基于时间密度的Web日志用户浏览行为分析

庄力可[1] 张长水[1] 勒中坚[2]

(清华大学智能技术与系统国家重点实验室 北京100084)[1] (江西财经大学计算机系 南昌330000)[2]

摘 要 本文针对Web日志中用户会话识别阈值问题,给出一种基于时间密度的频度分析方法。文中首先将基于时间间隔参数刻度的用户访问频度定义为一个随机向量,给出了随机向量的切尾算法;然后建立频度与IP用户的相关矩阵,矩阵的列为访问频度,矩阵的行为用户IP,矩阵中的每一个值为某一时间间隔的访问频度。通过列向量的聚类分析,对不同类别用户的访问行为进行探讨。最后,对会话识别的阈值进行参数估计,并通过抽样对阈值进行检测和参数修正。

关键词 Web日志挖掘,时间间隔,频度分布,随机向量,会话阈值

Analysis of Browsing Behaviour in Web Log Based on Time Density

ZHUANG Li-Ke[1] ZHANG Chan-Sui[1] LE Zhong-Jiang[2]

(State Key Lab of Intelligent Technology and Systems, Tsinghua University, Beijing, 100084)[1]

(Department of Computer, Jiangxi University of Finance and Economics, Najchang, 330000)[2]

Abstract Facing the threshold of session recognize in Web log mining, a frequency analysis method based on time interval is introduced. First, the visitor frequency of user based on scale parameter of time interval is defined as a random vector. The cut-tail algorithm for random vector is also given. Second, a frequency-user IP relevant matrix is set up, where frequency is taken as row and user IP is taken as column, and each element's value of this matrix is the user's visitor frequency on the time interval. The different IP users are classified by measuring similarity between column vectors and the browsing behaviour is also discussed. Finally, the parametric estimation and test of threshold of session recognize are given by further sampling.

Keywords Web log mining, Time interval, Frequency distribution, Random vector, Threshold of session

1 引言

随着Internet的迅速发展和广泛应用,Web服务器应用到很多领域,而每个Web站点都收集了海量的Web日志数据。Web日志中隐藏着大量有用的信息,发现并利用这些信息是Web站点管理者的迫切愿望。如何有效地发现这些信息,是我们面对的一个课题。一些数据挖掘技术已经引入到Web日志分析过程中。Web挖掘分为内容挖掘、访问信息挖掘和结构挖掘。通过Web挖掘,对总的用户访问行为、频度、内容的分析,可得到群体用户的访问行为和模式,以改进Web服务的设计。本文探讨的是对日志访问信息中频度特性的挖掘的主线索,以此展开对用户会话模式的行为分析。

2 Web日志挖掘的基本原理与研究现状

2.1 Web日志挖掘的基本原理

Web日志挖掘是对海量存储日志数据采用各种方法和技术进行处理,以获取用户的网络行为,以便为站点管理者提供改进意见。Web日志挖掘的研究通常围绕以下三个部分[1]:数据预处理、事务识别、算法实施和模式分析。

数据预处理 Web日志记录了用户访问本站点的信息。Web日志主要包括以下信息:用户访问的IP地址、用户请求访问的时间、请求访问的方式、被请求文件的URL、返回数据和大小等。数据预处理的主要任务是,从原始日志文件中选取出供用户浏览模式发现算法使用的规范化数据,其结果将直接影响到算法处理结果的准确度与可信度。数据预处理包括数据净化、用户识别、会话识别和路径补充等过程。

事务识别 事务识别包括用户识别、会话识别和路径补充。用户识别是将用户和请求的页面相关联的过程,其中主要处理多个用户通过Proxy或Firewall访问站点的情况。在用户识别的过程中,不仅需要Web日志,还需要知道站点各个页面之间的拓扑结构。会话识别是将一个用户在一段时间内所有的访问请求进行分解,以得到用户会话。会话阈值设定是访问请求分解的关键。路径补充过程就是将本地或代理服务器缓存所造成的遗漏请求页补充完整。

算法实施和模式分析 挖掘算法实施是对事务识别的结果施用挖掘算法产生规则和模式的过程,包括一般的统计,如每页的访问数、最频繁访问的页面、每页的平均浏览时间等。还包括其他的一些挖掘结果,如序列模式、关联规则、聚类等。最后,模式分析是分析挖掘得到的规则和模式,提取有意义的、感兴趣的规则与模式作为挖掘结果。

2.2 国内外研究现状

目前国内外,已陆续有一些商业化Web日志分析工具[2~4]投入使用。这些工具主要注重统计页面的点击次数、用户在站点的停留时间和用户的URL等。对分析网络个体行

庄力可 博士后,研究方向:计算机网络、人工智能;**张长水** 教授,博导,研究方向:模式识别;**勒中坚** 教授,研究方向:计算机网络。

为和群体的行为特征分析，还存在一定的局限性。在学术界，对于日志挖掘的研究，试图从社会学和心理学的角度，解释数据统计规律和分布特征，以发现日志中隐含的关系，并进行相关的行为分析。在文[5]中 Chen 等人提出了最大前向引用序列 MFR 的概念，并用它将用户会话分割成一系列的事务，以期发现用户浏览模式。在文[6]中 Han 等人则根据 Web 日志建立数据立方体，然后对数据立方体进行数据挖掘。在文[7,8]中 OLAP. Minnesota 大学的 Web Miner 系统提出了一种通用的 Web 日志挖掘的体系结构，自动从 Web 日志中发现用户关联规则和序列模式。在文[9]中 F. Bonchi 等人探讨建立一个智能的 Web Caching，来对日志数据仓库进行挖掘。在文[10]中 Xiao 等人介绍了一种决策树剪枝算法来有效挖掘用户的浏览行为模式。在文[11～17]中还对日志数据的采样及算法实现进行一系列相关的研究工作。

2.3 面临的困难和本文所需解决的具体问题

综上所述，Web 日志挖掘可以分为以 Han 为代表的基于数据立方体的方法和以 Chen 为代表的基于 Web 事务的方法。这两类方法均要进行用户识别和会话识别，而用户识别和会话识别要受到本地浏览器缓存、防火墙和代理服务器等的影响，因此用户识别和会话识别成为 Web 日志预处理阶段挖掘过程问题的关键点和难点。其面临的主要困难不仅在于本地和各级浏览器缓存、防火墙和代理服务器的存在，而且由于 COOKIE 使用条件的约束和相同 IP 不同用户的集中请求行为。本文从基于时间密度的用户频度分析入手，以国家旅游局官方网站的日志文件为研究样本，在后面几节分别分析和研究了用户会话识别中的以下3个问题。

(1) 如何根据用户的频度特征给 IP 用户分类?

(2) 不同类别的用户的行为特征的区别和规律。

(3) 如何根据 IP 用户的分类设定会话阈值?

3 数据准备与预处理

3.1 数据背景

针对本文提出的问题，我们采用了国家旅游局 WWW.CNTA.GOV.CN 网站的 Web 服务器日志数据作为本文的研究基础数据。CNTA 作为目前国家旅游局政府网站，是目前国内最大官方旅游网站，包括中文、英文、日文多种语言版本。访问用户具有广阔的地域性，日志分别记录在每天的文本文件中。本文选取的 Web 样本数据集为2001年10月1日至2001年10月9日的文本日志文件，文件共占7.8GB 字节。测试数据从2001年10月14日至2001年12月31日的文本日志文件中抽样获取。

3.2 数据库建立与数据导入

CNTA 网站的日志文本格式，存有大量多嵌入式文件，并有超长的非法字符存在。我们在 SQL2000的环境下建立日志处理数据库 WLDB，将文本日志文件转换成数据库格式，并将相应的日期格式转换后，对日志记录进行预处理。本文采用 SQL2000中的数据导入格式，将 CNTA 中样本数据集的文本日志文件逐一导入数据库 WLDB 中，导入前对日志文本记录中存在超长链接的非法字符作相应的预处理后，才能将日志文件导入 WLDB 中，本文共导入日志记录34,958,471条。

3.3 数据净化

数据净化就是删除 Web 日志中与访问用户行为无关的数据。日志文件记录了用户访问页面上的所有各类文件，包括大量的 GIF,SWF,CSS,ICO,JPEG,JPG,CGI 等各类嵌入式文件，这些文件均以一条记录存在于 WLDB 数据库中。在通常的日志文件净化过程中，以上各类嵌入式文件全部删除。但 CNTA 上的部分旅游信息以静态图片形式存在，诸如地图、旅游景点线路等，而用户访问旅游网站会经常浏览这类文件，这与其它商业网站会有较大的区别。因此，我们在数据净化过程中保留了这类文件，以便更准确地反映用户的浏览行为。WLDB 经数据处理后，记录总数由34,98,471条记录减至547,279条，有效日志记录占总数的15.64%。

4 用户行为分析

4.1 IP 用户的频度分布

WLDB 经数据净化后生成547,279条访问请求和21,077个 IP 用户。我们首先对 IP 用户数量与访问请求数量之间关系进行研究。根据其分布特性，选取适合的 IP 特征用户作为研究用户行为的样本。图1给出了 IP 用户数量与其访问请求频度的关系图。

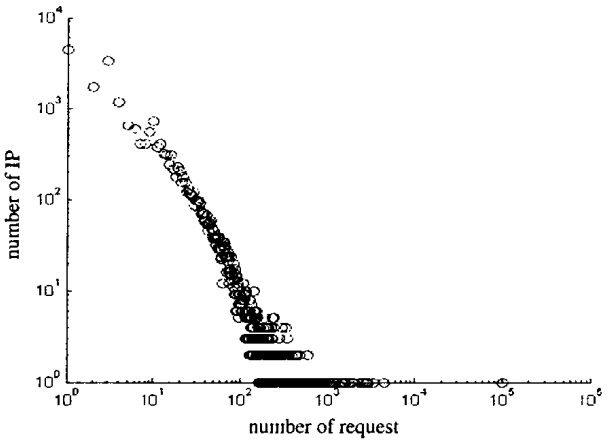

图1 IP 用户访问频度分布图

在图1中，X 轴表示用户访问请求的频度，Y 轴表示 IP 用户的数量。关系图中任一点表示具有某一访问频度的 IP 用户数。在图1中我们已分别对经 X 和 Y 轴取对数，其分布特征与财富分布的 Power-law 现象十分吻合，即大量的 IP 用户只发生少量的访问请求，而少量的用户却存在大量的访问请求。为此，我们进一步对 WLDB 中的数据进行了详细的统计分析。我们注意到，在9天样本数据集中有大量 IP 用户访问请求次数极少的情况，平均每天访问请求次数1～2次，其中在9天内 IP 用户请求频数低于4，占所有 IP 用户的56.06%，尤其是仅发生一次访问请求的 IP 用户多达4471个，占 IP 用户总数的21.21%，这个比例与 Margaret 等在文[18]中21%的统计结果非常接近。这种现象表明，网上的浏览行为具有一定偶然性，IP 用户数量并不能直接反映网站的访问频度。与之相反的是，大量的用户请求却来自部分少量的 IP 用户，其中访问频度最高的一个 IP 用户其访问请求的频数占到了总数的18.56%，也表明不同 IP 的请求行为存在巨大的差异。

我们将 WLDB 数据库中的数据分解，生成 IP 的访问频度分布函数，并根据频度分布函数的概率分布，抽取20个样本作为 IP 频度分布的样本集。采样的 IP 地址和访问请求频数如表1所示。

4.2 定义随机向量

前文已经阐述用户的行为特征描述，要经过用户识别、会话识别和事件识别和模式发现几个阶段。我们对已抽样产生

的20个 IP 用户的行为分析，要解决的关键问题是会话阈值的设定。会话阈值涉及到一个会话事务是否结束，另一个会话事务是否开始。在现有国内外文献中，没有对该问题进行深入的探讨。用户访问请求事务模式挖掘的过程中，会话阈值按经验值设为20分钟至1小时不等。但在相关文献报道中没有说明会话阈值选取的依据，也没有探讨阈值变化与事务挖掘之间的关系。本文通过频度随机向量的概率分布，对 IP 的实际访问频度和会话阈值进行探讨。我们将 IP 在单位时间轴的访问频度定义为时间密度维 Ω 上的一维随机变量 X，$\{(x_k, p_k), k=1,2,\cdots\}$称为 X 在 Ω 上的分布列，其中$\{x_k, k=1,2,\cdots\}$是 X 的所有可能值，$p_k=P(X=x_k)$表示 X 在 k 时的概率。N 个 IP 用户的访问频度，则定义为在同一分布刻度 Ω 上 N 个一维随机向量 $X_1,\cdots,X_n$。分布列中 k 值为 IP 访问频度的最大间隔 M，N 为抽样 IP 样本的个数。

表1 抽样 IP 用户访问频度表

ip	frequency	ip	frequency	ip	frequency	ip	frequency
128.242.109.15	2420	203.198.23.25	1302	203.93.168.71	2839	61.138.138.33	1573
168.160.224.20	1931	203.204.71.242	2994	203.93.175.243	1467	61.144.225.137	2749
193.128.60.250	4473	203.93.144.146	1056	204.166.111.9	1770	61.152.128.21	1429
202.109.116.25	1045	203.93.153.30	1319	209.73.162.112	3465	61.177.3.254	1219

我们首先求出每个 IP 样本的 M 值，确定时间密度 Ω 上的刻度单位（以秒为单位），然后给每个样本 IP 对应的 X 随机向量分布列 x_k 赋值。本文中所有样本的 M 理论值为777,600，实际计算值 MAX(M)=321,717。因为，IP 用户可能存在隔天访问的情况，当 M>86400时，M 的计算值并不能作为 IP 用户实际会话阈值取值判别依据。如果 M 的计算值不作判别而设置阈值过大，反而会引起对会话事务长短作出错误的判断，尤其是对 Proxy 的 IP 用户。在本文的实际计算中，每个 IP 的 M 值不同，因此其对应随机向量的分布列长短也不相同。为使 IP 对应的随机向量 X 在相同维数空间内聚类。我们给出了随机向量分布列的切尾算法。该算法的设计思想是根据随机向量的概率分布 $p_k=P(X=x_k)$切去数列的尾部，参数 $\alpha(0<\alpha<0.1)$为尾部数列概率所允许的最小值，参数 $h(1<h<k)$为数列切尾后剩余数列的个数。其实现过程是：取任一样本对应的随机向量 $X_i\{x_{ik}, i=1,2,\cdots,k=1,2,\cdots\}$，计算 X_{ik}的的概率分布 $p_{ik}=p(X_i=x_{ik})$，从数列尾部开始逐一向前取值，当 p_{ik}取到概率不小于 α 的数列 $p_{ir}(1<r<k)$时，令参数 $h_i=r$，则 $h=\mathrm{Max}(h_i)$。采用切尾算法后产生一维的随机向量列 $X_i\{x_{ih}, i=1,2,\cdots,h=1,2,\cdots\}$。本文实际计算中，切尾参数 α 取值为0.5%，h 取值为897。

4.3 随机向量的聚类分析

我们对随机向量列 X_i 进行聚类，聚类矩阵 Y 由 M 个一维的随机向量 X 构成。聚类矩阵维数 N 即为一维随机向量 X 分布列切尾后的 h 值。聚类样本的个数 M 为一维随机向量 X_i 的个数。聚类前，我们将 X 向量标准化为 Z 向量。

$$Z=\frac{(X-\overline{X})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}}$$

其中 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$

在本文中我们采用马氏距离 d_{ij}和相似系统 c_{ij}来衡量 IP 用户访问频度类别的亲疏程度。

$$d_{ij}=\sqrt{(x_i-x_j)^T V^{-1}(x_i-x_j)}$$

$$c_{ij}=\frac{\sum_{k=1}^{n}(X_{ki}-\overline{X}_i)(X_{kj}-\overline{X}_j)}{\sqrt{[\sum_{k=1}^{n}(X_{ki}-\overline{X}_i)^2]\cdot[\sum_{k=1}^{n}(X_{kj}-\overline{X}_j)^2]}}$$

我们在聚类过程中，为了对比聚类的效果，类间距离分别采用最短距离法、类平均法和重心法和沃德法等，再根据相似系数 c_{ij}来衡量其取聚类的效果。本文在 MATLAB6.5的环境下分别采用不同的方法运算后，其中采用类平均法中类间距离 D_{pq}效果最好，c_{ij}值为0.9928，重心法次之，c_{ij}值为0.9923。图2是 IP 用户聚类后生成的聚类树图。

$$D_{pq}=\frac{1}{n_p n_q}\sum_{i\in G_p, j\in G_q}^{n} d_{ij}^2$$

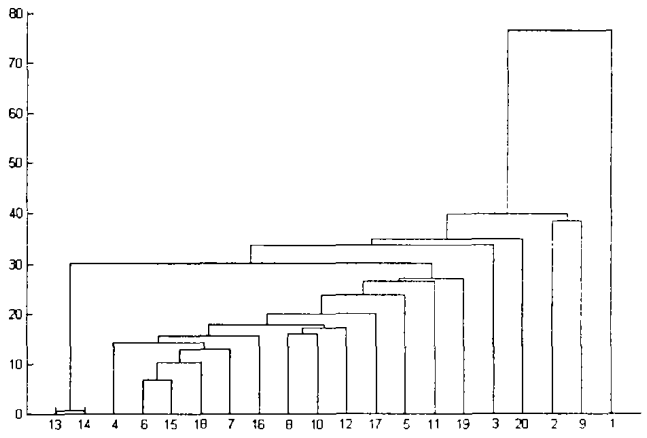

图2 IP 用户聚类树图

4.4 对聚类后特征类 IP 用户的分析

我们对随机向量经 X_i 聚类后结果进行观察，在时间密度 Ω 上它们频度呈现几种态势的分布特性。具有 Proxy 的 IP 用户和非 Proxy 用户，在 Ω 上的频度分布存在显著的差异，其中 Proxy 的 IP 用户大体呈幂律分布，幂律分布中 ρ 参数与 Proxy 中的实际用户数量相关；另一部分具有单独 IP 的用户呈近似正态分布。我们根据不同类别频度分布特征，可以分别置其会话阈值。在本文我们选择聚类后第一类样本 IP01=203.93.109.55作为阈值设置的研究对象。

在图3左图中，我们可以观察到，样本 IP01对应的 X_1向量在同一时间密度 Ω 上与 X 列向量和的 Power-law 分布相比，IP01样本的频度分布特性呈近似正态分布。我们对 IP01的分布形态进行了进一步的研究，并根据其分布特征来估计会话阈值 η。我们保留 IP01向量分布列中的非零项，采用 normfit 函数来估计 μ、σ 值，我们取 alpha 参数为0.01，给出99%的置信区间的估计值。结果为：$\mu=11.3188$，置信区间为[5.8196, 16.8181]；$\sigma=24.7310$。置信区间为[21.3758, 29.2230]。IP01用户99.97%的频度分布在$[\mu-3\sigma, \mu+3\sigma]$区间内，则 IP01的阈值 η 等于99.97%概率分布置信区间的下界 $\mu+3\sigma$：即阈值 $\eta=85.5188$，η 值99.5%的置信区间为[69.947, 104.4871]。

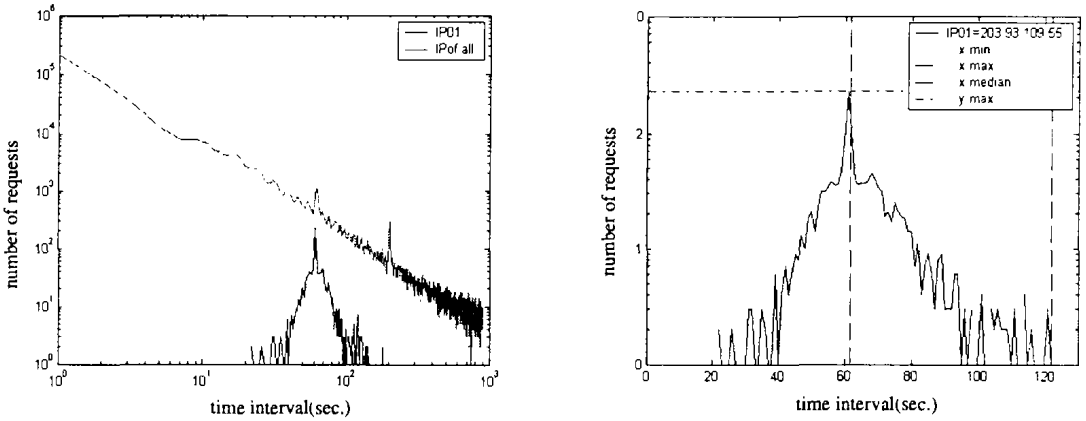

图3 IP01样本基于时间密度的频度分布

5 样本检测与参数修正

为了检测IP01阈值设定的估计，我们抽取其它时间段的IP01数据进行验证。我们在2001年10月11日至2001年12月31日范围，按日志记录的日期随机抽取6个的日志文件，并将其日志文件转换导入WLDB数据库中，经过数据预处理和净化，其中有2样本为空值，即当天未发生访问请求。将另4个测试样本经相同运算后生成4个一维随机向量，与IP01组成一个新的随机向量列，按相同切尾参数我们可以观察到新增的IP01具有同型的概率分布，如图4所示。由于4个测试随机向量频度分布的时间间隔最大值小于85.5188，因此直接减去切尾算法α%的偏差，其分布概率99.5%的置信区间下界分别为70,85,75,50，满足IP01样本阈值η的置信区间。由于样本数量局限性，在IP01正态分布的参数估计中，实际分布99.5%的时间最大值为137，与我们设定的η值下界还存在偏差，我们新增4个样本的随机向列，重新求和对IP01的分布进行参数估计。$\mu=28.4204$，置信区间为[15.6831，41.1576]；$\sigma=61.2020$。置信区间为[53.3571，71.5126]。则新增样本后，阈值η85.5188修正为89.6224，η值99.5%置信区间的下界为255.6954<137，满足实际IP01用户最大的访问请求分布的间隔。

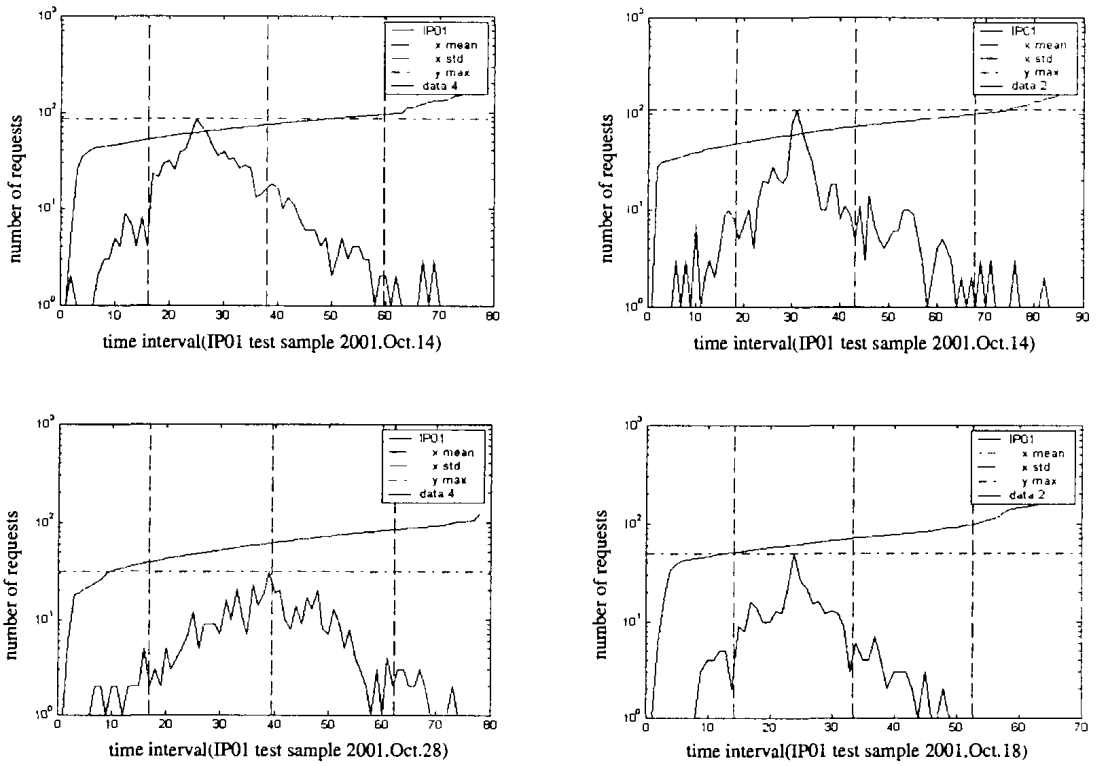

图4 IP01测试样本的频度分布

结论与展望 通过以上的研究和实验，我们得到以下几点结论，并明确未来开展研究的内容。

(1)通过对频度随机向量的聚类分析，可以对用户的请求频度进行数学描述和概率运算，并得到Proxy与单独用户频度分布的不同态势，单独IP用户的访问频度呈现近似正态分布的特性，下一步对其它态势的分布进行研究，对其它类型网站的日志进行挖掘，探讨这些频度特性是否是一种普遍存在的常态。

(2)会话识别阈值的设置，应按不同IP用户的频度分布特性来设置。本文对样本IP的阈值参数估计表明，目前文献中设定经验阈值普遍偏大，对会话识别及后续行为模式算法实现具有一定局限性。

参 考 文 献

1 Burton M C, Walther J B. A Survey of Web Log Data and Their Application in Use-Based Design. In: Proc. of the 34th Hawaii

Intl. Conf. on System Sciences - 2001
2 1SoftwareInc. Webtrends. http://www.webtrends.com,1995
3 OpenMarketInc. OpenmarketWebreporter. http://www.openmarket.com,1996
4 NetGenesisCorp. Netanalysisdesktop. http://www.netgen.com, 1996
5 Chen M S, Park J S, Yu P S. Data mining For Path traversal patterns in a Web environment. In: Proc. of the 16th Int'l Conf. on Distributed Computing Systems. HongKong, 1996. 385～392
6 Zaiane O R, Xin M, Han J. Discovering Web access patterns and trends by applying OLAP and datamining technology on Weblogs. In: Proc. of Advances in Digital Libraries Conf. Santa Barbara, CA, 1998. 19～29
7 Cooley R, Mobasher B, Srivastava J. Grouping Web page references into transactions formining World Wide Web browsing patterns. Department of Computer Science, University of Minnesota: [Tech Rep: TR97-021]. 1997
8 Mobasher B, Jian N, Han E, et al. Web mining: Pattern discovery from World Wide Web transactions. Department of Computer Science, University of Minnesota: [Tech Rep: TR96-050]. 1996
9 Berendt B, Spiliopoulou M. Analysis of navigation behaviour in web sites integrating multiple information systems. The VLDB Journal, 2000, 9: 56～75
10 Giannotti F, Gozzi C. Characterizing Web User Accesses: A Transactional Approach to Web Log Clustering. In: Proc. of the Intl. Conf. on Information Technology: Coding and Computing (ITCC. 02)
11 Adamic L A, Bernardo. The Nature of Market in the World Wide Web, Xeror Research Certer, May 1999
12 Zhong Su1, Qiang Yang. Correlation-Based Web Document Clustering for Adaptive Web Interface Design. Knowledge and Information Systems, 2002, 4: 151～167
13 Ozmutlu H C, Amanda Spink B. Analysis of large data logs: an application of Poisson sampling on excite web queries. Information Processing and Management, 2002, 38: 473～490
14 Tauscher L, Greenberg S. Revisitation patterns in world wide web navigation. In: Proc. of CHI97, 1997. 399～406
15 Ester M, Kriegal H P, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Simoudis E, Han J, Fayyad UM, eds. Proc. of thesecond intl. conf. on knowledge discovery and data mining. AAAI Press, Menlo Park, CA, 1996. 226～231
16 Ester M, Kriegal H P, Sander J, Wimmer M, Xu X. Incremental clustering for mining in a data warehousing environment. In: Gupta A, Shmueli O, Widom J, eds. Proc. of 24rd intl. conf. on very large data bases, New York. Morgan Kaufmann, San Mateo, CA, 1998. 323～333
17 Savasere A, Omiecinski E, Navathe S. An efficient algorithm for mining association rules in large databases. In: Dayal U, Gray PMD, Nishio S, eds. Proc. of 21st intl. conf. on very large data bases, Zurich, Switzerland. Morgan Kaufmann, San Mateo, CA, 1995. 432～444
18 Bonchi F, Giannotti F. Web log data warehousing and mining for intelligent web caching. Data & Knowledge Engineering, 2001, 39: 165～189

(上接第68页)

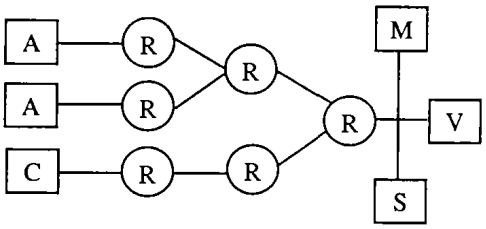

图4 实验网络模型

结论 PPL 对网络和路由器产生的开销都比较低，而且支持逐增配置。例如：在现实环境中，引入这些组件到一个管理域，如一个企业内联网，首先在域中实现回溯功能。如果相邻域也配置了回溯系统，域之间通过其 manager 交换回溯信息来实现跨网络的回溯。

通过我们的仿真研究，可以很明显地看到，路由器上存储量的增长率降低了，同时又没有降低逐跳(hop by hop)回溯的性能。

PPL 对当今协议没有改变，克服了 PPM 方案的一些问题。在 PPM 方案中，路由器必须放一些信息到分组中，这样会产生大量计算并增加了算法的复杂性，而且需对 IP 协议进行一些修改。

回溯所需要的时间与分组在缓冲器中的驻留时间有直接关系。当概率 P 值较小时，分组的驻留时间就会变长，被回溯分组的丢失率就会降低。

PPL 对骨干网更有意义，因为骨干网上的通信量非常高，更有必要采取有效的方法来降低对存储空间的要求。

但是，我们的方法也有一些弱点限制了它在一些情况下的可用性和有效性。用这种方法回溯由单包引发的攻击[6]很困难甚至不可能。该方法是否可有效地用于大规模 DDOS 攻击，还有待于验证。如何决定概率 P 的值也是一个有待于在实践中解决的问题。

参 考 文 献

1 Savage S, et al. Practical Network Support for IP Traceback. In: Proc. 2000 ACM SIGCOMM, ACM Press, New York, Aug. 2000. 295-306. Aavailable online at: http:// www. cs. washington-edu/ homes/ savage/ traceback. html.
2 Baba T, Matsuda S. Tracing Network Attacks to Their Sources. IEEE Internet Computing, 2002. 20～26
3 Song D X, Perrig A. Advanced and Authenticated Marking Schemes for IP Traceback. IEEE INFOCOM 2001
4 Bellovin S M. Security Problems in the TCP/IP Protocol Suite. Computer Communications Review, 1989, 9(2): 32～48
5 Microsoft Corporation. Stop 0A in tcpip. sys when receiving out of band (OOB) data. Available Online at: http://support. microsoft. com/sup-port/kb/articles/Q143/4/78. asp
6 Snoeren A C, et al. Single-Packet IP Traceback. IEEE/ACM Transactions on Networking, 2002, 10(6): 721～734