文章编号: 0455-2059(2012)04-0085-06

面向第三方服务平台的隐私保护

姜文广, 孙宇清

山东大学 计算机科学与技术学院, 济南 250101

摘 要: 提出了面向第三方服务平台的隐私保护模型, 支持用户定义个性化隐私策略, 以满足不同用户的隐私偏好, 并实施相应的访问控制, 有效保护用户数据. 设计实现了支持这一隐私保护模型的中间件, 并应用到山东省制造业信息化服务平台系统中.

关键词: 隐私保护; 访问控制; 第三方服务平台

中图分类号: TP311 **文献标识码:** A

Personalized privacy protection for third-party service platform

JIANG Wen-guang, SUN Yu-qing

School of Computer Science and Technology, Shandong University, Jinan 250101, China

Abstract: A privacy protection model for the third-party service platform was established in which users are allowed to define privacy policies according to their preferences and the privacy policy can be enforced by the system via access control modules. A middleware was designed to support the proposed privacy protection model and to apply it to the Shandong manufacturing information platform. Experiments showed that the proposed model and enforcement mechanism could effectively protect user privacy and satisfy their personalized privacy requirements.

Key words: privacy protection; access control; third-party service platform

随着互联网应用普及, 以淘宝、Amazon为代表的电子商务网站, 以Facebook、人人网为代表的社交网络网站等第三方服务平台迅猛发展并聚集了大量的用户. 在第三方服务平台上, 用户存储了大量的隐私数据, 如注册信息、上传的文件、交易信息等[1]. 这些隐私信息面临着来自于各方的威胁, 如攻击者采用信息检索或数据挖掘等技术, 收集用户在平台中的个人隐私信息及活动情况; 恶意的商家也可能在未经用户允许的情况下私自传播、滥用和篡改用户信息从而损害用户的利益. 若隐私信息不能得到恰当保护, 将阻碍电子商务、社交网络等第三方服务平台的发展.

现有的面向第三方服务平台的隐私保护技术主要分为两类: 一类在保护隐私数据的完整性和保密性上, 服务商可以采取加密、防火墙等方式, 防止攻击者窃取隐私信息; 另一类在收集用户隐私偏好和对海量用户行为分析以获得用户隐私信息上, 现有的技术主要有P3P[2](platform for privacy preferences), EPAL[3], XACML[4] 等. P3P 的工作原理是在服务商收集用户的隐私信息时, 将其隐私政策公布在网站上, 以提示用户是否同意其收集和使用隐私信息. 如果服务商的隐私策略同用户的P3P 中设定的标准相符, 用户可浏览该站点, 否则不允许访问. 在第三方可信基础上, 用户可将数据存放在第三方平台, 而P3P 技术不能阻止用户之间的隐私泄露, 它仅提供了用户和服务器端如何进行数据交换的机制, 并不能结合用户数据特征实施个性化的隐私保护, 也不能提供灵活、方便的隐私策略机制以保证用户之间的细粒度的访问控制. EPAL[3] 是一种基于XML 的形式化语言,

收稿日期: 2012-06-05

基金项目: 国家自然科学基金项目(61173140); 山东大学自主创新基金项目 (交叉基金 2010JC010); 山东省自然科学基金项目(Y2008G28); 中国科学院计算机系统结构国家重点实验室开放基金项目(ICT-ARCH200904)

作者简介: 孙宇清(1967−), 女, 山东即墨人, 教授, 博士, e-mail: sun_yuqing@sdu.edu.cn, 研究方向为协同计算、系统安全和隐私保护, 通信联系人.

它允许企业直接用标记语言来定义其隐私保护策略和规则. 与P3P不同的是它不仅仅用来与用户的客户端交换隐私保护的信息, 同时也定义企业内部对信息处理的规则来达到隐私保护的目的, 适用范围更广, 描述能力也更强. 缺点是与现有系统的整合较为繁琐, 需要作较多改动, 相应的工具也比较缺乏. XACML[4]是一种通用访问控制策略语言和执行授权策略的框架, 虽然应用较为广泛, 但不能有效地支持个性化隐私策略定义与实施, 同时与系统的整合较为繁琐. 此外在防止隐私泄露方面还有希波克拉底隐私数据库(Hippocratic databases)[5−12], 它实施在数据库层次且策略表达能力较差, 无法依据数据特征进行个性化访问控制. 因此, 实施支持用户隐私偏好的访问控制对用户隐私保护至关重要.

1 面向第三方服务平台的隐私保护模型

第三方平台(简称服务器)是指独立于数据拥有者和数据使用者的可信服务提供方, 如电子商务平台可以为交易双方提供网上交易、洽谈服务等. 在使用服务过程中, 用户需要将注册信息、上传数据、交易数据等存放在第三方平台上, 以方便他人授权访问.

1.1 基本概念

用户是指在第三方服务平台注册并使用平台服务的个体或组织, 也称为主体, 用户集合用U表示. 根据用户使用网络平台目的的不同, 一个用户既可以作为数据拥有者(服务使用者)访问系统, 也可以作为访问请求者访问系统. 用户以服务使用者身份访问系统时, 可以使用平台的策略管理服务对隐私策略进行操作; 用户以访问请求者身份访问系统时, 系统会根据隐私策略进行访问控制选择被访问用户相应的数据内容进行呈现.

数据是用户提供的个人隐私信息或是在使用第三方平台服务过程中产生的隐私信息, 也称为客体, 数据集合用R表示. 例如用户注册信息、电子商务平台中交易信息、社交网络用户的上传文件等.

动作是对数据的操作类型, 例如读、写、修改、查看等操作, 所有动作集合表示为A.

定义1 约束是授权过程中各项属性必须满足的条件, 用布尔表达式描述. 约束包括主体属性约束、客体属性约束和环境属性约束等. 约束集合用CT表示.

例如: 工资若大于5 000元可以描述为salary > 5 000. 企业的所有权结构为国有控股可以描述为Ownership ="国有控股".

定义2 策略是用户根据自身隐私偏好定义的一系列访问权限的规则, 目的是防止用户对数据的非授权访问. 策略描述的内容包括用户、数据、动作以及约束. 策略的形式化定义为policy: (Owner, res, a, ct), 其中Owner$\in U$, 表示数据拥有者, res$\in R$表示被访数据, $a\in A$, 表示动作, ct表示策略约束, 其中约束间的关系默认为与, 策略间的约束关系为或(约束为或关系可用不同策略表达). policy: (Owner, res, a, ct)表示数据拥有者u要求用户必须满足约束条件ct才能对数据r执行动作a.

例如: 雷沃集团(LOVOL)定义一条隐私策略, 它许可注册资金大于20万, 所有权结构为国有控股的用户查看其交易信息, 那么策略可以表示为policy: (LOVOL, Transaction information, Read, Registered capital > 200 000 & Ownership ="国有控股").

1.2 模型

为了支持用户个性化隐私保护需求, 本文提出了第三方平台个性化隐私保护模型(personalized privacy protection model for third party, P3M), 支持用户定义隐私策略并依照用户的隐私策略实施访问控制. 如图1所示, 它由以下部分构成:

策略管理: 为用户提供策略定义机制, 并转化为系统可读的策略定义形式, 存储在隐私策略库中. 此外还为用户提供策略更改、删除, 以及查看策略等功能.

身份认证: 获取访问请求者的身份信息并确定其身份, 进而得到系统对其开放的服务使用权限.

策略评估: 针对用户访问他人隐私数据的请求, 系统依据被访问者的隐私策略作出访问控制决策. 所有用户的策略存储在数据库中, 每当有请求发生时, 将请求与被访问者制定的策略进行评估, 若请求满足策略, 那么评估结果为许可, 否则为拒绝.

隐私数据库: 用于存储在使用第三方平台服务过程中产生的隐私信息, 例如用户注册信息、电子商务平台中交易信息、社交网络用户的上传文件等.

隐私策略库: 用于存储用户自定义的隐私策略, 其策略形式在2.3节中描述.

1.3　用户隐私策略定义和访问控制流程

用户隐私策略定义: 首先用户以服务使用者的身份登入第三方服务平台, 根据自己的隐私偏好定义隐私策略, 描述相关隐私信息如身份信息、私人数据等对哪些请求者许可访问, 然后服务系统把隐私策略存储到用户的隐私策略库中. 同时服务系统为用户提供策略查看、删除、修改等功能, 满足对隐私保护的动态性需求. 用户策略定义过程在图 1 中由数字虚线标示.

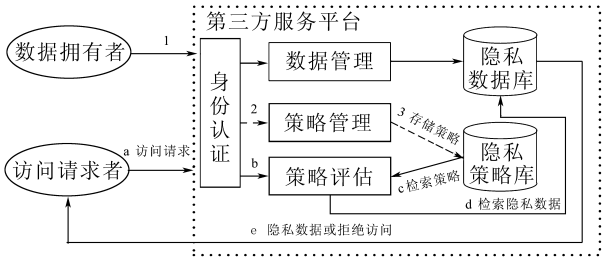

图 1　第三方服务平台隐私保护模型

Figure 1　Privacy protection model on third-party service platform

访问请求处理包括三个过程: 请求预处理, 策略评估, 返回结果. 系统对访问请求者进行身份认证后, 将其请求转化为具体应用环境下的访问控制请求(即将被访问者的属性信息合并到请求中形成新的访问请求), 然后将请求与策略数据库中被访问者的隐私策略进行评估, 最后根据评估结果检索相应的数据信息反馈给访问者或者拒绝访问者的请求. 访问请求处理过程在图 1 中由字母实线标示.

2　支持隐私偏好的访问控制

为了实施基于个性化隐私策略的访问控制系统, 需要在传统的访问控制系统中进行两方面改进: 隐私策略的管理和基于隐私策略的访问控制的实施. 本文将从以下几个方面进行论述.

2.1　扩展原有数据

当用户根据自己的隐私偏好定义个性化隐私策略时, 数据的保护范围应为现有的数据库信息. 而现实数据中, 用户属性、用户数据可能存放在不同表中. 为了灵活地定义隐私策略, 将用户制定的策略转化为系统设定的形式存储, 为用户提供选择范围动态生成策略定义交互页面, 需要将用户信息进行扩展, 并将隐私策略与用户扩展信息的相关字段进行映射建立隐私策略库.

为此分别设计了用户属性描述信息与数据库元数据映射表(用户信息映射表) 以及用户数据描述信息与数据库元数据映射表(数据信息映射表), 用户信息映射表描述内容包括用户属性所在的表, 用户属性所在的字段, 用户属性信息描述, 用户属性取值集合, 评估函数等. 表示为

[Table_name, Column_Name, Description, Value, function].

类似地, 建立数据信息映射表描述内容, 包括用户数据所在的表, 用户数据所在的字段, 用户数据信息描述, 表示为

[Table_name, Column_Name, Description].

在建立起用户信息映射表、数据信息映射表以及隐私策略表后, 将用户信息映射表与隐私策略表的主体字段关联, 将数据信息映射表与隐私策略表的客体字段关联进而构建起隐私策略表, 其映射关系如图 2 所示.

2.2　约束评估

在隐私策略中, 用户定义许可访问需要满足的约束, 当用户提出访问请求时, 需要将请求中的访问者用户属性与被访问者策略中的用户属性约束进行评估, 从而对满足约束的用户实施相应的授权. 而策略中约束种类、约束内容以及对属性的评估形式不同, 例如, 在约束中定义"Registered capital > 200 000, Ownership = "国有控股"表示"公司注册资金大于 10 万元, 所有权结构为国有控股". 上述约束中的两个条件需要用不同的函数来判断, 因此需要设计不同评估函数 function, 如取值比较函数、匹配函数等. 本文实例中用到的函数功能描述如表 1 所示. 因此在上述的例子中, 对于约束 Ownership = "国有控股", 判断所有权结构是否为国有控股可以利用字符串匹配函数 equals() 来进行约束评估; 对于约束条件 Registered capital > 200 000, 判断注册资金是否大于 200 000 元可以利用数值比较函数 isGreater() 来进行约束评估.

表 1　评估函数功能描述

Table 1　Summary of evaluation function

函数名称	函数功能
equals(a, b)	将字符串 a 与字符串 b 比较
Equalsint(a, b)	将数值 a 和数值 b 比较
isInRange(a, b, c)	判断数值 a 是否在 [b, c) 区间内
isGreater(a, b)	判断整数 a 是否大于整数 b
isSmaller(a, b)	判断整数 a 是否小于整数 b

2.3　隐私策略的存储

策略数据库的设计需要满足三个要求: 灵活的表达策略, 按规定的形式存储策略和便于解析策略. 每条策略必须包含策略的基本元素, 如包括

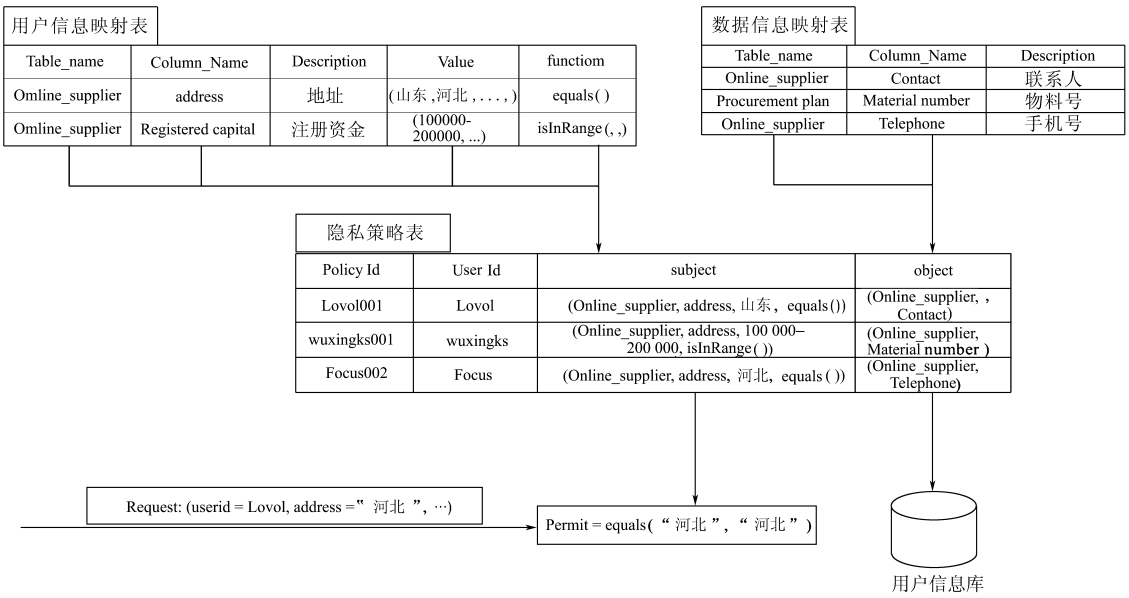

图 2 映射示意图

Figure 2 The mapping tables

用户、数据、动作以及约束等. 以用户属性为约束的访问控制为例, 将隐私策略表描述为

[UserId, PolicyId, SubAttribute, Res, Act],

其中每个元素含义以及形式如下所示.

策略中数据元素Res表示该条策略要访问的数据信息, 它的存储应该结合系统固有的属性字段, 采用Res = (Table_name, Column_Name)的形式, Table_name, Column_Name表示数据所在的表以及字段同时用于访问请求许可时返回数据信息; 策略中用户元素SubAttribute表示策略的限制是对于访问请求者的属性限制. SubAttribute的存储形式为SubAttribute = (Table_name, Column_Name, Atrributevalue, function), 其中Table_name, Column_Name指的是请求者的限制属性所在的数据库表和字段, Atrributevalue表示策略对用户限制属性的取值, function指定评估函数, 由用户信息映射表得到, 用于将策略的约束设定值与请求值进行比较. Act表示动作, Act ∈ {read, write, ···, }, 默认为read, 即访问操作; User Id表示用户的身份, 在系统中用惟一的身份标示符表示; PolicyId表示策略号用来标示用户的策略信息. 将策略的各个元素分开存储不需要额外的策略标示来区分它们, 便于进行访问控制.

策略表的作用是存储用户定义的隐私策略, 系统按如下过程得到用户定义的隐私策略: 在建立起用户属性描述信息与数据库元数据映射表以及用户数据描述信息与数据库元数据映射表后, 将它们中的每条记录(包括Table_name, Column_Name, descript以及function)查询出, 并以自然语言选项的形式提供给用户, 供用户选择. 由于在用户、数据信息映射表中隐私信息与元数据数据库字段以及function已经建立了映射, 因此在用户选择完成后, 所选择的内容返回给服务器即得到系统规定的隐私策略的形式. 图2所示为山东省制造业信息化服务平台隐私策略与用户, 数据信息映射表的相关字段进行映射的示意图.

2.4 访问控制

访问控制的实施应具有动态性, 将用户定义的隐私策略与用户数据绑定进行策略评估, 进而选择满足用户隐私策略的数据内容呈现, 实现动态的访问控制. 首先设计了上下文处理器处理访问请求. 上下文处理器作用是查询请求者的各个属性以及环境信息. 系统根据请求者的身份利用上下文处理器在用户信息库中查询出它的属性信息, 组成新的请求用于策略评估. 例如在基于主体约束条件的访问控制中, 平台用户雷沃集团想要访问用户福克斯汽车制造商的地址(address)信息, 系统会取得用户雷沃集团的身份信息再利用上下文处理器, 进而获得其属性信息和福克斯汽车制造商的身份信息等组成新的请求.

完成请求处理后, 将访问请求与数据拥有者的隐私策略进行绑定, 系统根据请求中被访问者的身份信息查询策略数据库中的策略, 通过其身份信息将访问请求与隐私策略关联进行策略评估.

策略评估是将请求中的请求者属性与策略中的属性约束进行比较、匹配等操作. 为此需要利用评估函数, 即系统提取策略约束中的function字段并利用评估函数将请求者属性与策略的属性进行比较. 对于满足策略的请求, 服务系统提取策略中的数据元素进而在用户数据库中检索得其数据集.

当请求者访问另一用户的隐私信息时, 系统会根据请求者的身份信息调用上下文处理器获取其属性信息, 根据被访问者的身份信息获取其策略, 然后将属性信息与策略中用户约束条件利用策略中的评估函数进行比较. 如果满足则查询隐私策略中的数据(客体)信息, 将结果集返回给用户, 否则提示用户访问被拒绝.

例如, 雷沃集团(Lovol)想要访问福克斯汽车制造商(Focus)的地址信息, 如图2所示, 服务系统会检索出雷沃集团的属性信息形成新的请求, 然后与策略表中福克斯汽车制造商的策略进行评估, 将评估结果返回.

3　隐私保护中间件的实现与应用

本文以山东省制造业信息化服务平台为背景, 设计实现了个性化隐私保护平台. 山东省制造业信息化服务平台是为胶东半岛汽车及汽车配件中小企业产品的信息服务、异地数字设计、协同研究开发, 提供的一个完备的和可靠的网络化先进制造技术支撑平台. 扩展后的山东省制造业信息化服务平台体系架构图如图3所示, 其中白色部分为扩展的功能部分.

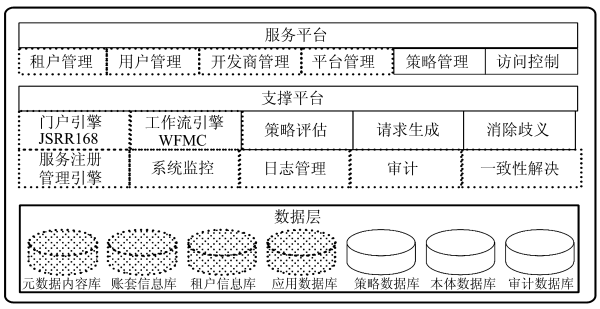

图 3　扩展后体系架构

Figure 3　Extended system architecture

本文主要扩展了数据层、支撑平台以及服务平台实现个性化隐私保护. 首先扩展数据层, 增加策略数据库, 本体数据库以及审计数据库用于存储策略, 消除语义歧义以及安全审计. 在扩展数据库的基础上扩展了支撑平台实现策略评估、请求生成以实现访问控制以及策略管理的功能, 同时为策略提供一致性检查有效防止策略冲突, 在扩展本体数据库的基础上实现了消除语义歧义功能, 在扩展审计数据库的基础上实现安全审计功能. 最后在服务平台中增加用户策略管理、访问控制和安全审计的服务.

本文依据第三方服务平台隐私保护模型, 设计实现了隐私保护中间件系统(图4), 针对山东省制造业信息化服务平台现有的用户数据, 实现了交互式隐私策略定义、策略管理并依据隐私策略实施访问控制. 实例系统如图4所示. 其中, 图4a为个性化隐私策略定义, 用户根据隐私需求定义隐私策略; 图4b为访问控制实例, 据隐私策略将可访问的数据予以显示, 不可访问的数据系统予以屏蔽; 图4c为策略管理实例, 用户可对策略进行查

a 个性化隐私定义

b 访问控制

c 策略管理

图 4　山东省制造业信息化服务平台个性化隐私保护实例系统

Figure 4　Snapshots of personalized privacy protection system applied in Shandong province manufacturing information platform

看修改以及删除等操作. 这一系统能够支持个性化的隐私偏好策略, 并按照其制定的隐私策略实现访问控制的功能平台, 当用户访问平台资源时, 系统自动依据被访问者的隐私策略实施访问控制功能.

4 结束语

针对第三方平台隐私保护的需求, 提出了第三方服务平台的隐私保护模型, 通过扩展原有数据以及据隐私策略动态访问控制实现了隐私保护模型中间件, 通过提供用户定义隐私策略, 为用户提供细粒度的访问控制机制解决了第三方平台上用户之间隐私泄露的问题并应用到山东省制造业信息化服务平台中. 本文的方法灵活、通用、有效、易扩展, 便于部署在第三方平台上. 未来工作将扩展第三方平台隐私保护框架, 利用本体技术对多异构域中的安全策略及主客体属性等附加语义信息, 更好地保护异构域间操作的安全性, 同时建立起安全审计机制, 实现对历史访问行为的监控与追踪, 做到更为完整的保护.

参考文献

[1] OASIS. eXtensible access control markup language 2(XACML) version 2.0[S/OL]. (2005-2-1)[2012-03-12]. http://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-core-spec-os.pdf.

[2] W3C. Platform for privacy preferences (P3P) project: enabling smarter privacy tools for the Web[S/OL]. (2007-10-7)[2012-03-12]. http://www.w3.org/P3P/.

[3] Paul Ashley, Satoshi Hada, Günter Karjoth, et al. Enterprize privacy authorization language 1.1 specification[R/OL]. (2003-11-14)[2012-03-12]. http://www.zurich. ibm. com/ security/enterprise-privacy/epal/Specification/.

[4] Yasin Laura-Silva, Walid Aref. Realizing privacy-preserving features in hippocratic databases [C/OL]//Data Engineering Workshop. Istanbul Turkey, 2006: 198−206. (2006-12-22)[2012-03-12]. http://www. cs. purdue. edu/research/technical_reports/2006/TR%2006-022.pdf.

[5] Hong Yuan, Lu Shuo, Liu Qian, et al. Preserving privacy in E-health systems using hippocratic databases[C]//2008 32 nd Annual IEEE International Computer Software and Applications Conference. Washington: IEEE Computer Society Press, 2008: 692−697.

[6] Ghani Norjihan Abdul, Sidek Zailani Mohamed. Privacy-preserving in Web services using hippocratic database[C]//2008 International Symposium on Information Technology. Kuala Lumpur: K & L Press, 2008, 1: 1−5.

[7] Kanna Al-Falahi, Yacine Atif, Said Elnaffar. Social networks challenges and new opportunities[C]//Green Computing and Communications (GreenCom), 2010 IEEE/ACM Int'l Conference on & Int'l Conference on Cyber, Physical and Social Computing. New York: IEEE Computer Society Press, 2010: 804−808.

[8] 马晓君, 孙宇清. 基于隐私保护的社会网络用户偏好分析模型[J]. 计算机科学, 2011, 38(11): 207−211.

[9] 马晓君, 孙宇清, 刘发朋. 社会网络中的隐私保护[J]. 中国计算机学会通讯, 2011, 7(1): 52−56.

[10] 时杰, 朱虹, 周淳, 等. 关系数据库细粒度访问控制设计与实现[J]. 华中科技大学学报: 自然科学版, 2010, 38(10): 35−38.

[11] Claudio A A, Sabrina De C V. An XACML-based privacy-centered access control system[C]//Proceeding WISG '09 Proceedings of the First ACM Workshop on Information Security Governance. New York: IEEE Computer Society Press, 2009: 49−58.

[12] Gina Kounga, Marco Casassa Mont, Pete Bramhall. Extending XACML access control architecture for allowing preference-based authorisation[C]//Trust, Privacy and Security in Digital Business Lecture Notes in Computer Science. Berlin: Spring-Verlag, 2010: 153−164.

(责任编辑: 张 勇)