分类,并对当前已存在的针对不同安全数据进行可视化的工具及技术进行总结,使得能够将已成熟的数据安全可视化技术用于大数据安全的研究,最后对未来的发展趋势进行了展望。

关键词:大数据;安全可视化;安全数据
中图分类号:TP309 **文献标志码**:A **文章编号**:1000-582X(2016)02-071-11

Big data and security visualization

XIANG Hong[a,b], *ZHANG Yu*[b], *HU Haibo*[a,b]
(a. Key Laboratory of Dependable Service Computing in Cyber Physical Society;
b. School of Software Engineering, Chongqing University, Chongqing 400044, P. R. China)

Abstract: The rapid development of IT technology opens an unprecedented age of big data. The massive amounts of information gives us more data. Meanwhile, all kinds of security problems follow. Data visualization techniques can help us to identify the possible security threats in the system roundly and timely, evaluate system security and ensure the security of infrastructure. In this paper, we first classified security data according to the feature of big data from data perspective; then, summarized the existing visualization tools and technology based on all types of security data so that we can use mature data security visualization technology for the researches of big data security; at last we outlined guidelines and directions for future studies.

Keywords: big data; security visualization; security data

“大数据”这个术语最早出现在 Michael Cox 等人[1]1997 年发表的文章中,称占据主内存、本地磁盘以及远程磁盘很大空间的大型数据集问题为大数据问题。2001 年 Doug Laney[2]正式提出大数据 3V(volume, velocity, variety)的概念。全球知名咨询公司 McKinsey & Company[3]在 2011 年 5 月 发布报告“Big data: The next frontier for innovation, competition, and productivity”提到:“大数据”就是指大小超出典型数据库软件的采集、储存、管理和分析等能力的数据集。此外,Gartner 公司[4]定义“大数据”是需要新处理模式才能

收稿日期:2015-09-12
基金项目:国家自然科学基金资助项目(61472054);中央高校基本科研业务费资助项目(106112014CDJZR098801)。
Supported by National Natural Science Foundation of China(61472054)and The Foundemental Research Funds for the Central Universities(106112014CDJR098801)
作者简介:向宏(1952-),男,重庆大学教授,博士生导师,主要从事数据分析,安全等方向研究,(Tel)13594361541;(E-mail)xianghong@cqu.edu.cn。

具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。2013年，由Mayer-Schönberger等人[5]编著的“Big data: A revolution that will transform how we live, work, and think”提出了“大数据”的4V特征：

1)Volume：即数据量大，数据每年都成几何级数增长，已从TB级上升到PB、EB乃至ZB级。

2)Variety：即数据类型众多，从数据结构的角度来看，大数据主要分为可以用二维表结构逻辑表达实现的结构化数据；像办公文档、文本、图片，音频，视频一类的非结构化数据以及虽不符合正式表结构或关系数据模型，但具有可分离出语义成分的标签或者其他标记，并形成记录或者字段的层次结构的半结构化数据[6]。

3)Velocity：即处理速度快，能够快速地从各类型数据中获取价值密度高的信息。

4)Value：即蕴含的价值高，大数据的数据规模很大，其中所蕴含的价值总量也是相当可观的，因此有大量潜在价值等待人们挖掘。

大数据带来更多的数据价值的同时，随之而来必须面临的是信息泄露和破坏等安全方面的问题。随着有用信息越来越多地被破坏，极大地增加了信息技术基础设施的脆弱性，导致它们非常容易被攻击，所以需要大量的关于网际安全(cyber security, CS)和信息保障(information assurance, IA)的战略、方法和工具用于保护这些重要的IT资产[7]。

1 网络安全可视化

数据安全可视化是一个非常年轻的术语[8]，这个概念还没有一个明确定义。网络安全可视化是数据安全可视化到目前为止研究最为广泛的一个方向，它能够利用人类视觉对模型和结构的获取能力，将抽象的网络和系统数据以图形图像的方式展现出来，帮助人们分析网络状况，识别网络异常或网络入侵行为，预测网络安全事件发展趋势[9]。

Richard A. Becher[10]在1995年第一次提出对网络数据信息可视化的概念，利用地图结合节点连接图对各地网络通讯的流量进行可视化展示。Girardin等[11]在1998年使用了多种可视化技术来分析防火墙日志记录，利用节点连接图，平行坐标轴等展示不同协议事件与活动之间的相关性。自2004年开始，每年IEEE都会举办网络安全可视化研讨会(The international symposium on visualization for cyber security, VizSec)[12]，标志着该研究领域的正式建立。会议重点是为网络安全领域探索有效可扩展的可视化界面，利用可视化技术更好地理解各种网络安全数据，以支持网络安全的分析和异常检测。2013年IEEE VizSec的程序委员会主席KWAN-LIU MA[13]带领的可视化与界面设计创新小组(visualization and interface design innovation group, VIDI)[14]的研究重点之一就包括网络安全可视化，自2003年开始就不断有相关论文发表。Raffael Marty[15]在2008年所著的《applied security visualization》一书中详细地介绍了安全日志数据可视化相关知识。而VAST challenge[16]自2011年起至2013年连续3年都采用了网络安全数据作为竞赛题目，推动着该领域呈现出一个新研究热潮。

就国内而言，哈尔滨工程大学、天津大学、北京邮电大学、吉林大学、北京大学和中南大学等研究机构的一些团队也开展了相关的研究[17]。安天实验室[18]是国内较早关注安全可视化的团队，专注于网络安全威胁态势、恶意代码源分布、感染情况方面的可视化研究。2014年在上海举行的互联网安全研讨会(ISF)[19]主题为安全可视化，会议讨论了可视化的方法与挑战，以及3D可视化的相关应用。

2 大数据与安全数据分类

数据是所有可视化的基础，没有数据，就没有可视化[15]。信息安全领域中，不同的数据源会产生不同类型的安全数据，如数据包、网络流量、BGP信息、时间序列数据、各种日志文件等，它们所能包含的信息是非常丰富的。将不同数据源的数据整合到一起，相互搭配进行可视化展示能够从多个角度来全面准确地监测分析一个网络事件，并且很好地体现当前网络及设备的数据传输、网络流量来源及流动方向、受到的攻击类型等安全情况。

大数据有4V特征，V不仅体现在时间上的Velocity，同样也体现在了空间上的Vast，形成信息安全数据的5V特征，正好映射到网络安全可视化中5种类型的数据集。

Volume：代表海量的数据规模，典型的案例为网络流量数据。虽然在小型信息系统产生的数据量达不到大数据所谓的 TB 级或 PB 级，但是，在城市级甚至国家级的网络监控以及网络态势感知过程中，数据库审计、各类防火墙、入侵检测系统及 web 服务器等在网络链路上产生的网络流量数据积累起来不可小觑。利用这些网络流量数据能够帮助网路安全分析人员快速发现端口扫描、蠕虫扩散以及拒绝服务攻击等安全事件。

Velocity：代表时间范畴，典型的案例为带有时间属性的时间序列数据。这一特征可理解为网络安全中快速的数据流转和动态的数据体系[20]。随着数据量的增大，网络安全检测对时间的即时性需求并没有因此有所改变，实时性一直是网络安全可视化需要解决的问题之一。而安全事件的产生及其原因是具有时间跨度的，因此需要可视化中带有时间序列的动态数据流作为可视化输入来帮助安全分析人员识别可疑的事件及行为。

Vast：代表空间范畴，典型的案例为 BGP 路由信息。网络安全事件以及攻击行为的产生不仅会涉及到时间的延续，也会涉及到空间位置的变化。BGP 路由器都会与周围的一个或者多个路由器相连接，建立连接关系之后，BGP 路由器之间将会相互交换路由信息。对 BGP 路由信息的可视化能够帮助网络安全分析员从空间上了解路由的路径变化以识别网络中的异常行为。

Variety：代表数据的多样性，典型的案例为各类网络日志数据。随着各类安全设备的使用，会产生防火墙、入侵检测、主机安全以及垃圾邮件等各种类型的日志数据。而安全事件与攻击行为产生的痕迹将会以各类日志的形式记录在不同的安全设备上，仅仅根据一两类的日志数据想要完整的描述安全事件是比较困难的，所以对多种日志数据的关联融合分析以及可视化能够帮助网络安全分析人员分析出事件关联，快速识别网络异常并发现不同的网络攻击模式。

Value：代表蕴含的价值高，典型的案例包括网络安全可视化中的其他类型安全数据，如数据包信息以及漏洞信息等。在网络安全中，随着安全事件越来越多样化、复杂化，除了以上 4 类数据信息外，还有很多其他类型的具有较高分析与可视化价值的安全数据，在这些海量安全数据信息中找到有价值的信息进行可视化分析能够帮助网络安全分析人员发现网络中更多未知的威胁，更好地维护网络及基础设施的安全。

针对以上对信息安全中安全数据的分类，表 1 列出了这些安全数据的部分数据来源以及其能够体现出的各种安全事件类型。

表 1　来自不同数据源的安全数据

Table 1　Security data from different data source

安全数据	对应的大数据的 V	数据源	事件类型
网络流量数据	Volume	网络流量记录，主机，交换机，路由器，服务器，虚拟专用网络，杀毒软件等	端口扫描，拒绝服务攻击，蠕虫病毒等安全事件
时间序列数据	Velocity	7 层应用程序上下文，机，交换机，路由器，服务器等	网络跟踪、活动等
BGP 信息	Vast	主机，交换机，路由器，服务器，7 层应用程序上下文，虚拟专用网络等	网络跟踪，网络事件，安全事件等
各类日志	Variety	防火墙，入侵检测系统，入侵防御系统，主机，交换机，路由器，服务器，应用程序数据库，安全管理平台等	分布式拒绝服务攻击，蠕虫病毒，复杂网络攻击等安全事件
其他（如漏洞信息，数据包等）	Value	漏洞扫描软件，数据包，交换机，主机等	网络事件、跟踪、活动等

3　网络安全可视化分类

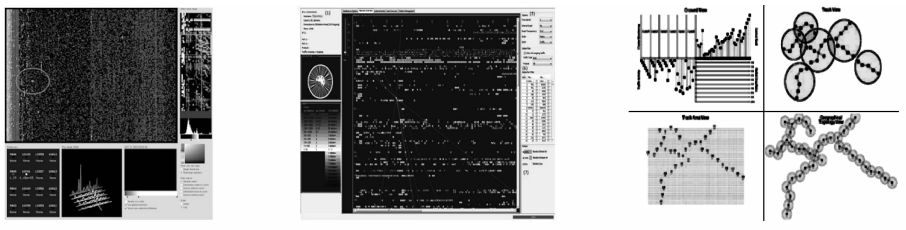
目前为止，已有从不同的安全角度对网络安全可视化技术进行分类的相关文献。2008 年，吕良福等人[21]针对不同的安全事件，利用基于网络数据流量、端口信息、入侵检测技术以及防火墙事件等不同的网络

安全可视化技术对当时的研究做出了分类总结。Shiravi 等人[8]在 2012 年根据主机/服务器监测、内部主机/外部 IP 监测、端口活动、攻击模式以及路由行为等需要解决的具体安全问题对网络安全可视化的系统进行了分类的介绍。2014 年，赵颖等人[17]则再次从网络安全问题和网络安全可视化方法 2 个角度，针对络监控、异常检测、特征分析、关联分析和态势感知五类应用，对已有的研究成果进行了系统的梳理，并对可视化的分类作了可视化的展示。

3.1 基于网络流量数据的可视化技术

由于端口扫描、蠕虫扩散以及拒绝服务攻击等安全事件在流量方面具有明显的一对一、一对多或多对一的特征，因此，此类攻击事件往往在流量方面出现明显的异常，显示网络流量可以帮助网路安全分析人员快速发现网络攻击，更好地防范和抵御网络入侵事件[21]。网络安全可视化在针对网络流量数据方面优势较大研发出的技术与工具非常多，针对网络流量的可视化可以用于体现端口的活性、网络表征扫描、监控大型 IP 空间、监控检测不同类型的网络异常等等。针对此类数据的可视化通常会利用点阵图、网格图、饼状图、折线图、节点连接图、散点图等对固定时间段内不同端口通过的流量以及总体变化趋势进行显示监测，找到有相似行为的设备，快速确定攻击模式，攻击来源以及受到网络攻击或感染病毒后将会被波及的设备与范围。视图中通常会用不同的颜色来代表流量的多少，但颜色的不同只能表示不同范围的数量，并不能精确地显示出流量及活动的数量。

2004 年 McPherson 等[22]人开发的 PortVis 利用网络流量数据，如图 1(a)，采用网格图对端口活性进行监测，帮助安全分析师识别出与端口紧密相关的如端口扫描、特洛伊木马等网络安全事件，流量越多，端口的活性越高，不同的颜色代表了不同的活性级别。Boschetti 等[23]人开发的用于网络监控和异常检测的可视化查询系统 TVi 利用了一些数学物理方面的知识如熵，欧几里得范数，结合网络跟踪情况，展示不同时间网络流量的概率分布，帮助用户进行异常检测，此工具还涉及到了地理位置的确定，可做成 3D 可视化，对识别物理位置有很大帮助。康斯坦茨大学研发的通过可视化检测异常流量来监控大型 IP 空间的工具 ClockView[24]，如图 1(b)，通过评估计算机软件处理不同情况的使用来监控网络并检测出异常流量，通过节点连接图和饼状图展示不同时间使用的网络流量情况来帮助用户识别相似的模型与拥有相同即时行为的主机。工具中的一个饼图代表一台主机，圆形根据不同时间分为了 24 等分的扇形，每个扇形的颜色表示流量的多少。2013 年马其顿大学信息系统部门研发的针对无线传感网络的网络流量监测的实时可视化工具 SRNET[25]，如图 1(c)，通过在地理拓扑图，轨道区域图等多个协调的视图上实时地查看网络信息流量数据包的漏传或误传来识别无线传感网络里的选择性转发攻击和干扰攻击，追踪网络攻击的来源以及危险程度，该工具较之前相似的工具的优势就是可以在一个单一视图中检测出很多网络攻击。针对网络流量数据的可视化工具还包括 3D 可视化，如 PGVis3D[26]技术综合利用了 IP Matrix 的 2D 和 3D 技术来表现网络内部和外部、内部和内部之间的流量状况。



(a) Port Vis (b) Clock View (c) Srnet

图 1 基于网络流量数据的可视化效果图

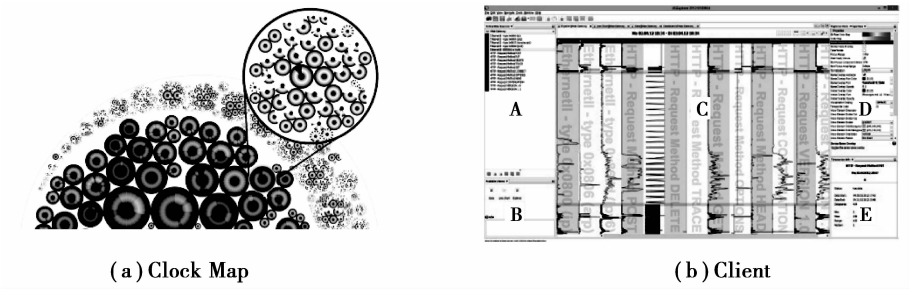
Fig.1 Visualization based on network flow data

3.2 基于时间序列数据的可视化技术

计算机网络监控需要从成千上万的计算机系统和网络设备中收集大量的时间序列数据。因此对时间序列的数据的可视化也十分重要。针对此类数据的可视化用能够展示基于层次关系的树图进行不同的数据维度的展示。

这类可视化技术在 Aigner 等[27]的书中可以看到较系统的概述，一个被称为 two-tone pseudo coloring[28]的

可视化技术使用了 2 个独立的颜色作为时间序列的每一个值，形成地平线图标与线图表进行比较，来强调关键或可疑的事件或行为。由 Fischer 等[29]人提出的可视化工具 ClockMap 对网络结构进行分级并且使用带有嵌入时间符号的圆形树图布局来代表它上下文中的许多指标，用此工具可以直观地探索从网络通信或系统监控应用程序中检索出来的分层时间序列数据，包括与基础设施相关的消息以及与互联网路由协议以及互联网的各种威胁，如图 2(a)，用语义缩放的方式可以在概述图与详细时间序列图之间进行切换。此工具采用了圆形树图的方式进行可视化，相对矩形树图来说虽然有空间的浪费，但是它能够很好地体现跨层次的数据结构，因此对时间序列数据来说是很好的，只是此工具用颜色代表数据值，并不能够较准确地展示出数值的多少。由 Maclachlan 等[30]人提出的可视化工具 LiveRAC 专注于网络安全角度的监控数据的分析。Best 等[31]使用的系统结合基于符号集合逼近的时间序列分析[32]对网络流量数据进行分析来发现异常序列从而提高网络安全性，并提供实时态势感知能力。Shafer 等人[33]也提供可视化系统来监控识别基于时间序列的异常机器。Kincaid[34]对于显示时间序列的数目具有扩展性，提供了一个压缩的可视化表示，能够很好的捕捉到很多时间序列的全局整体相似性。2013 年 Stoffel 等人[35]提出的工具，如图 2(b)，利用波动曲线图利用集成相似模型来分析应用程序，以及时间序列可视化的分析来可视化地识别大型数据集中的关联和异常，并确定安全相关联事件以帮助安全分析人员使用视觉探索，时间序列的功能和相似性搜索来检索相关数据，跟踪可疑事件并找到其根本原因。但是此工具主要是针对比较大型的数据集，普适性并不高，而且可视化效果较为单一，只有波动曲线图，且当曲线图太多的时候，看起来会比较混乱，要找图之间的相似性存在一定的困难。



(a) Clock Map　　(b) Client

图 2　基于时间序列数据的可视化效果图

Fig.2　Visualization based on time-series data

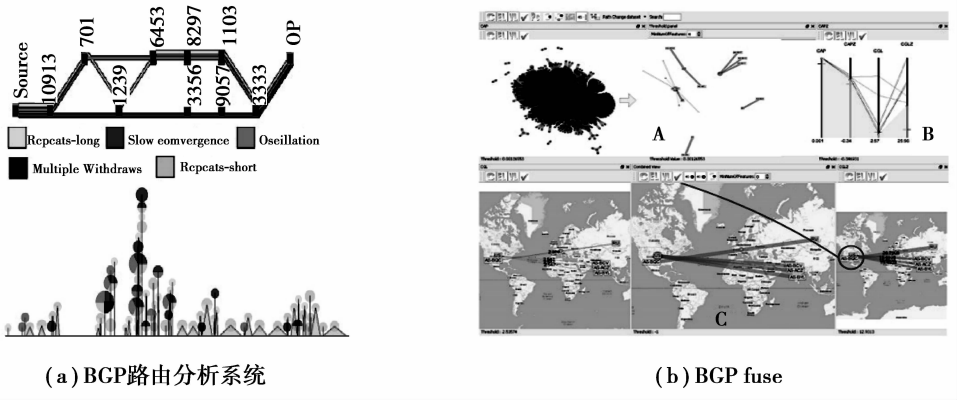
3.3　基于 BGP 的可视化技术

边界网关协议(border gateway protocol, BGP)在自治系统之间分发路由信息，是互联网路由基础设施的重要组成部分[36]。上一个核心的去中心化自治路由协议[36]。它是互联网最重要的协议之一，在 BGP 上的故障和攻击可能会导致全球范围内的链接损失，而通过对 BGP 的路径变化、通告，以及路由跟踪等信息的可视化可以帮助安全管理员及时进行网络的异常行为检测、分类，提早预防网络攻击。针对此类数据的可视化多会采用地图、节点连接图及平行坐标轴、饼状图分别对发生异常的设备进行地理位置定位，通过不同的属性及所占比例对异常事件进行分类，并追踪其路径。

2003 年 Teoh 等[37]提出的分析 BGP 通告的可视化工具如图 3(a)，通过对路径通告的可视化，采用节点图，树图及饼状图来对故障和异常事件进行检测分类。此外，Li 等[38]提出了一种基于签名的方法，使用柱状图识别每个特定类型的如路由器配置错误及蠕虫攻击的 BGP 异常行为表征，利用基于这些特征的 IF-THEN 阈值集合，使用折线图对异常的 BGP 事件进行检测和分类，但此方法的异常结果是静态的。Zhang 等[39]使用基于签名和统计的方法来进行 BGP 异常检测到后来可以对统计和签名参数进行调节，使得其折线图与柱状图的展示有了较之前更好的效果，之后 Teoh 等[40]又对此方法做了一个扩展来调节签名和统计的参数，用带有颜色的折线图取代了之前的黑白折线图，获得了更好的可视化效果展示。由 Fiacher 等[41]提出的可视化工具 VisTracer 通过 BGP 路由跟踪到垃圾邮件针对大型 traceroute 数据使用异常检测算法以支持端口分析师识别和分析可疑事件及其与恶意活动的关系。而 2013 年 VizSec 的一篇会议论文中[42]提出了 BGPfuse 如图 3(b)，使用一组边界网关协议的特性，能够量化每个路径变化程度的异常事件，并用可视化的方法来执行多个特性的有效融合。这是首个提到可视化与特征融合相结合的工具，利用独立特征图形之间

的结构相似性，采用节点连接图，平行坐标轴，世界地图来凸显可疑的BGP路径变更事件，通过观察多个特征来揭示可能参与多个事件的角色。当一个重要的信息系统面临世界各地的攻击的时候，可利用此工具推测相似的攻击路径来判断不同国家的黑客是否是同一个组织。只是此可视化工具针对的是以国家为单位的自治系统，范围较大，而且可视化的效果较为复杂，看起来会比较难以理解。

由于现今所有应用上BGP安全机制的缺乏，导致了各种针对BGP的攻击，对BGP的安全可视化研究越来越重要，也越来越迫切。



(a) BGP路由分析系统　　(b) BGP fuse

图3　基于BGP的可视化效果图

Fig.3　Visualization based on BGP

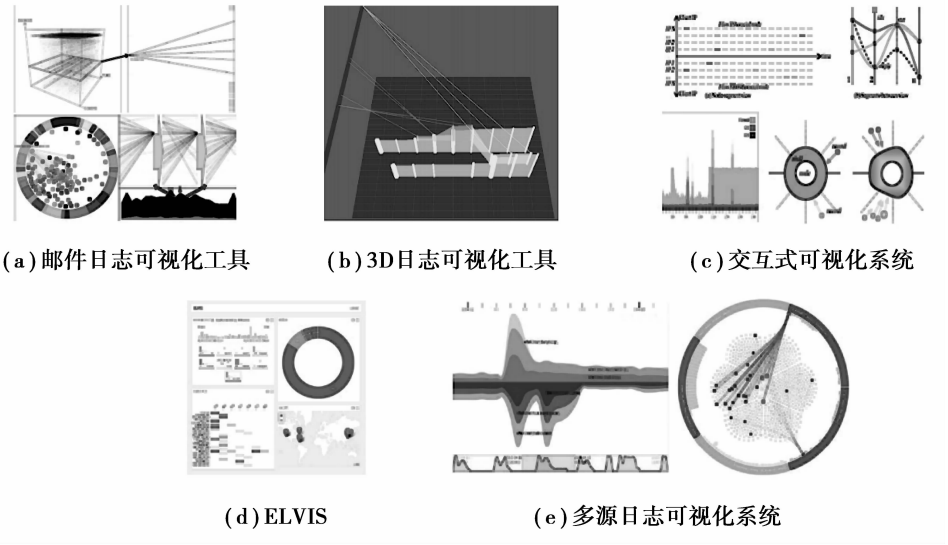
3.4　基于日志数据的可视化技术

网络安全会用到很多设施，如防火墙、入侵检测系统、漏洞扫描仪等，它们所产生的日志文件能够从很多不同的方面来进行安全事件的分析，从日志文件可以找到非常多有用的信息如时间、优先级、协议、源IP/端口、目的IP/端口等，这些信息对安全分析人员来说是非常有意义的。针对不同的日志文件的可视化技术不尽相同，对安全分析人员的帮助也体现在不同方面，因此，对日志文件的安全可视化研究也是很多的。由于日志数据的数量非常庞大，通常会采用堆叠图、折线图、直方图等显示出一定时间内，日志数据中不同活动的统计数据及变化趋势，采用散点图、平行坐标轴、地图等对日志文件中的IP、数据包和流量的传送路径进行定位。通过树图展示被防火墙拒绝或允许访问的端口或IP从而判断防火墙会保护或者拒绝的系统；以及以漏洞为基础，通过树图找出攻击者能够对网络进行攻击的所有路径，对网络安全性进行评估。通过环形图、雷达图、径向图或3D视图对复杂网络攻击所产生的一系列有关联的安全事件的类型、位置和时间进行显示，帮助用户快速识别异常、发现攻击模式和分析事件关联。

现如今大部分工具只是针对单数据源日志数据。2007年Muelder等人[43]研发的可视化工具，如图4(a)，通过邮件日志利用二分图，3D及动画的可视化方式来对垃圾邮件进行过滤，并且帮助用户寻找垃圾邮件的发送源和中转站。二分图能够根据邮件发送的源地址和目的地址分辨出类似僵尸网络的安全事件，动画和3D可视化能够展示出不同时间邮件发送的源地址和目的地址，且动画的展示效果明显比3D清楚明了。而环形图能够帮助用户识别出被当做了垃圾邮件中转站的地址。康斯坦茨大学研发的BANKSAFE[44]用来可视化监测警报和防火墙日志等安全数据集。ClockView[24]和PeekKernelFlows[45]用NetFlow日志来长时间监控大型IP空间。Nyarko等[46]人提出用对IDS日志文件进行三维可视化的技术来评估攻击影响，Takacla等[47]人提出用三维可视化系统来监控和审计系统日志。Anatoly Yelizarov等人[48]提出的对日志文件中记录的攻击事件进行3D可视化，如图4(b)，能够对复杂攻击事件进行检测识别，帮助安全分析人员直观的查看到主机受到复杂攻击时的状况。平面上不同颜色的圆柱代表了某段时间某一端口发生的不同简单攻击，圆柱的高度代表了攻击的严重程度，而一个复杂的攻击过程则由平面上连接起来的不同圆柱进行表示。文[49]中对大量的安全数据和日志信息的分析处理，网络数据以图形图像的方式表现出来，而这里的可视化也只是用简单的柱状图对事件日志进行查看管理。2013年Alsaleh等人[50]提出了第一个基于web的安全日志可视化，分析与相应的web服务器日志相关联的PHPIDS日志，使用PHPIDS从不同的IP地址检测出不同的网络攻击，构建交互式数据可视化，并为PHPIDS扩展建立可视化，绘制出相关的安全事件。这些日志

可视化工具可视化展示各不相同，都能达到各自的功能需求，但是都只是针对单一的一种安全日志进行可视化。少部分可视化工具能够处理不同数据源的日志文件。2012 年 Song 等人[51]提出的交互式可视化系统，如图 4(c)，使用数据同步器对防火墙和 IDS 日志进行实时的可视化，分别通过矩阵图展示不同时间的不同 IP 的流量情况，平行坐标轴展示流量经过的关键节点路径，堆叠直方图统计了同一系统中防火墙，IDS 及操作系统的网络事件以及基于图元的关键节点动画视图展示关键节点受到的来自不同 IP 的各类攻击及其受损情况，以此对关键基础设施的节点上相对可疑的网络活动进行分析推理。此工具的动画可视化效果简单易懂，能够让用户比较直观的看到网络攻击的过程以及节点的受损程度。此工具不仅能够让安全分析人员直观的看到节点安全状况的变化，对普通用户也能达到同样的效果。2013 年 Humphries 等[52]人开发的 ELVIS，如图 4(d)，能够把不同格式的日志文件加载到工具中，安全分析人员可以根据需要选定数据集中的特定字段，然后工具会根据自动选择的相关表述自动选择合适的饼状图、条形图、地图等视图对数据进行可视化展示，它能够处理尽量多不同格式的日志文件，并且遇到新类型日志文件的时候能够自动的扩展来进行可视化，但是不能可视化出不同类型的日志数据集之间的相互关联。2014 年波茨坦应用技术大学开发的视觉过滤器 LogSpider[53]展示了全日志的概述，结合焦点查询来搜索已知可疑术语，并在可视化和日志文件本身中突出显示出来。

而文献[54]提供的原型系统，如图 4(e)，通过数据融合技术对多源日志网络安全数据进行了统一格式的事件元组和统计元组的提取，并且设计了对比堆叠流图展示出不同时间，不同端口和 IP 的流量情况，雷达图展示了设备遭遇安全事件的整体情况，发生了某一类安全事件的设备以及某一特定设备遭受所有安全事件的情形，并根据多组时序数据的对比来监控端口的活动情况。



(a)邮件日志可视化工具　(b)3D日志可视化工具　(c)交互式可视化系统

(d)ELVIS　(e)多源日志可视化系统

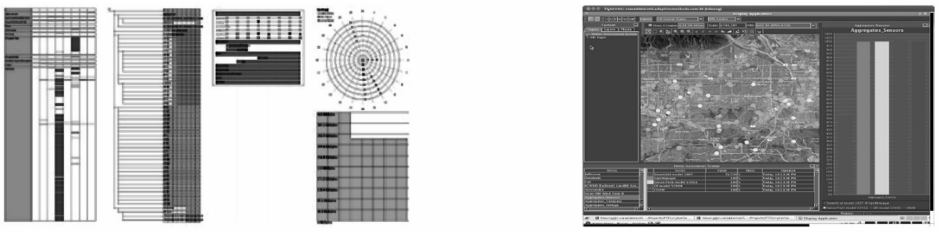
图 4　基于日志数据的可视化效果图

Fig.4　Visualization based on logdata

3.5　其他的安全可视化技术

随着各种安全事件的越来越多样化，数据越来越复杂，除了以上提到的数据种类可以可视化之外，其他的也能够用于可视化，因此开发出来的技术，工具与系统也多种多样。例如 InSeon Yoo 提出了利用计算机病毒的特征信息帮助安全分析人员发现并抵御嵌入在可执行文件中的病毒的自组织映射(self-organizing Maps)技术[55-56]。Ren 等[57]人在 2006 年提出了 Flying Term 技术利用堆叠图及脸谱图显示出相关 DNS 查询信息，帮助安全分析人员找出潜在的 DNS 攻击事件。2008 年的一篇学位论[58]中开发了可视化端口扫描检测系统 ScanViewer 运用安全可视化技术，通过分析处理网络数据包来发现攻击模式，有效地检测慢扫描、分布式扫描和各类 TCP 隐蔽扫描。用于数字取证的多视图工具 Change Link 2.0[59]，如图 5(a)，为更好地理解阴影数据如何随时间的改变而变化，利用树形图支持整个数据集的概述和理解目录树结构，了解各目录内

容和文件及元数据是怎样随时间变化而变化的。此工具在并行多画面的链接视图界面提供了一个概述图，展示了硬盘驱动器的根目录，以及此目录下的文件和目录子结构的数量，目录树视图展示出目录的存在与否，目录内容视图用以区分用户意图和自动进程的行为，元数据视图展示更改文件或目录的时间和大小信息，以支持随时间变化的文件和目录的简单浏览及检测。算是一种比较新的网络安全可视化方法，但是该工具的效果对于产生什么样的攻击，会造成什么样的后果都没办法表现出来。Matuszak W J 等人开发了一个原型网络态势感知可视化工具 CyberSAVe[60] 来对网络信任进行可视化，如图 5(b)，定义一个网络信任度，在智能电网系统中为网络信任提供算法，由信任计算、多维信任度，以及数学特性组成的信任模型计算整体的信任值，并且将其实时地展示在此工具的评估系统中，利用柱状图、条形图、时间历程曲线结合地图来检测对智能电网进行的各种类型的攻击。此工具仅能智能检测到单个恶意节点，并不能可视化恶意节点之间的关系，而且目前只适用于智能电网这一类的 SCADA 系统，要用到其他系统中还需要进行修改，局限性比较大。



(a) Change Link (b) Cyber SAVe

图 5 其他可视化效果图

Fig.5 Ohter visualization

4 结 语

从数据角度出发，结合大数据各大 V 的特征对安全数据进行分类，并对当前世界上已存在的针对不同安全数据的网络安全可视化的工具及技术进行总结，以帮助人们选择最合适的安全可视化技术挖掘出手中握有的大数据所蕴含的安全价值。到目前为止，网络安全可视化的研究越来越丰富，但依旧存在很多问题，特别是结合大数据及其安全研究，还面临着诸多挑战，包括但不限于：

1)大数据安全关联分析可视化数据更加着重于数据间的关联分析，而目前的网络安全可视化工具要么只能对比较单一的数据源进行可视化并进行较深入的分析研究，要么能够对较丰富的数据源进行可视化但是不能够展示出不同数据之间的关联，如何结合大数据的数据分析方法，将不同类型的安全数据之间的关联关系展示在用户面前是非常值得研究的。

2)大数据安全实时展示性永远都是网络安全可视化需要解决的问题，而因为某些安全事件或攻击的过程本身具有时间跨越性，使得本来是动态数据流的安全数据变成了静态数据块，且由于数据量太大成为了分布式大数据问题，而结合大数据，使用并行计算或者异步计算解决实时性问题将是一大研究热点。

3)可视化工具易用性差，即使是专业的分析人员要想熟练的使用都需要一定的时间对其进行熟悉，对一般用户来说更是这样，因此需要加强其易用性的研究。

4)由于理论知识的缺乏以及太强的主观性，到现在还没有一套完整的评价体系或标准能够对网络安全可视化效果进行有效的评估，如何建立一套完整的评价体系的研究也是必不可少的。

参考文献：

[1] Cox M, Ellsworth D. Application-controlled demand paging for out-of-core visualization[C]// Proceedings of the 8th conference on Visualization'97, October19-24,1997, Phoenix, AZ, USA: IEEE Computer Society Press,1997:235-244.

[2] Doug Laney. Application delivery strategies[M]. USA: META Group Inc,2011.

[3] Manyika J, Chui M, Brown B, et al. Big data: The next frontier for innovation, competition, and productivity[J].

McKinsey Global Institute，2011.

[4] Gartner Group Homepage.[EB/OL].[2015-03-26].http://www.gartner.com/.2016.

[5] Mayer-Schönberger V, Cukier K. Big data: A revolution that will transform how we live, work, and think[M]. USA: John Munay Publishers，2013.

[6] Buneman P. Semistructured data[C]// Proceedings of the sixteenth ACM Sigact-sigmod-sigart Symposium on Principles of Database Systems. USA: Association for Computing Machinery，1997：117-121.

[7] Measuring Cyber Security and Information Assurance: A State-of-the Art Report[M]. USA: Information Assurance Technology Analysis Center，2009.

[8] Shiravi H, Shiravi A, Ghorbani A. A survey of visualization systems for network security[J]. IEEE transactions on visualization and computer graphics，2012，18(8)：1313-1329.

[9] Fortier S C, Shombert L A. Network profiling and data visualization[C]// Proceedings of the 2000 IEEE Workshop on Information Assurance and Security, USA：IEEE, 2000.

[10] Richard A, Becker S G E, Allan R. Wilks, Visualizing network data[J]. IEEE Transactions on Visualization and Computer Graphics，1995，1(1)：16-28.

[11] Girardin L，Brodbeek D. A visual approach for monitoring logs[C]// Proceedings of Large Installation System Administration Conference. New York: Association for Computing Machinery Press，1998：299-308.

[12] VizSec Homepage.[EB/OL].[2015-03-26]. http://www.vizsec.org/.2014

[13] KWAN-LIU MA Homepage.[EB/OL].[2015-03-26]. http://www.cs.ucdavis.edu/~ma/.2010.

[14] VIDI Homepage.[EB/OL].[2015-03-26]. http://vidi.cs.ucdavis.edu/new.2014.

[15] Marty R. Applied security visualization[M]. Upper Saddle River: Addison-Wesley，2009.

[16] VAST Challenge Homepage.[EB/OL].[2015-03-26]. http://www.vacommunity.org/VAST+Challenge+2013.2013.

[17] 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.
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.(in Chinese)

[18] Antiy Labs Homepage.[EB/OL].[2015-03-26]. http://www.antiy.com/.

[19] Internet Security Forum(ISF) Homepage.[EB/OL].[2015-03-26]. http://isf.cisrg.org/.

[20] 潘柱廷. 安全大数据的 7 个 V——大数据基础问题与信息安全的交叉探究[EB/OL].[2015-03-26]. http://www.thebigdata.cn/JieJueFangAn/12951.html? utm_source=tuicool.2014.
PAN Zhuting. 7V of security big data-the cross inquiry between basic questions of big data and information security[EB/OL].[2015-03-26]. http://www.thebigdata.cn/JieJueFangAn/12951.html? utm_source=tuicool. 2014.(in Chinese)

[21] 吕良福，张加万，孙济洲，等. 网络安全可视化研究综述[J]. 计算机应用，2008，28(8)：1924-1927.
LV Liangfu, ZHANG JiaWan, SUN JiZhou, et al. Survey of network security visualization techniques[J]. Computer Applications，2008，28(8)：1924-1927.(in Chinese)

[22] McPherson J, Ma K L, Krystosk P, et al. Portvis: a tool for port-based detection of security events[C]// Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security. New York: Association for Computing Machinery，2004：73-81.

[23] Boschetti A, Salgarelli L, Muelder C, et al. Tvi: a visual querying system for network monitoring and anomaly detection[C]// Proceedings of the 8th International Symposium on Visualization for Cyber Security. USA: Association for Computing Machinery，2011：1.

[24] Kintzel C, Fuchs J, Mansmann F. Monitoring large ip spaces with clockview[C]// Proceedings of the 8th International Symposium on Visualization for Cyber Security. USA: Association for Computing Machinery，2011：2.

[25] Karapistoli E, Sarigiannidis P, Economides A A. SRNET: a real-time, cross-based anomaly detection and visualization system for wireless sensor networks[C]// Proceedings of the Tenth Workshop on Visualization for Cyber Security. USA: Association for Computing Machinery，2013：49-56.

[26] Le Malécot E, Kohara M, Hori Y, et al. Interactively combining 2D and 3D visualization for network traffic monitoring[C]// Proceedings of the 3rd international workshop on Visualization for computer security. USA: Association for

Computing Machinery,2006:123-127.

[27] Aigner W, Miksch S, Schumann H, et al. Visualization of Time-Oriented Data. Human-Computer Interaction Series[M]. USA: Springer,2011.

[28] Saito T, Miyamura H N, Yamamoto M, et al. Two-tone pseudo coloring: Compact visualization for one-dimensional data[C]// In Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization, INFOVIS'05. Washington, DC, USA: IEEE Computer Society,2005:23.

[29] Fischer F, Fuchs J, Mansmann F. ClockMap: enhancing circular treemaps with temporal glyphs for time-series data[C]// In M. Meyer and T. Weinkauf, editors, Proceedings of the Eurographics Conference on Visualization (EuroVis 2012 Short Papers). Vienna, Austria: IEEE,2012:97-101.

[30] McLachlan P, Munzner T, Koutsofios E, et al. Liverac: interactive visual exploration of system management time-series data[C]// In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'08. New York, USA: Association for Computing Machinery,2008:1483-1492.

[31] Best D M, Bohn S, Love D, et al. Real-time visualization of network behaviors for situational awareness[C]// In Proceedings of the Seventh International Symposium on Visualization for Cyber Security, VizSec'10. New York, USA: Association for Computing Machinery,2010:79-90.

[32] Keogh E, Lin J, Fu A. Hot sax: Efficiently finding the most unusual time series subsequence[C]// In Proceedings of the Fifth IEEE International Conference on Data Mining, ICDM'05. Washington, DC, USA: IEEE Computer Society,2005:226-233.

[33] Shafer I, Ren K, Boddeti V N, et al. Rainmon: an integrated approach to mining bursty timeseries monitoring data[C]// In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD'12. New York, USA: Association for Computing Machinery,2012:1158-1166.

[34] Kincaid R, Lam H. Line graph explorer: scalable display of line graphs using focus context[C]// In Proceedings of the working conference on Advanced visual interfaces, AVI'06, pages 404-411, New York, USA: Association for Computing Machinery,2006.

[35] Stoffel F, Fischer F, Keim D A. Finding anomalies in time-series using visual correlation for interactive root cause analysis[C]// Proceedings of the Tenth Workshop on Visualization for Cyber Security. New York: Association for Computing Machinery,2013:65-72.

[36] Kent S, Lynn C, Seo K. Secure Border Gateway Protocol (S-BGP)[J]. IEEE Journal in Communications,2000,18(4):582-592.

[37] Teoh S T, Ma K L, Wu S F. A visual exploration process for the analysis of internet routing data[C]// Proceedings of the 14th IEEE Visualization 2003 (VIS'03). USA: IEEE Computer Society,2003:69.

[38] Li J, Dou D, Wu Z, et al. An Internet routing forensics framework for discovering rules of abnormal BGP events[J]. ACM Sigcomm Computer Communication Review,2005,35(5):55-66.

[39] Zhang K, Yen A, Zhao X, et al. On detection of anomalous routing dynamics in BGP[C]// Networking 2004. Springer Berlin: Heidelberg,2004:259-270.

[40] Teoh S T, Zhang K, T seng S M. et al. Combining visual and automated data mining for near-real-time anomaly detection and analysis in BGP[C]// The Workshop on Visualization and Data Mining for Computer Security[S. L.]:IEEE,2004:35-44.

[41] Fischer F, Fuchs J, Vervier P A, et al. VisTracer: a visual analytics tool to investigate routing anomalies in traceroutes[C]// Proceedings of the Ninth International Symposium on Visualization for Cyber Security. New York: Association for Computing Machinery,2012:80-87.

[42] Papadopoulos S, Theodoridis G, Tzovaras D. BGP fuse: using visual feature fusion for the detection and attribution of BGP anomalies[C]// Proceedings of the Tenth Workshop on Visualization for Cyber Security. New York: Association for Computing Machinery,2013:57-64.

[43] Muelder C, Ma K L. Visualization of sanitized email logs for spam analysis[C]// Visualization, 2007. APVIS'07. 2007 6th International Asia-Pacific Symposium on.[S. L.]: IEEE,2007:9-16.

[44] Fischer F, Fuchs J, Mansmann F, et al. BANKSAFE: A visual situational awareness tool for large-scale computer

networks: VAST 2012 challenge award: Outstanding comprehensive submission, including multiple vizes[C]// Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on.[S. L.]: IEEE,2012:257-258.

[45] Wagner C, Wagener G, Dulaunoy A, et al. PeekKernelFlows: Peeking into IP flows[C]// Proceedings of the Seventh International Symposium on Visualization for Cyber Security. New York: Association for Computing Machinery,2010: 52-57.

[46] Nyarko K, Capers T, Scott C. Network intrusion visualization with NIVA, an intrusion detection visual analyzer with haptic integration[C]// In Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS'02). USA: IEEE,2002:277-284.

[47] Takada T, Koike H. Tudumi: Information visualization system for monitoring and auditing computer logs[C]// In Proceedings of the Sixth International Conference on Information Visualization. UK: IEEE,2002:570-576.

[48] Yelizarov A, Gamayunov D. Visualization of complex attacks and state of attacked network[C]// Proceedings of the 6th International Workshop on Visualization for Cyber Security, Octorber 11, 2009. Atlantic City, USA: IEEE,2009:1-9.

[49] 舒孝春.可视化入侵检测技术在校园网中的应用[J].电脑知识与技术,2011,5:18-19.
SHU Xiaochun, application of visualization intrusion detection technology in campus network[J]. Computer Knowledge and Technology,2011,5:18-19.(in Chinese)

[50] Alsaleh M, Alqahtani A, Alarifi A, et al. Visualizing PHPIDS log files for better understanding of web server attacks[C]// Proceedings of the Tenth Workshop on Visualization for Cyber Security. USA: Association for Computing Machinery, 2013:1-8.

[51] Song H, Muelder C W, Ma K L. Crucial Nodes Centric Visual Monitoring and Analysis of Computer Networks[C]// Cyber Security (CyberSecurity), 2012 International Conference on. USA: IEEE,2012:16-23.

[52] Humphries C, Prigent N, Bidan C, et al. ELVIS: Extensible Log VISualization[C]// Proceedings of the Tenth Workshop on Visualization for Cyber Security. USA: Association for Computing Machinery,2013:9-16.

[53] Stange J E, D M Landstorfer. Visual filter: graphical exploration of network security log files[C]// Proceedings of the Eleventh Workshop on Visualization for Cyber Security. USA: Association for Computing Machinery,2014:41-48.

[54] 赵颖,樊晓平,周芳芳,等.大规模网络安全数据协同可视分析方法研究[J].计算机科学与探索, 2014,8(7):848-857.
ZHAO Ying, FAN Xiaoping, ZHOU Fangfang, et al. Study on collaborative visual analysis of large scale network security data[J]. The Journal of Frontiers of Commputer Science and Technology,2014,8(7):848-857.(in Chinese)

[55] Labib K, Vemuri R. NSOM: A real-time network-based intrusion detection system using self-organizing maps[J]. Networks and Security,2002:1-6.

[56] Yoo I S. Visualizing windows executable viruses using self-organizing maps[C]// Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security. USA: Association for Computing Machinery,2004:82-89.

[57] Ren P, Kristoff J, Gooch B. Visualizing DNS traffic[C]// Proceedings of the 3rd international workshop on Visualization for computer security. USA: Association for Computing Machinery,2006:23-30.

[58] 吕良福.DDoS 攻击的检测及网络安全可视化研究[D].天津:天津大学,2008.
LV Liangfu. Research on DDoS attacks detection and related network security visualization techniques[D]. Tianjin: Tianjin Universiry,2008.(in Chinese)

[59] Leschke T R, Nicholas C. Change-link 2.0: a digital forensic tool for visualizing changes to shadow volume data[C]// Proceedings of the Tenth Workshop on Visualization for Cyber Security. New York: Association for Computing Machinery,2013:17-24.

[60] Matuszak W J, DiPippo L, Sun Y L. CyberSAVe: situational awareness visualization for cyber security of smart grid systems[C]// Proceedings of the Tenth Workshop on Visualization for Cyber Security. New York: Association for Computing Machinery,2013:25-32.

(编辑　侯　湘)