doi: 10.7690/bgzdh.2016.09.016

网络异常流量检测方法

张 楠，李洪敏，卢 敏，柯明敏
（中国工程物理研究院总体工程研究所，四川 绵阳 621900）

摘要：针对网络环境中可能导致网络流量异常的问题，归纳了当前主流的网络异常流量检测方法。通过阐明网络异常流量检测过程中的关键问题和核心技术，重点探讨了固定阈值、特征检测、统计分析、数据挖掘、云计算方法的原理和适用范围，对网络流量中异常流量的定义和分类进行了总结，介绍相关的研究进展，并分析异常流量检测的当前现状和发展前景。该分析结果为异常网络流量检测研究提供了参考依据，也为网络安全研究提供了支持。

关键词：网络流量；异常流量；检测；综述
中图分类号：TP393 **文献标志码**：A

Network Anomaly Flow Detection Method

Zhang Nan, Li Hongmin, Lu Min, Ke Mingmin
(*Institute of System Engineering*, *China Academy of Engineering Physics*, *Mianyang* 621900, *China*)

Abstract: Conclude the current anomaly network flow detection methods for the anomaly network flow problem in network environment. By introducing key problem and key technology of network anomaly flow detection, pay more attention on discussing principle and application area of constant-threshold, feature detection, statistical analysis, data mining, and cloud computing method. Conclude definition and classification of network anomaly flow, introduce related research, and carry out further analysis current situation and development prospect of anomaly flow. The analysis result provides a reference for the detection of anomaly network flow, and also provides support for network security research.

Keywords: network flow; anomaly flow; detection; overview

0 引言

随着大数据技术的发展，“互联网+”的思维模式正在向各行各业蔓延。网络信息化正在影响着人们的生活方式和工作模式，从政府机构到各型企业都需要网络信息化技术提供工作便利。网络的安全性和可靠性也就无时无刻不关系着业务运行。通过对网络流量的分析研究，可尽早发现网络中可能存在的网络攻击行为和网络故障，为网络安全管理和网络架构规划提供决策依据，进而保证网络的安全高效运行；为此，笔者对网络异常流量检测进行了分析。

1 网络流量

1.1 网络流量的定义

网络的结构和生活中的交通网络类似，网络流量的含义就是单位时间内网络传输的数据量[1]。根据传输网络数据的流量粒度、类型和时间等不同维度又可以对网络流量进行进一步划分，具体的划分方法依赖于用户的应用需求。网络中单一的数据包/帧是不具备具体含义的，只有它与具体的业务应用关联起来才有其现实意义。通常对特定线路和节点的网络流量检测，包含总流量、总帧流量、广播流量、错包率、丢包率和带宽占用比等指标，同时针对每个指标也有其对应的上行指标和下行指标[2]。此外，也可以检测这些指标的高峰值、低峰值和平均值等统计指标。

1.2 网络流量的采集方法

网络流量的采集是一套完整的工具系统，其中包含了对于硬件网络流量数据的获取、传输、存储、分析和展示等多个部分。在网络流量采集系统中，采集原始流量数据的方法是其核心技术。流量采集技术按照不同角度也有不同的分类方法，例如可按软硬件划分方法将其分为基于硬件和基于软件的方法，同时也可按照采集对象的不同将其分为基于主机和基于网络设备的方法。笔者列举几种主流的采集方法如下。

1.2.1 基于网络探针方法

网络探针往往被安装在距离路由出口较近的位置，能够检测到一定范围内的网络流量数据。该方法只能检测到一定范围内的网络流量，大多数情况下使用于局域网。由于网络探针不经过路由，所以网络探针的安装对于网络整体的带宽并无太大影响。该方法在流量检测和分析时对主机有较高的性能要求，这一点在 Internet 环境中表现尤为突出[3]。

收稿日期：2016-05-13；**修回日期**：2016-06-22
作者简介：张 楠(1990—)，男，陕西人，硕士，助理工程师，从事大数据与网络安全研究。

此外，安装网络探针的连接接口的传输速率、监测主机的缓存和数据处理能力都将影响着流量采集的准确性和效率。因此，该方法对于大数据传输流量的网络的适用性相对较差。

1.2.2　基于 Netflow 的方法

NetFlow 是由 Cisco 公司在 1996 年研发出用于分析网络数据包信息的流量轮廓的一套技术。目前 NetFlow 已成为了业界事实上的标准，它描述了路由器输出关于被路由套接字对(the routed socket pairs)统计信息的方法。目前 Cisco 的绝大多数路由器集成了该技术，Juniper、Extreme 以及其他硬件设备厂商也有集成该技术的路由器和交换机。Netflow 主要包含 3 部分[3]，即数据导出、数据采集和数据分析，如图 1 所示。许多学者也致力于基于 Netflow 方法的流量采集技术研究，例如：Zou 等[4]提出了一种可以采集更多数据的流量采集系统；李秉桓等[5]提出了基于 FPGA 和 DDRII SDRAM 硬件实现的 Netflow 流量采集探测器的设计方案。

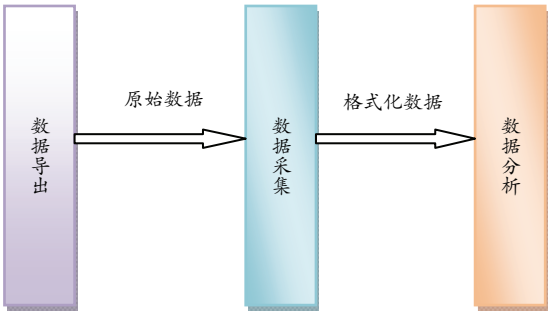

图 1　Netflow 架构图

1.2.3　基于 SNMP 协议的方法

SNMP 是基于 TCP/IP 协议簇的应用层协议，是一种简单网络管理协议[6]。在传输层是通过用户数据报协议(user datagram protocol，UDP)来实现的，现已成为行业中网络管理的一项标准。基于 SNMP 协议的网络流量采集方法构成的网络采集系统主要包括 MIB 管理信息库、SMI 管理信息结构以及 SNMP 管理协议本身。目前，该方法被广泛应用于网络管理中。

1.2.4　主机软件的方法

该方法就是在用户主机内安装一个流量监测的软件，软件通过监控网卡的通信数据包来实现流量监控。该方法的优势是可以提供更为详尽的网络流量数据资料，其劣势在于无法对网络设备和全局网络的流量进行监控，而且流量监测软件自身也要消耗一部分主机系统资源。

2　异常流量

2.1　异常流量的定义

异常是一个相对概念，是正常向对面或偏离正常一定偏移量的情形。网络流量的异常也就是偏离正常流量的情况。在网络环境中，“正常流量”不是一个特定的状态，而是会由于用户操作的变化、业务流的变化、管理实务的变化而不断更新的状态；因此，对于异常流量的判断也必须基于当前网络状态的正常基准值来确定。在日常网络环境中，异常网络流量的产生主要包含以下几种可能[7]：

2.1.1　网络操作

网络管理员对网络设备的操作可能会引起网络流量的波动，这种波动可能就是网络流量异常。网络管理员的操作可以分为正常操作和误操作 2 种。其中，正常操作主要包含数据备份、数据迁移、线路检修与更换和网络分流等操作；误操作是指网络管理员由于个人原因致使设备、链路、端口等发生异常阻塞和关闭等操作。

2.1.2　病毒传播

病毒在网络中传播过程伴随着网络带宽的消耗，最为典型的就是网络蠕虫病毒。病毒的快速蔓延在网络流量上的体现比较明显，但对于单个节点来说数据流量并不具备显著特征；因此，针对病毒传播，网络流量的变化必须从全网的角度观察并结合特定数据流分析才可以观察。这种流量异常对于网络服务质量和网络安全性都是一个巨大的威胁。

2.1.3　集中访问

这种类型的异常主要是由于某种业务的业务量剧增导致的网络流量峰值。现实中尤为常见的主要有学生选课、抢红包、购票等具体应用场景。这种拥挤导致的网络流量峰值如果在网络承载范围之内，那就仅表现出网络的暂时性拥堵；如果超过了网络的承载能力就可能导致网络瘫痪或者设备损坏。这种异常流量的检测相对来说难度较低，而且也易于发现和处理。

2.1.4　网络攻击

网络中某些攻击者为达目的发起的特殊网络操作会产生异常流量，比较典型的有暴力密码破解、DoS/DDoS 攻击、IP 欺骗等攻击行为。这些行为会在网络中表现出个别节点的发送和接收数据包的量出现异常峰值，对这种异常流量的检测有助于快速定位攻击者以避免对网络造成不必要的攻击伤害。

2.2 异常流量检测方法

异常网络流量检测是为了发现网络中可能存在的异常网络行为，及时发现网络中存在的攻击或异常操作，从而减少甚至避免对于单位造成的损失。近年来，针对异常流量的检测是一个重要的研究课题，除了阈值、特征检测、统计方法等传统方法以外，如大数据、人工智能、机器学习等新的理论和技术也被应用于异常流量的检测，详细介绍如下。

2.2.1 固定阈值方法

阈值检测方式是通过事先由网络管理员确定一个单位时间内的流量阈值，然后由管理员手动输入给检测系统，检测系统不断采集数据判定流量是否合法。如果采集到的数据流量值偏离了阈值，系统就会发出异常流量的报警。例如某系统针对某一链路设置 3 min 统计一次流量，设定固定流量阈值为 3 000 MB，如果某一时刻检测到前 3 min 的流量超过了 3 000 MB，就会产生报警。如图 2 所示，基准线以上的监测点第 3 min 和第 9 min 都会产生报警。

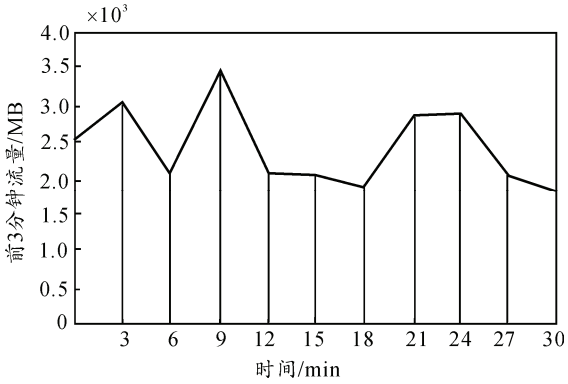

图 2 基于阈值的检测方法

这种检测原理和操作方法都十分简单，同时也具备高效和实时性强的特点，但对网络管理员提出了很高的要求。其一是要能够设定非常合适的阈值，如果阈值设置过高，则可能会对异常流量出现漏报的情况；而如果设置过低，则会大幅增加告警数量，管理员也很难从大量的告警中分辨出有价值的告警。其二是设定的流量阈值标准不是一成不变的，随着时间的推移、业务的变化、设备的增减等因素，网络中的流量阈值一直在变化，这就需要管理员及时更新阈值。对此，许多学者对此也做了研究，如曹敏等[8]提出了自适应阈值的阈值方法；杨雅辉等[9]提出了灵活的检测阈值及分裂值的计算方法。

2.2.2 基于特征的检测方法

该方法通过建立所有异常网络行为的特征库，将当前网络特征与异常特征库进行匹配，根据匹配结果判定当前网络数据是否正常。这种方法可以高效地检测出具有明显特征的已知网络攻击行为，如蠕虫病毒和 DDoS 攻击，而对于未知的网络攻击，该方法并不能很好地检测出异常流量。对此，很多学者也研究了基于正常行为特征库的误用检测[10]。此外，特征库的数量和更新速度也将影响检测的准确性。面对大量的特征库，特征匹配的速度对检测算法和机器性能提出了挑战。例如，Choi 等[11]提出了基于 Higuchi 分形维数和统计测量与滑动窗口操作结合提取特征值的两级检测方法用于异常流量行为的识别；Crovella 等[12]提出通过对数据抽象的非监督学习提取特征值从而检测异常流量的方法。

2.2.3 基于统计分析的方法

统计方法是针对已有历史数据记录通过分析得出一个判断的基准，再针对新的网络流量数据进行判断。这种检测具有较高的准确性和实效性，但会忽略网络流量自身的关联性和逻辑性。在利用统计方法解决异常流量检测问题的研究中，Zou[13]利用时间序列分析的方法来检测异常流量；Zare[14]使用马尔可夫(Markov)过程思想，利用 ARIMA 模型预测流量的方法检测异常流量；Cao 等[15]又利用这种方法成功检测 DoS 攻击引发的流量异常。此外，统计学中的 CUSUM 算法也被 Cheng 等[16]成功应用于检测 SYN Flood 造成的异常网络流量。

2.2.4 基于数据挖掘的方法

数据挖掘是一个知识发现的过程，是探究数据内部结构和外部关联的一种工具。数据挖掘包含多种分析方法，在异常流量检测方面主要应用的有聚类分析、关联分析、分类分析等。如 Bhardwaj 等[17]利用数据挖掘的方法，结合可视化图形展现网络的安全性，可以快速识别和检测受到攻击的设备；Othman 等[18]利用模糊关联规则成功检测了异常流量；薛静锋等[19]使用分类方法将网络流量分类，再利用决策树的方法检测异常流量；李天枫等[20]提出了一种基于大数据平台的大规模网络异常流量实时监测系统，通过对流量、日志等网络安全数据的挖掘检测网络异常流量。数据挖掘的方法对于检测异常流量有着很好的适用性，而且可以结合大数据技术处理包含更多数据量的流量数据，将来数据挖掘方法会朝着更智能化的方向发展，以期更精准地检测网络异常流量。

2.2.5 基于云计算的方法

云计算与其说是一种技术思路，还不如说是一

种按需服务的商业模式。云计算被应用于网络流量异常检测，主要是用于解决网络流量数据存储和实时快速分析的性能瓶颈问题。海量的网络流量监控数据的存储和访问难题可以通过云计算的方法高效地解决。网络流量数据清洗、整理、分析和检测过程都需要消耗大量的 CPU 和内存等性能资源，高性能的计算机或集群能为其提供高效的平台支撑，同时也按照业务需求满足不同规模的流量分析。著名的 Map/Reduce、Hadoop、Spark 等新兴技术也逐渐应用于网络流量分析，如 Aishwarya 等[21]将 Hadoop 应用于网络异常流量检测的过程中；方峰[22]将 Spark 引入到网络异常流量检测的过程中，成功地实现了实时监测 DDoS 攻击。

2.3　异常流量检测发展趋势

从文中分析的各种异常流量检测方法来看，异常流量检测技术需要更好地融合当今的先进技术，异常流量检测依然面临着以下问题：1) 异常流量的检测主要还是针对于已知攻击类型的检测，对于未知类型的异常流量存在判断不准确且定位难的问题；2) 异常流量检测的准确性一定程度上依赖于网络管理员的专业水平且需要人工干预，未来的流量检测必将是朝着更加智能化的方向发展；3) 由于目前网络应用逐渐趋于多样化，网络流量数据剧增，基于大数据、云计算、高性能计算等技术应用于网络异常流量检测会大幅度提升其效率和可靠性。

3　结束语

笔者从介绍网络流量数据的定义和采集方法出发，分析了网络环境中可能导致网络流量异常的情形，重点探讨了固定阈值、特征检测、统计分析、数据挖掘、云计算方法的原理和适用范围，介绍了相关的研究进展，最终分析了异常流量检测的当前现状和发展前景，为网络安全管理和网络安全的研究提供了一定的参考依据。

参考文献：

[1] 张宾, 杨家海, 吴建平. Internet 流量模型分析与评述[J]. 软件学报, 2011, 22(1): 115−131.

[2] 邱婧, 夏靖波, 吴吉祥. 网络流量预测模型研究进展[J]. 计算机工程与设计, 2012, 33(3): 865−869.

[3] 许晓东, 卞鹏, 朱士瑞. 基于 Netflow 的异常流量分离以及归类[J]. 计算机工程与设计, 2009, 30(21): 4818−4820.

[4] Zou R, Xu T, Hou H. An Enhanced Netflow Data Collection System[C] Instrumentation, Measurement, Computer, Communication and Control (IMCCC), 2012 Second International Conference on. IEEE, 2012: 508−511.

[5] 李秉桓, 董永吉. 一种 NetFlow 流量采集探测器的设计[J]. 计算机安全, 2012(2): 2−4.

[6] 薛蔡. SNMP 协议及其在网络管理中的应用[J]. 兵工自动化, 2003, 22(6): 40−42.

[7] 贾慧, 高仲合. 异常流量的分析与研究[J]. 计算机安全, 2010(7): 55−57.

[8] 曹敏, 程东年, 张建辉, 等. 基于自适应阈值的网络流量异常检测算法[J]. 计算机工程, 2009, 35(19): 164−166.

[9] 杨雅辉, 杜克明. 全网异常流量簇的检测与确定机制[J]. 计算机研究与发展, 2009, 46(11): 1847−1853.

[10] 杨智君, 田地, 马骏骁, 等. 入侵检测技术研究综述[J]. 计算机工程与设计, 2006, 27(12): 2119−2123.

[11] Ji S Y, Choi S, Dong H J. Designing a two-level monitoring method to detect network abnormal behaviors[C]. Information Reuse and Integration (IRI), 2014 IEEE 15th International Conference on. IEEE, 2015: 703−709.

[12] Crovella M, Lakhina A. Method and apparatus for whole-network anomaly diagnosis and method to detect and classify network anomalies using traffic feature distributions: U.S. Patent 8, 869, 276[P]. 2014-10-21.

[13] Zou B X. A real-time detection method for net-work traffic anomalies[J]. Chinese Journal of Computers, 2003, 26(8): 940−947.

[14] Zare M H. ARIMA model for network traffic prediction and anomaly detection[C]. Proceedings of ITSim International Symposiumon Information Technology, 2008: 1−6.

[15] Mei C X, University N, Nanjing. DoS Attack Detection Scheme for Sensor Networks Based on Traffic Prediction[J]. Chinese Journal of Computers, 2007, 30(10): 1798−1805.

[16] Cheng J, Lin B, Jianzhi L U, et al. Detection of SYN Flooding Attacks Based on Non-parametric CUSUM Algorithm[J]. Computer Engineering, 2006, 32(2): 159−161.

[17] Bhardwaj A K, Singh M. Data mining-based integrated network traffic visualization framework for threat detection[J]. Neural Computing & Applications, 2014, 26(1): 117−130.

[18] Othman Z A, Eljadi E E. Network anomaly detection tools based on association rules[C]. Electrical Engineering and Informatics (ICEEI), 2011 International Conference on. IEEE, 2011: 1−7.

[19] 薛静锋, 曹元大. 贝叶斯分类在入侵检测中的应用研究[J]. 计算机科学, 2005, 32(8): 60−63.

[20] 李天枫, 姚欣, 王劲松. 大规模网络异常流量实时云监测平台研究[J]. 信息网络安全, 2014(9): 1−5.

[21] Aishwarya K, Sankar S. Traffic analysis using Hadoop Cloud[C]. Innovations in Information, Embedded and Communication Systems (ICIIECS), 2015 International Conference on. IEEE, 2015: 1−6.

[22] 方峰, 蔡志平, 肇启佳, 等. 使用 Spark Streaming 的自适应实时 DDoS 检测和防御技术[J]. 计算机科学与探索, 2015(10): 41−46.