面向网络大数据的安全分析技术应用

汪来富，金华敏，刘东鑫，王帅

（中国电信股份有限公司广州研究院，广东 广州 510630）

摘　要：大数据分析技术的蓬勃发展，给安全行业带来了许多新的思路和发展机遇。从电信运营商视角，深入解析了面向 Netflow、DPI、DNS 等网络大数据资源的大数据安全分析平台的架构、技术实现机制等，并介绍了大数据安全分析产品的相关功能和应用场景。

关键词：大数据；安全分析；攻击检测

中图分类号：TP393.08　　**文献标识码**：A

doi: 10.11959/j.issn.1000-0801.2017061

Application of security analysis technology for network big data

WANG Laifu, JIN Huamin, LIU Dongxin, WANG Shuai

Guangzhou Research Institute of China Telecom Co., Ltd., Guangzhou 510630, China

Abstract: Big data technology and solutions have been continuously booming for several years, which has brought much innovation ideas and opportunities for security analysis. From the perspective of telecom operators, the architecture and key technology of the big data security analytic platform were analyzed, which were based on the network big data including Netflow, DPI, DNS and so on. At last, some related function and service scenarios of big data security analytic services were introduced.

Key words: big data, security analysis, attack detection

1　引言

在当前万物互联的时代，各类信息应用日益丰富，与安全相关的各类数据呈指数级增长趋势，数据来源丰富、内容更为多维、种类繁多。而具有目标性强、长期潜伏渗透特性的 APT（advanced persistent threat，高级持续性威胁）攻击更是让传统的安全分析技术防不胜防。因此，在当前不断发展的安全形势下，传统的基于特征匹配的安全防护技术难以起效，各类安全威胁更具杀伤力和逃避力。在此背景下，数据驱动安全已逐渐成为业界共识，而大数据技术的出现，则为其落地和发展奠定了技术基础。

大数据技术可实现大容量、低成本、高效率的数据分析能力，满足海量安全信息的处理和分析需求，将大数据技术应用于网络安全分析领域已日趋成熟，由此催生了大数据安全分析产业的快速崛起，并对网络安全技术发展带来深远的影响。大数据安全分析技术是指将大数据技术应用到网络和信息安全领域，通过采集、存储、挖掘和分析流量、日志、事件等与安全相关的各类网络行为数据，从更高视角、更广维度上发现异常、捕获威胁，实现对异常行为、未知威胁的早期检测和快速发现。与传统安全分析技术相

收稿日期：2017-01-13；修回日期：2017-02-28

比，大数据安全分析技术具有以下两个重要特征。

- 基于海量异构数据存储与快速计算处理能力，可拓展安全分析与监控数据源的广度和深度，有助于发掘更为隐蔽的安全威胁。
- 可在更长时间窗口内对多维度数据进行深度回溯和关联分析，有助于快速发现异常行为或未知安全威胁。

2　大数据安全分析应用

从应用主体的维度来划分，目前积极引入大数据安全分析技术的主力军包括互联网安全公司、传统安全厂商和电信运营商，因安全理念、原有产品体系以及对业务流、系统数据、日志等资源掌控能力的不同，其应用重点和技术实现也存在较大差异。

2.1　互联网安全公司

新兴互联网安全公司不受传统产品线的束缚，主要将大数据技术应用于威胁发现领域，在云端通过多纬度跨域分析、深度数据挖掘和人工智能技术对海量数据进行深度分析，以实时获取未知安全威胁的发展动态，通过构建完善的威胁特征库，提供对未知安全威胁的解决方案。互联网安全公司基于其拥有的海量样本库、日志以及与各类恶意行为相关的漏洞、网址、域名等信息，可以支持未知威胁发现所需的存储、搜索、挖掘、机器学习等资源。互联网安全公司一般将大数据安全分析平台作为基础安全能力平台，一方面为其他产品提供基础安全能力，另一方面也面向对 APT 攻击敏感的企业客户以及有特殊安全需求的政府机构或相关单位进行定制开发，以满足其个性化的高等级安全需求。

2.2　传统安全厂商

传统安全厂商，尤其是 SIEM/SOC 厂商，引入大数据安全分析技术的初衷是因其传统的集中化安全分析平台在处理、分析海量异构数据存在性能瓶颈，传统的基于规则和特征的分析引擎在未知安全威胁面前无能为力，因此其主要应用大数据安全分析技术对 SIEM/SOC 进行改造和重塑，侧重于提升 SIEM/SOC 安全分析平台的分析处理能力，以提供更具竞争力的整体安全解决方案。在技术实现上，传统安全厂商主要利用大数据的海量信息采集和处理能力，实现对海量异构数据的准实时分析、各类安全事件的快速回溯和取证以对安全报表的快速统计、查询和可视化呈现等。

2.3　电信运营商

电信网络作为关乎国计民生的基础通信设施，其自身安全保障及安全能力建设是国家网络空间安全战略的重要环节。在网络安全能力体系中，安全风险的快速检测和早期预警是提升基础通信设施安全防护水平的关键要素。随着网络应用的全社会化渗透，网络安全分析的数据规模将不断增大、数据来源也日益丰富，重点业务、关键网络节点更提出了实时性防护要求，而传统安全分析方法难以满足新形势下提出的安全检测需求，大数据安全分析技术则为解决这些问题提供了有利契机。

在电信运营商领域，数以亿计的客户量决定了其庞大的网络规模和不断拓展并日益复杂的业务系统，由此带来的是海量、异构、多变、低密度价值的网络大数据，其拥有全网 Netflow、重要链路 DPI、DNS 数据等重要的网络大数据资源，在开展大数据安全分析服务方面具有先天优势。引入大数据安全分析技术，保障电信网络运营安全，开拓新兴大数据安全分析和安全检测业务，是电信运营商健全网络安全防护能力、提升网络核心价值的必备选择。鉴于上述原因，电信运营商积极开展大数据安全分析技术研发实践，融聚大网多维安全数据资源，以期将丰富的网络大数据资源转化为强大的安全服务能力。

本文基于电信运营商视角，提出面向网络大数据的大数据安全分析平台，系统地阐述该平台的技术架构和主要功能模块技术实现机制，并简要分析基于该平台的大数据安全分析产品业务功能与应用前景。

3　大数据安全分析平台架构

该平台采用基于 Hadoop 平台的分布式存储与计算框架，实现对现网各类安全大数据的融合关联分析与可视化展示，通过建模分析 DDoS 攻击、僵尸网络/恶意域名、Web 攻击等安全事件及数据间的关联关系；实现对网络安全状况的深度感知，为企业用户提供网络安全分析及预警服务，并为网络运营管理提供可视化的安全分析工具及分析报表。

该平台采用分层架构，自下而上分为 4 层，依次是数据采集层、数据存储层、数据计算分析层和数据呈现层，具体架构如图 1 所示。

平台采集的数据源包括 Netflow、DPI、DNS 等多维大网数据，同时支持以 Flume/syslog 等方式采集的客户端数据。数据采集层主要完成数据的泛化处理和标准化处理，

图 1　大数据安全分析平台技术架构

然后进入数据存储层。数据计算层是平台的核心能力模块，按功能维度分为攻击溯源分析模块、僵尸网络/恶意域名分析模块、Web 安全分析模块和客户安全分析模块。数据呈现层主要完成分析结果的可视化呈现，分为管理门户和客户门户，其中管理门户主要为后台运维人员提供安全分析视图、可视化的安全分析手段和数据钻取工具；客户门户则是大数据安全分析产品的载体，客户通过登录自服务门户查看自身安全状况、安全事件等各类数据报表。

同时，平台基于上述各安全分析模块，可输出“肉鸡”/疑似“肉鸡”、CC 控制端/疑似 CC 控制端、恶意 URL 等数据分析结果，构建现网第一手威胁情报库资源，并可进行持续滚动分析和动态更新。

3.1　攻击溯源模块

攻击溯源模块实现的主要功能是对现网各类 DDoS 攻击进行深度挖掘、精细化分析和可视化呈现，功能框架如图 2 所示。其实现机制是通过采集大网路由器层面的 Netflow 数据和 DDoS 攻击事件等数据信息，结合网络拓扑信息、路由器接口信息等数据，通过数据关联和统计分析，实现对 DDoS 攻击流量的深度分析、精准溯源和可视化回溯。

该模块根据流量分析系统检测以及系统自身分析发现攻击，结合采集存储原始 Netflow 流量信息、路由器端口信息、路由拓扑信息、城域网 IP 地址库等多维参数进行关联分析，通过基于多元广度遍历算法，快速全景回溯攻击流量穿越路径及流量分布特征，可对互联网发生的网络攻击进行实时监测、溯源及攻击路径重演，解决了伪地址攻击溯源难题，并大幅提升攻击溯源分析效率，具体方案如图 3 所示。该技术方案监控范围大、智能性高、灵活快速，

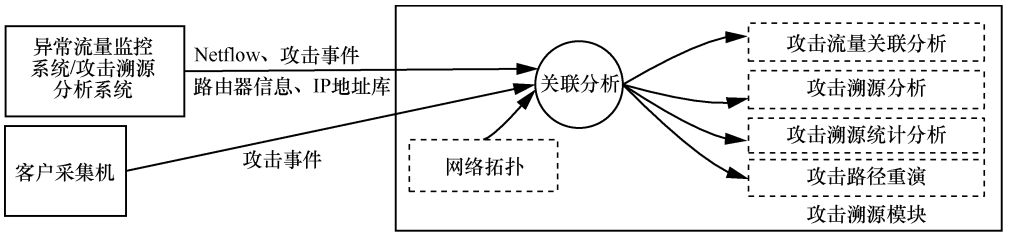


图 2　攻击溯源模块功能框架

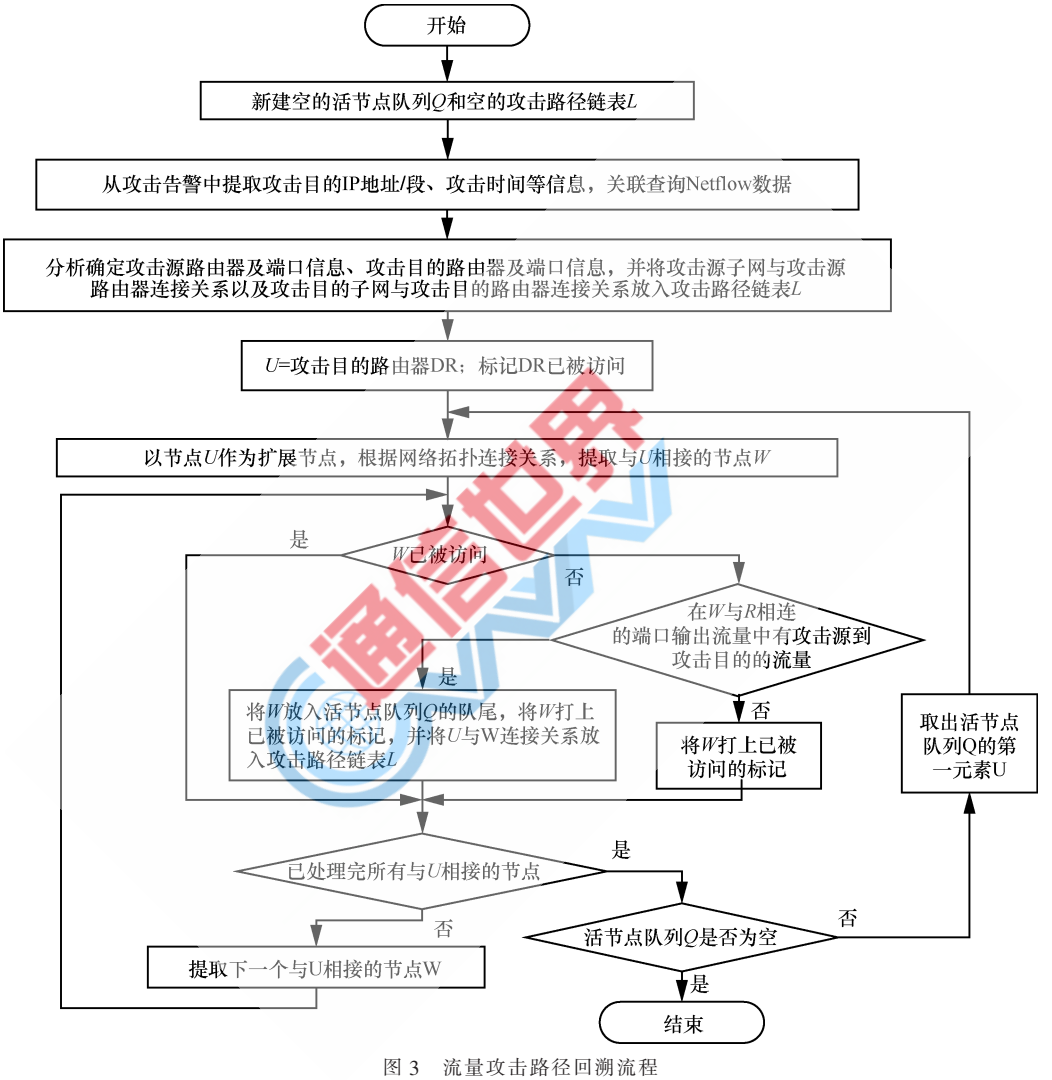


图 3　流量攻击路径回溯流程

不需过多人工参与分析攻击源和攻击路径，能够在攻击发起初期就进行攻击发现和抑制，不但能直观显示攻击源，而且可对攻击流量穿行路径进行可视化分析，有效提升 DDoS 攻击应急响应处理效率。

3.2　僵尸网络/恶意域名分析模块

从 Zeus、Cutwail 等著名僵尸网络的例子来看，一个大型僵尸网络的构建往往代价不菲，并且需要一定的时间。在僵尸网络生命周期的“传播—感染—加入—受控—攻击”等阶段，几乎都存在 CC 控制端和“肉鸡”的交互行为。在数据驱动安全的理念下，只要有一个保持及时更新、恶意 IP 地址/域名足够丰富的安全威胁情报库，就可对僵尸网络做有效的检测和控制。

在本系统中，僵尸网络的检测分析包括定位 CC 控制端的 IP 地址、发现 CC 控制端所使用的域名和定位“肉鸡”的 IP 地址。首先，从已部署的“僵木蠕检”测系统、攻击溯源系统和移动互联网恶意程序监控系统等安全系统中归

并生成相关的恶意 IP 地址、恶意域名等安全情报；进一步地，在监控链路中部署 DPI 系统、采集 Netflow 数据流；最后，根据已生成的安全情报信息对 DPI 日志、Netflow 数据流进行关联匹配，可以检测得到“肉鸡”和疑似“肉鸡”的 IP 地址列表。僵尸网络检测分析处理流程如图 4 所示。

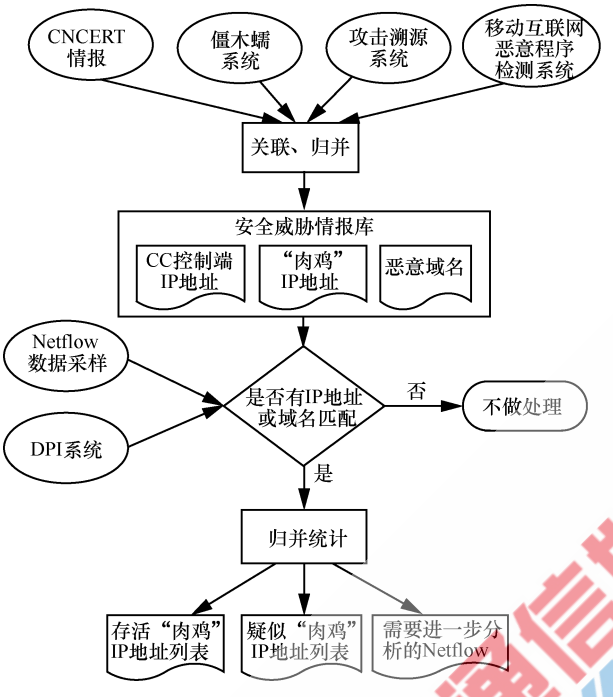


图 4　僵尸网络检测分析处理流程

其中，在 IP 地址或域名匹配检测中，当一个 Netflow 数据流中发现与 CC 控制端 IP 地址通信的流量，如果在一定的时间窗口（例如 30 min）内，Netflow 条数大于一定阈值 N，则 Netflow 里的另一个 IP 地址可以判断为存活“肉鸡”；如果 Netflow 条数小于阈值 N 而落在一个区间 $[M,N)$，则 Netflow 里的另一个 IP 地址可以判断为疑似“肉鸡”。值得注意的是，阈值 N 可以根据统计结果做设置，降低运维人员工作负担。相比之下，如果在一个 Netflow 数据流中发现“肉鸡”的 IP 地址，但是另一个 IP 地址既不是已知的“肉鸡”，也不是 CC 控制端，那么对于这个 IP 地址的判断应结合后续的 DNS 等数据做进一步分析。

DNS 数据分析已经成为僵尸网络检测的重要入口。为了躲避追踪、延长生命周期，大部分僵尸网络会采用 fast-flux 技术，频繁变换 IP 地址，而僵尸网络内部的通信可通过 DNS 查询，获取到最新、及时、有效的 IP 地址。这些关键网络行为特征总结如下。

- 域名频繁变换 IP 地址。
- 所频繁变换的 IP 地址地理归属地差异较大。
- 域名服务器有多个 IP 地址，且跨多个 ASN。
- whois 信息不完整。
- 域名的注册 E-mail 地址曾经以恶意域名注册人出现过。
- 域名通常较长并且字符随机。

基于以上关键特征定义，可以对 DNS 流量日志做挖掘分析，得到恶意域名、恶意 IP 地址等安全情报。依据 DNS 流量日志的分析结果，对僵尸网络检测分析中需要进一步分析的 Netflow 数据，提取与“肉鸡”通信的 IP 地址，可以在 DNS 流量日志中做进一步的匹配分析，以确定该 IP 地址是否为 CC 控制端或者其他“肉鸡”。

3.3　Web 安全分析模块

Web 安全模块的主要功能是对针对 Web 网站的攻击行为进行检测和统计分析。其实现机制是通过采集 DPI 数据，分离出 HTTP 会话文件，将文件数据通过元数据提取、数据分析、数据挖掘及事件呈现在具体检测方法上，主要通过行为特征库的匹配和 Web 入侵规则检测，基于正则表达式等方式，实现对各类 Web 攻击行为的检测和识别，并进行可视化呈现。

下面以 XSS 攻击检测为例阐述其具体实现机制。首先 XSS 跨站脚本攻击监测模块从数据仓库中提取元数据，对元数据的内容进行抽取、解析，从中获得待监测 Web 系统的 URL，并对其进行提取和还原；然后再结合 XSS 跨站脚本规则库去比对和发现该 Web 系统中的 XSS 跨站脚本漏洞；最终将所获得的 XSS 跨站脚本攻击分析结果队列存储到内存数据库中。此外，还需要结合如图 5 所示的控制流图（CFG 图）对 URL 页面进行静态分析检测，获知并保存当前 URL 页面所有存在的 XSS 漏洞位置信息，为下一步识别 XSS 跨站脚本有效攻击做准备。

对于 XSS 跨站脚本攻击同样也需要区别有效攻击和一般攻击。首先比较当前 URL 的 XSS 注入点和之前通过源码静态检测得到的该 URL 页面的 XSS 漏洞注入点，如果其 XSS 漏洞的注入点位置相同，则判定为一次有效的 XSS 跨站脚本有效攻击。然后再检测返回的响应报文，根据返回的响应报文反馈信息判断是否有注入点不相同的 XSS 漏洞攻击成功，若成功，则对应的 XSS 跨站脚本攻击也是一次有效的攻击。

3.4　用户安全模块

客户关联分析模块的主要功能是从大数据安全分析

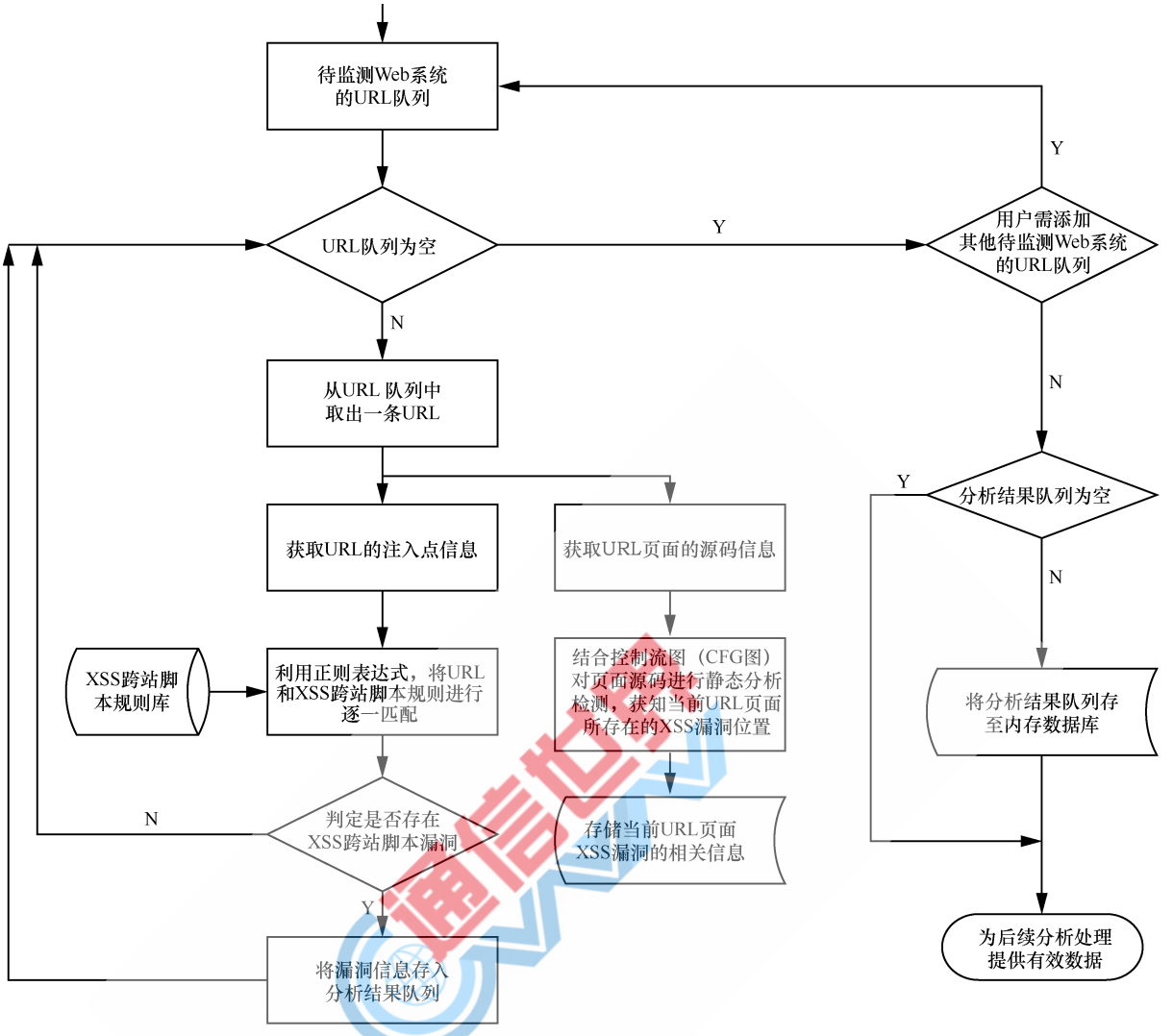


图 5　XSS 跨站脚本攻击监测分析流程

平台中分析、提取与客户资产相关的各类安全事件、网络行为日志和相关安全数据，其主要机制是将客户资产信息以及各类网络行为信息和平台分析出来的各类安全事件、分析结果和知识库进行自动关联和自动匹配，从而为用户安全状况分析和安全报表提供原始的数据资源。

该功能模块需要基于客户的监控 IP 地址、站点域名、报表查询条件等数据，从 DDoS 攻击统计数据、Web 应用层安全数据、“肉鸡”统计、CC 控制端统计等分析结果数据以及 DDoS 攻击事件、僵木蠕事件、恶意程序事件等原始事件数据中统计与客户相关的安全事件信息。例如，以客户 IP 地址为索引，从平台分析结果中检索其是否为“肉鸡”或 CC，是否发起过 DDoS 攻击，攻击时间段和攻击流量大小以及从原始流数据中检索其是否访问过恶意 URL 等。这些匹配的数据都将作为客户安全状况分析和安全报表具体的数据来源。

综上所述，该平台通过融聚电信大网数据资源与客户数据，基于大数据技术进行存储、挖掘与可视化展示，实现安全态势分析、安全威胁与异常检测等基础安全能力，并通过构建威胁情报库等方式，实现安全能力开放。

4　面向 SME 的大数据安全分析产品

当前以 FireEye 公司 Threat Analytics Platform 等为代表的大数据安全分析产品，主要面向政府、金融等高端目标客户，分析的数据源以客户自身网络侧的流量数据、日志数据为主，侧重于 APT 攻击检测和未知威胁发现，具有较高的技术门槛和商用门槛。

本平台依托大网 DPI、DFI、DNS 等海量数据资源，具有覆盖范围广、运行成本低等特点，具备强大的集约化优

势。基于该平台面向网络大数据的安全分析能力，通过和客户资产的关联匹配实现用户安全状况的快速感知，可为用户提供持续安全监测、安全预警和深度安全诊断服务。

(1)安全监测

基于大数据分析和威胁情报等技术，为用户提供长时间周期的持续安全监测服务，通过安全事件等信息的主动呈现，协助用户主动发现企业内网已经发生和正在发生的安全威胁。

(2)安全预警

基于大网 DPI、Netflow、DNS、僵木蠕等多维海量安全数据，进行自动挖掘分析，提供安全态势分析、安全威胁等预警服务。

(3)深度安全诊断

结合用户内网业务流、日志等数据，进行深度安全威胁分析和安全评估，提供专业安全分析报告和方案建议。

该产品依托运营商服务渠道资源优势，基于平台运营模式，可为全网用户提供低成本的网络安全体检服务。该产品主要面向企业用户，尤其是没有专业安全的运营团队和缺乏安全分析能力的中小企业用户(SME)，通过为其提供具有普遍服务性质的网络安全体检服务，一方面可以提升电信运营商传统宽带产品的用户黏性，另一方面也可带动其他专业安全服务的推介和推广，具有良好的市场发展空间。

5 结束语

数据驱动安全成为安全业界的发展共识，而大数据安全分析技术则是体现数据驱动安全这一理念最重要的技术应用形态，它将对安全产业产生非常深远的影响。在大数据蓬勃发展的时代，电信运营商拥有天然的大数据资产，随着技术壁垒的打破、管理模式的变革和越来越多的业务创新，大数据安全分析平台将成为运营商精细化安全管理和安全数据运营的重要支撑平台。

参考文献：

[1] 王帅, 汪来富, 金华敏, 等. 网络安全分析中的大数据技术应用[J]. 电信科学, 2015, 31(7): 139-144.
WANG S, WANG L F, JIN H M, et al. Big data application in network security analysis[J]. Telecommunications Science, 2015, 31(7): 139-144.

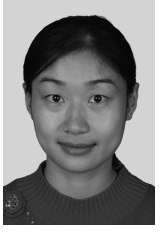
[2] 程学旗, 靳小龙, 王元卓, 等. 大数据系统和分析技术综述[J]. 软件学报, 2014, 25(9): 1889-1908.
CHENG X Q, JIN X L, WANG Y Z, et al. Survey on big data system and analytic technology [J]. Journal of Software, 2014, 25(9): 1889-1908.

[作者简介]

汪来富(1976–)，男，中国电信股份有限公司广州研究院高级工程师，主要研究方向为大数据安全、云计算安全、网络安全。

金华敏(1972–)，男，中国电信股份有限公司广州研究院高级工程师，主要研究方向为 IP 网、云计算、大数据安全、网络安全。

刘东鑫(1985–)，男，中国电信股份有限公司广州研究院工程师，曾获得 CCIE、CISSP 和 CISA 等认证，主要研究方向为网络与信息安全、大数据安全。



王帅(1979–)，女，中国电信股份有限公司广州研究院高级工程师，主要研究方向为大数据安全、云计算安全、网络与信息安全体系及攻防技术。