基于熵的网络异常流量检测研究综述①

李 蕊, 张路桥, 李海峰, 刘 恺

(成都信息工程大学 信息安全工程学院, 成都 610000)

摘 要: 网络流量异常检测及分析作为一种重要的网络监管控制手段, 是网络及安全管理领域的重要研究内容. 本文探讨了网络异常流量的种类, 简述了基于传统的异常检测方法在网络异常流量检测中的应用以及存在的问题. 针对基于信息熵、相对熵、活跃熵等熵值理论在网络异常流量检测中的研究, 阐述了基于熵值理论的异常检测在国内外的研究进展情况. 总结了当前基于熵值理论的异常检测研究工作中存在的问题及改进方向.

关键词: 网络安全; 异常检测; 异常流量; 熵理论; 信息熵

Survey of Entropy-Based Network Traffic Anomaly Detection Methods

LI Rui, ZHANG Lu-Qiao, LI Hai-Feng, LIU Kai

(College of Information Security, Chengdu University of Information Technology, Chengdu 610000, China)

Abstract: It is an essential means to detect and analysize the abnormal network traffic in network supervision. And it is also an important research topic in the field of network security management. At the beginning of this paper, we discuss some types of abnormal network traffic, and point out some problems while using traditional anomaly detection methods in network traffic anomaly detection. And then, we specify the latest research achievements of anomaly detection method based on entropy theory which apply information entropy, relative entropy, and active entropy theory to detect abnormal network traffic. Finally, we conclude some problems of anomaly detection methods based on entropy theory and the direction of improvement.

Key words: network security; anomaly detection; anomaly traffic; entropy theory; information entropy

1 引言

随着网络通信技术的迅速发展, 复杂多变的互联网应用层出不穷, DDOS、蠕虫、木马等网络安全问题日渐突出. 据统计, 2013 年我国境内感染木马僵尸网络的主机就有 1135 万个[1], NSFOCUS 曾分析提到, 2013 年 3 月发生的针对反垃圾邮件 Spamhaus 的 DDOS 攻击, 利用 DNS 反射形成了前所未有的 300Gbps 的峰值流量[2]. 网络异常流量检测及分析作为一种重要的网络监管控制手段, 是网络监控管理中的关键环节, 通过对流量进行监控和分析, 及时发现网络中异常现象, 对于维护网络正常运行有着重要意义, 一直以来都是网络安全学术界和工业界重点关注的研究方向, 已形成众多相关检测方法及系统.

流量异常检测技术源于网络安全中的入侵检测领域, 是网络监控管理中的关键一环, 通过对流量进行监控和分析, 及时发现网络中异常现象,流量的异常检测首先在对信息源建模分析的基础上, 勾画出检测对象的行为模式轮廓, 通过新数据样本和行为模式轮廓的对比来发现当前行为特征的偏离.

2 网络异常流量分类及检测方法

2.1 网络异常流量分类

网络异常流量是指对网络正常使用造成不良影响的网络流量, 网络异常流量根据其异常特征各有不同的表现方式. Barford 等人[3]通过观察总结网络流量行为, 在网络流层次使用统计特征的相似度将网络异常流量分为 3 大类, 网络操作异常、突发拥塞异常和网络滥用异常, 这也是目前对于网络异常流量较为普遍

① 收稿时间:2016-09-23;收到修改稿时间:2016-11-21 [doi:10.15888/j.cnki.csa.005807]

的一种分类方式.

2.2 网络异常流量检测方法

根据相关研究, 可以将网络流量异常检测方法概括为以下几种: 基于特征/行为的研究方法、基于信号处理的异常检测、基于统计的异常检测、基于机器学习的方法和基于数据挖掘的方法等. 苗甫[4]综合利用包层和流层特征, 提出基于流量指纹的恶意流量代码检测模型. 夏正敏[5]根据网络流量的非平稳及自相似特性, 提出一种非平稳网络流量自相似模型. 张伟[6]提出基于传输层会话行为统计特征的恶意流量识别方法. 刘肖琛[7]提出了一种聚类的检测算法, 用于检测恶意流量.

基于特征/行为的检测方法可以做到实时分析和检测, 但无法检测出未知的攻击类. 信号处理的方法对流量变化不明显的隐藏性攻击检测能力不足. 基于统计的研究不能确定产生异常的原因及异常属性, 适合离线的非实时的网络异常检测. 机器学习的方法在用户动态行为变化及单独流量异常检测方面有待改善.

为了更好的进行流量异常检测, 需要了解网络流量的分布规律, 研究表明, 网络流量具有自相似、长相关和重尾分布[8]等特征, 熵值可以用来描述此类特征, 因此利用熵值理论分析网络流量已成为近年来的一个研究热点.

3 基于熵的网络异常流量的检测研究

网络流量数据由无数离散信息源组成, 熵可以有效度量系统参数分布的变化情况, 描述长时间的随机过程以及网络流量在某些维度上的分布状况, 国内外的很多研究人员利用熵值理论提出了多种异常检测的方案. 基于熵的异常检测系统的主要思想是: 一旦有异常流量发生, 总体流量的熵值会随之发生变化, 通过熵值的变化检测出该异常. 常见的用于异常流量检查的理论有信息熵、活跃熵、相对熵等.

3.1 基于信息熵的异常检测研究

在信息论中, 通信中随机性的干扰是无法避免的, 信息熵代表了所含信息量的多少, 是对系统不确定性程度的一种度量, 反映了信息源信号出现几率. 我们假设系统处于几种不同状态, 每种状态出现的概率为 $p_i(i=1, 2, \dots, n)$则该系统的熵定义为:

$$E = -\sum_{i=0}^{n} p_i \ln p_i, 0 \le p_i \le 1 \& \sum_{i=0}^{n} p_i = 1 \quad (1)$$

熵具有极值性, 当各个信源为状态等概率分布时, 熵值最大, 并且等于信源输出状态数, 熵的这个极值性给出一个比较重要的信息, 即分布越平坦熵值越大, 反之则越小, 熵值的大小表明特征分布的集中或分散程度, 因而可以作为评价特征分布的一个指标.

基于信息熵理论的异常检测方法可以提供比传统流量分析更精细的结果, 识别能力更加优秀, 目前国内外也有很多相关研究. 文献[9]中陈锶奇等人利用信息熵的极值性, 结合 Netflow 数据流, 对各指标熵值进行相关性分析, 进而确定异常的存在; 文献[10]采用粗细粒度结合的思想, 利用信息熵进行分析, 可有效地检测出骨干网中的 DOS/DDOS 攻击; Fonseca 等[11]用信息熵的方法发现流量空间上的信息单元也存在长相关特性. 文献[12]采用时间作为单位形成统计元组, 在每个统计元组内分别计算各网络属性的信息熵, 根据正常网络状态下单位统计元组内流量信息结构的稳定性来判断网络是否发生异常. 文献[13]运用信息熵寻找显著特征, 根据显著特征进行级联分簇, 提出了一种基于信息熵的流量识别方法. 文献[14]对熵值理论进行研究, 发现基于信息熵的流量异常检测对于端口扫描效果很好, 而基于联合熵的检测算法则对 DOS/DDOS 攻击有较好的检测效果.

3.2 基于活跃熵的异常检测研究

活跃通信理论为认识和分析通信系统建立了一种新的标准, 根据刘衍珩等人在文献[15]中的假设, 将网络整体视为一个系统, 包含 n 个主机个体, 即 $S=\{s_1, s_2, \dots, s_n\}$, 系统中每台主机的一次动作记作一个动作单元. 当系统外主机访问系统内主机时, v 值加 1, 反之减 1. 定义状态集合 $U=\{\mu_1, \mu_2, \dots, \mu_n\}$为系统在经过 n 次动作后状态序列的集合. 从状态集合中可以得到其活跃度集合 $A=\{a_1, a_2, \dots, a_k\}$, 其中, a_i 表示系统经过规定时间或数量等尺度后, 状态 μ_i 出现的次数. 通过计算状态 μ_i 出现的概率 p_i, 可以得到各状态的活跃度概率集合 $P=\{p, p_2, \dots, p_k\}$. 对概率集合应用熵理论可得到活跃熵的定义为:

$$Ha = -\sum_{i=1}^{k} p_i \log p_i, \quad p_i = a_i / \sum_{j=1}^{k} a_j \quad (2)$$

文献[15]基于活跃通信理论将信息熵和网络流会话相关性结合起来, 使用访问流与应答流共同建立系统的活跃状态, 通过分析网络流量活跃熵值的变化实现对 DOS 攻击行为的检测. 文献[16]将活跃熵用于异

常流量检测, 记录分析进出系统的 NetFlow 流, 根据流量大小选择不同的尺度计算熵值, 降低误报率, 对端口扫描具有很好的检测效果. 文献[17]提出了基于活跃熵的 Web 应用入侵检测模型. 通过对 GET、POST 请求参数以及 HTTP 头部关键数据的熵值进行分析, 利用熵值变化来发现 Web 数据流中存在的攻击行为. 文献[18]在 IP 流层面分析系统活跃熵值, 对整个流量进行初步探析, 剔除正常流量, 利用多特征广泛权重最小二乘孪生支持向量机算法进行攻击确认.

3.3 基于相对熵的异常检测研究

相对熵是指两个随机序列之间距离的度量, 从统计学角度它是指两个随机序列之间的相似程度. 相对熵可定义为:

$$D(P\|Q)=\sum_{i=1}^{n}p_i\ln\frac{p_i}{q_i} \tag{3}$$

其中, $P=\{p_1, p_2, \ldots, p_n\}$, $Q=\{q_1, q_2, \ldots, q_n\}$. P 为当前时段某一测度的分布序列, Q 为上一时段该测度的分布序列. 相对熵具有一个重要性质: 当且仅当 $P=Q$ 时, $D(P||Q)=0$, 即当两个分布序列 P、Q 完全相同时它们之间的相对熵值为 0. $D(P||Q)$的值越小表示分布序列 P 和 Q 越相似, 反之则相差越大.

文献[19]针对 SIP 攻击中的 DOS 洪泛攻击, 提出了一种基于相对熵的异常检测算法, 用以反映网络流量中的动态变迁过程. Altaher 等人[20]提出了将相对熵理论应用于实时网络检测. 文献[21]基于相对熵理论, 利用网络流量的自相似性, 通过判断相邻时刻流量之间的相对熵值是否发生突变来进行 DOS 攻击检测. 文献[22]通过对网络流量进行分维和分层, 在网络流量的各个分析视图上采用相对熵对异常流量进行检测.

3.4 熵值理论与算法结合的异常检测研究

目前熵值理论以及机器学习、数据挖掘等算法被广泛的应用到网络异常流量检测中, 文献[23]使用后近邻算法作为分类器, 并使用相关信息作为补充特征, 在在小训练样本情况下获得较高识别性能. 文献[24]中 Wu 等人则提出一种相关的 KNN(k-Nearest Neighbor, k 最邻近)算法, 并应用于非平衡的流量识别中. 文献[25]采用条件熵和相异样本熵构建具有良好区分度的多维攻击检测向量, 并且采用多维无参 CUSUM 算法在高速骨干网环境下检测异常, 对于 DDOS 攻击响应速度快. 文献[26]在多个不同维度上采用熵度量流量数据的分布特征, 把熵值排列成多维检测向量, 采用一类支持向量机对检测向量进行分类. 文献[27]在 SDN 中将采集到的数据进行熵值计算, 判断异常流量的存在, 并利用 SVM 算法对不同攻击流量进行分类. 文献[28]提出了让基于信息熵的大规模网络流量异常检测, 不仅吸收了子空间方法的思想, 并结合了 K-means 分类方法. 文献[29]将网络流量幅值看作随时间变化的信号, 利用小波分析区分出背景流量和异常流量, 而后根据异常持续时间和信号频率的不同采用不同的方来检测攻击.

3.5 其他基于熵值理论的异常检测研究

很多研究人员, 将熵值理论进行改进创新, 提出了多种熵值理论. 文献[30]提出了一种基于非扩展熵的异常检测方法, 利用非广延熵以及 Lyapunov 指数来表示源、目的 IP 分散情况来检测攻击. 文献[31]根据 DOS 攻击发生时网络中的流量特性和 IP 熵特性, 提出了基于流量和 IP 熵特性的 DOS 攻击检测算法. 文献[32]通过分析信息熵在异常检测中的应用和条件熵的固有特性, 提出了一种基于条件熵和加权熵理论的异常检测方法. 文献[33]利用特征熵依据流量特征参数的分布变化来检测异常, 通过分析异常间隔的流量迭代地排除类似正常的流, 从而识别根源流. 文献[34]提出了一种基于交叉熵的异常检测及分类方法. 文献[35]将条件随机场用于 DDOS 攻击检测, 在相应条件下有效的改善了检测效率差的问题.

4 结语

本文首先对网络异常检测技术进行对比分析, 突出了基于熵值理论的网络异常流量检测在网络异常分析上的重要作用. 熵值理论可以提供比传统流量分析更精细的结果, 但已有研究工作还存在以下问题:

(1)已有的各种熵值算法在执行效率上有待进一步提高; (2)未充分利用熵值特性, 随机量化过程也需要改进; (3)如何动态、精确的计算阈值也是需要进一步研究的问题.

参考文献

1 国家互联网应急中心.2013 年互联网网络安全态势综述,2014.

2 NSFOCUS. Analysis of DDoS Attacks on Spamhaus and Recommended Solution. [2013-6-2].

3 Barford P, Plonka D. Characteristics of network traffic flow

anomalies. Proc. of the 1st ACM SIGCOMM Workshop on Internet Measurement (IMW'01). New York. ACM Press. 2001. 69–73.

4 苗甫,王振兴,张连成.基于流量统计指纹的恶意代码检测模型.计算机工程,2011,37(18):131–133.

5 夏正敏.基于分形的网络流量分析及异常检测技术研究[博士学位论文].上海:上海交通大学,2012.

6 张伟.基于传输层会话行为统计特征的恶意流量识别[硕士学位论文].南京:南京邮电大学,2014.

7 刘肖琛.基于大数据的网络恶意流量分析系统的设计与实现[硕士学位论文].北京:北京邮电大学,2014.

8 朱应武,杨家海,张金祥.基于流量信息结构的异常检测.软件学报,2010,21(10):2573–2583.

9 陈锶奇,王娟.基于信息熵理论的教育网异常流量发现.计算机应用研究,2010,27(4):1434–1436.

10 周颖杰,焦程波,陈慧楠,马力,胡光岷.基于流量行为特征的 DOS&DDOS 攻击检测与异常流识别.计算机应用,2013, 33(10): 2838–2841.

11 Fonseca N, Crovella M, Salamatian K.Long range mutual information. ACM SIGMETRICS Performance Evaluation Review, 2008, 36(2): 32–37.

12 严承华,程晋,樊攀星.基于信息熵的网络流量信息结构特征研究.信息网络安全,2014,(3):28–31.

13 吴震,刘兴彬,童晓民.基于信息熵的流量识别方法.计算机工程.2009,35(20):115–116,120.

14 崔锡鑫,苏伟,刘颖.基于熵的流量分析和异常检测技术研究与实现.计算机技术与发展.2013,23(5):120–123.

15 刘衍珩,付枫,朱建启,等.基于活跃熵的 DoS 攻击检测模型.吉林大学学报(工学版),2011,41(4): 1059–1064.

16 穆祥昆,王劲松,薛羽丰,等.基于活跃熵的网络异常流量检测方法.通信学报,2013,34(Z2): 51–57.

17 莫秀良,常畅,王春东.基于活跃熵的 Web 应用入侵检测模型.武汉大学学报(理学版),2014,60(6):543–547.

18 张明明,李玉峰,张鹏,等.大流量下一种基于活跃熵的 DDOS 攻击检测方法.计算机应用研究,2016,33(6): 2148–2151.

19 张晓月,胡访宇.基于相对熵的 SIP Dos 洪泛攻击检测算法和仿真.计算机应用,2015,24(1):135–138.

20 Altaher A, Ramadss S, Almomani A. Real time network anomaly detection using relative entropy. Proc. of the 8th International Conference on High-Capacity Optical Networks and Emerging Technologies. IEEE Press. 2011. 258–260.

21 涵秋,马艳,雷磊.基于相对熵理论的网络 Dos 攻击检测算法.电讯技术,2011,51(3):89–92.

22 张登银,廖建飞.基于相对熵的网络流量异常检测方法.南京邮电大学学报(自然科学版),2012,32(5):26–31.

23 Zhang J, Xiang Y, Wang Y. Network traffic classification using correlation information. IEEE Trans. on Parallel and Distributed Systems, 2013, 24(1):104–117.

24 Wu D, Chen X, Chen C, et al. On addressing the imbalance problem: a correlated KNN approach for network traffic classification. 8th International Conference Network and System Security. Xi'an, China. Springer. 2014. 138–151.

25 赵小欢,夏靖波,郭威武,等.基于多维信息熵值的 DDOS 攻击检测方法.空军工程大学学报(自然科学版),2013,14(3): 58–62.

26 郑黎明,邹鹏,韩伟红,等.基于多维熵值分类的骨干网流量异常检测研究.计算机研究与发展,2012,49(4):1972–1981.

27 王铭鑫,周华春,陈佳,等.一种 SDN 中基于熵值计算的异常流量检测算法.电信科学,2015,31(9):1–7.

28 王海龙,杨岳湘.基于信息熵的大规模网络流量异常检测.计算机工程,2007,33(18):130–133.

29 严俊龙,李铁源.基于 SVM 的网络安全风险评估模型及应用.计算机与数字工程,2012,40(1):82–84.

30 Ma XL, Chen YH. DDoS detection method based on chaos analysis of network traffic entropy. IEEE Communications Letters, 2014, 18(1): 114–117.

31 杨君刚,王新桐,刘故箐.基于流量和 IP 熵特性的 DDOS 攻击检测方法.计算机应用研究,2016,33(4):1145–1149.

32 范晓诗,李成海.加权条件熵在异常检测中的应用.计算机应用研究,2014,31(1):203–205.

33 徐倩,程东年,张建辉,等.基于特征熵的异常流识别技术.计算机科学,2012,39(12):38–41.

34 颜若愚,郑庆华.使用交叉熵检测和分类网络异常流量.西安交通大学学报,2010,44(6):10–15.

35 陈世文,邬江兴,黄万伟.融合规则的条件随机场 DDOS 攻击检测方法.计算机工程与应用,2013,49(17):9–11.