基于 ELK 流量分析平台在高校网络安全管理中的应用

秦 锐[1] 袁毅哲[2] 秦道祥[3]

[1](同济大学电子与信息工程学院 上海 200092)

[2](同济大学软件学院 上海 200092)

[3](同济大学信息化办公室 上海 200092)

摘 要 网络安全设备普遍存在误报率高、难以验证的问题,对于已发生网络安全技术事件,也缺乏追踪和溯源的手段。基于 ELK 流量分析,可以实时采集校园网的流量并对数据进行分类。在 ELK 数据平台进行分布式存储,从多个维度建立分析视图,以不同的业务场景和图形方式进行展示。该方案可以发现校园网中的访问攻击并相互印证,为网络安全防御体系建设提供新的技术方案。同时可以提供访问日志审计为网络故障处理提供支持,提升校园网运维和信息安全管理水平。

关键词 ELK 网络安全 流量分析 大数据

中图分类号 TP391 **文献标识码** A DOI:10.3969/j.issn.1000-386x.2019.06.057

APPLICATION OF ELK TRAFFIC ANALYSIS PLATFORM IN UNIVERSITY NETWORK SECURITY MANAGEMENT

Qin Rui[1] Yuan Yizhe[2] Qin Daoxiang[3]

[1](*School of Electronics and Information Engineering*, *Tongji University*, *Shanghai 200092*, *China*)

[2](*School of Software Engineering*, *Tongji University*, *Shanghai 200092*, *China*)

[3](*Information Office*, *Tongji University*, *Shanghai 200092*, *China*)

Abstract The problem of high false alarm rate and difficult to verify exists in network security equipment. There is also a lack of traceability for the network security technology event has occurred. Based on ELK traffic analysis, we could collect and classify the traffic of campus network in real time, store the data in ELK data platform in a distributed way, build analysis views from multiple dimensions, and display them in different business scenarios and graphics. This scheme can discover the access attacks in the campus network and prove each other, and it provide a new technical solution for the construction of network security defense system. It also provide access log audit to support network fault handling and improve the level of campus network operation and information security management.

Keywords ELK Network security Traffic analysis Big data

0 引 言

高校信息系统建设发展了多年,积累了大量的师生、教学、科研、管理方面的数据,并且网站的访问量高,网络攻击者趋之若鹜,攻击每时每刻都在发生。随着对安全管理要求的提高,在校园网上部署防火墙、下一代防火墙、Web 防火墙、流量控制、防病毒软件、VPN、IPS 等安全软件和设备。这些安全防护设施在日常的管理中能发现各种各样的攻击并能够对部分攻击进行拦截和防护,对于高校网络安全管理起到一定的帮助。但是,由于高校的安全设备都是不同时间购买,由多个品牌组成,且这些设备的防护规则库和防护策略是由各厂家自行定义,它们独立工作、信息不能共享,安全设备误报率高的也是大家的共识,对于发现的大量的问题多数高校管理者会进行忽略处理,导致网

收稿日期:2018-09-22。**秦锐**,硕士生,主研领域:网络,分布式计算。**袁毅哲**,硕士生。**秦道祥**,高工。

络安全事件时有发生。如何验证安全设备发现的安全问题是摆在高校网络安全管理者面前的难题。此外网络安全设备配置的存储有限,安全厂家默认只对触发防护规则的行为进行记录,一旦发生网络安全技术事件,不能为追踪和溯源提供充分的信息。校园网管理部门需要一种高效、灵活的数据分析方式,为问题定位、故障处理、攻击分析、信息安全事件溯源提供更好的支持。本文提出的基于流量分析的 ELK 大数据分析平台,可以对校园网的流量进行实时监控,发现校园网中的异常攻击行为,为高校网络安全管理提供了很好的技术支持。

1 ELK 流量分析平台架构

ELK 是 ElasticSearch、Logstash 和 Kibana 三个开源工具组合,支持实时数据存储、检索和分析。ELK 平台能处理多种格式数据,配置简便,集群分布式部署易于扩展,检索性能高效,数据分析可视化易于操作。ELK 提供一整套大数据分析方案,软件之间相互配合使用,完美衔接,是高校经费不足的情况下进行大数分析工作的首选。本文把校园网数据中心的流量和校园网出口的流量镜像出来,用 Ntopng 工具进行采集,通过 Logstash 传送到 Elasticsearch 进行分布式存储,用 Kibana 进行数据分析和展示。

1.1 平台使用工具介绍

Ntop 是一款流量分析工具,可以实时显示网络流量和活动主机,识别多种通信协议,监视网络吞吐量、传输字节数和数据包,使用深度包检测技术,对网络中的数据包进行分析,通过 Mysql、ElasticSearch、LogStash 进行数据导出。Ntop 是一款收费软件,教育用户可以申请免费的 license,本文实际部署的是 Ntopng,以下简称为 Ntop。ElasticSearch 简称“ES”,是一套基于 Apache Lucene 搜索引擎库之上,提供存储、搜索、分析数据三大功能,它是开放的、支持全文搜索、可扩展的分布式集群系统。Logstash 是用来接收、过虑、清洗数据的管道,它支持包括系统日志、Syslog、网络流量、文件、Web 信息等多种格式,并能够以多种方式输出数据,具有实时传输能力的数据收集引擎。Kibana 是一个基于 Web 的图形界面,用于搜索、分析和可视化存储在 ES 数据的工具。它利用 ES 的 REST 接口来检索数据,允许用户创建自己的数据的定制仪表板视图,支持查询和数据过滤,并生成各种维度表格、图形。

1.2 平台架构设计

高校用户数量多,每天的日志量多在数亿条以上,如何对日志收集、归档、存储、分类、汇总、多维度检索、统一管理与访问,是规划 ELK 平台必须考虑的问题。本方案是对数据中心和校园网出口流量采集分析,根据 ELK 和 Ntop 软件特点,使用 6 台服务器搭建 ELK 平台。因为计划使用在 ELK 平台上的功能对于服务器的 CPU 性能要求不高,本次使用数据中心淘汰下的 Dell R710 服务器,其中 2 台 Ntop 服务器分别对两处流量进行采集和传输,3 台部署 ES 集群,1 台安装 Kibana 用于可视化操作与查询。由于校园网出口流量上下行在峰值 6 Gbit/s,Ntop-WAN 服务器配置了 16CPU、32 GB 内存、Intel 82599ES 10-Gigabit 网卡,满足高带宽处理的要求。系统架构如图 1 所示,ES 集群随时可以支持新设备扩展,虚线框起来部分下一步计划把网络设备、信息系统 Syslog、业务应用的日志、网页爬虫等数据接入到平台里,进行数据关联分析。

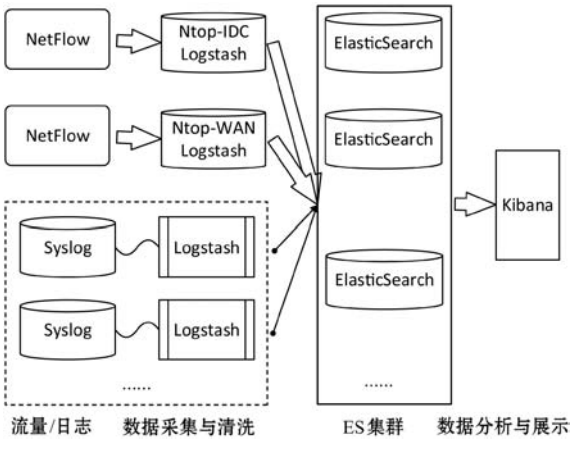

图 1 ELK 平台架构

1.3 平台实施与运维

ELK 与 Ntop 可以在多种 Unix 平台及 Windows 操作系统下安装,本文以 Ubuntu 16.04 Server 为操作系统环境进行部署。ELK 使用 Java 环境,6 台服务器须安装 Oracle JDK8。本次 ELK 安装的版本为 6.2.2。

流量采集工具 Ntop 安装需要注意几点。首先要把服务器的网卡启用,接收流量的网卡设置成“promisc”模式,流量接入时网卡灯闪烁。Ntop 的安装可参考官网的安装文档,使用命令行方式安装,之后需要修改 ntopng.conf 里的网卡、监控流量的 IP 地址段、默认端口号、文件路径等配置。在启动 Ntop 服务之前,需要先启动 redis 服务。之后可以通过 Web 方式(http://IP:Port)进行访问,默认登录用户名和密码均为 admin,可在登录后进行修改,并对输入 Licence 对 Ntop 进行激活。此外,Ntop-WAN 服务器为了解决数据流量大丢包的问题,需要安装 nProbe、配置网卡多队列。

ElasticSearch、Logstash 和 Kibana 三个软件官网教程非常详细,本文不在赘述。但在部署的过程中有很

多的问题需要注意。首先,根据 ElasticSearch 的系统分区与数据分区建议设置数据卷 LVM,对 jvm. options 里的默认内存配置也很重要。Logstash 与流量采集的 Ntop 工具部署在同一台服务器上,需要在 ntopng. conf 文件添加“ - F = logstash;127. 0. 0. 1;tcp;5510”配置,logstash 输入的“Index Patterns”配置分别为:ntopng - IDC - %{ + YYYY. MM. dd}、ntopng - WAN - %{ + YYYY. MM. dd}。Kibana 建议修改 Web 访问的默认端口号。由于流量与日志信息都属于敏感信息,且数据量巨大,网络安全问题是 ELK 平台必须引起重视;本方案中 6 台服务器防火墙都配置了限制访问的 IP 和端口策略,使用 nginx 的反向代理给 Kibana 配置登录认证。

2　数据分析与展现

ES 里接收到来自 Ntop 采集的网络流量记录里除了日志产生时间、源 IP、目的 IP、源端口、目的端口、协议类型、传输字节数、传输字节数等信息,还包括 DNS 请求、访问域名、URL、HTTP_CODE、HTTP_METHOD、MAC 等信息。ELK 平台每天接收的日志数亿条数据,如何从 TB 级的数据里发现网络中的异常访问和攻击,快速地检索到关键数据,有针对地展示才是我们搭建平台的目的。

2.1　黑客攻击行为分析

“不知攻,焉知防”,做好网络安全工作,需要了解黑客和他们攻击过程和方法,才能做好有价值的分析。通常黑客攻击的过程是漏洞扫描、漏洞攻击、获得权限、保持连接、消除痕迹等几个阶段,他们常用的攻击手法有端口扫描、网页爬虫、系统或应用指纹探测、暴力破解、缓冲区溢出、木马后门攻击等。攻击一般是针对某个服务器进行,利用某种漏洞,通过某个端口或 URL 实现。因此 IP 地址、端口、URL 等信息在攻击各个阶段都是黑客们需要获取和利用的关键信息,对重点保护区域 IP 的连接数、端口、传输数据包等信息进行统计和分析是发现攻击或异常问题的重要突破口。

2.2　校园网业务应用分析

根据 ES 存储的流量数据,结合校园网业务的特点,本文从校园网整体情况、重点服务器及黑客攻击监控等几方面建立分析视图。校园网整体监控,重点关注数据中心出口、校园网出口流量实时趋势,各种应用流量分类排序、端口访问排序、服务器上传与下载量排序等,从全局角度监控校园网承载情况。校园网的重点服务器一般分为三类:一是校园网关键基础设施如 DNS、DHCP、上网认证服务器、统一身份认证等用户使用校园网必须使用的服务;二是邮件、人事、学工、教务、科研、OA 等存储师生敏感信息和重要数据的信息系统;三是主页、招生、就业、留学生、网站群等对校外开放的、有影响力、访问量大的网站。通过从目的 IP 和源 IP 两个维度进行统计,对重点服务器监控可以发现黑客攻击的蛛丝马迹。黑客攻击角度比较广泛,可以从端口、Web 访问进行分析,比如服务器 22、3389 端口的长连接,135、445、1080 等黑客经常利用端口的服务器连接,31、555、666、1025 等木马常用端口服务器连接;Web 访问方面,一般黑客用自动化工具访问页面,HTTP_CODE 的会报“404”出错代码。

2.3　建立分类视图与仪表盘展示界面

Kibana 有多个功能的菜单,数据可视化常用“Discovery”、“Visualize”、“Dashboard”三个菜单,它们也是实现数据可视化的三个步骤。

1)数据源的处理。一般直接从“Index Patterns”里取用数据源,如本文的“ntop-IDC”,有些可视化需要进行处理后才能达到更好的效果。“Discovery”菜单界面主要用于通过搜索请求,过滤结果,获取字段值的统计情况并通过柱状图进行展示,搜索语法非常直白,支持布尔运算符、通配符与字段筛选,也是我们排除问题进行日志查询的常用界面。比如对邮件服务器的监控,要对访问邮件系统 IP 和域名的访问信息进行全部搜集,在“Discovery”里用正则表达式“(IPV4_DST_ADDR:("邮件 IP") OR IPV4_SRC_ADDR:("邮件 IP")) OR HTTP_HOST:("邮件域名")”检索,得到访问信息后保存“Server_Mail”查询结果。除重点服务器外,对端口、HTTP_CODE 等其他需要预处理字段也可处理,比如对数据中心服务器的数据库服务器监控,就可以通过访问 1433、1521、3306、11211、9200 等端口源 IP、目的 IP 进行查询并将得到结果进行保存。

2)数据可视化视图。“Visualize”菜单界面主要用于将某个“索引”或查询结果保存。在创建中,有直方图、柱形图、折线图、散点图、饼图、地图和数据统计表等多种图表可供选择。如何选择合适图形对可视化的展现非常重要,除了对图形进行研究以外,还须熟悉业务应用。比如区块图来可视化多个不同序列的总体情况、折线图来比较不同序列、饼图来显示每个来源对总体的贡献等,这几类图形适合于对校园网全局情况进行监控;柱形图作为一个通用图形,可以用对重点服务器、端口等进行连接数统计,用来展示异常访问和攻击行为。可视化的工具栏里聚合构建器(Aggregation Builder)需要结合业务进行分析,配置可视化要用到

“Metrics”和“Buckets”聚合工具,其中 Buckets 的效果类似于 SQL GROUP BY 语句。关于如何可视化官网有详细示例,仅以出口流量监控来说,区块图是监控的流量通用的方法,单击“Visualize”菜单界面,在“Basic Charts”里选择“Area”,之后数据源选择“ntopng-IDC”出口流量监控:Metrics 以“OUT_BYTES”的统计为聚合,“Buckets”的 X-Axis。

3)建立 Dashboard。为了让展示平台有特点,可以制作 Markdown 与 Metric 视图。根据之前规划的校园网整体情况、重点服务器及黑客攻击监控三类把“Visualize”保存的结果添加相应的分类中即可,以便统一展示,如图 2 所示。

图 2 流量数据分析可视化展示

2.4 校园网监控分析分析与问题处理案例

ELK 平台搭建完成后,分别建立了全局、端口、应用业务、重点服务器监控等多个 Dashboard,有效地对校园网的流量进行实时监控,准确地发现网络中重要业务应用,检查出校园网中异常行为。下面从端口、数据中心服务器、重要业务保障等角度发现的网络安全问题,数据采集以过去 24 小时期间,主要展现的是源和目的服务器连接数倒排序。

2.4.1 源与目的端口分析

首先对所有源与目的端口进行排名,Top5 是 53、161、80、443、25,对应的应用是 DNS、邮件、WEB 应用,与我校实际情况相符。其次分别对服务器常用运维的远程管理、数据库、文件上传等端口分类进行分析。由于在校园网出口开启了限制 21、22、3389、1433、1521、3306、9200、11211 等重要端口的外网访问策略,未发现校外 IP 通过这些端口连接数据中心的记录。TELNET、FTP 等业务目前高校只有少量业务在使用,安全问题相对较少。数据库进行限制 IP 的访问控制策略,也未发现异常访问行为,监控到反常的现象是数据中心服务器有主动利用 22、3389 外连现象,经查是运维人员使用该服务器做跳板机远程管理现校外设备,通过数据中心的防火墙做访问策略消除了此类的安全隐患。对黑客常用扫描 38 个端口及木马多数使用 54 个端口分类分析,监控到除了正常的网络运维监控系统、漏扫工具外,某日有个服务器的 445 端口异常访问,排查发现该服务器中了“WannaCry”病毒在进行内网扫描,后对服务器整改,及时排除了数据中心的安全隐患。

2.4.2 应用服务器监控分析

从数据中心服务器 IP 连接数、二级域名访问量、数据中心服务器的流量等几个维度进行分析,然后再对不同的业务系统,比如 DNS 服务器、DHCP 服务器、上网认证服务器、统一身份认证服务器等业务,进行具体监控,可以为安全事件发生后及时地提供数据综合分析。这里分享几个安全事件案例。案例一:某天夜间数据中心流量图显示有一小时波峰的高于平常 5 倍,检查服务器流量排行的 TOP,开源软件服务器有大流量,有一台 IP 在 30 分钟内发生 40 万次连接的请求,后经核实是该 IP 使用的第三方软件一个 bug 所致。案例二:某日监控到邮件服务器的 TOP10 被俄罗斯的一个 C 段的 IP 各有 80 万次左右的连接,根据邮件系统管理员排查,这些 IP 对邮件用户进行口令暴击力破解,通过在防火墙上加上攻击 IP 黑名单阻止这些地址访问,避免攻击危害。案例三:上网认证服务器显示部分校园网用户日访问 5 000 次以上,这与正常用户日均 10 次左右出入太大,经查这些主机均感染病毒或被置入木马,后及时通知相关老师进行处理消除了隐患。此类监控须对业务熟悉,然后根据数据统计分析情况分进行判断。

2.4.3 重要业务保障及校外特殊 IP 监控

学校选课、选房、高考招生系统有大量集中访问业务,关键时期都发生过服务器宕机的事件。ELK 平台上线后通过对这些业务服务器的访问进行排名,能及时发现异常访问的 IP 并进行限制,保证了今年的选课、选房、高考招生工作平稳有序的开展。2018 年很多服务器被爆出挖矿行为,根据中国科技大学共享的矿池 IP 进行分析,发现我校有台服务器与矿池 80.82.70.187 有异常连接,经查该服务器中了门罗币挖矿病毒:xmrig,由于及时发现并做了处理避免了更严重的安全事件发生。

3 结 语

ELK 数据分析平台开源项目部署方便,采用分布式架构易扩展,容易接收其他数据,没有其他基础软件

依赖,基本无需写代码,使大数据处理变得容易。通过 ELK 平台对采集到的流量数据进行分析,可以对校园网的流量、重要业务系统进行监控,能掌握出口数据中心的带宽利用率、趋势,能发现网络中异常攻击,验证安全设备发现的问题,对网络信息安全技术事件的追踪溯源提供数据支撑。另一方面 ELK 数据分析平台的日志查询功能,可以帮助信息系统部署和运维中的问题快速定位,降低故障维修的平均时间,提高故障检测精度,通过分析把问题消灭在萌芽的状态,在校园网网络安全管理、IT 运维中发挥积极的作用。

本方案采集的数据量巨大,但仅限于网络流量的信息,对网络攻击分析和安全事件的取证和溯源同样存在局限性。与安全厂商的态势感知分析平台相比,缺少规则库、沙箱等,不能识别恶意的 IP、恶意的域名,不能对网络中的可疑文件进行分析。计划下一步把服务器的系统日志、应用日志,交换机、安全设备的日志等传输到 ELK 平台上,结合多个数据源的海量数据进行联动分析,通过科学算法,做好校园网实时监控和风险预警。这样可以更科学地发现网络攻击和信息安全技术事件,从而提升安全威胁及事件的洞悉和感知能力,为网络安全管理决策提供数据支撑,更好地为校园网信息化保驾护航。

参 考 文 献

[1] 高凯. 大数据搜索与挖掘及可视化管理方案[M]. 3 版. 北京: 清华大学出版社,2017.

[2] 饶琛琳. ELK Stack 权威指南[M]. 2 版. 北京:机械工业出版社,2017.

[3] 曾恒. 基于 ELK 的网络安全日志管理分析系统的设计与实现[D]. 北京: 北京邮电大学,2017.

[4] 360 企业安全研究院. 走进安全:网络世界的攻与防[M]. 北京: 电子工业出版社,2018.

[5] 姚攀, 马玉鹏, 徐春香. 基于 ELK 的日志分析系统研究及应用[J]. 计算机工程与设计, 2018, 39(7): 2090-2095.

[6] Lei X F, Wang Z, He Y Z. The Data Management and Real-time Search Based on Elasticsearch[C]//2015 4th International Conference on Computer, Mechatronics, Control and Electronic Engineering. 2015.

[7] Sung J S, Kwon Y M. Performance of ELK stack and commercial system in security log analysis[C]//Malaysia International Conference on Communications, 2017.

[8] Prakash T, Kakkar M, Patel K. Geo-identification of web users through logs using ELK stack[C]//2016 6th International Conference—Cloud System and Big Data Engineering (Confluence). IEEE, 2016.

[9] Dharur S, Swaminathan K. Efficient surveillance and monitoring using the ELK stack for IoT powered Smart Buildings [C]//International Conference on Inventive Systems and Control, 2018.

[10] Al-Mahbashi I Y M. Network security enhancement through effective log analysis using ELK[C]//2017 International Conference on Computing Methodologies and Communication (ICCMC), 2017.

[11] Jati G, Hartadi B, Putra A G, et al. Design DDoS attack detector using NTOPNG[C]//International Workshop on Big Data & Information Security. IEEE, 2017.

[12] Kortebi A, Aouini Z, Delahaye C, et al. A platform for home network traffic monitoring[C]//Integrated Network & Service Management. IEEE, 2017.

[13] Chang N, Lan A, Liao M, et al. ELK delaminate improvement methodology on Cu pillar interconnect BOP structure [C]//Electronic Components & Technology Conference. IEEE, 2014.

(上接第 291 页)

[3] Sendik O, Cohenor D. Deep Correlations for Texture Synthesis[J]. Acm Transactions on Graphics, 2017, 36(4):1.

[4] Otori H, Kuriyama S. Data-embeddable texture synthesis [C]//Proceedings of the 8th international symposium on Smart Graphics. Springer-Verlag, 2007:146-157.

[5] Otori H, Kuriyama S. Texture synthesis for mobile data communications [J]. IEEE Computer Graphics and Applications, 2009, 29(6): 74-81.

[6] 张新鹏, 钱振兴, 李晟. 信息隐藏研究展望[J]. 应用科学学报, 2016, 34(5):475-489.

[7] Wu K C, Wang C M. Steganography using reversible texture synthesis[J]. IEEE Transactions Image Processing, 2015, 24(1): 130-139.

[8] Zhou H, Chen K J, Zhang W M, et al. Comments on "Steganography Using Reversible Texture Synthesis"[J]. IEEE Transactions on Image Processing, 2017, 26(4):1623-1625.

[9] 朱桂斌, 曹长修, 胡中豫, 等. 基于仿射变换的数字图像置乱加密算法[J]. 计算机辅助设计与图形学学报, 2003, 15(6):711-715.

[10] 罗海波, 葛斌, 王杰, 等. 整合神经网络置乱图像的动态自反馈混沌系统图像加密[J]. 中国图象图形学报, 2018, 23(3):48-63.

[11] 柏森, 曹长修. 一类基于行列式计算思想的图像置乱加密算法[J]. 计算机工程与应用, 2002, 38(8):37-39.

[12] Efros A A, Freeman W T. Image quilting for texture synthesis and transfer[C]//Proceedings of the 28th annual conference on Computer graphics and interactive techniques. ACM, 2001:341-346.