DOI: 10.13718/j.cnki.xdzk.2017.07.023

基于流量特征的登录账号密码暴力破解攻击检测方法①

魏琴芳[1]，　杨子明[1]，　胡向东[2]，
张　峰[3]，　郭智慧[3]，　付　俊[3]

1. 重庆邮电大学 通信与信息工程学院，重庆 400065；2. 重庆邮电大学 自动化学院，重庆 400065；
3. 中国移动通信有限公司研究院，北京 100033

摘要：针对暴力破解通过尝试用户所有可能的账号与密码组合来远程登录他人的信息设备或系统，使网络安全面临重大风险的问题，提出一种基于流量特征的远程登录暴力破解检测方法，通过获取通信流量的统计特征，基于进程数量过滤明显的攻击行为；利用数据包特征对数据进行深度分析和再检测. 实验测试结果表明，该方法能识别出针对 TELNET,SSH,FTP 和 RDP 等协议的单机或分布式暴力破解行为，并能取得不低于 98%的检测准确率.

关　键　词：暴力破解；流量特征；字典攻击；网络安全

中图分类号：TN915.08　　**文献标志码**：A　　**文章编号**：1673-9868(2017)07-0149-06

随着移动互联网的快速发展，人们越来越多地使用远程连接来管理云服务、智能设备和计算机等系统中的资源和数据，在带给用户极大使用便利的同时，却也面临着黑客通过暴力破解登录账号密码从而远程窃取用户信息或控制用户设备等风险. 所谓暴力破解，就是攻击者通过尝试所有可能的账号与密码组合远程登录他人的信息设备或系统，进而获得用户的全部使用权限，控制用户主机或系统、窃取用户资料或发动其他攻击，使网络安全面临重大风险的行为.

如 2014 年苹果公司 icloud 泄露大量用户照片的事件就是攻击者利用暴力破解发动的脚本网络攻击；同年，我国铁道部网站 12306 遭遇撞库(暴力破解的一种)攻击，近 10 万条用户信息泄露. 到目前为止，暴力破解作为威胁网络安全的主流攻击手段之一，发展出了更多新型的攻击方式，严重威胁着网络安全.

1　国内外研究现状

近年来，国内外专家、学者相继提出了一系列针对暴力破解攻击的检测方法，包括基于日志审计、基于机器学习、基于流量的暴力破解检测方法等[1-8]. 这些研究成果从不同角度分析或提出了提高暴力破解攻击识别准确率的方法. 但是随着用户对隐私保护的重视以及通信协议内容加密的普及，传统的日志审计方法将无法满足安全需求. 在基于流量特征的攻击识别中，目前的研究重点主要针对单个协议，但真实场景中针对多种协议的攻击是并存的，这也增加了暴力破解行为的识别难度.

本文针对现有暴力破解检测方法的不足，从通信流量的角度分析攻击行为，提出一种基于流量特征的暴力破解检测方法，能有效提高检测率.

① 收稿日期：2016-12-13
基金项目：教育部-中国移动联合研究基金项目(MCM20150202)；国家自然科学科学基金项目(6117029)；重庆市教委科研项目(KJ1602201).
作者简介：魏琴芳(1971-)，女，云南曲靖人，高级工程师，硕士研究生导师，主要从事无线通信系统、信息安全等研究.

2 登录账号密码暴力破解流量特征

2.1 进程特征

应用进程是网络通信的终点，两台网络主机间可能存在不止一个通信进程，在TCP/IP协议族中，运输层使用IP地址和协议端口号来区分主机的不同进程. 通过对大量实验数据分析可知，正常的远程连接中两台主机间的通信进程数维持在一个左右，少有波动；但暴力破解攻击中，攻击者为加快攻击速度，会同时开启多个进程进行攻击；以SSH协议为例，图1是攻击及正常通信进程数对比图.

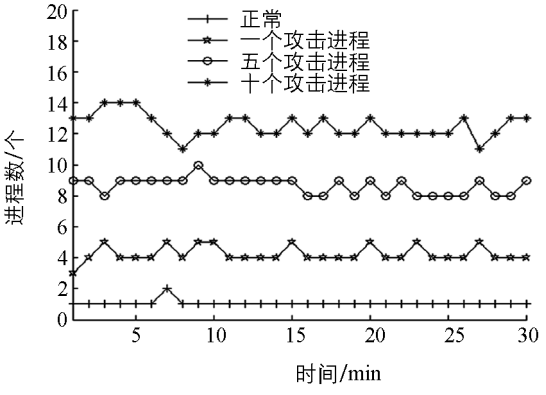

图1 攻击及正常通信进程数对比

由图1可见，即使只使用一个进程发动攻击，1 min里也会出现两个或更多通信进程. 这是因为在验证账号密码时，如果连续验证失败次数过多，两台主机间的通信会自动断开，攻击者想要继续攻击就必须再开启一个新的进程来链接目标主机，攻击者使用了脚本软件，攻击频率比人工输入快很多，所以在攻击过程中会不断开启新端口. 而正常通信中，验证通过后，通信进程是长时间稳定存在的，图中正常链接出现的一次波动是因为用户进行了退出再链接操作. 除了SSH协议以外，TELNET,FTP,RDP通信中也存在这种统计规律. 同时，在分布式暴力破解攻击中，主机发动攻击的频率非常低，会出现每分钟进程数为1或者更低的情况. 此时单凭进程数无法判定异常行为的存在，需要进一步分析数据包特征.

分析主机间的通信进程数量需要对通信的数据包进行统计分析，步骤如下：

步骤1 使用抓包工具抓取流经网卡的数据包，以min为单位保存为pcap文件 $N_i(i=1,2,3,\cdots)$.

步骤2 解析 N_i 中的五元组(源IP地址、源端口、目的IP地址、目的端口和应用层协议).

步骤3 保留 N_i 中含有有用协议的数据包，将数据包分组、相同源IP到目的IP的数据包分为一组.

步骤4 根据端口号统计每个组内的进程数 P_i.

2.2 数据包特征

所有在因特网上传送的数据都是以分组(即IP数据报)为传送单位的，计算机将数据封装成数据包时，会在包头加入一些信息. 通过分析大量的正常数据包和异常数据包包头，发现异常数据在包的平均大小和进程发包数上有很强的统计规律，而正常用户通信中不存在这种规律，可以通过统计数据包包头信息来分析是否存在统计特征作为判定暴力破解的依据. 以暴力破解SSH协议为例，本文使用了Hydra，Bruter，Ncracker 3种暴力破解软件进行试验，并和正常用户数据进行对比，图2是不同情况下数据包平均大小随时间的变化情况.

从图2可以看出，在攻击情况下，每分钟包的平均大小会因为使用的攻击软件不同而发生变化，但是使用同一种攻击软件时数据包大小不会出现太大波动；而正常通信中，数据包的大小是随机变化的、没有统计规律，波动很大. 类似地，通过大量的实验分析发现，这种特征也出现在每分钟进程平均发包数中，以暴力破解SSH为例，图3是不同攻击情形每分钟进程的平均发包数.

除了SSH协议暴力破解，TELNET,RDP,FTP协议的暴力破解也存在类似统计规律. 可以通过分析多个连续统计周期中数据的波动状况作为判断暴力破解的依据.

分布式的攻击其实就是多个主机以低攻击频率进行攻击，样本可能出现不连续的情况. 如果去除样本间的不连续(即通信进程为0的情形)，依然可以通过统计规律发现攻击行为.

判断数据波动首先要对流量进行统计处理，判断步骤如下：

步骤1 使用抓包工具抓取流经网卡的数据包，以min为单位保存为pcap文件 $N_i(i=1,2,3,\cdots)$.

步骤2 解析 N_i 中的五元组(源IP地址、源端口、目的IP地址、目的端口和应用层协议).

步骤 3　保留 N_i 中含有有用协议的数据包，将数据包分组，相同源 IP 到目的 IP 的数据包分为一组.

步骤 4　根据包头信息计算出每个分组的包的总大小 T_i，包数 B_i 并计算每个分组的包的平均大小 $\bar{t}_i=\frac{T_i}{B_i}$ 和进程的平均发包数 $\bar{b}_i=\frac{B_i}{P_i}$.

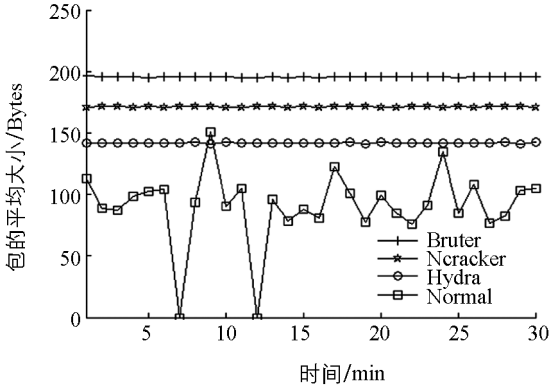

图 2　包的平均大小变化随时间变化图

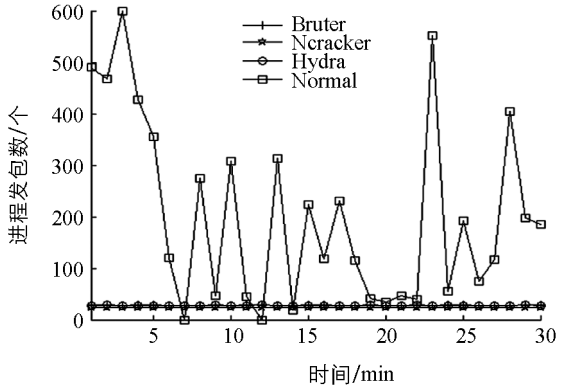

图 3　进程平均发包数变化图

3　检测方法的构建

3.1　进程特征检测

由以上分析可知，正常连接中主机间通信进程数 P_i 少且稳定，其数学期望 $E(P_i)$ 近似为 1；暴力破解行为存在的可能性和进程数成正比，可以根据进程数将通信分为 2 个等级：$1\leqslant P_i<5$ 时，不确定是否异常，需要做进一步检测；$P_i\geqslant 5$ 时为暴力破解攻击. 通过如此分级可以过滤大量明显的攻击数据包，减少计算量，加速检测效率. 分布式攻击中会出现 2 个样本间不连续的情况，可以只保存线程不为 0 的数据样本，这样在逻辑上数据样本依然是连续的，可以进行下一步数据包的特征检测.

3.2　数据包特征检测

由前述分析可知，正常流量中每分钟包的平均大小 $\bar{t}_i$ 和进程平均发包数 $\bar{b}_i$ 波动大，异常流量中 $\bar{t}_i$，$\bar{b}_i$ 波动非常小. 通过计算标准差可以判断数据波动大小，标准差 σ 为

$$\sigma=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{X})^2} \tag{1}$$

其中：N 为样本个数，$\overline{X}$ 为样本均值，标准差越小代表数据波动幅度越小，一组样本的 σ_t 小于阈值 σ_T 且 σ_b 小于阈值 σ_B 时认为发生了暴力破解行为，否则为正常数据. 容易看出，N 越大 σ 越小时识别结果的准确率越高，但是随着 N 的增大密码被破解的可能性也越大，需要找到一个合适的 N 既可以保证准确率也要保证被破解的可能性较低. 在 $1\leqslant P_i<5$ 时才会进入数据包的特征检测，此时的攻击频率约为 10～50 次/min，那么 1 min 内破解最简单的 6 位数纯数字密码的可能性为 0.001%～0.005%，通过实验分析，当 $N=5$ 时，可以很好保证正确率，攻击的成功率仅为 0.005%～0.025%，而且随着密码强度的增加，被破解的可能性随之减小. 检测时，对当前及前 4 个样本计算标准差，不满 5 个数据不计算. 具体步骤如下：

1）将数据包预处理，得到当前分钟每个分组的 $\bar{t}_i$ 和 $\bar{b}_i$.

2）遍历当前分组的历史样本，不满 5 个时不进行计算，只保存样本，否则计算分组的 σ_b 和 σ_t. 每个分组只保存 5 个样本(当前样本加上前 4 个样本)，当分组有新的样本保存进来时，删除最老的样本.

3）σ_b 和 σ_t 与阈值 σ_B，σ_T 比较，都满足小于关系就判定为存在暴力破解行为，否则不存在异常.

4）判定存在异常后，不再保存新来的样本，而是使用罗曼诺夫斯基准则来判断后续样本是否为攻击行为，$\bar{t}_i$，$\bar{b}_i$ 都被判定为异常则样本为异常，否则不存在异常.

5) 被判定为异常或非异常后，当前检测结束，准备进入下1 min的检测. 循环1)－5).

确定了存在暴力破解行为后，还需要判断后续的数据是否为异常数据. 为了减少计算量，后续的数据样本使用罗曼诺夫准则来判断是否为异常数据. 本文使用了5个样本来判定暴力破解行为的发生，样本数量较少，所以按 t 分布的实际误差分布范围来判断后续数据是否和暴力破解数据无明显差异较为合理. 判断步骤如下：

1) 确定样本某个特征的标准差 σ_0.

2) 判断罗曼诺夫斯基准则

$$|x_j-\overline{x}|>K(n,\alpha)\sigma_0 \tag{2}$$

是否成立，其中：x_j 为待测样本的值，$\overline{x}$为样本均值(不包含待测样本)，$K(n,\alpha)$为 t 分布的检验系数(通过查询罗曼诺夫斯基准则检验系数表获得)，n 为样本个数(本文 $n=5$)，α 为显著水平.

3) 如果式(2)不成立，认为 x_j 依然为异常数据，否则，x_j 为正常数据.

可以看出，式(2)中 K 值过小会造成识别正确率下降，K 值过大容易造成正确数据的误判. 本文检验了 $\alpha=0.01(K=6.53)$与 $\alpha=0.05(K=3.56)$时的检测率，发现正确率和误判率没有太大差异. 考虑到正常数据波动是随机的，K 的增加对正确数据的误判影响更大，所以本文中取 $\alpha=0.05$.

3.3 检测方法的实现

根据上述方法，采用Python语言实现暴力破解攻击的检测系统. 该系统可用于发现单机暴力破解和分布式暴力破解行为. 输入为pcap格式的数据包，输出为包含暴力破解行为的IP组. 检测流程如图4所示.

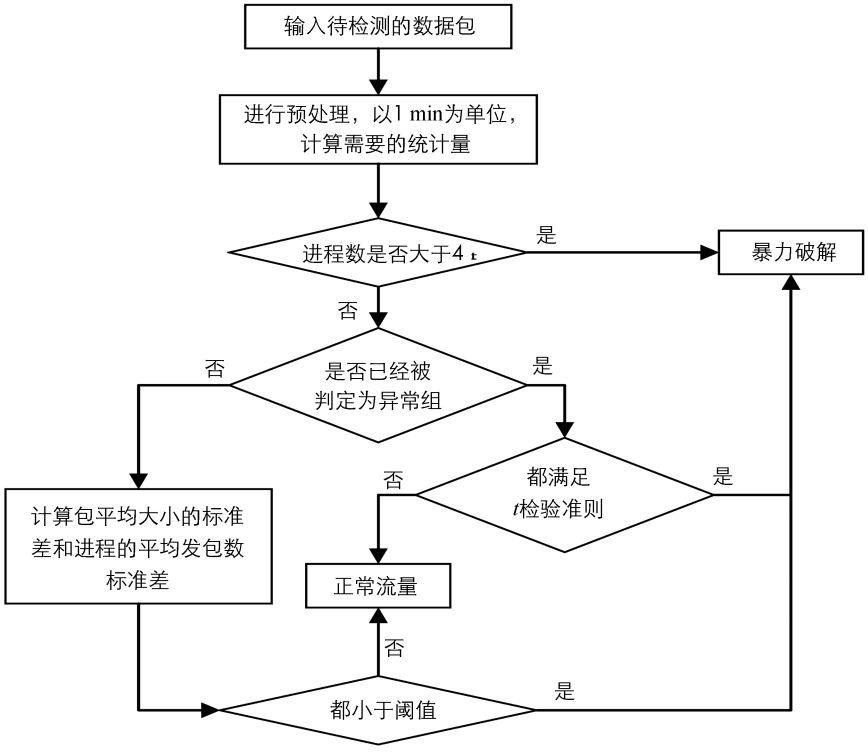

图4 暴力破解行为检测流程

4 实验结果与分析

为了验证暴力破解识别系统的效果，本文搭建了模拟环境来进行测试. 整个实验环境由16台PC机以及1台交换机组成，其中1台主机作为被攻击的服务器，4台主机作为攻击者，其余为正常用户. 测试持续通信1 080 min，期间一共进行了44次攻击，针对TELNET,FTP,RDP,SSH的远程登录账号密码暴力破解攻击各10次，分布式攻击4次(SSH,FTP,TELNET,RDP)，每次攻击持续10 min，总计攻击440 min. 得到的测试结果如表1所示.

表 1　测试结果

通信类型	判定为攻击的时长 M'/min	判定为正常的时长 L'/min
攻击通信时长 $M=440$ min	424	16
正常通信时长 $L=640$ min	5	635

定义暴力破解识别算法的性能评价指标如下：

1）准确度(Accuracy)

$$ACC=\frac{T_{L\to L'}+T_{M\to M'}}{T_{L\to L'}+T_{M\to M'}+T_{M\to L'}+T_{L\to M'}} \tag{3}$$

2）误报率(False Positive Rate)

$$FPR=\frac{T_{L\to M'}}{T_{L\to L'}+T_{L\to M'}} \tag{4}$$

3）漏报率(False Negative Rate)

$$FNR=\frac{T_{M\to L'}}{T_{M\to L'}+T_{M\to M'}} \tag{5}$$

其中：$T_{M\to M'}$是被正确检测出包含攻击的时长；$T_{L\to M'}$是不包含攻击被错误判断为包含攻击的时长；$T_{M\to L'}$是包含攻击被错误判断为不包含攻击的时长；$T_{L\to L'}$是被正确检测出不包含攻击的时长.

结合性能指标定义，基于测试结果可以容易地计算得出检测方法的各项性能指标如表 2 所示.

表 2　性能指标

误报率/%	漏报率/%	准确度/%
0.8	3.6	98.1

经过分析，测试结果中 5 个误报(有 5 min 的正常数据被误判为攻击数据)是因为在正常通信中有 5 个连续的正常数据包的平均大小以及进程平均发包数波动极小，且与阈值相比较都满足小于关系，造成了误判；16 个漏报(有 16 min 攻击数据没有被发现，被误判为正常通信数据)存在于 3 次 TELNET 协议攻击检测中，这是因为在 TELNET 协议的攻击中，其进程平均发包数在少数情况下会出现较大的波动从而影响了标准差和阈值的比较以及罗曼诺夫斯基准则的判断造成漏报.

5　结　论

本文通过对大量暴力破解流量的分析，总结出表征异常数据的流量特征以及描述方法，提出基于流量的登录账号密码暴力破解攻击检测方法，并完成了检测方法的系统开发与实现. 检测系统利用标准差和罗曼诺夫斯基准则提取数据的流量特征，结合通信进程的统计规律，实现对登录账号密码暴力破解攻击的检测和判定. 实验测试结果证明了本文基于流量特征的检测方法对暴力破解判定的有效性，可以识别 TELNET，FTP，RDP，SSH4 种典型协议的单机或分布式攻击，能取得较高的检测准确度和较低的误报率.

参考文献：

[1] VIZVÁRY M，VYKOPAL J. Flow-Based Detection of RDP Brute-Force Attacks [C]//Proceedings of 7th International Conference on Security and Protection of Information. New York：IEEE Computer Society Press，2013：131－137.

[2] THAMES J L，ABLER R，Keeling D. A Distributed Active Response Architecture for Preventing SSH Dictionary Attacks [C]//IEEE Southeast Conference. New York：IEEE Computer Society Press，2008：84－89.

[3] VYKOPAL J，PLESNIK T，MINARIK P. Network-Based Dictionary Attack Detection [C] //2009 International Conference on Future Networks. New York：IEEE Computer Society Press，2009：23－27.

[4] JAEGER D，USSATH M，CHENG F，et al. Multi-Step Attack Pattern Detection on Normalized Event Logs [C]//2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing. New York：IEEE Computer Society Press，2015：390－398.

[5] NAJAFABADI M M，KHOSHGOFTAAR T M，KEMP C，et al. Machine Learning for Detecting Brute Force Attacks at the Network Level [C]// 2014 IEEE International Conference on Bioinformatics and Bioengineering. New York：IEEE Computer Society Press，2014：379－385.

[6] HELLEMONS L, HENDRIKS L, HOFSTEDE R, et al. SSH Cure: A Flow-Based SSH Intrusion Detection System [C]//IFIP International Conference on Autonomous Infrastructure, Management and Security. Berlin: Springer, 2012: 86－97.

[7] VYKOPAL J. Flow-Based Brute-Force Attack Detection in Large and High-Speed Networks [D]. Czech: Masaryk University, 2013.

[8] ABDOU A R, BARRERA D, VAN OORSCHOT P C. What Lies Beneath? Analyzing Automated SSH Brute Force Attacks [C]//International Conference on Passwords. Berlin: Springer 2015: 72－91.

[9] CREECH G, HU J. A Semantic Approach To Host-Based Intrusion Detection Systems Using Contiguous and Discontiguous System Call Patterns [J]. IEEE Transactions on Computers, 2014, 63(4): 807－819.

[10] KHEIRKHAH E, AMIN S M P, SISTANI H A J, et al. An Experimental Study of SSH Attacks by Using Honeypot Decoys [J]. Indian Journal of Science and Technology, 2013, 6(12): 5567 - 5578.

[11] ALSALEH M, MANNAN M, VAN OORSCHOT P C. Revisiting Defenses Against Large-Scale Online Password Guessing Attacks [J]. IEEE Transactions on Dependable and Secure Computing, 2012, 9(1): 128－141.

[12] LEE J K, KIM S J, PARK C Y, et al. Heavy-Tailed Distribution of the SSH Brute-Force Attack Duration In a Multi-User Environment [J]. Journal of the Korean Physical Society, 2016, 69(2): 253－258.

[13] SATOH A, NAKAMURA Y, IKENAGA T. A New Approach to Identify User Authentication Methods Toward SSH Dictionary Attack Detection [J]. IEICE Transactions on Information and Systems, 2015, 98(4): 760－768.

[14] JONKER M, HOFSTEDE R, SPEROTTO A, et al. Unveiling Flat Traffic on the Internet: An SSH Attack Case Study [C] //2015 IFIP/IEEE International Symposium on Integrated Network Management (IM). New York: IEEE Computer Society Press, 2015: 270－278.

[15] SATOH A, NAKAMURA Y, IKENAGA T. A Flow-Based Detection Method for Stealthy Dictionary Attacks Against Secure Shell [J]. Journal of Information Security and Applications, 2015, 21: 31－41.

A Remote Login-Focused Brute-Force Attack Detection Methods Based on Network Flow Characteristics

WEI Qin-fang[1], YANG Zi-ming[1], HU Xiang-dong[2], ZHANG Feng[3], GUO Zhi-hui[3], FU Jun[3]

1. *School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;*
2. *School of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;*
3. *Research Institute of China Mobile, Beijing 100033, China*

Abstract: As one of the main attack means of threatening network security, brute-force attack makes network security encounter large risk by trying all possible combinations of the user's account and password to remotely log in someone's equipment or system. In this paper, a remote login-focused brute-force attack detection method based on network flow characteristics is proposed, which filters out those obvious attacks based on process number by gaining the statistical features of overload, and makes deep analysis and re-detection based on the statistical features of packets. The results of an experiment show that the proposed method can distinguish the single or distributed brute-force attacks in remote login targeted TELNET, SSH, FTP and RDP, and has achieved a detection accuracy of no less than 98%.

Key words: brute-force attack; characteristics of flow; dictionary attack; network security

责任编辑 张 枸
崔玉洁