浅析大数据安全分析技术与在网络安全领域中的应用

胡佳茵

2020 年，黑客入侵 SolarWinds 公司的一款知名网络监控软件（Orion），“轻松”访问了 SolarWinds 公司的 18000 个客户计算机网络，其中包括多个美国政府机构网站在内，导致美国政府自身因网络防御能力不足而受到强烈谴责。据安全和漏洞分析机构 Risk Based Security 数据显示，截止 2020 年上半年，数据泄露事件导致 360 亿条记录在网上曝光；据 IBM 数据显示，数据泄漏的平均成本为 386 万美元、发现数据泄漏的平均时间为 207 天、从发现到处理一次数据泄露平均周期为 280 天；据思科披露，全球分布式拒绝服务攻击（DDoS）攻击总数到 2023 年将高达 1540 万。类似惊人的数据比比皆是，揭示了网络安全问题所带来的危害性。传统网络安全措施感知与预测能力有限，远远落后于日趋复杂的网络环境，而基于大数据内涵与特性所形成的大数据安全分析技术能够解决网络安全问题。

该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，能够解决传统网络安全问题。本文着重介绍大数据安全分析技术中的三个常用手段，即安全可视分析技术、安全事件关联分析技术和用户行为分析技术。

安全可视化分析技术有助于加强网络态势感知

1、安全可视化分析技术概述

数据可视化技术可以协助分析者洞悉数据背后所蕴藏的信息，而网络安全数据以可交互的图形图像的方式表现出来，分析人员借此观察网络安全数据中隐含的信息，从而感知网络安全问题。数据可视化技术能够对某一领域数据进行全周期研判和多维可视分析，辅助用户掌握不同阶段数据变化规律，因此数据可视化工具功能愈发完善、研究理论愈发扎实。

从学者研究层面讲，国内外诸多研究学者早已对这一交叉领域展开研究，通过不断丰富理论研究，进一步降低分析人员的认知难度，分析挖掘问题的效率得以提升。Richard A. Becher 于 1995 年首次提出网络数据信息可视化的概念，即对网络流量状况进行可视化，通过利用地图结合点连接图对各地网络通讯流量进行可视化展示。航天工程大学赵立军提出了围绕“熵”的堆叠条形图设计方法和基于平行坐标的安全可视化方法，并将总图和细节有机结合，进一步降低了分析人员的认知困难。西南科技大学吴亚东提出了异构树网络安全数据组织方法，并设计了三维多层球面空间可视化模型，不仅增强了分析系统的可交互性，也提高了网络安全数据的分析效率。

从业内聚焦层面讲，专业组织定期举办行业研讨会，通过集中探讨方式不仅使该领域得到重视，也为构建网络安全良性生态提出优化解决方案。自 2004 年起，电气与电子工程师协会（IEEE）每年都会举办网络安全可视化研讨会（The international symposium on visualization for cyber security，VizSec），这是专注于安全领域可视化技术的国际会议。会议期间，多位专家针对越来越复杂的网络环境、庞大的网络数据和出其不意的攻击手段提出相关见解。

从市场产品层面讲，市面上出现了一批创新性的可视化系统和工具，帮助用户提高威胁监测效率，进行全方位态势监控。以美国Palantir公司为例，其推出的 Palantir Gotham 软件通过对结构化数据和非结构化数据的高度集中分析，能够得到可视化的关联线索信息。该款软件最吸引市场的两个亮点在于，1）这是一款情报分析软件，核心逻辑为是采用动态本体管理器的思想。2）并非一个单纯的数据解决方案，其强大的灵活性和兼容性，能够满足任何机构对海量数据的处理需求。

经分析：1）当前数据可视化产品有别于传统产品，不再是简单的数据提取和抓取，而是将行业特点和可视化技术相结合，打造一个集数据监控、采集、智能分析等功能于一体的平台，从而实现精细化管理、智能研究与判断；2）数据可视化是一场用户与数据的“视觉对话”，但是以往的可视化工具需要通过完成极其复杂的编码过程才能实现可视化目标，当下很多诸如 Data Focus 这样的数据可视化工具可以帮助用户省去编码过程，直接实现无代码可视化。

2、加强网络态势感知，增强网络安全防御能力

网络安全态势感知（network security situational awareness，NSSA）通常被定义为在大规模网络环境下，对影响网络态势的安全要素进行察觉提取、理解评估和预测研判。该技术的出现，将网络安全的事后补偿机制转变为事前预测与评估，将网络攻击造成的风险尽可能降到最低。具体来说，网络安全要素提取是从单一要素与多源数据两个层面采集网络环境中的相关要素状态、属性和动态等信息；网络安全态势理解基于要素提取，并处理海量数据，旨在帮助分析者和决策者以更高维的视角理解网络状态，且不被复杂的表象所困惑；网络安全态势评估态是在获取、变换及处理历史和当前态势数据序列的基础上，通过建立数学模型，探寻演变规律，对未来发展趋势和状况进行推理。

在网络安全态势感知的全过程中，安全可视化分析技术起到了重要作用，不仅提高了人机交互的能力，还能提升研究人员的感知能力，从而进一步挖掘海量数据背后的网络态势。研究学者目前将两种逻辑与可视化平台相结合，并进行了相关的理论研究，一是打造以事件和流量作为基本粒度构造可视化平台，如针对多变量数据创建了基于像素图的网络安全可视化平台，利用直观数据直接指出可疑活动区域以帮助调查未知网络安全事件，还可以采取网络安全警报和流量解释为集合的方式，建立网络集成可视化系统；二是引入排序分析思想，可将基于 Lucene 的改进版搜索结果排序算法结合多种可视化方法，融合到前后端分离的网络安全数据可视化系统的开发设计中，或打造一个自顶向下的网络流量时序分析流程模型，以该模型为架构，设计并实现了一个多视图合作的网络流量时序数据可视分析原型系统。

安全事件关联分析技术有助于防止 APT 攻击

1、安全事件关联分析技术概述

网络安全事件呈指数级增长，且相互之间的关系错综复杂。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，寻找内在关系，去掉冗余信息后还原攻击事件始末，旨在及时发现网络攻击者的入侵行为。

近些年来，国内外研究学者提出了多种网络安全事件关联分析方法，如根据关联类别，可以划分为基于属性特征的安全事件关联分析技术，回归事件本身，分析属性特征之间的关联特性，并以此为基础配置关联策略，根据事件属性对其进行匹配检测；基于逻辑推理的安全事件关联分析技术，从事件关联关系入手，通过运用一定的理论知识进行推测，从而有效解决复杂问题；基于概率统计的安全事件关联分析技术，基于事件发生概率和统计数据角度出发，对所预警的众多信息关系以概率形式进行描述，从而揭示网络安全事件的时序和因果关系；基于机器学习的安全事件关联分析技术，通过数据挖掘和机器学习的方法训练数据集，生成事件关联规则，最后得到新型攻击事件模式，具有可实时运行的特点。

上述安全事件关联分析方法所应用的场景不同，但均能不同程度地将相对孤立的低层级网络安全事件数据集进行关联整合，通过过滤、聚合等手段发掘这些事件真实的关系，最终生成高层级的安全场景。

2、有效检测 APT 攻击，让网络攻击有迹可循

近几年等 APT（高级持续威胁）攻击活动发生了多个维度的变化：一是与现实社会关联，不再是单纯技术型攻击行为；二是与黑色产业挂钩，利益链条错综复杂；三是交叉学科，网络攻击日益复杂化。针对上述特点，学者认为扩大空间和时间范围是检测 APT 攻击最有效的技术解决方案之一。由于安全事件关联分析可以对海量网络安全数据进行深度关联分析，因此在检测 APT 攻击方面具有明显优势。若对安全事件的攻击时序和地域没有进行充分关联分析，那么研究人员针对还原、预判和防御 APT 攻击活动就缺乏了强有力的依据支撑。

截至目前，市场上的开源安全事件关联分析软件大致可以分为单纯日志监控软件和完整的安全管理功能软件两类。据网络安全研究人员表示，尽管 APT 攻击活动所使用的恶意代码变种速度快、传播范围广，但究其根本，恶意代码背后的命令控制通道通信模式并未发生频繁变化。因此，针对相对固化的命令控制通道通信模式，开源安全事件关联分析软件均会制定程序脚本，并制定相应的规则，再读取某个文件输入流后，根据规则进行相应的匹配，从而查找与规则相适应的输入流。为了提升关联分析效率和整体安全检测能力，很多开源安全事件关联分析软件还会添加一些额外的工具，除了日志分析，还提供文件完整性检查、策略监控、实时报警、联动响应等多种功能，可进一步适应多场景。此外，这些开源软件还可叠加威胁情报、攻击链追溯等功能建立分析模型，从而完成安全事件状态评价、安全事件趋势预测、安全事件风险预警等。

用户行为分析技术有助于检测网络异常情况

1、用户行为分析技术概述

用户行为分析（user behavior analytics，UBA）主要是用于防止“内鬼”，即检测出内部威胁，后被 Gartner 顶级科技峰会后将该技术更名为用户和实体行为分析（user and entity behavior analytics，UEBA），并预测 2022 年 UEBA 技术会应用于 80% 网络安全检测活动中。该技术通过使用高级数据分析用户的行为，并建立用户的行为基线，以便发现用户的异常行为。所谓基线，其主要包括用户活跃时间、使用服务类型、使用服务的频率等等。该技术最早应用于电商领域，通过统计和分析用户访问电商网站的相关数据，形成用户的线上购物基线，探究用户的消费偏好，最终实现精准营销。

随着企业员工流动性增大，导致企业内部威胁愈发严重，如一名花旗银行员工被辞退后，出于报复心理，恶意删除了北美地区控制中心的 10 个路由器上的相关数据；一名 Google 自动驾驶汽车研发部门的员工，在离职前违规下载项目工程文件，并依靠该商业机密创业；曾就职于马印航空下属某研究中心的两名员工，在离职前盗窃了数百万乘客的信息，再将这些信息提供给新就职公司。《2020 Securonix 内部威胁报报告》指出，研究人员根据 300 多个案件分析后发现，从泄漏的内容看，62% 的内部威胁涉及到敏感数据；从泄漏的手段看，44% 的数据泄漏通过电子邮件完成、16% 的通过云存储网站完成、10.7% 的由员工下载后自行带走、8.9% 依靠移动介质拷走。总体来说，云窃取数据威胁增加，且距离正式离职的前半个月或两个月为内部威胁发生的高峰期。

2、全方位异常检测，做网络隐形守护者

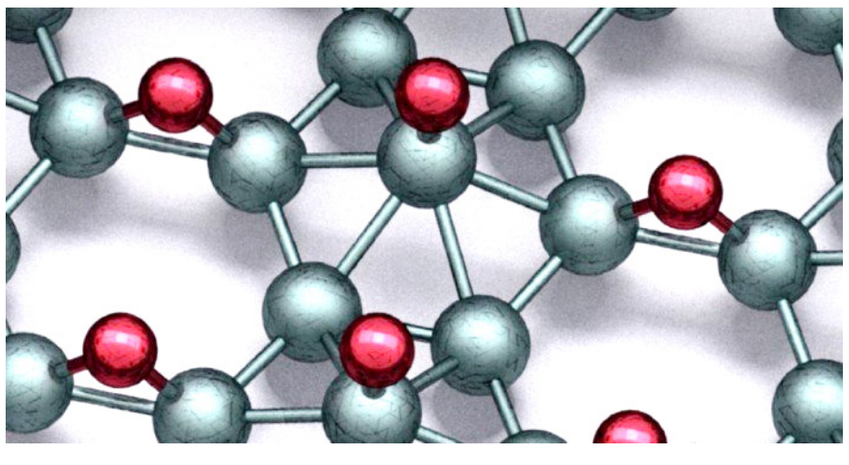
网络异常检测内容包括流量突变、设备失效、越权资源访问、可疑主机等，其原理为构建检测模型，对违背策略或偏离正常行为模式的行为进行判定。近些年来，网络用户行为分析极大提升了当前网络异常检测的准确率。

从国内外研究现状来看，对于检测网络用户异常行为现有大量的研究成果。国内学者陈兴蜀、胡洋瑞、夏景明、梁杰等学者针对提升检测率而提出了多个检测方法，如改进版 k-means++ 余弦聚类算法、结合高斯混合模型聚类算法和随机森林的网络入侵检测方法、基于原始流量数据包的深度学习方法、无监督聚类算法等，这些方法有效结合了多种技术，漏检率和误检率不仅有效降低，检测结果以可视化形式呈现，大幅提升了检测工作效率。国外学者致力于基于大数据的网络用户异常行为检查平台的研究，从 Hadoop 检测转变为 Spark 技术，从而识别更隐蔽的网络异常情况。当然，国外不少学者基于 Spark 技术创新了理论，W. He 构建了异常行为分析框架，经过数据挖掘、机器学习与统计分析技术三步骤，快速分析大规模日志数据，准确识别异常情况；H. Zhang 提出一种实时分布式随机森林网络入侵预防系统，具有实时检测和速度快等特点；Dobson A 利用 Spark 库实现传统机器学习算法以及多层感知器深度学习算法的网络入侵检测。

当前，用户行为分析被广泛应用于高校网络建设。众所周知，高校网络访问频繁、应用繁多，保存了从教职工到学生、从上网记录到教学记录等庞大信息。通过用户行为分析，可优化网络防御系统，及时预警异常行为，提升校园网络安全性能。

结束语

本文阐述了大数据安全分析技术中的安全可视化分析技术、安全事件关联分析技术和用户行为分析技术这三个手段的基本概念，并基于维护网络安全介绍了上述三个手段的理论研究、应用情况、发展趋势等。尽管当前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一定应用，但网络安全形势依旧不容乐观，如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等。因此，为了加强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，大数据安全分析技术应需不断完善，才能有效检测网络威胁，保证网络安全稳定运行。



科学家发现新型材料氢化硼烯

二维材料科学的世界是一个令人激动的世界，新的进展不断开辟新的可能性，其中最受关注的莫过于神奇材料石墨烯。不过现在来自美国的科学团队找到了一个新的玩家—氢化硼烯，他们认为这种材料在未来先进的电子技术中会得到广泛应用。

这种新材料是西北大学、佛罗里达大学和阿贡国家实验室的研究人员共同合作的成果，是从厚度只有一个原子的硼片开始的。这种单层硼烯于 2015 年首次合成，此后形成了其他新型材料的基础，如球形笼状的硼磷烯。

由于具有很好的强度和灵活性，在从电池到传感器技术的一切领域中，硼烯都有广泛的应用前景，但它也存在非常多的缺点。其中最大的难点就是很难生产，而且会与开放的空气发生负面反应，很快就会氧化，变得非常不稳定，而且很容易改变形状。

科学家已经找到了克服这些障碍的方法。这种新型材料是通过在银基底上生长硼烯，然后再将其暴露在氢气中，从而产生复杂的、两原子厚的硼原子和氢原子排列。这种结构是通过计算机视觉和扫描隧道显微镜成像的结合来确认的。

西北大学材料科学与工程教授 Mark Hersam 表示：“单层硼烯本身就存在非常多的问题。但当我们将硼吩与氢气混合时，产品突然变得更加稳定，对于在新兴的纳米电子学和量子信息技术领域的使用更具吸引力”。

科学家们列举了一些硼烯可能产生影响的领域，包括太阳能电池、电池和电子设备。他们表示，它的特性可以特别适合于依靠光来获得卓越性能的电子产品，这个领域被称为光电子学，包括从更高效的计算机芯片到先进的 LED 显示器的一切。

阿贡纳米材料中心的纳米科学家 Pierre Darancet 表示：“从我们的结果来看，真正令人鼓舞的是，我们发现银基底上的硼烯纳米片相当稳定，和单层硼烯不同。这意味着它应该很容易与其他材料集成，用于构建新的光电子器件，即光与电子相结合的设备”。

（逸文）