RepTrust:P2P环境下基于声誉的信任模型

杨 超 刘念祖
(上海立信会计学院数学与信息学院 上海 201620)

摘 要 声誉是一种新型的信任建立机制,在P2P网络中,利用基于声誉的信任机制可以孤立劣质节点,选择更为可信的节点进行交互,以规避和减少风险。针对目前研究中存在的与网络结构耦合过紧、防止联合欺诈和声誉时效性处理不佳等问题,提出了一种新的基于声誉的信任模型RepTrust。模型采用与获取路径甚至评价者rater无关的声誉信息计算信任度,利用基于大多数原则的算法过滤不实声誉,利用衰减因子处理声誉的动态性。仿真实验表明模型在信任度计算、对付欺诈和动态声誉处理等方面效果明显。

关键词 P2P,声誉,信任模型,联合欺诈,动态性
中图法分类号 TP393.08 **文献标识码** A

RepTrust:Reputation Based Trust Model in P2P Environment

YANG Chao LIU Nian-zu
(School of Mathematics and Information,Shanghai Lixin Commerce University,Shanghai 201620,China)

Abstract Reputation is a new method to build trust. In Peer-to-Peer networks,trust mechanism based on reputation can isolate inferior peers and recommend excellent peers in order to evade potential risk and provide better service. Some shortcomings were found in related research such as overly dependency on network structure,neglect of collusion cheat and disregard of dynamic reputation change. Aiming at these shortcomings,RepTrust,a novel trust model based on reputation was presented. In this model,reputation unrelated with raters and routes through that it been gotten was used to compute trust. Mendacious reputation was filtrated and an attenuation factor was introduced into to amend the dynamic reputation. According to the simulation experiment,this model is effective in trust computing,dealing with collusion cheat and dynamic reputation.

Keywords Peer-to-Peer,Reputation,Trust model,Collusion cheat,Dynamic

1 引言

P2P(Peer-to-Peer)计算模式已广泛应用于资源共享、对等协作等多个领域,提供了很多便利。同时,由于系统开放、动态和匿名等特点,网络中不可避免会地存在一些恶意行为,如传播病毒、提供劣质服务、传播虚假信息甚至欺诈等,参与交互的节点就不得不面对这些潜在的风险,使P2P网络的安全性、可用性降低,这个问题已成为制约P2P计算广泛应用的瓶颈。如何在交互的节点间建立信任关系,规避或降低交互风险已成为计算机领域的热点课题。

类似于社会网络,信任可以有效地改善开放式系统中节点间交互的效率。在传统网络环境下(如Internet),信任关系的建立主要是通过认证技术来实现的,拥有数字证书的节点被认为是可信任的。然而这种基于第三方的信任机制很难在P2P环境下实施,这主要是基于以下两点考虑:(1)认证技术需要一个认证中心,这一方面违背了P2P技术无控制中心的原则,另一方面在大规模网络中认证中心将面临巨大开销,会成为新的瓶颈,甚至造成单点失效。(2)在P2P环境下要求所有节点进行身份和服务认证可操作性差(在电子商务中这一点已经被证明),而且与P2P技术的开放匿名原则相悖。因此对于快速发展的P2P网络,开发新的分布式信任管理模式非常必要。

在社会交往中,信任建立在双方足够了解的基础上,而这种了解可以是直接的,即通过双方的直接交往了解对方,也可以是间接的,即通过他人渠道了解对方。在信任的建立方面,P2P网络与社会网络具有很大的相似性:(1)节点在选择交互对象时具有充分的自主权。(2)交互完成后,节点间可以进行满意度评价,这种评价我们称之为"声誉"。声誉可以反映节点的可信程度。(3)通常情况下,其他节点之间的声誉评价,对于第三方节点选择交互对象是有参考价值的,即优先选择声誉好的节点作为交互对象。(4)声誉评价机制可以同时起到激励优质节点和惩罚、孤立劣质节点的作用。基于声誉的信任模型正是基于此提出的,并已成为P2P环境下建立信任的重要方法。

到稿日期:2010-04-23 返修日期:2010-08-05 本文受上海市教育委员会科研创新项目(10YZ193),上海高校选拔培养优秀青年教师科研专项基金(slx08017)资助。

杨 超(1975—),男,博士,讲师,主要研究方向为信任管理、贝叶斯网,E-mail:yangchao@lixin.edu.cn;刘念祖(1955—),男,教授,主要研究方向为图像处理、决策。

2 相关工作

按照体系结构来分，基于声誉的信任模型可以分为集中式和分布式两种。在早期电子商务中(如 eBay，Amazon)应用的声誉系统都采用集中式结构，系统中存在一个声誉中心负责收集所有的交易评价(rating)，然后根据这些评价计算交易者的声誉得分，并向所有用户公开。这些系统在改善交易效率、规避交易风险方面效果明显，然而，如前所述，这种集中式且过于简单的声誉系统不能满足 P2P 计算环境的要求，但是为对等环境下基于声誉的信任管理提供了有益的参考。目前基于声誉的信任模型主要采用分布式体系结构，即声誉信息被分布于各个节点上。由于在 P2P 环境下，节点是完全自治的(autonomous)，因此声誉信息不能存储于目标节点(ratee)上，只能由声誉评价者(rater)负责存储。在分布式体系中，声誉的发现和获取是信任模型必须要解决的关键技术[1,2]。在模型 P2PREP[3,4]中，节点的声誉信息由交互对象(rater)来计算和存储。在该模式下，节点在交互前首先需要对整个系统广播针对目标节点的声誉咨询请求。收到请求后，所有拥有目标节点声誉信息的节点反馈声誉给请求者。请求者根据这些声誉反馈综合计算目标节点的信任度，并最终决定是否与之交互。

根据声誉的建立方式，信任模型使用的声誉可大致归纳为 3 类：基于网络拓扑信息的声誉、基于交互成败的声誉和基于交互质量的声誉。第一类声誉建立方法的典型代表是 NodeRanking[5]，该方法主要应用于结构化 P2P 系统中，利用节点间的拓扑信息按照一定的搜索或路由计算节点的声誉。基于拓扑信息的声誉假设：节点被越多节点指向则越好或者说具有越好的声誉。这类方法的优势是声誉求解相对容易，不需要任何的交互评价信息，但它也有一个明显的缺点：单纯依赖拓扑信息，而忽略了节点的个性。P2P 环境下节点提供的服务千差万别，交互需求方的兴趣也各有不同。按照统一规则求解的声誉不可能真实地反映这些差异，也就不可能反映出交互双方对于交互的真实感受和评价。第二类声誉建立的方法相对更为常见，它根据节点参与交互的成败情况计算声誉，即交互成功的次数占全部交互次数的比例。该方法的优势是声誉计算方法简单，可操作性强。但该方法不太适合 P2P 环境。对等环境下的交互不只是电子交易，还有很多的其他服务形式，如存储服务、查询服务、计算服务等，这些服务的过程和质量很难简单地用成败来界定。与第二类声誉建立方式有些相似，基于交互质量的声誉建立方式也要求节点根据实际交互过程中的自身体验给出打分评价。与第二类声誉建立方式不同的是，在基于交互质量的声誉建立方式中，声誉评价的依据不再是简单的交互成败，而是交互节点的服务质量。这种声誉评价在表达上与第二类声誉建立方式也有差别，不再使用离散的数据，而用一个位于特定区间(如[−1，1])内的连续值表示声誉，是对声誉细粒度的刻画，能够更准确地反映服务质量[6,7]。Xiong L 等人曾经尝试将这种基于服务质量的声誉细分，提出了基于多因子的声誉表达方式，包括满意度、可信度、交互背景、处理方式和全面信任度 5 个方面[8,9]。考虑到交互节点的多样性，任何既定的评价标准都无法准确描述每个交互过程，因此更多人坚持使用单因子的声誉评价，即用单一值综合地表示交互体验。

在计算信任时使用的声誉信息有两种：直接声誉和推荐声誉。直接声誉是指信任计算中采用的声誉都来自于真实的交互体验，如上所述，这些声誉信息被存放在各个与目标节点(ratee)有过真实交互的节点(rater)处，主体节点获得这些声誉信息后直接利用它们计算目标节点的声誉和信任，计算过程与找到这些信息的路径(包括声誉的评价者 rater)无关。推荐声誉是指信任计算中采用的声誉都来自于中间节点的推荐，这种推荐关系甚至是迭代的，在计算目标节点声誉时，推荐者的可信度也会被作为影响因子考虑在内，即声誉计算与推荐者、推荐路径密切相关[10,11]。基于推荐声誉的信任模型存在如下缺点：(1)迭代路径越长越脆弱，而且声誉在传递过程中面临越多的联合欺诈风险。(2)要求整条推荐路径上所有节点之间都要有良好的信任关系，否则根据 BBK 方案[12]，不管推荐路径的长短，最终得到的声誉都是差的，即目标节点不可信，推荐也就失去了意义。但这样的路径在实际中很难找到。(3) 基于推荐声誉信任模型与 P2P 网络拓扑结构耦合性太强，缺乏普适性。

此外，基于声誉的 P2P 信任模型还需要处理好两个问题：(1)防止联合欺诈。联合欺诈是指多个节点为了自身利益或者夸大目标节点的声誉，或者诋毁目标节点[13]。联合欺诈将严重影响声誉计算的准确性，最终影响信任模型的效果。(2)声誉反映了节点的历史表现，具有时效性，而节点的服务质量具有动态性，因此其声誉应该及时反映出节点的变化。针对这两个问题，目前的解决方案都或多或少地存在一些缺陷。例如文献[14]利用赋予直接声誉较大权重的方法防止联合欺诈，而在大规模的 P2P 网络中，直接声誉是极少的。文献[15]给直接声誉设置基于时间的衰变函数，试图解决声誉信息的时效性问题，但在推荐声誉中未做相应处理。同时其利用全局声誉防止联合欺诈的方法也值得商榷。

借鉴前人的研究成果，同时针对目前研究中存在的问题，提出了一种新的面向 P2P 环境的基于声誉的信任模型——RepTrust。

3 RepTrust

3.1 模型描述

首先结合相关文献，对信任模型的几个重要概念给出描述性定义。

定义 1 声誉(也称为声誉评价)R 是交互结束后，节点根据交互体验给予交互伙伴的一种评价(rating)，反映了评价者(rater)对被评价者(ratee)的满意程度。声誉是对过去行为的评价，具有客观性。在 RepTrust 中，采用基于交互质量的声誉评价，声誉 $R\in[-1,1]$，大于 0 表示正面声誉评价，小于 0 表示负面声誉评价。声誉采用分布式管理，存储于评价者处。

RepTrust 模型采用的声誉是一维数据，即用单一值表示声誉。

定义 2 直接声誉 R_d 是指主体节点通过与目标节点的直接交互而得到的声誉，该声誉值保存于主体节点，无需从网络中搜索。在大规模 P2P 网络中，直接声誉通常是罕见的。

定义 3 间接声誉 R_i 是指主体节点从网络中搜索到的关于目标节点的声誉。其是第三方节点与目标节点交互后给出的声誉评价。本模型采用基于交互声誉的信任计算模式，

即主体节点获得间接声誉信息后直接利用它们计算目标节点的声誉和信任，计算过程与获得这些信息的路径(包括声誉的评价者 rater)无关。

定义 4 合成声誉 $\bar{R}$ 是依据直接声誉和间接声誉，通过一定的算法综合求解出的用于表达 ratee 声誉状态的声誉值。

定义 5 信任 Tr 是指在无法监控对方行为的情况下，rater(施信方)对于 ratee(受信方)将要按照自身期望选择行为方式的可能性的主观预测。信任取决于 rater 的知识和经验，同时与交互上下文有关，具有主观性。在基于声誉的信任模型中，目标节点的声誉是主体节点选择是否信任对方并与对方交互的关键依据。所有信任本质上都是主观的。信任度是信任的量化，表示相信的程度，有时也简称信任。在 RepTrust 中，目标节点的信任度是根据直接声誉和间接声誉综合求解得到的，$Tr\in[0,1]$。0 表示完全不信任，1 表示完全信任。

在 RepTrust 模型中，主体节点积极地收集有关目标节点的直接声誉和间接声誉，对目标节点的信任度进行评估，而目标节点一直处于被动状态。工作过程描述如下：在主体节点 A 与目标节点 B 交互前，首先评估 B 的信任度，如果以前曾经多次与 B 交互过，则依靠本地的直接声誉以算术平均的形式求出合成声誉和信任度。在大规模 P2P 网络中这种情况十分罕见，因此多数情况下 A 需要在网络中搜索关于 B 的间接声誉。最后根据本地的直接声誉和搜索到的间接声誉计算 B 的合成声誉和信任度。该流程如图 1 所示。R_d 表示 A 对 B 的直接声誉，R_s 表示 B 的间接声誉。$f(R_d)$ 是直接声誉是否充足的检查函数，如果有充足的直接声誉，则 B 的合成声誉和信任度可以完全根据直接声誉计算，否则需要利用直接声誉和间接声誉综合计算 B 的合成声誉和信任度。若 B 的信任度大于设定信任阈值，则与之交互，否则重新选择其他交互对象。交互结束后，A 给予 B 声誉评价，并保存于本地。

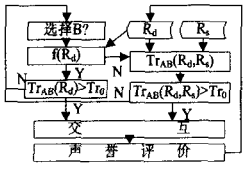

图 1 RepTrust

3.2 间接声誉搜索

RepTrust 信任模型中的声誉搜索机制是在 Gnutella 协议[16]基础上改进得到的。搜索过程包括 3 个步骤。

第一步，服务搜索。节点 A 广播服务请求，系统中能够提供该服务的节点 Peers 进行响应，形成备选节点集。

第二步，声誉搜索。节点 A 产生一对临时公钥 PK、私钥 SK 对，作为声誉搜索的钥匙。针对一个候选节点广播声誉查询请求，任何第三节点可以回应该请求，并用公钥加密声誉信息，保护候选节点的隐私。

第三步，用私钥 SK 解密声誉，获得间接声誉。

该搜索方法存在缺陷，声誉数据在传递过程中容易被篡改，这是目前所有 P2P 信任模型面临的共同问题。在 RepTrust 信任模型中，不实的声誉信息由合成声誉计算算法识别和过滤。

3.3 合成声誉计算

如上所述，信任模型必须处理好联合欺诈和声誉动态性这两个问题，才能保证信任管理的有效性。直接声誉充足情况下的合成声誉计算多采用算术平均的形式求解，算法比较简单，这里不再讨论。

在无直接声誉或者直接声誉稀少的情况下，RepTrust 信任模型不区分直接声誉和间接声誉。

3.3.1 不实声誉过滤

假设 1 声誉信息中客观声誉评价明显多于由联合欺诈造成的不实声誉评价。这是多数原则成立的前提。

算法思想：通常我们都简单地以数据的大小为依据判断数据是否属实，例如奥运会上体操比赛的评分机制：去掉一个最高分和一个最低分，然后用剩下的数据求算术平均。在 RepTrust 信任模型中，更多地关注声誉评价出现的频率，或者说是分布，将声誉的取值范围$[-1,1]$分为若干个区间，一个声誉数据就是对其对应区间的投票，根据大多数原则，落入获得支持最少的区间内的声誉评价，可以视为不实声誉。同时，声誉计算的结果应该出现在获得最多支持的评价单元当中。具体算法(记为 RTA1)描述如下：

1. 把声誉范围$[-1,1]$划分为若干区间：$D_i=[R_i-\varepsilon, R_i+\varepsilon]$。$R_i$ 为区间中心，ε 为区间半径。

2. 统计各区间内声誉评价出现的次数。$h_i=|R|=|\{R|(R_i-\varepsilon)\leqslant r<(R_i+\varepsilon)\}|$。

3. 寻找声誉数据最为集中的连续的若干个(记为 n)区间。

4. 以落入该连续区间内的声誉评价作为有效数据，求其算术平均，作为目标节点的合成声誉。

RTA1 算法简单，易操作，考虑到不实声誉评价必然会偏离客观声誉评价的主体，因此该方法可以滤除由联合欺诈造成的不实声誉信息。

3.3.2 声誉时效性处理

定义声誉队列结构 Q，该队列有如下属性：①队列中的声誉数据满足条件 $t(R)\geqslant t-T$，t 表示当前时间，T 表示有效时间帧长度，时间帧随时间的推移自动向前移动，以保证当前时间始终在时间帧头部，如图 2 所示。②队列头部为最新声誉评价数据，队列尾部为最旧声誉，当 $t(R)\leqslant t-T$ 时，声誉从队列尾部删除。第 i 个声誉评价的衰减因子定义为：

$$\lambda_i=\begin{cases}0, & t(R)<t-T\\ 1-\dfrac{t-t_i(R)}{T}, & t\geqslant t(R)\geqslant t-T\end{cases} \tag{1}$$

图 2 声誉队列与时间帧

由式(1)可知，越新的声誉评价具有越大的衰减因子，有效时间帧外的声誉衰减因子为 0，即不再用来计算目标节点的合成声誉。用衰减因子修正 RTA1 得到 RTA2，二者的区别在于步骤 2：统计各区间内声誉评价出现的次数。$h_i=\sum\lambda_i$，与 λ_i 对应的声誉评价 R_i' 满足条件$(R_i-\varepsilon)\leqslant R_i'<(R_i+\varepsilon)$。

以有效声誉数据，按式(2)计算目标节点的合成声誉。

$$R=\frac{\sum R_i\times h_i}{\sum h_i} \tag{2}$$

3.4 信任计算

主体节点对目标节点的信任度与目标节点的合成声誉正

相关，但两者却不能等同视之，因为(1)含义不同，声誉是对过去交互的评价，具有客观性。而信任则是对未来的预测，是主观的。(2)如前所述，两者的值域也不相同。数量关系可以表达为：

$$Tr=\frac{R+1}{2} \tag{3}$$

3.5 信任阈值

主体节点最终是否信任目标节点，不但取决于目标节点合成声誉的好坏，还受到信任阈值的影响。当目标节点的信任度高于信任阈值时，主体节点才会信任它，与之交互。主体节点会根据自身情况，交互背景设置信任阈值。信任阈值的计算可采用如下式子：

$$Tr_0=k\times\frac{risk+cost}{risk+cost+gain} \tag{4}$$

式中，*risk* 代表交互风险，即一旦交互失败给主体节点带来的损失，*cost* 代表交互代价，*gain* 代表交互成功后的收获，k 为主体节点的谨慎指数，$0<k<1$。

4 实验与分析

利用仿真实验评测所提出的模型及算法。实验环境为100M 局域网，微机配置为 Pentium D 3.40GHz /1G，WINXP，用 VC++6.0 实现。实验模拟了 100 个节点，所有节点提供文件下载服务。100 个节点配置了大小不等的文件，下载节点以文件的下载时间长短评价提供文件的节点，所以文件小的节点，容易获得高的声誉评价和信任度。设最大的下载时间为 t_{max}，则下载时间为 t 的节点获得的声誉评价可表示为式(5)。这里所有声誉为正值。交互时节点间随机组合，不考虑信任度的大小，否则配置了大文件的节点将很难获得足够的声誉评价。

$$R=\frac{t_{max}-t}{t_{max}} \tag{5}$$

RTA 算法中的相关参数设置如表 1 所列。

表 1 RTA 算法中的参数表

参数	参数说明	值
ε	区间宽度的一半	0.1
n	计算声誉时保留的区间个数	6
T	时间帧长度	35

4.1 实验 1：模型的有效性

算法的有效性可以通过算法的收敛效果和稳定性来评价。从 100 个节点中按照下载文件的大中小，各取一个节点作为考察对象，得到的结果如图 3 所示。

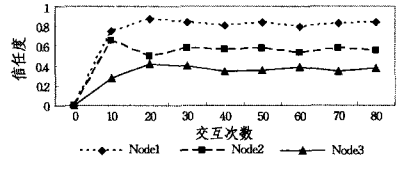

图 3 模型有效性

由图 3 可以看出，交互次数小于 30 次时，由于数据量偏少，使得信任度不稳定，随着交互次数的增加，信任度基本平稳。说明模型有效。

4.2 实验 2：对联合欺诈的处理效果

随机选择 15 个节点，调整其声誉评价策略，使其给出的声誉评价偏离正常值。这里使用诋毁策略，使声誉比正常值偏低 30%，即：

$$R=0.7\times\frac{t_{max}-t}{t_{max}}$$

选取一个节点分无诋毁(C1)、有诋毁无过滤(C2)、有诋毁有过滤(C3)3 种情况进行观察，得到的结果如图 4 所示。

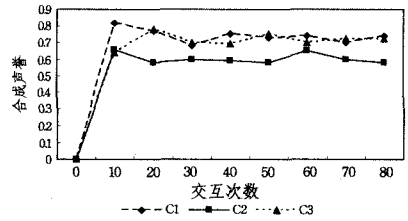

图 4 联合欺诈的处理效果

由图 4 可以看出，C2 明显偏离 C1，即不实声誉评价对于合成声誉计算产生了显著影响。C3，C1 比较接近，说明增加过滤后，不实声誉对合成声誉计算的影响明显减小。算法对防止联合欺诈有效。另外，即使同时存在诋毁和夸大，算法依然有效。

4.3 实验 3：对声誉动态性的处理效果

选择一个观察节点，每 10 次交互后将其下载文件做删减，仿真一个声誉不断改善的节点。分别用 RTA1 和 RTA2 求解合成声誉，对变化过程中每种大小的文件做 80 次下载，计算声誉的算术平均作为静态合成声誉，得到的结果如图 5 所示。

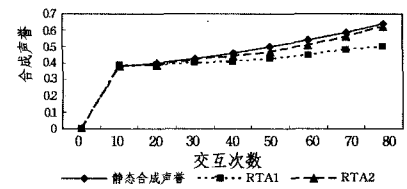

图 5 声誉动态性的处理效果

可以看出，RTA1 合成声誉逐步偏离静态合成声誉，即如果不考虑声誉的动态性，合成声誉和信任度的计算结果将明显滞后于节点状态的变化。算法 RTA2 处理声誉信息的动态性有明显效果。

结束语 基于声誉的信任模型是 P2P 环境下规避风险、提高交互效率的重要机制。目前对这类模型的研究还处于起步阶段，存在一些有待解决的问题，如与 P2P 拓扑结构耦合过于紧密，普适性差，对不实声誉信息和声誉动态性的处理不够令人满意等。提出了一种新的信任模型——RepTrust。模型采用交互声誉计算信任度，与获得这些声誉的路径甚至声誉的评价者 rater 无关，模型重点考虑了由于联合欺诈所带来的不实声誉的处理以及声誉动态性的处理。仿真实验说明模型有效。

未来的工作将重点研究声誉的多维表达以及声誉评价稀疏等问题。

参 考 文 献

[1] Kamvar S D, Schlosser M T. EigenRep: Reputation management in P2P networks[C] // Proceedings of the 12th International WWW Conf. Budapest, 2003: 123-134

[2] 窦文，王怀民，贾焰，等. 构造基于推荐的 peer-to-peer 环境下的 Trust 模型[J]. 软件学报，2004，15(4)：571-583

[3] Damiani E, Vimercati S D C, et al. Managing and sharing servents' reputation in P2P systems[J]. IEEE Transactions on Knowledge and Data Engineering, 2003, 15(4): 840-854

[4] Damiani E, Vimercati S D C, et al. A reputation-based approach for choosing reliable resources in peer-to-peer networks[C]// Proceedings of the 9th ACNM Conference on Computer and Communications Security(CCS'02). ACM, 2002: 207-216

[5] Jøsang A, Ismail R. The beta reputation system [C] // Proceedings of the 15th Bled Conf. on Electronic Commerce. Bled, Slovenia, 2002: 708-721

[6] Zhou R, Hwang K. PowerTrust: A robust and scalable reputation system for trusted P2P computing[J]. IEEE Transaction on Parallel and Distributed systems, 2007, 18(5)

[7] Wu Xu, He Jing-xia, Xu Fei. An enhanced trust model based on reputation for P2P networks[C]// IEEE International Conf. on Sensor Networks, Ubiquitous, and Trustworthy Computing. Taichung, Taiwan: IEEE Press, 2008: 67-73

[8] Xiong L, Liu L. A reputation-based trust model for P2P ecommerce communities [C] // Proceedings of IEEE International Conf. on Electronic Commerce, 2003

[9] Xiong L, Liu L. PeerTrust: Supporting reputation-based trust for peer-to-peer communities[J]. IEEE Transactions on Knowledge and Data Engineering, 2004, 16

[10] Srivatsa M, Xiong L, Liu L. Trustguard: countering vulnerabilities in reputation management for decentralized overlay networks[C]// Proceedings of WWW. 2005

[11] Singh A, Liu L. Trust me: Anonymous management of trust relationships in decentralized P2P systems[C]// Proceedings of P2P'03. 2003

[12] Beth T, Borcherding M, Klein B. Valuation of trust in open networks[C]// Proceedings of Computer Security-ESORICS 94. 1994: 3-14

[13] Dellarocas C. Mechanisms for coping with unfair ratings and discriminatory behavior in online reputation reporting systems[C]// Proceedings of the 21st International Conf. on Information Systems(ICIS). Brisbane, Australia, 2000, 12

[14] 毕方明,等. 面向对等网络的主观逻辑信任模型[J]. 计算机工程与应用, 2009, 45(33): 99-102

[15] 田祥宏,等. P2P 环境下的一种混合式信任模型[J]. 计算机工程与应用, 2009, 45(31): 73-76

[16] http://www.guntella.com, 2003

(上接第 122 页)

定理 2 假设同态加密方案具有语义安全性,则协议 2 在半诚实模型中具有安全性。

证明:

1) 正确性

由定理 1 可知,如果 $b_\ell \in A$,当且仅当 $f_\ell=0$,则 B 包含于 A,当且仅当 $f=\sum_{\ell=1}^{m} f_\ell=0$。

2) 隐私性

证明同定理 1 隐私性证明类似,略。证毕。

这里需要指出的是协议 1 和协议 2 的判定泄漏了数据集 A 和 B 的势,并且协议 2 尽管采取了对中间结果的置换,阻止了用户获取两个数据集的交集,但仍然泄漏了交集的势。这些问题是协议 1 和协议 2 美中不足的地方。

4.2 效率

1) 协议 1 的效率

协议 1 需要一轮通信,通信量为 $o(kmn\tau)$ 比特;用户 U 做 $o(m)$ 次加密,$o(mn)$ 次解密;参与的每个服务提供者 SP_{l_j} ($1\leqslant j\leqslant k$) 做 $o(mn)$ 次指数运算和乘法运算。

2) 协议 2 的效率

由于协议 2 中的置换代价可以忽略,因此协议 2 与协议 1 的效率相等。

结束语 本文讨论了基于数据库外包的分布式环境下的隐私匹配问题,给出了一个具体的隐私匹配协议(协议 1),在此基础上提出了一种判断用户数据集是否包含于数据所有者的数据集的安全计算协议(协议 2)。并在半诚实模型下,证明了两个协议的安全性。

下一步的工作将在半诚实模型或者恶意模型下,研究基于数据库外包的分布式环境下的其他运算问题。

参 考 文 献

[1] Hacigümüs H, Mehrotra S, Iyer B. Providing database as a service[C]// Proceedings of the 18th International Conference on Data Engineering. San Jose, USA, IEEE Computer Society Press, 2002: 29-40

[2] Agrawal D, Abbadi A E, Emekci F, et al. Database management as a service: challenges and opportunities[C]// Proceedings of the 2009 IEEE International Conference on Data Engineering. Washington, DC, USA: IEEE Computer Society Press, 2009: 1709-1716

[3] Freedman M J, Nissim K, Pinkas B. Efficient private matching and set intersection[J]. Lecture Notes in Computer Science, 2004, 3027: 1-19

[4] Oleshchuk V A, Zadorozhny V. Secure multi-party computations and privacy preservation: results and open problems[J]. Telektronikk, 2007, 2: 20-26

[5] Yao A C. How to generate and exchange secrets[C]// The 27th IEEE Symposium on Foundations of Computer Science. Toronto, Ontario, Canada, IEEE Computer Society Press, 1986: 162-167

[6] Yao A C. Protocols for secure computations[C]// The 23th IEEE Symposium on Foundations of Computer Science. Chicago, USA, IEEE Computer Society Press, 1982: 80-91

[7] Kissner L, Song D. Privacy-preserving set operations[J]. Lecture Notes in Computer Science, 2005, 3621: 241-257

[8] Sang Y, Shen H, Tan Y, et al. Efficient protocols for privacy preserving matching against distributed datasets[J]. Lecture Notes in Computer Science, 2006, 4307: 210-227

[9] 李顺东,司天歌,戴一奇. 集合包含与几何包含的多方保密计算[J]. 计算机研究与发展, 2005, 42(10): 1647-1653

[10] 李荣花,武传坤,张玉清. 判断集合包含关系的安全计算协议[J]. 计算机学报, 2009, 32(7): 1337-1345

[11] Ye Q, Wang H, Tartary C. Privacy-preserving distributed set intersection[C]// Proceedings of the 3th International Conference on Availability, Security and Reliability. Barcelona, Spain: IEEE Computer Society Press, 2008: 1332-1339