

P2P 文件共享系统用户行为特性研究 *

陈宝钢 许 勇 胡金龙 张 凌

(华南理工大学广东省计算机网络重点实验室 广州 510641)

摘 要 利用含有用户请求和下载信息的 P2P 文件共享系统 Maze 的日志数据,对 P2P 文件共享系统中的用户行为进行了分析,论证了一些关于用户行为特性的已有观点,并从新的角度考察了一些行为特性,最后对行为的一些相关性进行了比较和研究。相关性分析的结果表明,用户的流量大小与其活跃时间具有很强的相关性,而带宽高低对流量及用户请求次数的影响并没有绝对的关系。

关键词 P2P 文件共享系统, Maze, 用户行为, 特性, 相关性

Research on User Behavior Characteristics of P2P File Sharing Systems

CHEN Bao-Gang XU Yong HU Jin-Long ZHANG Ling

(Communication and Computer Network Laboratory of Guangdong Province, South China University of Technology, Guangzhou 510641)

Abstract In Maze p2p file sharing system, log data record users' request and download information. With the log data, we analyze some of the users' behaviors. In our study, we verify some opinions about the characteristics of users' behaviors, and demonstrate some characteristics from new directions. At last, we compare and explore some of the correlations of the user's behaviors. The result of correlations shows that high correlation exists between traffic volume of users' and active time of users' while users' bandwidth have little influence on users' request numbers and traffic volume.

Keywords P2P file sharing system, Maze, User behaviors, Characteristics, Correlations

1 引言

P2P 文件共享是 P2P 网络技术发展最为迅速的应用,已成为承载了大部分网络流量的互联网主流应用。由于 P2P 文件共享系统是由用户组成的自组织、自治的系统,因而在 P2P 文件共享系统中,用户的行为特性对系统具有很大影响和作用。

文[1,2]发现 Gnutella 和 Kazaa 文件共享系统的大部分用户是贪婪的,不上传或共享文件。文[3]发现 Kazaa 系统用户对完成下载是耐心的,而随着在系统中停留时间的增长,用户会减少对系统的使用,并且发现了用户对系统中文件的获取具有 fetch-at-most-once 的特点。文[4]发现 Gnutella 和 Napster 系统中用户的异构性及参与系统的用户之间缺乏合作性。文[5]发现 FastTrack, Gnutella 和 DirectConnect 系统用户连接其他用户的数量、网络流量、在线时间、平均带宽的分布都是高度偏斜和重尾的,但都不适用 zipf 分布来模型化。文[6]发现 eDonkey, BitTorrent 等系统中用户的连接时间中有很多空转的等待时间,用户之间大量的连接是非正常终止的,流量的“日模式”影响并不明显。文[7]发现 Gnutella 系统中用户的共享文件数与其网络带宽没有绝对依赖关系,大多数用户的共享文件类型与其查询文件类型是相似的。此外,文[8]研究了 Kazaa 系统超级点的生命期和邻居点的选择特性,文[9]研究了 eDonkey 系统用户信号流量和下载流量的不同特点等。

以上研究从多方面描述了 P2P 文件共享系统中用户的行为特性,但是研究大多采用的是插入测量平台收集部分流

量信息的方法,具有一定的局限性,无法完整地获得系统的运行信息,而且很少反映用户行为之间的相关性,一些研究在有些观点上还存在分歧,如文[5]和文[6]对于用户流量是否明显具有“日模式”具有不同的观点等。

2 Maze 系统及其日志

Maze 系统是一个集中式的 P2P 文件共享系统,已有注册用户 300 万,最高同时在线用户数超过 10 万。由于 Maze 系统的客户端实现了在结束一个下载或上传后向日志服务器报告其结果的功能,因而 Maze 系统可以记录整个系统的所有在线用户的文件请求和下载情况,这对准确理解和分析 P2P 文件共享系统的用户行为和特点提供了极大帮助。

我们收集了连续 7 天的 Maze 系统用户活动日志(2005 年 12 月 10 日到 2005 年 12 月 16 日)。这些数据按用户的请求时间来排列,记录了所有在线用户对文件的请求和下载情况。具体如表 1 中所列。

表 1 日志信息汇总

| 日志持续时间 | 7 天 |
|----------|-----------|
| 活动用户 | 126,289 |
| 请求下载次数 | 6,177,011 |
| 请求数据量 | 581.26TB |
| 实际传输量 | 130.28TB |
| 有下载的请求数 | 5,677,718 |
| 没有下载的请求数 | 499,293 |
| 每秒传输的数据量 | 215.41MB |

* 国家 863 计划资助项目(2005AA121560);国家 973 计划资助项目(2003CB314805);国家下一代互联网 CNGI 资助项目(CNGI-04-12-2A)。陈宝钢 博士研究生,研究方向为 P2P 计算和网络测量;许 勇 副教授,研究方向为网络流量分析;胡金龙 讲师,研究方向为网络监控和管理;张 凌 教授,博士生导师,研究方向为网络体系结构。

3 用户行为特性分析

3.1 用户请求和下载的时间特性

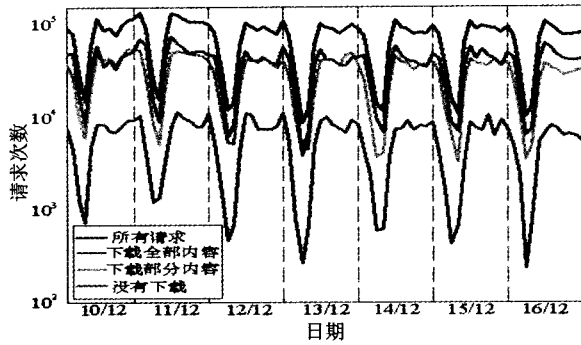


图1 用户请求数变化

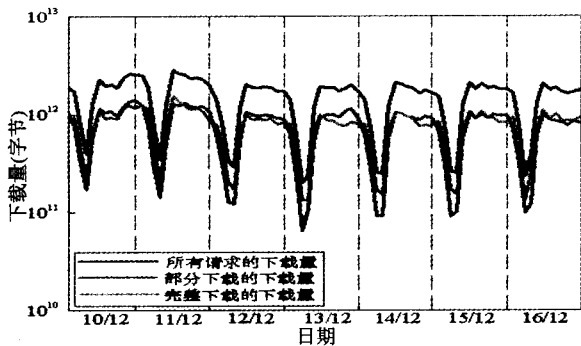


图2 用户下载流量变化

我们首先分析用户提出的下载请求数与时间的关系。根据日志中的下载结果,把请求分为下载了全部文件内容、只下载了部分文件内容以及没有下载文件内容的请求。我们发现,所有用户请求、只下载了部分文件内容的请求、下载了整个文件的请求以及没有下载任何文件内容请求的数量变化趋势几乎是一样的,明显地具有以一天为周期的波动模式。请求数在凌晨(4:00~8:00)达到最低,而通常又在中午 12:00 左右和夜间 12:00 左右达到最高,在中午 12:00 和晚上 12:00 之间的波动比较平缓。同样,对于用户下载量的变化,我们把所有请求的下载量、只下载了部分请求文件的下载量以及完整下载了文件内容的下载量的时间变化趋势进行比较,发现其变化的趋势也基本如此。这表明用户的请求数和下载量的变化与用户使用网络的习惯是有关系的,用户的流量变化具有明显的“日模式”。

3.2 用户的 fetch-at-most-once 行为

Krishna P. Gummadi 发现,94%的时间内,一个 Kazaa 系统用户请求一个同样的文件至多一次;99%的时间内,用户请求一个同样文件至多两次。他把这种现象称为 fetch-at-most-once 行为^[3]。还发现,由于存在 fetch-at-most-once 行为,Kazaa 系统中文件的请求分布并不服从 zipf 分布,最流行文件的请求数远没有按照 zipf 曲线预测的值高,其分布的曲线明显比 zipf 曲线扁平。他认为由于系统中文件内容是不变的,因而 Kazaa 系统用户一般只获取文件一次,因此流行文件的最大请求数被限制。

我们在 Maze 系统中也取大于 100MB 的 rm 和 rmvb 格式的视频文件来分析,发现被完整下载的 100MB 以上的视频

文件流行度分布和上述结论较为接近,而所有大于 100MB 文件的请求分布头部并没有那么扁平,最多的文件下载请求次数和最高的被完整下载的文件请求次数之间相差近 10 倍。一些流行文件请求下载失败次数很多,导致对该文件大量重复的下载请求,但是从趋势上看两种分布都不服从 zipf 分布。文件被完整下载的请求分布在尾部还具有一些重尾分布特征,在文件排名的后部有大量下载次数相同的文件。

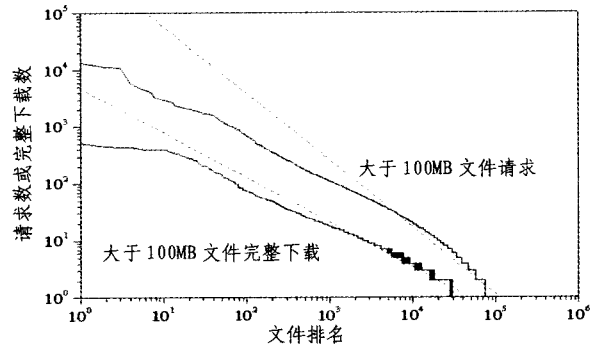


图3 Maze 系统文件流行度分布

3.3 用户的 free riding 行为

在 P2P 文件共享系统一些用户只下载文件而却从不共享任何文件或者只共享没有用的文件。这些用户被称为 free rider,而只下载共享资源却从不提供任何有用的共享文件的行为称为 free riding^[1]。

Eytan Adar^[1]和 Daniel Hughes^[2]的分析认为 free rider 大量存在,少量用户提供了大部分的共享文件。我们发现,Maze 系统中 70%的活动用户从不上传文件,10%的活动用户从不请求下载文件,20%的活动用户既上传了文件也请求下载文件。从不上传文件的用户占了所有用户下载量的 64%,既上传文件也下载文件的用户占据总下载量的 36%,没有请求下载文件的用户上传次数占上传文件总次数的 23%,既请求下载也上传文件用户的上传次数占上传文件总次数的 77%。为了文中分析方便,我们把只下载而不上传的用户称为 free rider 用户,只上传而不下载的用户称为 server 用户,而把既上传又下载的用户称为 servent 用户。

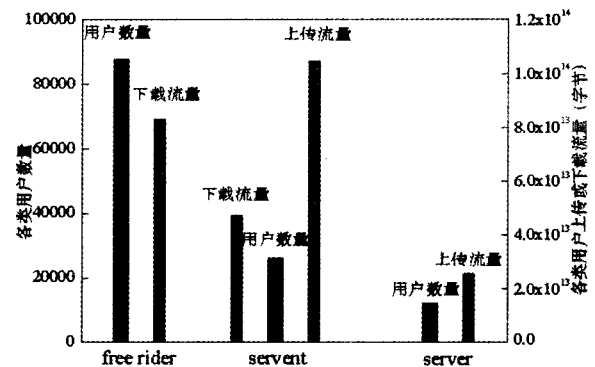


图4 Maze 系统的不同用户数量和流量

对于共享文件数量与 free riding 行为的关系,Eytan Adar^[1]认为用户的共享文件数与是否存在 free ring 行为无关,且查询主要是集中在某些特殊的或流行的文件。我们也发现,Maze 系统中文件被请求次数排名在前 1%的视频文件的被请求次数占了所有被请求视频文件总次数的 29%,79%的视频文件下载请求集中在前 20%的被请求文件中。

3.4 用户的带宽

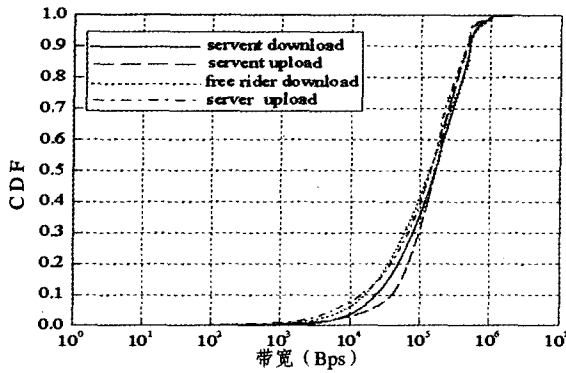


图5 用户上传和下载带宽分布

为了解 Maze 系统中用户的网络状况,我们对用户的带宽进行计算。设用户为 U_i , $S(U_i)$ 和 $D(U_i)$ 分别为其每次的上传量或下载量及每次上传或下载的持续时间, $StartTime(U_i)$ 和 $FinishTime(U_i)$ 分别为其每次传输过程开始和结束时间, $\sum S(U_i)$ 为其上传或下载总量, $\sum D(U_i)$ 为其上传或下载的持续时间总和。带宽计算公式为

$$D(U_i) = FinishTime(U_i) - StartTime(U_i)$$

$$B(U_i) = \frac{\sum S(U_i)}{\sum D(U_i)}$$

对于 free rider 用户及其带宽的分布, Eytan Adar^[1] 认为用户的网络连接速度并没有影响 free rider 的分布。但 Daniel Hughes^[2] 发现不同带宽等价类的用户结点所产生的 QueryHit 消息的值不一样, 因此认为在网络连接速度上 free rider 并不是一致分布的, 低带宽结点更可能是 free rider。在图 5 中, 可以看到三类用户的下载带宽和上传带宽分布在大部分范围内并没有什么差别, 在高带宽和低带宽的分布上的趋势也是近似的。我们认为这些数据可以支持 Eytan Adar 的用户网络连接速度并没有影响 free rider 的分布的观点。

3.5 用户的上传流量和下载流量

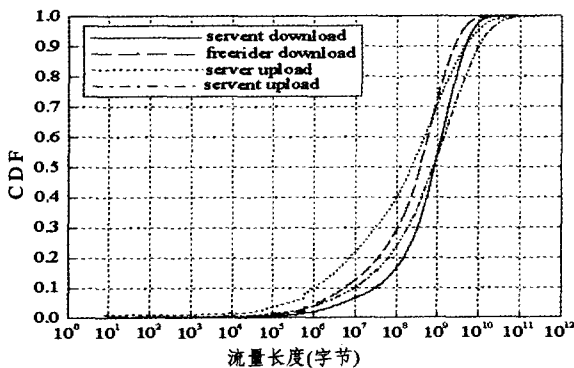


图6 用户上传和下载流量分布

我们研究三类用户上传和下载的流量分布。在图 6 中, 我们看到三类用户的上传和下载流量的分布具有较大区别。在 server 用户中, 46% 的用户上传流量大于 1GB, 而在这 46% 的用户中又有 55% 的用户下载流量超过 1GB, 70% 的用户下载流量大于 500MB, 这 46% 的用户占有整个 server 用户上传量的 97%, 占整个系统上传量的 78%, 占 server 用户下载量的 60%, 占整个系统下载量 22%。Free rider 用户有 27% 下载量超过 1GB, 他们占 free rider 用户下载量的 79%,

占整个系统下载量的 50%。而 29% 的 server 用户上传量超过 1GB, 这些用户的上传量占 server 用户总上传量的 93.4%。从这些结果看到系统中的流量主要是由少部分用户的上传和下载造成的。这些用户上传或下载流量很大, 它们构成了系统运行的动力。系统中上传和下载流量都很大的 server 用户, 对于保障系统正常运行起了很大作用, 这些用户应该是系统设计时被关注的对象。

表 2 用户请求次数、被请求次数和传输流量

| 分类 | | 数据 | | |
|------|------------|---------|----------|--------|
| | | 所有 | <1file | ≥1file |
| 请求次数 | free-rider | 44 | 19 | 54 |
| | server | 85 | 47 | 109 |
| 被请求数 | server | 110 | 7 | 144 |
| | server | 184 | 9 | 211 |
| 下载量 | free-rider | 945.2MB | 534.67MB | 1.09GB |
| | server | 1.81GB | 1.39GB | 2.06GB |
| 被下载量 | server | 2.07GB | 130.23MB | 2.71GB |
| | server | 4.01GB | 193.69MB | 4.61GB |

同时我们也发现, 在每类用户中的下载量和请求数也是有区别的。如表 2 所示, 完整地下载了一个以上文件的用户, 其请求数和下载量明显地多于从没在一次传输中完整地下载了一个文件的用户。我们认为这可能是因为当用户在下载完了一个文件后, 很可能会继续搜索新的文件来下载。而如果他的多次请求都没有得到完整的下载或者下载排队时间很久, 他很可能会放弃搜索并离开系统。

3.6 用户的活跃时间

P2P 文件共享系统中的用户, 具有三种状态: 离线, 没有连接到系统中; 不活跃状态, 连接到了系统中但没有在系统中有任何上传和下载行为; 活跃状态, 用户连接到了系统中并且在系统中有上传或下载行为。由于用户实际在线情况很难精确掌握^[4], 只有通过研究用户在系统中的活跃部分才可以更准确地把握系统中用户的行为特点和作用。

定义一个会话为用户提出下载请求后在对方的下载队列中排队和结束下载之间的时间段, 把传输过程看作用户的下载开始和下载结束之间的时间。经过计算我们发现, 用户的单个会话时间并不长, 平均值在 10min 左右。

使用用户平均会话时间和用户平均数据传输时间来量化研究活跃时间特点。为准确计算时间长度, 要减掉每个用户在同一时间内的多个会话重叠的部分和数据传输重叠的部分。以 $S_1, \dots, S_i, \dots, S_j, \dots, S_n$ 表示用户在一段时间内的每个交迭或连续的会话, T_i 表示每段连续会话时间的长度, T 表示用户在一周的测量时间内的所有会话时间, 则有

$$StartTime(T_i) = \min_{S_i \in T_i} StartTime(S_i)$$

$$FinishTime(T_i) = \max_{S_j \in T_i} FinishTime(S_j)$$

$$T_i = \max_{S_j \in T_i} FinishTime(S_j) - \min_{S_i \in T_i} StartTime(S_i)$$

$$Total\ Session\ Time(T) = \sum T_i$$

结果如表 3 和表 4 所列。可以看出, server 用户和 server 用户在系统中的活动时间较长, 是系统得以继续运转的主要力量。同时我们也看到, 能够一次请求就完整地传输一个文件的 server 用户和 server 用户其会话时间都要多于其他用户。这很可能是由于这些用户是长期在线的用户, 请求下载的用户会更多地对其提出下载请求, 而这些下载请求又能够被完成, 这就增加了这些用户的活跃时间。

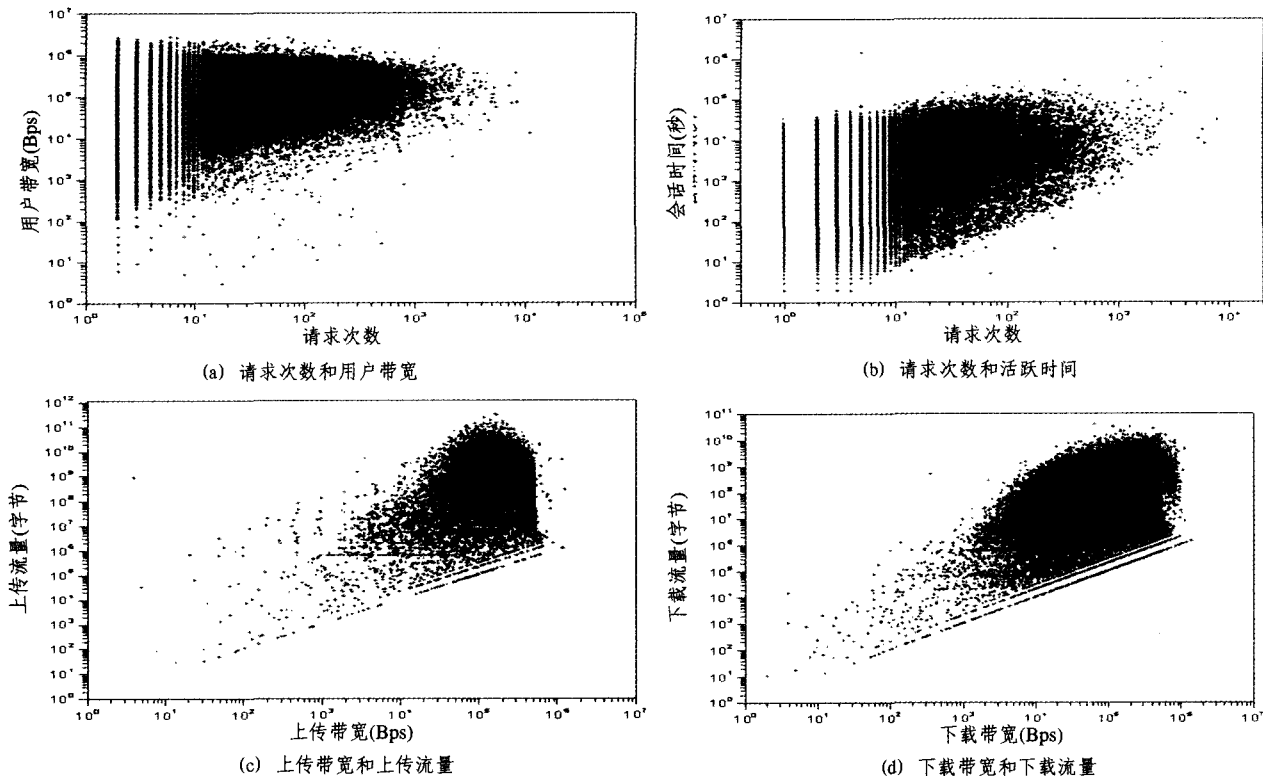


图 7 测量数据的散点图

表 3 会话时间和传输时间长度

| 分类 | | 数据 | 均值 | |
|------|------------|----|-------|-----------|
| | | | 所有(秒) | ≥1file(秒) |
| 测量期间 | server | | 16288 | 20776 |
| | free-rider | | 3851 | 4116 |
| 会话时间 | server | | 20370 | 24529 |
| | free-rider | | 16099 | 20538 |
| 长度 | server | | 3788 | 4046 |
| | free-rider | | 18638 | 22030 |
| 传输时间 | server | | | |
| | free-rider | | | |
| 长度 | server | | | |
| | free-rider | | | |

表 4 会话时间和传输时间所占的比例

| 分类 | | 数据 | 均值 | |
|-------|------------|----|--------|--------|
| | | | 所有 | ≥1file |
| 会话时间 | server | | 0.1885 | 0.2405 |
| | free-rider | | 0.0446 | 0.0476 |
| 占测量期间 | server | | 0.2358 | 0.2839 |
| | free-rider | | 0.0438 | 0.0468 |
| 时间比例 | server | | 0.1863 | 0.2377 |
| | free-rider | | 0.2157 | 0.2550 |
| 传输时间 | server | | | |
| | free-rider | | | |
| 占测量期间 | server | | | |
| | free-rider | | | |
| 时间比例 | server | | | |
| | free-rider | | | |

4 用户行为的相关性

表 5 数据之间的相关性

| 测量变量 (x, y) | 相关性 | |
|----------------|--------|--------------|
| | (x, y) | (logx, logy) |
| 请求次数, 下载带宽 | -0.018 | 0.151 |
| 请求次数, 活跃时间 | 0.233 | 0.577 |
| 活跃时间, 上传流量 | 0.874 | 0.906 |
| 活跃时间, 下载流量 | 0.555 | 0.828 |
| 上传带宽, 上传流量 | -0.053 | 0.435 |
| 下载带宽, 下载流量 | 0.149 | 0.509 |

间的关系,我们研究一些行为的相关性。由于对测量值取对数可以减少其取值范围太大对相关分析所造成的影响,因此我们考察原始数据的相关性,并分析对数转换后的数据的相关性。

从表 5 看到,请求次数和用户下载带宽具有很低的正相关,而请求次数和活跃时间、上传带宽和上传流量、下载带宽和下载流量具有中等的正相关,活跃时间和上传流量、活跃时间和下载流量则获得了较高的正相关。通过分析散点图(图 7),我们发现,请求次数较少的用户的下载带宽分布在很大范围内,高带宽用户的请求次数也在很大范围内变化,但是请求次数最多的用户很可能是带宽高的用户。在请求次数和活跃时间的散点图上,我们看到请求次数越多的用户很可能活跃时间也越长,但是请求次数少的用户的活跃时间在一个很大的范围内变化。上传流量和上传带宽的散点图表明,上传带宽越高,则更有可能上传更大的流量,但上传流量最多的用户并非带宽最高的用户,并且高带宽用户的上传流量在一个很大范围内变化。同样,下载流量和下载带宽的散点图也表明下载带宽越高的用户更可能下载更多的流量,但高带宽用户的下载流量在一个很大的范围内变化。实际上,相关性计算已表明,上传或下载流量的多少同用户活跃时间有着很高的相关性,而同带宽高低并无必然关系。

结束语 通过分析 Maze 系统的数据,我们证实 P2P 文件共享系统中用户的请求和传输流量具有很强的时间特性,大多数用户是自私的用户,各类用户带宽分布较为一致,也发现用户流量及其活跃时间的分布不是均匀的,完整传输过一个以上文件的用户的活跃时间、流量、请求数或被请求数都高于别的用户。此外,我们还发现下载带宽与用户的请求量之间相关性很低,并且用户在系统中传输流量大小和其带宽并

以上分析论述了用户的行为特性。为进一步了解行为之

(下转第 142 页)

确。如果正确,用户就确定服务器知道 K ,从而认证服务器的身份。用户签名 $H_3(y, a, r, Uid)$,也就是 $v = \text{Sig}_U(H_3(y, a, r, Uid))$ 。用会话密钥加密 V 和公钥证书得到 $e = E_K(v, \text{Cert}U)$,将加密结果 e 发送给服务器。

4)解密服务器通过解密 $(v, \text{Cert}U) = D_k(e)$,获取 e 。当公钥证书 $\text{Cert}U$ 被检验是合法的,使用公钥 $v = \text{Sig}_U(H_3(y, a, r, Uid))$ 来验证用户的签名。

4 基于虹膜生物识别的 DRM 系统分析

4.1 性能

基于虹膜识别技术的 DRM 系统具有以下性能:

- 1)系统具备独特的设计和不同的技术分类。能够清楚表示出所执行指令和数字内容保护的详细内容;
- 2)系统拥有识别和监测非法用户的功能;
- 3)在线数字权利保护方案与离线的是一致的,并且具有统计分析功能;
- 4)系统具有钥匙管理功能。虹膜识别技术提供了一种保护合法用户获取访问权限的简单方法。

4.2 安全性分析

本文方案的安全性主要从以下方面考虑:

- 1)每次认证过程中,用户随机选择 $u \in Z_p$,并计算 $a = g^u \bmod p$,不同的 u 保证了协议认证消息的新鲜性,能够防止攻击者重放攻击;
- 2)尽管 $a = g^u \bmod p$ 是以明文形式发送给服务器,而任何人要从 a 中计算出 u 是计算上不可行的,这将面临着求解离散对数的困难性,从而能够实现对 u 的保密性;
- 3)如果攻击者对第 1 条消息进行恶意篡改,或者冒充合法用户发起认证过程,尽管在服务器端无法识别这种攻击,但当用户收到第 2 条消息后,通过验证可以发现这种攻击;
- 4)如果攻击者对第 2 条和第 3 条消息进行恶意篡改,那么由于第 2 条和第 3 条消息使用了数字签名保护,因此,攻击者试图篡改消息而不被发现是不现实的。

通过以上分析,可以发现该认证协议充分考虑了消息的保密性、完整性和新鲜性,从而说明该协议是安全可行的。

结论 本文在虹膜识别技术和现有的 DRM 系统基础上,提出了一种安全的 DRM 系统模型,这种方法可安全地进行数字内容的分发。详细介绍了一种 AIRS 系统,这种系统用于对保护的内容建立接入控制。与现有文献中的方法不同

的是本文提出的模型也给出了用于虹膜生物识别的认证协议,确保了系统的安全性。AIRS 促进了接入控制与内容之间的相互作用。本文提出使用 AIRS 识别机理来进行数字版权管理,这个思想的目的是要引入数据内容的保护措施,以便于只有经允许的用户能够存取制定的数据和登记他们的虹膜代码。如果未经允许的用户想要进入数字内容环境,AIRS 将起禁止作用。因此,使用这种具有不变性的 AIRS,能够公平地使用数字内容。

参考文献

- 1 Federrath H. Scientific evaluation of DRM systems. [Online]. Available at: <http://www.inf.tu-dresden.de/~hf2>
- 2 Hartung F, Ramme F. Digital rights management and watermarking of multimedia content for m-commerce applications. *IEEE Commun. Mag.*, Nov. 2000, 38: 78~84
- 3 Jain A K, Bolle R, Pankanti S. *Biometrics-Persona Identification in Networked Society*. Norwell, MA: Kluwer, 1999
- 4 Ortega-Garcia J, Gonzalez-Rodriguez J, Simon-Zorita D, et al. From biometrics technology to applications regarding face, voice, signature and fingerprint recognition systems. In: D. D. Zhang, ed. *Biometric Solutions for Authentication in an E-World*, Norwell, MA: Kluwer, July 2002. 289~337
- 5 Zhang D D. *Biometrics Solutions for Authentication in an E-World*. Norwell, MA: Kluwer, July 2002
- 6 Waller A O, Jones G, Whitley T, et al. Securing the delivery of digital content over the Internet. *Electron. Comm. Eng. J.*, Oct. 2002, 14(5): 239~248
- 7 Federrath H. Scientific evaluation of DRM systems. [Online]. Available at: <http://www.inf.tu-dresden.de/~hf2>
- 8 Ortega-Garcia J, Bigun J, Reynolds D, et al. Authentication gets personal with biometric. *IEEE signal processing magazine*, March 2004. 50~62
- 9 FAN Kefeng, MO Wei, WANG Meihua, et al. Human identification Technique based on Iris Feature Watermarking. *Chinese Journal of Electronics*, 2006(2): 251~256
- 10 Wildes R P. Automated Iris Recognition: An Emerging Biometric Technology. *Proceedings of the IEEE*, 1997, 85(9): 1348~1363
- 11 Daugman J. Recognizing persons by their iris patterns. In: K. Jain, et al., eds. *Biometrics-Persona Identification in Networked Society*, Norwell, MA: Kluwer, 1999. 103~121
- and Networking, San Jose, CA, 2002. 156~170
- 5 Sen S, Wang J. Analyzing peer-to-peer traffic across large networks. *IEEE/ACM Trans on Networking*, 2004, 12(2): 219~232
- 6 Plissonneau L, Costeux J L, Brown P. Analysis of peer-to-peer traffic on ADSL [C]. In: Proc. of the PAM 2005. LNCS 3431, Heidelberg: Springer-Verlag, 2005
- 7 Goh S T, Kalnis P, Bakiras S. Real Datasets for File-Sharing Peer-to-Peer Systems [C]. In: 201-213, 10th International Conference, DASFAA 2005, Beijing, China
- 8 Liang J, Kumar R, Ross K W. The FastTrack overlay: a measurement study [C]. *Computer Networks: The International Journal of Computer and Telecommunications Networking*, 2006, 50(6): 842~858
- 9 Tutschku K. A measurement-based traffic profile of the eDonkey filesharing service [C]. In: Proc. PAM'04, Juan-les-Pins, France, Apr. 2004. 12~21

(上接第 125 页)

无必然的关系,而是由其活跃时间来决定的。

致谢 本文所用数据由北京大学网络实验室肖明忠老师提供,在此表示非常感谢!

参考文献

- 1 Adar E, Huberman B A. FreeRiding on Gnutella [R]. www.hpl.hp.com/research/idl/papers/gnutella/gnutella.pdf
- 2 Hughes D, Coulson G. FreeRiding Riding on Gnutella Revisited: The Bell Tolls? [C]. *IEEE Distributed Systems Online*, 2005, 6(6)
- 3 Gummadi K P, Dunn R J, Saroiu S, et al. Measurement Modeling, and Analysis of a Peer-to-Peer File-Sharing Workload [C]. In: Proc. SOSP, Bolton Landing, NY, Oct. 2003
- 4 Saroiu S, Gummadi P, Gribble S. A measurement study of peer-to-peer file sharing systems [C]. In: *SPIE Multimedia Computing*