

一种基于用户会话的异常检测方法

曾永忠^{1,2}, 张 帅¹, 马忠权²

(1. 四川大学 计算机学院, 四川 成都 610065;
2. 西昌卫星发射中心, 四川 西昌 615000)

摘 要:随着网络技术的发展,人们对网络的依赖性越来越强,但同时网络攻击给网络用户造成了严重的信息泄露和巨大的经济损失。如何从浩瀚的用户访问信息中发现对网站具有恶意攻击行为的用户就成为了 Web 服务管理者亟需解决的重要问题。对 Web 服务日志的深入分析后,发现攻击访问用户与正常访问用户在访问 Web 服务时形成的日志记录具有不同的特征。通过特征提取并且进行必要假设后,利用朴素贝叶斯分类算法构建异常检测分类模型,取得了较好的检测效果。

关键词:Web 日志;数据预处理;访问行为;贝叶斯;有监督学习

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2014)07-0141-04

doi:10.3969/j.issn.1673-629X.2014.07.035

An Anomaly Detection Method Based on Session

ZENG Yong-zhong^{1,2}, ZHANG Shuai¹, MA Zhong-quan²

(1. College of Computer, Sichuan University, Chengdu 610065, China;
2. Xichang Satellite Launch Center, Xichang 615000, China)

Abstract: With the development of network technology, the dependence of the network is more and more strong, but simultaneously the network attack cause serious leak of information and huge economic losses. How to find out the attacker from vast user access information is an important issue needs to be solved for Web service administrator. After the deep analysis of Web service log, find that there are different in character between abnormal and normal access users. By feature extracting and making some necessary assumptions, build anomaly detection model using Naïve Bayesian classifier with a good detection effect.

Key words: Web log; data-preprocessing; access behavior; Naïve Bayesian; supervised learning

0 引 言

随着计算机技术的发展和网络的普及,网络给人们带来了极大的生活便利,但同时网络攻击也给用户造成严重的损失。网络攻击特别是针对 Web 应用的网络攻击使入侵检测成为日益关注的焦点。异常检测是入侵检测的一种,它通过建立用户的正常行为模式库,通过被监测用户的实际行为模式与正常行为模式之间的比较和匹配来检测异常。但入侵方法呈现多样化趋势,使得单靠传统的检测技术已难以满足要求,基于智能技术的异常检测正成为研究的一个重要方向。文献[1]通过建立一种双层机制来检测针对 Web 数据库系统的入侵。文献[2]提出了一种基于机器学习的变长命令序列检测模型。文献[3]介绍了针对 HTTP

流量的异常检测模型的设计与实现。文献[4]描述了一种简单高效的针对 SQL 注入的检测方法。文献[5]提出了一种两阶段的异常检测机制。文献[6]介绍了一种基于弱估计的异常检测方法。文中则提出了一种基于用户会话的异常检测方法。该方法首先将 Web 访问用户分成正常访问用户和异常访问用户,通过有监督学习得到朴素贝叶斯分类模型,然后利用模型将待检测用户会话预测为正常会话或异常会话。

1 数据预处理及相关工作

Web 日志挖掘中数据预处理是指剔除 Web 日志中不能反映用户访问信息的无效记录,并根据需要把原始日志数据转换成便于挖掘的数据形式。数据预处

理一般包括:数据清理、用户识别、会话识别以及事务识别等几个阶段^[7]。下面介绍文中的数据预处理过程。

数据清理:由于 HTTP 协议是一个无连接协议^[8],用户每下载一个文件,它都会在日志中增加一条记录。用户访问网站时,网页中嵌入的一些图片等资源也作为用户请求而形成日志记录,但这些日志记录并不是用户的访问请求,从而影响了用户对用户访问行为的分析处理,因此需要将 URL 中后缀名为:.jpg,.jpeg,.cgi,.gif,.js,.css,.swf 等资源请求作为无效日志记录而删除。

用户识别:就是识别每一条日志记录的访问用户。由于本地缓存、代理服务器和防火墙的存在,使得准确识别用户十分困难。通常的用户识别通过 IP 地址、cookie、注册用户等方法来提高用户识别的准确率^[9]。但最为通用的方法是通过 IP 地址来区别不同的用户。文中正是以 IP 地址作为用户的唯一标识。

会话识别:会话是指用户在访问 Web 服务器过程中,从进入网站到离开网站期间对网站一系列的浏览行为^[10-11]。可以将用户会话定义^[12]为 $s = \langle \text{userID}, \text{RS} \rangle$,其中,userID 为用户标识,RS 为用户在一段时间内浏览的 Web 页面的集合。最常用的会话识别方法是时间阈值,包括页面访问时间阈值和持续访问时间阈值。对于前者指的是,给用户在单个网页停留时间设置一个时间阈值 δ ,如果连续的两个页面请求间隔小于 δ ,则两个请求属于同一会话,否则后一个请求是一个新会话的开始。 δ 一般取 10 min^[13];而持续访问时间阈值,是指给定固定的时间阈值 θ ,从会话起始时间 t_0 开始,若当前请求时间 $t - t_0 \leq \theta$,则请求同属一个会话,否则为不同的会话。 θ 一般取 30 min^[13]。

2 基于用户会话的异常检测

2.1 基于会话的异常检测模型

基于会话的异常检测模型主要包括两个阶段。选择合适的用户会话,并根据会话情况将其分成正常会话和异常会话两类,并作为有监督学习的训练集数据,然后采用朴素贝叶斯分类方法对训练集数据进行分类模型的构建;第二阶段利用第一阶段得到的分类模型对用户会话进行分类。检测模型结构如图 1 所示。

2.2 用户会话特征选取

要对用户会话分为正常会话和异常会话两类,首先要明确正常访问行为和异常访问行为的概念。

正常访问行为是指用户通过 Web 服务器授权的合法接口访问服务器中允许访问的资源,符合人们日常生活中浏览网站的行为。不符合正常访问行为的访问则为异常访问行为。

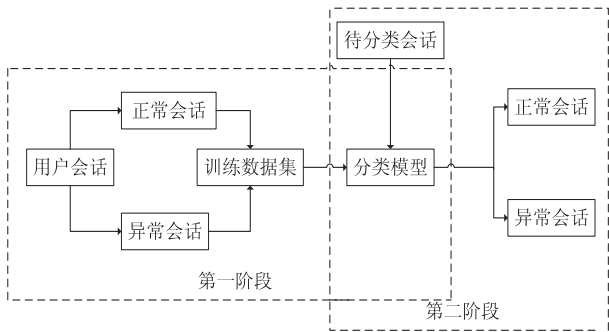


图 1 基于用户会话的异常检测模型

以四川大学某 apache 服务器日志文件为例进行 Web 日志特征的分析。该文件记录了从 2012 年 5 月 4 日至 2012 年 10 月 18 日访问服务器时形成的日志记录,包含了 30 048 个 IP 地址访问服务器时形成的 2 143 681 条日志记录,数据清理后得到 733 653 条有效记录,173 074 个会话,无论是时间跨度还是数据量方面都满足分析的基本要求。

通过对日志记录的深入研究发现,正常访问行为与异常访问行为产生的日志记录在某些特征上具有明显的区别。

1) 会话点击数:也就是在单个会话内,用户向服务器发送的资源请求次数,反映了用户在该会话中点击网页的频率。统计发现单个会话中请求数小于等于 10 的会话高达 168 333 个,占 173 074 会话总数的 97.26%,而会话请求数在 (0,30] 区域内的会话数更是高达 99.6%,如图 2 所示。这说明绝大多数用户在访问网站时,只浏览了少量的网页。这也符合人们浏览网站的习惯。通常,正常访问用户在浏览网页时,一般都是通过网站首页,逐层从导航页面进入自己感兴趣的信息页面或下载所需资料,在获取自己所要信息后就离开网站。

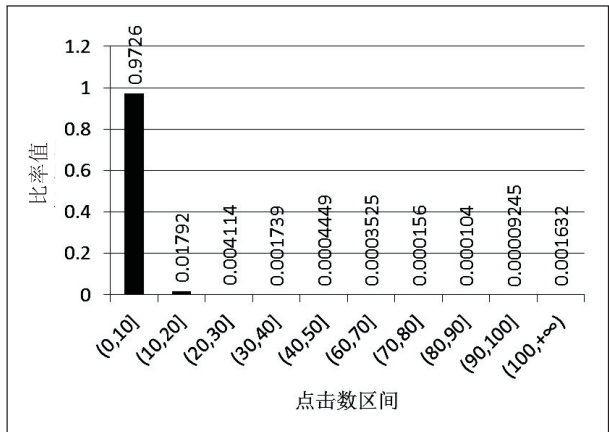


图 2 用户会话点击数分布

2) 日志记录中状态码为“200”所占的比率:在用户的单个会话中,状态码为“200”的日志记录条数所占的比率。状态码“200”表示用户向服务器发送的访问请求得到正确响应,服务器成功返回用户请求的资

源^[11]。因此该比率值的大小直接反映了用户访问网站时得到正确响应的程度。对于正常访问行为,如果会话中点击数比较高,那么其请求成功的比率应该比较高的,因为作为正常访问用户,你肯定忍受不了在点击网站 100 次的情况下只有一次打开了网页。对样本数据进行统计中,请求成功率大于 0.9 的会话数为 133 228 个,去除所有请求为 robots.txt 文件的 11 548 个会话,占总会话数的 82%。

3) 日志记录中状态码为“404”所占的比率:日志记录中状态码为“404”表示用户所请求的资源不存在^[11]。对样本数据进行统计发现,该比率小于 0.1 的会话数为 156 554 个,占会话总数的 90.4%。这充分说明对于正常的用户访问,返回状态码为“404”的日志记录是较少的。这也符合人们浏览网站的习惯,对于正常的网站浏览用户,如果频繁出现打不开的网页,就会放弃对网站的继续访问。

4) robots.txt 文件的访问:robots.txt 是一个协议文件,其功能是规定服务器中允许或拒绝被搜索的资源。对于正常访问行为,用户并不会发出对该文件的请求甚至绝大多数访问用户根本就不知道该文件的存在,因为无论是网站的内部还是网站之外的页面都没有指向 robots.txt 文件的链接^[14],因此若日志中包含有对该文件的请求,则可能是由攻击用户通过扫描软件发送的请求。样本数据中,未产生 robots.txt 文件请求的会话数为 154 354 个,占总会话数的 89.2%。

3 朴素贝叶斯分类

朴素贝叶斯分类是一种基于贝叶斯定理的统计学分类方法。它可以预测类隶属关系的概率,如一个给定的元组属于某一个特定类的概率。该分类法是一种准确度较高的分类器。它可以与决策树和经过挑选的神经网络分类器相媲美。即使用于大型数据库,也能达到令人满意的准确率^[15]。

3.1 贝叶斯定理

设 X 是数据元组,在贝叶斯术语中, X 看作“证据”,通常, X 用 n 个属性集的测量值描述。假设 H 为给定的数据元组 X 属于某个特定的类 C 。在贝叶斯分类中,希望确定给定“证据”或观测数据元组 X ,假设 H 成立的概率为 $P(H|X)$ 。也就是在给定数据元组 X 的条件下, X 属于类 C 的概率。

贝叶斯定理如式(1)所示。

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \tag{1}$$

式中, $P(H|X)$ 是后验概率,是指在条件 X 下, H 的后验概率; $P(H)$ 是先验概率;同样 $P(X|H)$ 是条件 H 下 X 的后验概率。

3.2 朴素贝叶斯分类过程

(1) 设 D 是训练元组和它们相关联的类标号的集合。通常,每个元组用一个 n 维属性向量 $X = \{x_1, x_2, \dots, x_n\}$ 表示,用于描述对元组的 n 个测量。

(2) 假定有 m 个类 C_1, C_2, \dots, C_m 。朴素贝叶斯分类法,就是预测给定的元组 X 属于类 C_i 。要使元组 X 属于 C_i 则必须使 X 具有最高后验概率的类。也就是,当且仅当式(2)成立时,数据元组 X 属于 C_i 。

$$P(C_i|X) > P(C_j|X) \quad 1 \leq j \leq m, j \neq i \tag{2}$$

$P(C_i|X)$ 最大的类 C_i 称为最大后验假设。根据贝叶斯定理,可得

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \tag{3}$$

式中, $P(X)$ 对所有类来说为常数,所以只需要 $P(X|C_i)P(C_i)$ 最大即可。

(3) 为了预测数据元组 X 的类标号,对每个类 C_i ,计算 $P(X|C_i)P(C_i)$ 。当且仅当公式(4)成立时,数据元组属于 C_i 类,从而完成数据元组的分类。

$$P(X|C_i)P(C_i) > P(X|C_j)P(C_j) \tag{4}$$
$$1 \leq j \leq m, j \neq i$$

4 实验和结果分析

数据分类是一个包含两个阶段的过程。第一阶段是学习阶段或训练阶段,在这一阶段要完成分类模型的构建;第二阶段是分类阶段,就是使用第一阶段构建的分类模型预测给定会话的类标号。

4.1 构建分类模型

以第 2 节提及的日志文件为分析对象,如前所述,日志文件预处理后识别出 173 074 个用户会话。以基于会话的四个特征为会话元组属性,并设会话点击次数为 clicknum 属性,会话中状态码“200”所占比率为 ratio200 属性,状态码“404”所占比率为 ratio404 属性,会话中对 robots.txt 请求情况为 robots 属性。并设 C_1 为正常会话类标号, C_2 为异常会话类标号。选取前 150 000 个会话为训练集数据,并根据会话实际情况关联为相应的类标签。

经分析发现 clicknum ≤ 20 时,均为正常会话,而概率会话比率又高达 99.05%,致使正常会话类别 C_1 的先验概率过高而无法进行正常分类。因此在构建模型时作如下假设:当 clicknum ≤ 20 时,均为正常会话。所选训练集 150 000 个会话中 clicknum > 20 的会话数为 1 271,其中有 1 254 个会话标记为正常会话,17 个会话标记为异常会话。根据训练元组计算每个类的先验概率 $P(C_1) = 0.986\ 62$, $P(C_2) = 0.013\ 38$,进而得到模型构建所需的条件概率值 $P(x_i|C_1)$ 和 $P(x_i|C_2)$,如表 1 所示。

表 1 条件概率值

	x_i	$P(x_i C_1)$	$P(x_i C_2)$
clicknum	$20 < \text{clicknum} \leq 40$	0.608 59	0.3
	$30 < \text{clicknum} \leq 300$	0.251 39	0.4
	$\text{clicknum} > 300$	0.140 02	0.3
ratio200	$\text{ratio200} \leq 0.1$	0.095 46	0.85
	$0.1 < \text{ratio200} \leq 0.5$	0.042 96	0.1
	$\text{ratio200} > 0.5$	0.861 58	0.05
ratio404	$\text{ratio404} \leq 0.1$	0.902 94	0.05
	$0.1 < \text{ratio404} \leq 0.5$	0.015 91	0.05
	$\text{ratio404} > 0.5$	0.081 15	0.9
robots	$\text{robots} = 0$	0.972 13	0.684 21
	$\text{robots} \neq 0$	0.027 87	0.315 79

由公式(4)以及表 1 各条件概率值可计算待分类会话属于正常会话 C_1 和异常会话 C_2 的概率值分别为:

$$P_1 = P(x_1 | C_1)P(x_2 | C_1)P(x_3 | C_1)P(x_4 | C_1)P(C_1)$$
$$P_2 = P(x_1 | C_2)P(x_2 | C_2)P(x_3 | C_2)P(x_4 | C_2)P(C_2)$$

如果 $P_1 > P_2$, 则将该会话分类为正常会话; 反之 $P_1 < P_2$, 则分类为异常会话, 从而完成异常检测模型的构建。

4.2 分类预测

要将构建的模型应用于实践, 首先需对该模型的准确率进行评估。评估模型的数据称为检验集, 它与训练集数据一样与由会话元组以及与它们相关联的类标号组成。选取剩余的 23 074 个会话作为分类模型的检验集, 经统计 $\text{clicknum} > 20$ 的会话有 369 个, 其中 362 个会话标记为 C_1 类, 7 个会话为 C_2 类, 经分类模型预测得到的分类结果如表 2 所示。

表 2 分类预测结果

类	标记数	预测正确数
C_1	362	360
C_2	7	7

由表 2 可知, 该模型预测准确率高达 99.46%, 高准确率使模型具有实用价值。将模型应用于 apache 服务所产生的日志文件异常检测。对结果为异常会话的日志进行分析, 发现具有明显的异常特征: 会话中在同一时刻产生的日志条数除去图片等无效日志记录后依然能够达到几十条, 甚至达到上百条, 而在整个会话中的请求次数更是高达 6 432 条, 而对于正常的用户访问根本无法做到; 且状态码为“404”的比率高达 96.54%; 另外从日志中也可看出, 该用户对同一目录下的所有文件发出访问请求, 这种行为只有使用扫描

工具才能做到, 可以将其视为对服务器进行漏洞扫描, 显然这是异常行为。

实验证明, 进行文中所描述的会话特征提取后, 使用朴素贝叶斯分类算法能很好地从日志记录中预测具有异常行为的 Web 访问用户。

5 结束语

文中通过对 apache CLF 格式日志的深入分析, 提取了基于用户会话的四种日志特征, 会话点击数、会话中两种状态码所占比率以及会话中是否包含对 robots.txt 文件的访问请求。以四种特征作为数据元组属性建立有监督学习的训练集数据, 构建了基于朴素贝叶斯方法的分类模型。

实验证明, 该模型都能较好地检测出日志中异常会话。但该模型也存在一定的缺陷, 对用户会话进行分析时, 发现点击次数小于 20 的会话在训练数据集中比率超过 99%, 而且均属于正常访问会话, 如果把这部分数据也作为训练数据则无法构建分类模型, 因此在构建分类模型时大胆假设了 $\text{clicknum} \leq 20$ 的会话均为正常会话, 这同时也致使文中所构建模型只适应于点击次数大于 20 的用户会话。

参考文献:

[1] Shu Wenhui, Daniel T T H. A novel intrusion detection system model for securing Web-based database systems[C]//Proc of 25th annual international computer software and applications conference. [s. l.]:IEEE,2001:249-254.

[2] 孙宏伟,田新广,李学春,等. 一种改进的 IDS 异常检测模型[J]. 计算机学报,2003,26(11):1450-1455.

[3] Kirchner M. A framework for detecting anomalies in HTTP traffic using instance-based learning and k-nearest neighbor classification[C]//Proc of 2nd international workshop on security and communication networks. Karlstad:IEEE,2010:1-8.

[4] Lee I,Jeong S,Yeo S,et al. A novel method for SQL injection attack detection based on removing SQL query attribute values[J]. Mathematical and Computer Modelling,2012,55(1):58-68.

[5] Das D,Sharma U,Bhattacharyya D K. A Web intrusion detection mechanism based on feature based data clustering[C]//Proc of IEEE international advance computing conference. [s. l.]:IEEE,2009:1124-1129.

[6] Zhan J,Oommen B J,Crisostomo J. Anomaly detection in dynamic systems using weak estimators[J]. ACM Transactions on Internet Technology,2011,11(1):16-33.

[7] 陈红丽,李春生,张明. Web 日志挖掘中数据预处理方法研究[J]. 科学技术与工程,2012,20(8):1928-1930.

GR-15)。在训练模型评估与验证阶段,对所选定的训练模型实施十倍交叉验证法,并记录各模型对每一个类别捕捉率和整体准确率,如表 4。

表 4 十倍交叉验证法捕捉率及整体准确率结果 %

	DT-RF-30	Bag-GR-20	RF-GR-15
Normal	94.2	93.5	92.5
DoS	100	99.9	99.9
Probe	98.4	95.9	96
U2R	82.5	90.4	89.5
R2L	67	87.5	85.1

该算法与使用传统专家系统的入侵检测系统性能比较如表 5。

表 5 该算法与使用传统专家系统的入侵检测系统性能比较

类别	评估指标	传统专家系统	文中算法
Normal	Recall	99.35%	99.52%
	F-measure	86.63%	85.34%
DoS	Recall	97.4%	97.43%
	F-measure	98.11%	98.64%
Probe	Recall	78.23%	79.22%
	F-measure	84.1%	86.15%
U2R	Recall	10.53%	16.67%
	F-measure	18.6%	27.54%
R2L	Recall	12.56%	13.72%
	F-measure	21.95%	24.73%
Accuracy		93.068%	96.127%
Cost		0.213	0.208 3

由表 5 可以看出,文中算法的整体分类准确率提升到 96.127%,而成本降低为 0.208 3。

3 结束语

文中算法提出孤立点滤除、多特征选取、类别完全加权算法来提升入侵检测系统的检测性能,实验表明该算法可成功改善网络异常入侵检测的分类效能。未来若能通过赋予各分类模型不同的成本值,在最后推

论时实施权重投票,应可获得最佳的分类效果。

参考文献:

[1] Lu Huibin,Xu Gang. A new intrusion detection method based on data mining[J]. Microprocessors,2006,27(4):58-60.

[2] Li Hanguang,Ni Yu. Intrusion detection technology research based on apriori algorithm[C]//Proc of 2012 international conference on applied physics and industrial engineering. Hong Kong:[s. n.],2012:1615-1620.

[3] Wu Suyun,Yen E. Data mining-based intrusion detectors[J]. Expert Systems with Applications,2009,36(3):5605-5612.

[4] 李 睿,肖维民. 基于孤立点挖掘的异常检测研究[J]. 计算机技术与发展,2009,19(6):168-170.

[5] 罗 敏,阴晓光,张焕国,等. 基于孤立点检测的入侵检测方法研究[J]. 计算机工程与应用,2007,43(13):146-149.

[6] 黄 斌,史 亮,姜青山,等. 基于孤立点挖掘的入侵检测技术[J]. 计算机工程,2008,34(3):88-90.

[7] Kamal A H M,Zhu Xingquan,Pandya A,et al. Feature Selection with biased sample distributions[C]//Proceedings of the IEEE international conference on information reuse and integration. Las Vegas,NV:IEEE,2009:23-28.

[8] Kamal A H M,Zhu Xingquan,Pandya A S,et al. Feature selection for datasets with imbalanced class distributions[J]. International Journal of Software Engineering and Knowledge Engineering,2010,20(2):113-137.

[9] 赵晓峰,叶 震. 基于加权多随机决策树的入侵检测模型[J]. 计算机应用,2007,27(5):1041-1043.

[10] 王鹏英,黄 海,黄晓平. 基于加权特征筛选的入侵检测系统[J]. 计算机科学,2012,39(1):89-91.

[11] Tsai Chih-Fong,Hsu Yu-Feng,Lin Chia-Ying,et al. Intrusion detection by machine learning;a review[J]. Expert Systems with Applications,2009,36(10):11994-12000.

[12] 王 骐,王青萍. 一种基于特征的入侵检测模块的优化布置算法[J]. 计算机仿真,2011,28(6):136-140.

[13] 夏永祥,史志才. 基于 GPU 和特征选择的 SVM 入侵检测模型[J]. 计算机工程,2012,38(8):111-113.

+++++
(上接第 144 页)

[8] 刘加伶,范 军. 基于用户访问树的 Web 日志挖掘数据预处理[J]. 计算机科学,2009,36(9):154-156.

[9] 李 志. 基于 Web 服务器日志挖掘的数据预处理技术研究[D]. 成都:电子科技大学,2012.

[10] 顾兆军,李晓红,王 伟,等. Web 日志挖掘中的会话识别方法研究[J]. 计算机技术与发展,2012,22(4):45-49.

[11] 杨 楠. 基于关联规则 Apriori 算法的 Web 日志挖掘[D]. 成都:成都理工大学,2012.

[12] 方 杰,朱京红. 日志挖掘中的数据预处理[J]. 计算机技

术与发展,2010,20(4):17-20.

[13] Spiliopoulou M,Mobasher B,Berendt B,et al. A framework for the evaluation of session reconstruction heuristics in Web-usage analysis[J]. INFORMS Journal on Computing,2003,15(2):171-172.

[14] Stevanovic D,Vlajic N,An A. Detection of malicious and non-malicious Website visitors using unsupervised neural network learning[J]. Applied Soft Computing,2013,13(1):698-708.

[15] Han Jiawei,Kamber M. 数据挖掘概念与技术[M]. 范 明,孟小峰,译. 北京:机械工业出版社,2012.

作者：[曾永忠](#), [张帅](#), [马忠权](#), [ZENG Yong-zhong](#), [ZHANG Shuai](#), [A Zhong-quan](#)
作者单位：[曾永忠, ZENG Yong-zhong\(四川大学 计算机学院, 四川 成都 610065; 西昌卫星发射中心, 四川 西昌 615000\)](#), [张帅, ZHANG Shuai\(四川大学 计算机学院, 四川 成都, 610065\)](#), [马忠权, A Zhong-quan\(西昌卫星发射中心, 四川 西昌, 615000\)](#)
刊名：[计算机技术与发展](#) 
英文刊名：[Computer Technology and Development](#)
年, 卷(期)：2014(7)

参考文献(15条)

1. [Shu Wenhui; Daniel T T H](#) [A novel intrusion detection system model for securing Web-based database systems](#) 2001

2. [孙宏伟; 田新广; 李学春](#) [一种改进的 IDS 异常检测模型](#) 2003(11)

3. [Kirchner M](#) [A framework for detecting anomalies in HTTP traffic using instance-based learning and k-nearest neighbor classification](#) 2010

4. [Lee I; Jeong S; Yeo S](#) [A novel method for SQL injection attack detection based on removing SQL query attribute values](#) 2012(01)

5. [Das D; Sharma U; Bhattacharyya D K](#) [A Web intrusion detection mechanism based on feature based data clustering](#) 2009

6. [Zhan J; Oommen B J; Crisostomo J](#) [Anomaly detection in dynamic systems using weak estimators](#) 2011(01)

7. [陈红丽; 李春生; 张明](#) [Web日志挖掘中数据预处理方法研究](#) 2012(08)

8. [刘加伶; 范军](#) [基于用户访问树的Web日志挖掘数据预处理](#) 2009(09)

9. [李志](#) [基于Web服务器日志挖掘的数据预处理技术研究](#) 2012

10. [顾兆军; 李晓红; 王伟](#) [Web日志挖掘中的会话识别方法研究](#) 2012(04)

11. [杨楠](#) [基于关联规则Apriori算法的Web日志挖掘](#) 2012

12. [方杰; 朱京红](#) [日志挖掘中的数据预处理](#) 2010(04)

13. [Spiliopoulou M; Mobasher B; Berendt B](#) [A framework for the evaluation of session reconstruction heuristics in Web usage analysis](#) 2003(02)

14. [Stevanovic D; Vljajic N; An A](#) [Detection of malicious and non-malicious Website visitors using unsupervised neural network learning](#) 2013(01)

15. [Han Jiawei; Kamber M; 范明; 孟小峰](#) [数据挖掘概念与技术](#) 2012

引用本文格式：[曾永忠, 张帅, 马忠权, ZENG Yong-zhong, ZHANG Shuai, A Zhong-quan](#) [一种基于用户会话的异常检测方法](#)[期刊论文]-[计算机技术与发展](#) 2014(7)