ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.17, No.4, April 2006, pp.939–948
DOI: 10.1360/jos170939

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

基于信任的 P2P 真实性查询及副本管理算法*

李治军+, 廖明宏

(哈尔滨工业大学 计算机科学与技术学院,黑龙江 哈尔滨 150001)

P2P Authenticity Query and Replica Management Algorithm Based on Trust

LI Zhi-Jun+, LIAO Ming-Hong

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

+ Corresponding author: Phn: +86-10-86413213, Fax: +86-10-86412241, E-mail: lizhijun_os@hit.edu.cn

Li ZJ, Liao MH. P2P authenticity query and replica management algorithm based on trust. *Journal of Software*, 2006,17(4):939–948. http://www.jos.org.cn/1000-9825/17/939.htm

Abstract: For the information sharing Peer-to-Peer (or P2P) systems, the document security is an important metric to evaluate the performance, so this paper concentrates on the optimization of document security in file sharing P2P system. For the highly autonomous P2P systems, because the document security in P2P systems mainly depends on two aspects: the security of the documents' carrier and the mechanisms related to the document, such as the replica management, the improvement of the document security can not depend on the improvement of the peers' security, but rely on the mechanisms related to the document. A query protocol sensitive to the document security is designed first in this paper. Based on this protocol, the mechanisms related to the document can be formally described as functions, and the improvement of the system document security can be transformed into the mathematical analyses on the function space. Derived from the results of mathematical analyses, a set of algorithms for replica managements are designed, aiming at improving the document security. In ideal situation, this set of algorithms can achieve the optimization for document security seen from the theoretical analyses, and in realistic systems, the algorithms can obtain good effects, approach to the optimal level. The algorithms are verified by lots of experimental results.

Key words: peer-to-peer network; document authenticity; query protocol; replica management; trust

摘 要: 文档安全性对于信息共享 Peer-to-Peer(或 P2P)系统而言是一项重要的性能指标,以 P2P 系统的文档安全性优化为目标.P2P 系统的文档安全性主要取决于两方面的因素:其载体的安全性和文档相关机制的构造,如副本管理等.对于 P2P 这样高度自主的分布式系统而言,文档安全性的提高无法依赖于结点安全性的提高,而应依靠对文档相关机制的控制来实现.首先设计了一个对文档安全性敏感的查询协议,以该查询协议为基础,与文档相关的机制就可以形式化地表述为函数,而系统文档安全性的提高就转化为函数空间上的数学分析.基于函数分析的结果,设计了一套旨在提高文档真实性的副本管理算法集合.理论分析的结果表明:在理想情况下,该算法集合可达到文档真实性的优化.对于实际系统,经过大量的模拟实验结果验证,该算法集可以获得良好的效果,

* Supported by the Scientific Research Foundation of Harbin Institute of Technology under Grant No.HIT 2002.74 (哈尔滨工业大学基金)

Received 2004-10-10; Accepted 2005-08-24

接近优化水平.

关键词: P2P 网络;文档真实性;查询协议;副本管理;信任
中图法分类号: TP393 文献标识码: A

目前,peer-to-peer(简称 P2P)系统正被广泛地应用于信息共享、分布式计算以及网络数据库等诸多领域中.P2P 系统以其能够充分利用网络资源、良好的自适应性、扩展性好、可靠性高等诸多优点在上述领域中发挥着重要的作用[1].但目前许多 P2P 底层技术还不是很成熟,其中表现得最为显著、并直接影响 P2P 发展的是其安全问题.另外,由于 P2P 系统的自治性和动态性,将导致 P2P 上的安全问题复杂而难解[2].

目前的 P2P 系统主要以信任管理为基础来实现系统安全,其基本思想是结点依据相互的信任程度 Tr_{AB} 来决定结点间的通信[3−5].目前的信任管理主要集中在信任模型的建立和信任值的计算上,而对在信任管理基础上如何构造上层安全系统的研究还很少,且提出的机制也较为简单,如文献[3]提出的 XRep 协议,其基本原理就是在名誉收集的基础上选择最好的 servent 来访问其资源.Yao W 等人在文献[4]中提出了一种基于贝叶斯网构建的信任值计算模型,但在此基础上如何控制查询信息的流向,文献[4]只是简单地选取信任值高的结点进行信息转发.文献[5]给出的是一种基于角色的信任管理机制,同时文中还给出了如何基于信任管理断开和恶意结点的连接,但文中给出的方法很简单,就是和信任值过低的结点断开连接.综合目前的 P2P 信任管理系统可以看出,上述的安全机制都很简单,几乎都可归结为基于信任值的高低排序来决定数据包的转发.与本文思想有些类似的是 Neil D 等人的工作[6],但文献[6]的研究集中在如何通过一些控制策略来抵抗 DoS 攻击上,且文中的策略是零散的,没有一个统一的模型,所以其上的理论分析也不具有整体性.与已有的 P2P 的安全系统相比,本文设计的以信任管理为基础的复杂安全模型具有如下明显区别于其他系统的特点:(1) 本文给出的信息流控制要复杂得多,主要分成 3 个层次,融合信息流向的控制以及文档副本管理等多个方面;(2) 本文提出的模型是以整体模型的数学分析结果为基础,从理论上可以达到优化水平;(3) 同时,本文还考虑了大量的与实际环境相关的因素.

另一方面,现在的 P2P 系统多以文档共享为主.对于 P2P 文档共享平台而言,其文档安全性的提高是一个基本目标.P2P 文档安全性一方面依赖于其载体的安全性;另一方面取决于和文档相关的机制.由于 Peer 行为具有高度自主性,所以通过结点安全性的提高来提高文档安全性是不适宜的,而应该通过文档相关机制的控制来提高文档安全性.本文研究如何通过文档相关机制来实现系统级的文档安全,并以文档真实性作为研究焦点.

本文选取文档真实性作为研究内容是因为文档真实性是目前 P2P 系统中一个重要的安全特性,且目前开展的相关研究还较少,并都存在一定的局限性[7−9].如 PIPE[7]框架集中在文档复制以提高系统可用性和真实性,但仅通过副本增加来实现系统真实性显得粗糙,本文给出的机制较 PIPE 要复杂得多.另外,文献[8]中给出了 P-Grid 结构,并针对电子商务平台上的基本安全问题(包括文档真实性)进行了分析,与本文相比,P-Grid 结构要僵硬得多,可靠性也要差一些.其中,EigenTrust[9]和本文类似,也给每个结点赋一个全局信任值,并基于该值的分布式计算来标识并隔离恶意结点,其效果取决于信任值的准确程度,对于 P2P 而言,全局信任值的精确计算是不可能的.与 EigenTrust 不同,本文只是针对全局信任值进行理论分析,再加上过滤机制的存在,本文对信任值的依赖程度要低得多.另外,虽然本文以真实性作为主要研究对象,但本文的机制也可应用在其他文档安全特性上.

1 基于信任的 P2P 文档真实性安全模型

本文设计的基于信任的 P2P 文档真实性安全模型如图 1 所示.该安全模型主要分为 4 个层次:基础层、信息流动控制层、文档副本管理层和信息过滤层.其中基础层的主要任务是基于信任管理来向上层反映结点的真实性,信息流动控制层实现了结点间信息流向控制,文档副本管理算法实现诸如副本复制、删除及移动等文档管理,信息过滤层的目的是通过对信息的过滤进一步提高安全性.其中,除基础层以外的各层次都是本文的研究内容.基础层直接采用已有的机制,由于该安全模型是分层设计的,对于不同基础层,上层文档管理机制可以直接应用.其中基础层中的信任管理,可选取目前常见的任何信任管理机制[3−5].

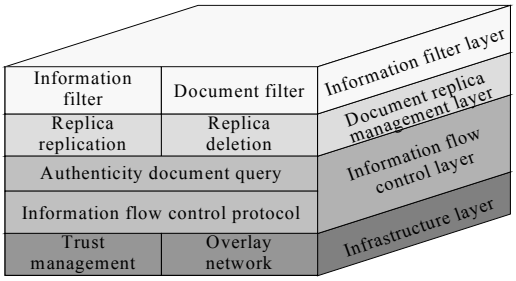


Fig.1 The security model for the P2P document authenticity based on the trust

图 1 基于信任的 P2P 文档真实性安全模型

1.1 真实性查询控制

1.1.1 信息流动控制协议

定义 **1**. 信息流动可定义为一个二元组$(M,v_1v_2\ldots v_n)$.

信息流动$(M,v_1v_2\ldots v_n)$表示信息 M 由结点 v_1 产生,流经 $v_2\ldots v_{n-1}$,最终到达结点 v_n.

对信息流动的控制通过上述信息流动上的规则而体现.就系统真实性而言,可定义如下信息流动控制规则.

① 信息波及范围随信息真实性递增的规则

$$\forall i\in\{2,3,\ldots,n\}\ \gamma(M)\geq\gamma(v_i) \tag{1}$$

其中,$\gamma(v)$为结点 v 的真实性,表现为 v 以往所提供文档的真实性的综合度量,同时也是该结点将来提供文档的可能真实值.式(1)定义的规则将保证真实程度越低的信息对系统中结点的影响越小.但在规则(1)中,使值$\gamma(M)$对于结点序列上的所有结点都具有统一的意义是较为困难的,所以常用下面的规则来代替规则(1).

② 结点序列真实性递增规则

$$\gamma_{v_2}^{v_1}\leq\gamma_{v_3}^{v_2}\leq\ldots\leq\gamma_{v_{n-1}}^{v_{n-2}}\leq\gamma_{v_n}^{v_{n-1}} \tag{2}$$

其中,$\gamma_{v_k}^{v_{k-1}}$ 是 v_k 对 v_{k-1} 的真实性认识.

1.1.2 真实性文档查询

将上述信息流动控制加入查询过程中,就可以形成真实性限制查询协议.这里,定义两种真实性查询协议:

① 令 Q 为定义 1 中的 M.首先定义$\gamma(Q)$=1,然后应用式(1)的规则来控制查询包的转发.假定真实性属于[0,1],所以规则(1)恒真.此时,真实性查询和一般查询一样.根据这一特点,可将其命名为完全性查询.

② 首先定义$\gamma(Q)$为 v_1 对 v_2 的真实性认识,然后应用规则(2)来控制查询请求包的流向.该查询算法可以保证查询应答结点的真实性一定高于查询请求结点,从而保证了返回真实性较高的文档.同时,真实性越高的结点其查询返回的信息就越真实,能够激励结点对自身信息的真实性进行审查,由其特点可称其为限制性查询.

两种算法各具优缺点,其定性比较见表 1.可以看出,完全性查询适合于系统真实性较高的情况,此时可以获得更多的查询结果,而限制性查询适合于真实性较低的系统,此时可以获得真实性较高的结果,减少了无用的查询网络负载以及非真实文档的传播.

Table 1 The qualitative comparison of two query algorithms

表 **1** 两种文档查询算法的定性比较

	All query	Limited query
Difficulty of implementation	Call existent query algorithm directly	Trust management plus comparison of authenticity when transmit
Query load	More query packages	Less query packages
Influence of error in authenticity	Small	Big
The ratio of authenticity for document response	Low	High
The ratio of document response (the ratio of all document response)	High	Low

1.2 文档副本管理算法集(DRMAS)

文档副本管理算法集(document replica management algorithms set,简称 DRMAS)是真实性安全模型中的核心部分,其基本内容是文档副本的移动.副本移动的直观想法就是将真实性高的文档副本移到信息密集的地方,副本的移动是通过副本增加和副本删除的配合来实现的.增加文档副本的基本方法是沿查询应答的返回路径进行副本复制.文档副本的减少可以用两种方法:一是通过生命期来控制文档在结点上的存在时间;二是结点可以对非真实副本进行检测并删除.文档副本增加算法如图 2 所示,副本删除算法如图 3 所示.

```
Algorithm: the document replica replicate algorithm (DRRA)
Input: (P,D) /* P is the set of all peers, D is the set of all documents*/
Output: None
  p∈P, d∈D, t∈P
  p→P:Q(d)/* p send a query for document Q(d) to the system */
    /* q1,q2,…,qk is the peer sequence of peers passed by Q(d), d is stored at t */

  if   adopt all query strategy
    for q∈pk,…,p2,p1
      if   the authenticity of d is not lower than the authenticity of q
        replicate the replica of d at q /* the authenticity of d is γq^t*/

  if adopt limited query strategy
    d is replicated along the path pk,…,p2,p1
```

Fig.2 The document replica replicate algorithm

图 2 文档副本复制算法(DRRA)

```
Algorithm: the document replica delete algorithm (DRDA)
Input: (P,D) /* P is the set of all peers, D is the set of all documents*/
Output: None
  p∈P, e∈E
  if e is the periodical event
    for d∈D   the life of d is decreased
      if   the life of d is 0    delete d

  if e is the event detecting d is a inauthentic document    delete d
```

Fig.3 The document replica delete algorithm

图 3 文档副本删除算法(DRDA)

1.3 信息过滤算法

信息流的控制以及副本的移动可以实现信息流和文档依据其真实性进行重新分布,但是仅仅这样还不能达到系统真实性的最优情况,还需要在这一分布基础上进行过滤.过滤的基本思想是在信息流量和文档真实性接近相随的基础上实现过滤,以某一真实性为阈值(过滤点)进行过滤.过滤的引入使得在理论上文档的真实性可以达到优化水平.另外,由于过滤算法的特点,将大幅度降低底层真实性刻画误差的影响.

过滤算法主要包含两个方面:一是依据结点真实性对查询请求进行信息过滤,算法如图 4 所示;二是以结点真实性为基础进行文档过滤,算法如图 5 所示.

```
Algorithm: the information filtrate algorithm(IFA)
Input: (P,D) /* P is the set of all peers, D is the set of all documents*/
Output: None
  p∈P,d∈D,t∈P
  p→P:Q(d) /* p send a query for document Q(d) to the system */
/* q1,q2,…,qk is the peer sequence of peers passed by Q(d), d is stored at t*/
  if qk figures that the authenticity of t is higher than some threshold tv
    qk sends Q(d) to t
  else
      qk returns the message of failure of Q(d)
```

Fig.4 The information filtrate algorithm

图 4 信息过滤算法(IFA)

```
Algorithm: the document filtrate algorithm(DFA)
Input: (P,D) /* P is the set of all peers, D is the set of all documents*/
Output: None
  p∈P,d∈D,t∈P
  p→P:Q(d) /* p send a query for document Q(d) to the system */
  /* q1,q2,…,qk is the peer sequence of peers passed by Q(d), d is stored at t*/

  for q∈pk,…,p2,p1
    if the authenticity of d is higher than some threshold tv
      replicate a replica of d at q with probability of the authenticity of d
        /*suppose the authenticity of d is less than 1*/
```

Fig.5 The document filtrate algorithm

图 5 文档过滤算法(DFA)

其中,信息过滤将使得影响系统真实性的文档集中在真实性较高的那一部分上,而文档过滤将提高真实性高于过滤点的文档的真实性,两种过滤的配合可以形成理论上的优化条件.信息过滤和文档过滤的核心在于过滤点 t_v(对于两种过滤,过滤点是相同的)的确定上.

2 P2P 文档真实性安全模型的理论分析

2.1 用函数表示系统中文档的真实性

为了进行理论分析,用函数来表示系统中文档的真实性.该文档真实性函数定义为Γ:$D\to[0,1]$.其中 D 为全部文档的集合;$\Gamma(d)$为文档 d 的某种安全性程度,此处是 d 的真实程度.为了分析方便,同时在不影响分析结果的前提下,可以对函数Γ进行适当的调整.由于本文主要考察的是文档分布对真实性的影响,所以$|D|$值是不影响分析结果的,但由Γ的定义可以看出,$|D|$值的不同将会影响函数的变化,所以可对$|D|$值进行归一化处理.另一方面,文档 d 副本数量的多少也会影响系统的真实性.例如,如果 d 的真实性很高,且 d 的副本个数也很多,那么 d 对系统真实性的贡献就较大.基于上述两方面原因,对函数Γ的自变量在考虑文档 d 副本数量的基础上对$|D|$归一化,具体方法是:令区间$[x,x+dx]$为文档 d 对应的自变量范围,其中 dx 为文档 d 副本个数在系统文档总数(包含副本在内)中所占的比率,区间$[x,x+dx]$上的Γ值等于 d 的真实性.且针对不同的文档,x 可以任意选取,但各归一化区间$[x,x+dx]$互不相交.这样处理后,D 集所有元素的归一化区间将覆盖区间$[0,1]$,这样,真实性函数就成为$\Gamma(x)\in[0,1]$,且 $x\in[0,1]$.

显然,$\Gamma(x)$满足 Lebesgue 可测且非负,仿照 p 范数,可以定义$\Gamma(x)$的 1 范数(式(3))和 2 范数(式(4)).

$$\left\|\Gamma(x)\right\|_1=\int_0^1\Gamma(x)\mathrm{d}x \tag{3}$$

$$\left\|\Gamma(x)\right\|_2=\sqrt{\int_0^1\Gamma^2(x)\mathrm{d}x} \tag{4}$$

从系统真实性角度出发,$\|\Gamma(x)\|_1$ 描述的是系统中所有文档的平均真实性.而$\|\Gamma(x)\|_2$ 反映了引入某种信息流动控制后系统的真实性,具体地说就是使结点接收信息数量随其真实性大小相应变化时的系统真实性.

定理 **1**. 对于$\Gamma(x)$恒有$\|\Gamma(x)\|_1\le\|\Gamma(x)\|_2$.

证明: $\int_0^1\Gamma(x)\mathrm{d}x)^2-\int_0^1\Gamma^2(x)\mathrm{d}x=\int_0^1\int_0^1\Gamma(x)\Gamma(y)-\frac{\Gamma^2(x)+\Gamma^2(y)}{2}\mathrm{d}x\mathrm{d}y\le 0$. □

由定理 1 可以看出,采用了适当的文档相关机制后,系统的真实性就会在绝对真实性的基础上有所提高,这就给本文的研究找到了理论依据.

2.2 用函数表示系统结点上的信息分布

同样地,也用函数来表示结点上的信息分布.该函数定义为ω:$P\to R^+$,其中 P 为 Peer 集.由于文档会在结点上备份,所以需要定义函数Ω:$2^P\to R^+$.Ω的计算方式有很多种,本文采用如式(5)所示的方式,其中 $H=\{p_1,p_2,\ldots,p_k\}$.

$$\Omega(H)=\Sigma\omega(p_i) \tag{5}$$

2.3 文档真实性和信息分布的相伴结构

基于上述函数定义,系统运行时文档 d 的真实性就是下面的泛函值:

$$\Gamma^r(d)=\int_{P_d}\gamma(d_p)\omega(p)\mathrm{d}p \tag{6}$$

其中 P_d 为存放 d 副本的结点集;$\gamma(d_p)$为结点 p 处 d 文档副本的真实性.式(6)描述的是系统运行时文档 d 的真实性,体现了相伴的微观结构.而系统运行时所有文档的真实性均值,即相伴的宏观结构,就是下面的泛函值:

$$A[\Gamma,\Omega]=\int_0^1\Gamma(x)\Omega(x)\mathrm{d}x \tag{7}$$

由式(6)和式(7)可以看出,文档真实性和信息分布从微观和宏观角度都具有完全一样的相伴结构,而泛函就是该相伴结构的具体表现.本文的理论目标就是使式(7)的泛函值 A 达到最大.

2.4 文档真实性和信息分布相伴结构的最大值分析

对式(7)的分析需要分为两种情况:

(1) 对系统的绝对真实性不加限制

此时,A 会随着文档真实性的增加而不断增大,直到$\Gamma(d)=1$.但由于 P2P 系统的高度自治性,依靠对结点真实性的提高来提高系统真实性是不适合的.所以应在$\|\Gamma(x)\|_1=C$(定值条件)下,通过文档机制的构造来使 A 最大.

(2) 限制系统的绝对真实性为定值

定理 **2**. 在定值条件下,当$\Gamma(x)$和$\Omega(x)$满足式(8)时,式(7)的泛函值 A 达到最大.

$$\Gamma=\chi(\Omega)\vee\Omega=\chi(\Gamma) \tag{8}$$

证明:实际上,系统中的信息总量也可以认为是定值,即$\Omega(x)$也满足定值条件.由于式(7)对于$\Gamma(x)$和$\Omega(x)$是完全对称的,所以无论固定$\Gamma(x)$来讨论信息流的运动,还是固定$\Omega(x)$来讨论真实性分布的变化,其结论完全一样.

① A 为有界线性泛函

将$\Gamma(x),\Omega(x)$中的一个固定后,泛函 A 显然是线性的,又由 Hölder 不等式可知,有泛函值 A 有界.

② 泛函 A 一定存在最大值

显然,上述函数空间在 p 范数定义下形成距离空间,现在证明在 L^1 空间上的集合 $D=\{f(x)\in L^1|0\le f(x)\le 1\}$为闭集,设 D'为 D 的导集,可证$\forall f\in D'$必有 $0\le f\le 1$,又因 L^1 完备,所以 $f\in D$,故 D 为闭集.

③ 用变分法进行分析

将函数的基本条件 $0\le\Gamma(x)\le 1$ 加入,应用欧拉方程有$\Omega(x)+\lambda_1(x)+\lambda_2(x)+\lambda_3=0,\lambda_1\omega_1=0,\lambda_2\omega_2=0$.由于$\|\Gamma(x)\|_1=C$,不能有$\omega_1=0$ 或$\omega_2=0$,所以只能有$\lambda_1=\lambda_2=0$.当$\lambda_1=\lambda_2=0$ 时,有$\Omega(x)=-\lambda_3$,其中λ_3 为常数.而信息分布函数是一个任意函数,得出矛盾.这表明极值不能取在内部而只能取在边界上. □

由式(8)的形式可以看出,相伴结构体现的是信息分布和文档真实性之间的相互过滤结构.

3 P2P 文档真实性安全模型评价参数的定义及相关算法的简要分析

3.1 文档真实性安全模型上的若干评价参数

(1) 查询真实率(QAR)

QAR 是评价系统真实性的基本参数指标,是真实的查询结果在全部查询结果中所占的比率.

(2) 查询成功率(QSR)

QAR 将没有返回结果的查询也算作真实的查询.QSR 定义为返回真实结果的查询在全部查询数中的比率.

(3) 查询数据包数(QPN)

QPN 定义为单个查询引起的数据包数,反映了查询造成的网络负载.

(4) 查询平均跳数(QAH)

QAH 参数反映了查询的效率.

3.2 真实性查询控制协议和文档管理算法的分析

3.2.1 未引入真实性机制前的分析

在未引入真实性机制之前,系统的真实性完全依赖于底层覆盖网结构.由于本文提出的真实性路由协议是以 Gnutella[10]为基础的,所以这里我们首先对 Gnutella 进行分析.假定结点数为 N,广播半径为 r,TTL 为 L,α为任一结点应答某文档的概率,β为该文档的真实率.假设 Gnutella 沿返回路径上备份文档.此时,完成一次查询后β将变为

$$\beta_{new}=\beta_{old}\cdot\frac{N\alpha_{old}\beta_{old}+\overline{D}}{N\alpha_{old}+\overline{D}}+(1-\beta_{old})\cdot\frac{N\alpha_{old}\beta_{old}}{N\alpha_{old}+\overline{D}}=\beta_{old} \tag{9}$$

显然,对于 Gnutella,系统的文档真实性不发生变化.

3.2.2 DRMAS 的分析

(1) 应用 DRRA 和 DRDA 实现副本移动

副本移动的基本目标是实现信息流分布和文档真实性大小的相随,该相随关系可以表示为函数 $W_d(a)$,函数 $W_d(a)$定义为真实性为 a 的文档波及范围.显然,相随的基本条件是 $W_d(a)$递增.当 $W_d(a)$为递增的线性函数时,相伴效果最好.如果假定文档被删掉的时刻在其生命周期 T_{ed}上均匀分布,那么,任何时刻真实性为 a 的文档被删掉的副本数为 $W_d(a)/T_d$,形式化描述就是 $W'_d(a)=W_d(a)-W_d(a)/T_{ed}$.算法 DRRA 中文档副本的增加是沿查询返回路径进行的.假定查询请求包路由经过的结点序列为 $s,p_1,p_2,\ldots,p_U,t$,当结点真实性在[0,1]上满足均匀分布时,如果从 t 处获得真实性为 a 文档,DRRA 算法将导致该文档副本数增加 Ua.此时有 $W'_d(a)=W_d(a)+Ua/T_{qd}$,其中 T_{qd}为 d 的查询周期.综合后得到式(10):

$$W'_d(a)=W_d(a)-W_d(a)/T_{ed}+Ua/T_{ed} \tag{10}$$

式(10)表明,DRRA 和 DRDA 的组合基本上可以实现信息流动和文档真实性的相随.同时,以该式为基础,也可以确定 DRDA 中的参数 T_{ed}.由于系统存储能力有限,所以文档副本不能无限制地增加.在假定系统中所有文档(包括副本)的总数 M 基本保持不变的情况下,有

$$M/T_{ed}=\int_0^1 Ua/T_{qd}\,\mathrm{d}a\Rightarrow T_{ed}=\frac{2MT_{qd}}{U} \tag{11}$$

在式(11)中,U 通常约为查询请求 TTL 的一半,一般是 5~10,所以 T_{ed}是 T_{qd}的 $0.25M$ 倍.该结果表明,系统中文档的查询频率越小,系统的存储加大,都使得 T_{ed}变大,且 T_{ed}比 T_{qd}大得多.

(2) 过滤算法的分析

过滤算法 IFA(DFA)的核心是过滤点 t_v的确定.显然应该满足信息总量不变的条件,所以应有

$$\max\{\Omega(x)\}\cdot\left|\{x\,|\,\Gamma(x)\geq t_v\}\right|=\int_0^1\Omega(x)\mathrm{d}x \tag{12}$$

4 模拟实验及结果

4.1 模拟实验平台

本文选取 P2P 文件共享系统作为模拟实验平台,并以 Gnutella 作为底层查询协议,底层覆盖网依照 power-law 原则进行组织[11].共享系统包含若干类文档,每类文档用一个关键字进行标识.初始化时,每类文档拥有相同数量的副本,且均匀地随机分布在各 Peer 上,各类文档的真实性在[0,1]上均匀分布.查询模拟是按轮(run)进行的.每轮有 20%的结点发起查询,一轮查询完后进行下一轮.

模拟实验平台的若干基本参数及其初值设定见表 2.

Table 2 Definition and value of symbols in the simulation platform

表 2 模拟实验的符号说明和参数设定

Description	Symbol	Value
Total number of nodes	N	1000
Total number of document type	DN	200
Replica number for document	RN	50 (initial)
Time to life for query	TTL	10
Power-Law $p(k)=k^{-\tau}$	τ	2
The query flood radius	r	Min (k,5)

4.2 实验结果及分析

4.2.1 真实性查询控制的影响

表 3 给出了真实性查询控制对系统的影响.可以看出,引入文档备份的 Gnutella 对系统真实性没有贡献.采用真实性查询控制(这里应用的是完全性查询)以后,系统真实性有了一定的提高,但幅度不大,平均提高 10%,但 QSR 下降得很多.这说明,单纯的信任管理以及简单信息流控制是不能取得良好效果的.

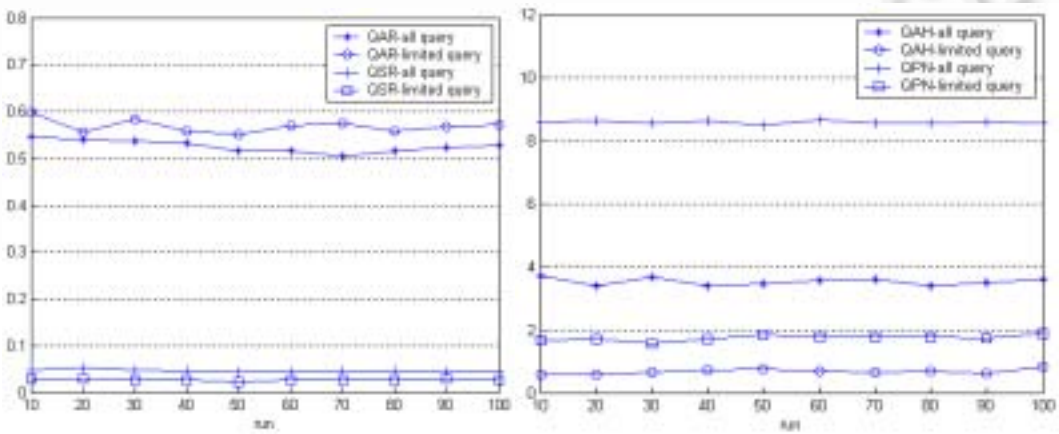
Table 3 Influence of authenticate query control comparing with Gnutella

表 3 真实性查询控制较 Gnutella 对系统性能的影响

	Gnutella without replica			Gnutella with replica			Authenticate query control		
	run=10	*run*=50	*run*=100	*run*=10	*run*=50	*run*=100	*run*=10	*run*=50	*run*=100
QAR	0.469	0.470	0.470	0.469	0.463	0.475	0.528	0.512	0.527
QSR	0.193	0.196	0.195	0.219	0.316	0.373	0.047	0.043	0.050
QPN	9.313	9.535	9.563	9.378	7.907	6.325	8.655	8.553	8.634
QAH	4.545	4.423	4.491	4.347	3.519	2.728	3.556	3.552	3.518

4.2.2 两类真实性查询的比较

图 6 对两类真实性查询机制进行了实验比较.可以看出,两种真实性查询控制的 QSR 基本相同,而限制性查询会在一定程度上提高 QAR.另外,限制性查询较完全性查询而言,查询效率要高得多,QAH 的比例约为 1:6.



(a) The comparison in QAR and QSR (b) The comparison in QAH and QPN

(a) 比较 QAR 和 QSR (b) 比较 QAH 和 QPN

Fig.6 Comparison of two authenticate query controls

图 6 两种真实性查询控制的比较

4.2.3 DRMAS 的影响

(1) DRMAS 对文档波及范围的影响

图 7 给出了 P2P 拓扑结构均匀时 DRMAS 对文档波及范围的影响.由图可以看出,随着查询的进行,DRRA 算法基本保证了 $W(d)$随文档真实性的线性变化.其中,真实性高于 0.5 的部分整体上相随,但存在振荡.这是因为真实性较高的文档副本变化较快,模拟引入的不确定因素体现得较为明显.这也是过滤算法引入的原因之一.

(2) DRMAS 对文档绝对真实性的影响

先将系统中的文档分为若干组(组个数为 G),一组内的文档其真实性在[0,1]上均匀分布.图 8 给出了实验结果,可以看出,在 DRMAS 下,系统真实性有较大幅度的提高,400 轮查询后,QAR 能提高 20%以上(对于各 G 值).另外还可以看出,G 越大,真实性提高的速度越快.这是由于组内文档数较少时,可以较快地将真实程度高的文档选取出来.总的看来,经过较长时间后,DRMAS 的相随机制可以使文档真实性大幅提高.

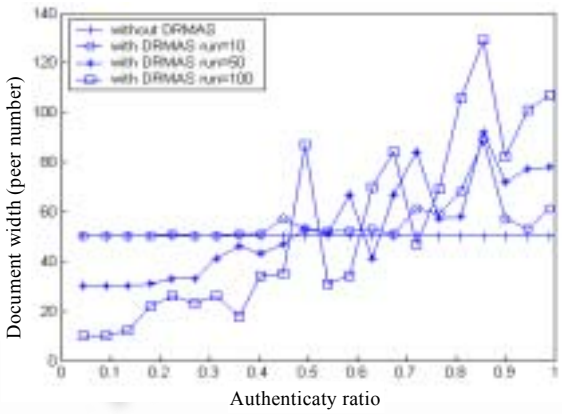


Fig.7 The influence on $W(d)$ of DRMAS

图 7 DRMAS 算法对文档波及范围的影响

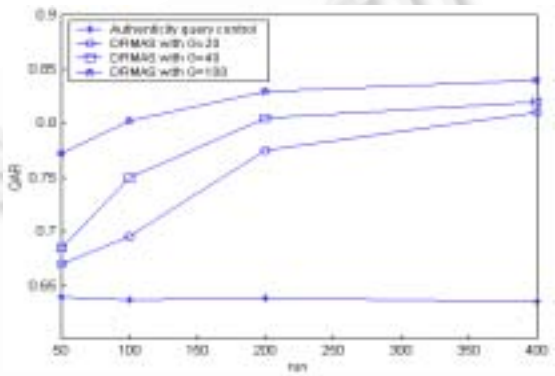


Fig.8 The influence on QAR of DRMAS

图 8 DRMAS 算法对 QAR 影响

(3) 过滤机制对系统真实性的影响

由图 9 可以看出,加入过滤机制后系统的真实性可以进一步提高,在系统运行一段时间后,QAR 约为 90%,模拟实验中设置文档的最高真实性副本的真实性为 0.95,也即基本达到了真实性的最大值.

(4) 系统异构对 DRMAS 的影响

图 10 给出了系统异构对 DRMAS 的影响.虽然这里只考虑了覆盖网的异构,但其他情况是类似的.可以看出,异构的覆盖网可以加速系统真实性提高的速度,且系统的真实性提高过程表现得较为稳定和平滑.

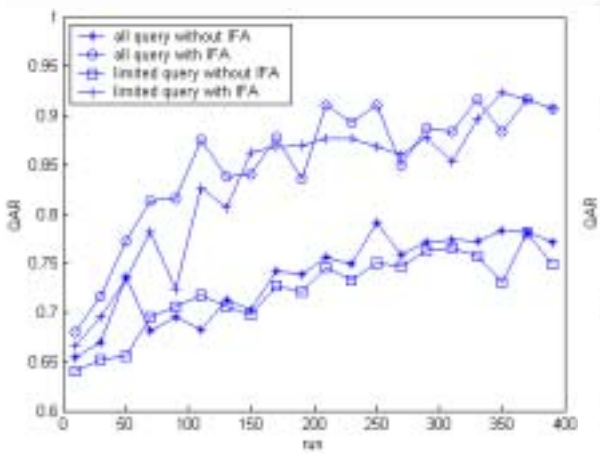


Fig.9 The QAR after the adoption of IFA

图 9 过滤机制引入后的 QAR

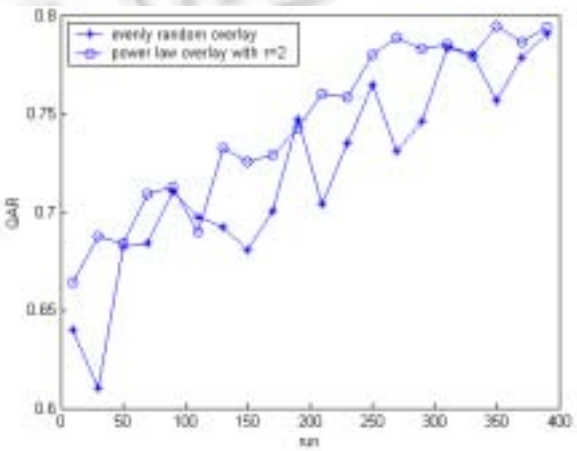


Fig.10 The influence on DRMAS of heterogeneity

图 10 系统异构对 DRMAS 的影响

5 结束语

信任管理是目前 P2P 安全研究的核心工作,同时也是 P2P 安全系统构造的基础部分.但目前以信任管理为基础的 P2P 安全系统主要集中在信任的建立和管理上,而在信任基础上的安全机制显得较为简单,通常只是简单地依据信任值的大小决定查询的转发.我们认为这样的简单机制是不能取得最好效果的,所以本文意在依托于数学分析结果构造一套以信任为基础的复杂安全模型,包含信息流动控制、文档管理以及过滤等 3 个层次.

本文对这一模型的安全性进行数学分析,并证明了这一机制在理想情况下可以达到优化水平.本文给出了大量的模拟实验,对文中提出的安全模型进行了验证.实验结果表明:引入简单的信任管理机制后,安全性仅提高 10%;而在加入文档管理机制并基本达到信息分布和文档真实性相随时,安全性进一步提高近 20%;如果在相随基础上再加入过滤机制,安全性会再提高 10%,基本达到安全性能够达到的最大值.另外,P2P 的异构性将加快安全性提高的速度,这表明本文提出的模型在实际系统中效果要好一些.

致谢 在此,我们向对本文的工作给予支持和建议的同行,尤其是哈尔滨工业大学计算机科学与技术学院王义和教授、匡正教授表示感谢.

References:

[1] Andy O. Peer-to-Peer: Harnessing the Power of Disruptive Technologies. O'Reilly & Associates, Inc., 2001. 3−159.

[2] Neil D, Hector GM, Beverly Y. Open problems in data-sharing peer-to-peer systems. In: Proc. of the 9th Int'l Conf. on Database Theory. Siena: Springer-Verlag, 2003. 1−15.

[3] Ernesto D, Sabrina DC, Stefano P, Pierangela S, Fabio V. A reputation-based approach for choosing reliable resources in peer-to-peer networks. In: Proc. of the 9th ACM Conf. on Computer and Communications Security. Washington: ACM Portal Press, 2002. 207−216. http://www.comp.nus.edu.sg/~ooibc/courses/cs6203/reputation-based-approach.pdf

[4] Yao W, Julita V. Trust and reputation model in peer-to-peer networks. In: Proc. of the 3rd IEEE Int'l Conf. on Peer-to-Peer Computing. Linkŏping: IEEE Computer Society, 2003. 150−158.

[5] Khambatti M, Dasgupta P, Ryu KD. A role-based trust model for peer-to-peer communities and dynamic coalitions. In: Proc. of the 2nd IEEE Int'l Information Assurance Workshop. Charlotte: IEEE Computer Society, 2004. 141−154.

[6] Neil D, Hector GM. Query-Flood DoS attacks in Gnutella. In: Proc. of the 9th ACM Conf. on Computer and Communications Security. Washington: ACM Press, 2002. 181−192.

[7] Brian FC, Mayank B, Neil D, Sergio M, Hector GM. Authenticity and availability in PIPE networks. Future Generation Computer Systems, 2005,21(3):391−400.

[8] Anwitaman D, Manfred H, Karl A. Beyond "Web of trust": Enabling P2P e-commerce. In: Proc. of the IEEE Int'l Conf. on Electronic Commerce. Newport Beach: IEEE Computer Society, 2003. 303−312.

[9] Sepandar DK, Mario TS, Hector GM. The EigenTrust algorithm for reputation management in P2P networks. In: Proc. of the 12th Int'l Conf. on World Wide Web. Budapest: ACM Press, 2003. 640−651.

[10] Gnutella. http://www.gnutella.com/. 2005.

[11] Michalis F, Petros F, Christos F. On power-law relationships of the Internet topology. In: ACM Conf. of the Special Interest Group on Data Communication. Cambridge: ACM Press, 1999. 251−262.

李治军(1977－),男,内蒙古伊盟人,博士生,讲师,主要研究领域为 P2P 系统,操作系统安全,分布式系统.

廖明宏(1966－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为人工智能,嵌入式操作系统,传感器网络.