软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2016,27(8):1948−1963 [doi: 10.13328/j.cnki.jos.004945]

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

面向服务组合的用户隐私需求规约与验证方法*

彭焕峰[1,2], 黄志球[1], 范大娟[2], 章永龙[3]

[1](南京航空航天大学 计算机科学与技术学院,江苏 南京 211106)

[2](南京工程学院 计算机工程学院,江苏 南京 211167)

[3](扬州大学 信息工程学院,江苏 扬州 225127)

通讯作者: 黄志球, E-mail: zqhuang@nuaa.edu.cn

摘 要: 用户向 Web 服务组合提供隐私数据时,不同用户有自身的隐私信息暴露需求,服务组合应支持用户隐私需求的可满足性验证.首先提出一种面向服务组合的用户隐私需求规约方法,用户能够定义隐私数据及不同使用情境的敏感度,采用敏感度-信誉度函数明确可以使用隐私数据的成员服务,简化隐私需求的同时,提高了隐私需求的通用性.为了验证服务组合是否满足用户隐私需求,首先通过隐私数据项依赖图(privacy data item dependency graph,简称 PDIDG)描述组合中隐私数据项的依赖关系,然后采用隐私开放工作流网(privacy open workflow net,简称 POWFN)构建隐私敏感的服务组合模型,通过需求验证算法验证服务组合是否满足用户隐私需求,从而能够有效防止用户隐私信息的非法直接暴露和间接暴露.最后,通过实例分析说明了该方法的有效性,并对算法性能进行了实验分析.

关键词: 信誉度;服务组合;隐私保护;隐私开放工作流网;隐私数据项依赖图

中图法分类号: TP311

中文引用格式: 彭焕峰,黄志球,范大娟,章永龙.面向服务组合的用户隐私需求规约与验证方法.软件学报,2016,27(8):1948−1963. http://www.jos.org.cn/1000-9825/4945.htm

英文引用格式: Peng HF, Huang ZQ, Fan DJ, Zhang YL. Specification and verification of user privacy requirements for service composition. Ruan Jian Xue Bao/Journal of Software, 2016,27(8):1948−1963 (in Chinese). http://www.jos.org.cn/1000-9825/4945.htm

Specification and Verification of User Privacy Requirements for Service Composition

PENG Huan-Feng[1,2], HUANG Zhi-Qiu[1], FAN Da-Juan[2], ZHANG Yong-Long[3]

[1](College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

[2](College of Computer Engineering, Nanjing Institute of Technology, Nanjing 21167, China)

[3](College of Information Engineering, Yangzhou University, Yangzhou 225127, China)

Abstract: Users have different privacy information disclosure requirements when they submit private data to service composition, and the composition should support the verification of users' privacy requirements. This paper puts forward a flexible method for users to produce privacy requirement specifications. Users can define the sensitivity of private data and its usage in different situations, and restrict the member services that can use private data with sensitivity-reputation function. The simplification and universality of the privacy requirements can be improved by using this method. The process first establishes privacy data item relations by using the privacy

* 基金项目: 国家自然科学基金(61272083); 国家高技术研究发展计划(863)(2015AA015303); 中国博士后科学基金(20110491411); 江苏省博士后科研计划(1101092C)

Foundation item: National Natural Science Foundation of China (61272083); National High-Tech R&D Program of China (863) (2015AA015303); China Postdoctoral Science Foundation (20110491411); Jiangsu Planned Projects for Postdoctoral Research Funds (1101092C)

收稿时间: 2014-09-21; 修改时间: 2015-04-08; 采用时间: 2015-11-21; jos 在线出版时间: 2015-12-21

CNKI 网络优先出版: 2015-12-22 14:59:18, http://www.cnki.net/kcms/detail/11.2560.TP.20151222.1459.002.html

data item dependency graph (PDIDG), then models the service composition with privacy open workflow net (POWFN), and at last, makes sure whether service composition meets the user's privacy requirements by privacy requirements verification algorithm. An example is provided to illustrate the effectiveness of the method, and experiment analysis on the performance of the verification algorithm is carried out at the end of paper.

Key words: reputation; service composition; privacy protection; privacy open workflow net; privacy data item dependency graph

Web 服务作为一种基于 Internet 的崭新分布式计算模型,适合作为一种独立而开放的实体在互联网环境中发布和使用[1].用户为使用服务提供的功能,需要提供必要的个人隐私信息,但由于 Web 服务开放、动态和自治的特点,隐私信息一旦被收集,用户就难以控制服务如何使用和暴露这些信息[2].随着用户隐私信息侵犯案例的增加,隐私信息保护问题越来越受到用户的关注,特别是在服务组合的情况下,服务提供者通过将服务进行组合以形成粒度更大的服务,从而实现复杂的业务逻辑[3].用户隐私信息是通过服务组合暴露给成员服务,由于用户与成员服务之间缺乏隐私信息使用的相关协议,因此难以保证在组合执行过程中,隐私信息能否按照用户的意愿进行暴露和使用[4].

用户对隐私信息有着自身的保护需求,对具体隐私数据项的敏感程度也不同,且对隐私数据项组合使用时敏感度更高,例如同时使用身份证号码和姓名时,比单独使用更担心隐私信息的泄露.随着用户越来越重视隐私信息的保护,用户更倾向于选择在隐私保护方面信誉度更高的服务.用户为使用服务组合提供的功能,需要向组合提供隐私数据,这类数据称为直接隐私数据.组合将直接隐私数据提供给某成员服务的行为,称为直接隐私暴露.组合执行过程中会产生新的数据,而有些新产生数据可能会依赖于直接隐私数据,这类数据称为间接隐私数据.组合将间接隐私数据提供给某成员服务的行为称为间接隐私暴露.为提高竞争力,服务组合提供者需要构建隐私敏感的服务组合模型,以能够支持用户隐私需求的验证,且用户隐私需求验证算法能够同时检测非法的直接隐私暴露和间接隐私暴露.

本文主要工作和创新点主要如下:

(1) 提出一种基于隐私保护信誉度的用户隐私需求规约方法.采用此方法,用户能够定义隐私数据项及其组合使用时的敏感度;同时,可以指定隐私数据使用情境的敏感度,并通过敏感度-信誉度函数明确可以使用隐私数据的服务.与其他隐私需求规约方法相比,用户不必指定组合中使用隐私数据的成员服务,从而具有更好的灵活性与通用性,且简化了需求的复杂度.

(2) 通过隐私数据项依赖图描述组合中隐私数据项的依赖关系,使用带隐私语义的开放工作流网构建隐私敏感的服务组合模型,最后,通过隐私需求验证算法验证组合是否满足用户的隐私需求.该方法不但能够检测隐私信息的非法直接暴露,而且能够检测隐私信息的非法间接暴露.

1 相关工作分析

服务计算及云计算中,用户隐私保护研究可以分为面向数据和面向使用行为两类:前者通过对隐私数据进行加密、匿名、扰动等方法对隐私信息进行保护;后者主要关注隐私数据使用行为的分析与约束,包括用户隐私需求规约方法、服务对隐私需求的实施、服务隐私策略与用户需求协商及演化等研究内容[5].本文的研究属于后者.

对用户隐私需求规约方法及实施等方面的研究,许多组织提出了相应的规范与技术框架,例如,W3C 组织提出隐私偏好平台(platform for privacy preferences,简称 P3P)[6]来定义服务提供者的隐私策略,并引入隐私偏好描述语言(a P3P preference exchange language,简称 APPEL)[7]以方便用户定义其隐私需求.但 P3P 与 APPEL 主要针对 Web 站点定义隐私策略及用户隐私偏好,并不能直接用于面向服务组合的隐私需求规约.OASIS 组织提出的可扩展访问控制标记语言(extensible access control markup language,简称 XACML)[8]是一种通用的访问控制策略语言和执行授权策略框架,但主要关注对服务提供者的隐私数据应用隐私保护策略.针对用户使用在线服务时如何表达其隐私需求,研究者从各国或组织对个人隐私数据的法律或指导原则出发提出相应的用户隐私需求规约方法.例如,文献[9]根据经济合作与发展组织(organization for economic co-operation and development,

简称 OECD)提出的个人隐私数据保护原则,从隐私数据的收集者、使用目的、保留期限等方面定义了用户隐私需求的元模型.但由于未对使用隐私数据的服务信誉度做出限定,因此很难避免隐私数据暴露给不可信的服务.针对如何保证服务组合按照用户需求使用隐私数据这一研究问题,从隐私需求实施的阶段可以分为组合阶段、运行前检查、运行时监控这 3 类实施方法.组合阶段是指在进行服务组合时需满足用户隐私需求,例如,文献[10]提出了一种同时满足功能需求和隐私需求的服务组合算法,且该算法能够保证用户隐私信息的最小暴露.运行前检查是指用户在使用组合前,首先验证用户需求与组合隐私策略是否匹配,只有匹配才开始业务流程的执行.例如,文献[11]将用户隐私需求与 BPEL 的元素建立映射关系,在此基础上验证 BPEL 是否满足用户需求.运行时监控是指在组合执行过程中,监控服务对隐私数据的使用行为是否满足用户需求.例如,文献[12]提出了可信第三方隐私代理服务(privacy proxy service,简称 PPS)的概念,由隐私代理服务存储用户隐私数据,通过代理向成员服务暴露隐私数据,从而在运行时控制组合中成员服务对隐私数据的使用行为.上述研究都只能避免隐私数据的直接暴露,并不能将保护需求延伸到间接隐私数据的保护.针对此问题,有研究者采用信息流技术来保证隐私信息流的安全,信息流技术可以有效防止隐私信息的直接暴露和隐式暴露,从而更好地保护用户隐私信息.信息流静态分析方法为用户隐私数据标注机密等级,为成员服务标注安全等级,需要分析组合的所有可能执行路径,只有满足安全策略的信息流才是被允许的,典型的研究如文献[13].信息流动态分析的方法同样为隐私数据与成员服务添加标签,在运行时检查是否满足信息流安全策略,以决定是否释放数据给使用者,典型研究如文献[14–16].这些研究基于信息流安全策略计算服务流程新产生数据的机密等级,并制定访问规则,以保证服务组合信息流的安全,虽然能够有效防止用户隐私信息的非法暴露,但要求新产生数据的机密等级不低于输入数据的机密等级.而有些新产生数据并不包含隐私信息,或者机密性会降低,由于未细粒度地构建降密策略,新产生数据仍然被赋以高的机密等级,因此限制过于严格,从而影响实际应用.

本文针对用户如何描述隐私暴露需求与验证服务组合是否满足用户需求这两个关键问题开展研究.文献[17]关于网络隐私问题的研究指出,隐私数据的敏感度因人而异;同时,Hayes 等人对用户选择在线服务时考虑因素的研究发现,服务的信誉度排在第一位[18].因此,为保证隐私数据释放给更为可信的服务,用户可以根据自身隐私数据的敏感度设定服务的信誉度阈值,在保证隐私数据释放给可信服务的同时,由于不需要在需求中指定具体的服务,从而在简化隐私需求的同时,提高了隐私需求的通用性.已有研究者对此开展类似研究工作,例如,文献[19,20]提出使用隐私策略矩阵对成员服务的隐私数据使用权限进行规约,通过比较隐私策略中敏感数据被释放的信誉度阈值与服务信誉度的大小,决定是否对成员服务的隐私数据使用行为进行授权,从而细粒度地保护用户的隐私数据.但在构建服务组合模型时,缺乏现有工具的支持.文献[21]提出一种基于隐私数据敏感度及服务信誉度的服务组合隐私保护框架,但用户隐私需求的验证工作由客户端完成,服务组合需要将隐私数据的使用信息反馈到客户端,这会在一定程度上暴露服务组合的内部细节.更重要的是,这些方法均不支持隐私数据项组合的暴露需求,也不能避免非法的间接隐私数据暴露.文献[17]的研究同时指出,用户隐私信息的敏感度与使用情境也有关系.因此,本文的方法与现有研究相比具有如下特点:

(1) 用户在隐私需求中可以定义隐私数据项组合的敏感度及不同使用情境的敏感度,通过隐私数据使用综合敏感度对服务使用隐私数据的行为进行综合度量,采用敏感度-信誉度函数明确可以使用隐私数据的服务,在便于统一用户隐私需求的同时,能够保证组合将隐私数据释放给可信的成员服务.

(2) 通过隐私数据项依赖图建立隐私数据项的依赖关系,采用依赖因子表示隐私数据项之间的依赖程度,结合直接隐私数据项的敏感度计算组合执行过程产生的间接隐私数据的敏感度,将用户的隐私需求扩展到间接隐私数据的保护,从而能够有效防止用户隐私信息的非法直接暴露和间接暴露.

(3) 由于开放工作流网(open workflow net,简称 OWFN)便于分析系统与外界的信息交互行为,且开源工具 BPEL2oWFN 可以将 BPEL 自动生成 OWFN 模型.对 OWFN 进行隐私语义扩展得到隐私开放工作流网(privacy open workflow net,简称 POWFN),对 BPEL 构建的 POWFN 模型既有现有工具的支持,也便于分析组合对成员服务的隐私暴露行为.

2 隐私需求规约方法

随着Web服务的普及,网络中出现了大量的Web服务应用,对于提供相同功能的多个服务,用户往往倾向于熟悉且信誉度较高的服务.随着隐私信息侵权案例的增加,用户越来越关注隐私信息的保护,相应的权威组织应对软件服务在个人隐私信息保护方面进行综合评价,评价的结果作为用户考虑隐私保护时选择服务的一个依据[22].服务的隐私保护信誉度越高,用户越信任该服务,且会赋予越多的隐私数据使用权限.

定义 1(服务隐私保护信誉度). 服务隐私保护信誉度是服务在保护用户隐私数据的能力、历史表现等多方面的综合评价指标.将服务隐私保护信誉度量化到区间[0,1],数值越大表示信誉度越高,1表示具有最高信誉度.

服务隐私保护信誉度的评定超出本文的范围,在此不做进一步讨论.为简化表述,后文中将服务隐私保护信誉度简称为信誉度.对当前运行服务的隐私保护信誉度具有客观、普遍认可的评价机制是本文提出的隐私需求规约方法的前提.

参照Weible对个人信息敏感度的定义[23],给出用户隐私数据敏感度的定义.

定义 2(隐私数据敏感度). 隐私数据敏感度是指在某一特定情形下,用户对提供隐私数据感到顾虑的程度.将隐私数据敏感度量化到区间(0,1],数值越大表示敏感度越高,1表示最高敏感度.

用户背景的不同会对数据的敏感度产生影响,且向同一服务暴露的隐私数据越多,隐私信息泄露后的危害越大.因此,用户根据自身的隐私保护偏好,应不但可以对单独的隐私数据项设定敏感度,同时也能对多个数据项的组合设定敏感度.

定义 3(数据敏感度声明). 数据敏感度声明 $dStatement=(dGroup,s)$是一个二元组,其中:$dGroup=\{d_1,d_2,...,d_n\}$ $(n\geqslant 1)$是隐私数据项的有限集合;s表示隐私数据项集合 $dGroup$ 的敏感度,且 $s\in(0,1]$.

例如,$dStatement_1=(\{name\},0.3)$表示用户认为姓名是隐私数据,且其敏感度声明为 0.3;$dStatement_2=(\{name,phone\},0.6)$表示服务若同时使用姓名和电话号码两个隐私数据项,用户声明这一隐私数据项组合的敏感度为0.6.值得说明的是,如果 $dStatement_2$ 中声明$\{name,phone\}$的敏感度小于等于 0.3,则 $dStatement_1$ 与 $dStatement_2$ 这两个数据敏感度声明之间存在冲突.由于冲突检测算法相对简单,本文不再详述.

用户对隐私数据使用情境也是敏感的,本文从服务使用数据的目的、保留期限两个方面描述用户的敏感度,提出的方法很容易扩展到其他方面.

服务保留隐私数据的时间越长,用户对服务非法泄露或使用其隐私信息越担忧.通过期限敏感度函数,量化用户对隐私数据保留期限的敏感度.

定义 4(期限敏感度函数). $\delta=fRetention(sr),sr\in SR,SR$ 是保留期限的有限集合,$\delta\in(0,1]$,δ称为期限敏感度系数,δ越大,表示用户的敏感程度越高.

SR 由用户自由定义,例如,可以定义 $SR=\{[0,0],(0,10],(10,100],(100,+\infty)\}$,单位是天,表示将服务保留隐私数据的期限分为4个区间段;$fRetention((10,100])=0.4$表示若保留数据期限在10到100天区间内,则敏感度系数为0.4.

用户对服务使用其隐私数据的目的也是敏感的,例如,用户的身份证号码用于银行支付服务及网站调查分析两种使用目的,用户可能对网站调查分析这一使用目的更为敏感,通过目的敏感度函数,量化用户对隐私数据使用目的的敏感度.

定义 5(目的敏感度函数). $\mu=fPurpose(p),p\in P,P$ 是隐私数据使用目的的有限集合,$\mu\in(0,1]$,μ称为目的敏感度系数,μ越大,表示用户的敏感程度越高.

例如,$fPurpose(individual\_analysis)=0.5$ 表示用于个人分析时用户的使用目的敏感度系数为 0.5.使用目的集合 P 应有统一的标准,集合 P 的确定超出了本文的研究范围,且限于篇幅,因此不做进一步讨论.

服务组合在接收了用户隐私数据后,需要调用成员服务完成业务功能,即成员服务接收了服务组合释放的隐私数据,执行其相应的功能,并可能保留隐私数据一定的时间.下面给出成员服务使用用户隐私数据的形式化描述.

定义 6(隐私数据使用). 成员服务对用户隐私数据的使用可以形式化表示为一个三元组 $dUse=(dGroup,$

purpose,*retention*),其中,*dGroup* 是使用的隐私数据项的有限集合,*purpose* 表示隐私数据的使用目的,*retention* 表示隐私数据的保留期限.

成员服务使用隐私数据的目的、保留期限反映了数据的使用情境,隐私数据、使用目的、保留期限一起构成了成员服务的一次隐私数据使用行为.

定义 7(隐私数据使用综合敏感度). 隐私数据使用综合敏感度是用户对成员服务使用其隐私数据行为的敏感度综合度量,计算公式如下:

$$s=((1+\delta+\mu)\times dGroup.s)/3 \tag{1}$$

隐私数据使用综合敏感度由 3 部分组成:*dGroup.s* 代表使用的隐私数据项集合的敏感度;δ由成员服务保留数据的期限及用户定义的期限敏感度函数确定,$\delta\times dGroup.s$ 表示用户对 *dGroup* 的期限敏感度;μ由成员服务使用数据的目的及用户定义的目的敏感度函数确定,$\mu\times dGroup.s$ 表示用户对 *dGroup* 的使用目的敏感度.为便于统一处理,对这 3 部分的计算结果取平均,故 *s* 的取值范围为(0,1].由于做了归一化处理,所以隐私数据使用综合敏感度比隐私数据敏感度要小.

服务的信誉度越高,用户就越信任该服务,可以释放敏感度更高的数据,这也是本文提出的用户隐私需求规约方法的核心.用户通过定义敏感度与服务信誉度的对应关系,确定成员服务是否可以使用隐私数据,即服务组合根据成员服务使用的隐私数据、目的、保留期限,通过公式(1)计算隐私数据使用综合敏感度,然后比较用户定义的敏感度与服务信誉度的对应关系,以此来决定是否释放用户的隐私数据给成员服务,从而实现了服务组合按照用户需求暴露和使用其隐私数据这一目的.敏感度与服务信誉度的对应关系通过敏感度-信誉度函数定义.

定义 8(敏感度-信誉度函数). *r*=*fSense_Repute*(*s*),*s* 表示隐私数据使用综合敏感度,且 $s\in(0,1]$;*r* 表示服务的信誉度,且 $r\in[0,1]$.

例如,0.4=*fSense_Repute*(0.5)表示隐私数据使用综合敏感度为 0.5,若本次隐私数据的使用行为被允许,则成员服务的信誉度应大于等于 0.4.

定义 9(用户隐私需求). 用户隐私需求 *UPR*=(*DS*,*fRetention*,*fPurpose*,*fSense_Repute*)是一个四元组,其中,*DS* 是数据敏感度声明的有限集合,*fRetention* 是期限敏感度函数,*fPurpose* 是目的敏感度函数,*fSense_Repute* 是敏感度-信誉度函数.

从用户隐私需求的定义可以看出,用户只需给出关心的隐私数据的敏感度声明,通过期限敏感度函数定义数据保留期限方面的敏感度,通过目的敏感度函数定义隐私数据使用目的方面的敏感度;最后,通过敏感度-信誉度函数定义隐私数据使用综合敏感度与服务信誉度的对应关系,就可以灵活地表达自身的隐私需求偏好,不必针对不同的服务组合定义多个隐私保护需求,从而简化了隐私需求的复杂度,且定义的隐私需求更具通用性.

3 隐私需求验证

3.1 面向隐私保护的数据依赖模型

用户在隐私需求中定义了直接隐私数据的敏感度及暴露需求,为将用户的隐私需求扩展到间接隐私数据的保护,需构建面向隐私保护的数据依赖模型,用于描述隐私数据的依赖关系,并计算间接隐私数据的敏感度.

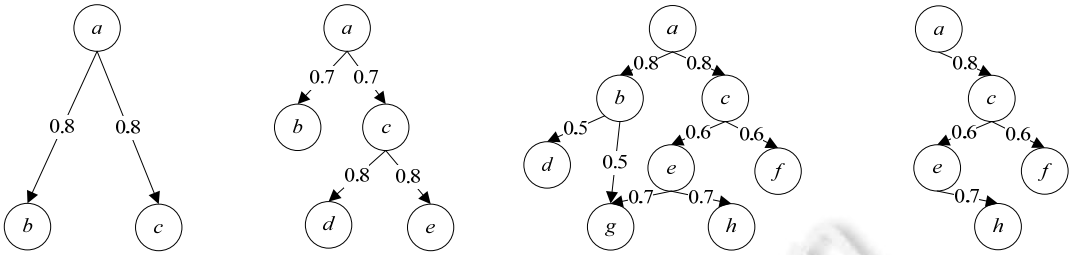
定义 10(直接隐私数据项). 用户在与服务组合交互过程中提供的涉及隐私信息的数据项.

定义 11(间接隐私数据项). 在服务组合流程执行过程中新产生的数据项,其在一定程度上依赖于直接隐私数据项,且能够间接暴露用户隐私信息.

定义 12(隐私数据项直接依赖). *dDep*=(*S*,*d*,λ)是一个三元组,*S* 是隐私数据项的有限集合;*d* 为一个原子数据项,且输出数据 *d* 依赖于输入数据项集合 *S*;λ为依赖因子,取值范围为(0,1].则关系 *dDep* 表示数据项 *d* 直接依赖于数据项集合 *S*,依赖因子为λ,λ的值越大,表示依赖程度越高,依赖关系也可以记作 $d\xrightarrow{\lambda}S$.

隐私数据项集合 *S* 中可能包含多个数据项,为简化问题,以集合 *S* 为整体设置依赖因子,而不是针对 *S* 中的每个数据项设置依赖因子.图 1(a)为隐私数据项直接依赖的例子,隐私数据项 *a* 依赖于隐私数据项集合{*b*,*c*},依

赖因子为 0.8,表示数据项 *a* 的计算输入中包括 *b* 和 *c* 两个隐私数据项,且在特定上下文中数据项 *a* 能够在一定程度上暴露 *b* 和 *c* 的内容.



(a) 隐私数据项直接依赖 (b) 隐私数据项间接依赖 (c) 隐私数据项依赖图 (d) 约简后的隐私数据项依赖图

Fig.1 Privacy data item dependency graph

图 1 隐私数据项依赖图

定义 13(隐私数据项依赖链). 设有隐私数据项直接依赖关系集合 $chainD=(dDep_1,dDep_2,...,dDep_n)$,其中,$n\geqslant 1$,如果任意相邻的两个依赖关系 $dDep_i,dDep_{i+1}$,其中,$1\leqslant i\leqslant n-1$,存在 $d_{i+1}\in S_i$,则集合 *chainD* 构成一个隐私数据项依赖链.

定义 14(隐私数据项间接依赖). 设有隐私数据项依赖链 $chainD=(dDep_1,dDep_2,...,dDep_n)$,其中,$n\geqslant 1$,对任意依赖关系 $dDep_i,dDep_j,i<j$,则隐私数据项 d_i 间接依赖于隐私数据项集合 S_j,记作 $d_i\rightarrow S_j$.

在图 1(b)中,由于存在依赖链 $\{a\xrightarrow{0.7}\{b,c\},c\xrightarrow{0.8}\{d,e\}\}$,则存在间接依赖关系 $a\rightarrow\{d,e\}$.隐私数据项间接依赖仅表示隐私数据项与某数据项集合之间具有间接依赖关系,具体依赖程度由依赖链中的直接依赖关系确定,因此不需要标注间接依赖的依赖因子.

定义 15(隐私数据项依赖图). 隐私数据项依赖图(privacy data item dependency graph,简称 PDIDG)=(*V*,*E*,*S*,*W*)是一个带权有向无环图,其中,

- *V* 是隐私数据项的有限集合;
- *E* 是边的集合;
- *S* 称为 PDIDG 的敏感度函数,即对于每个结点 $v\in V$,均有取值范围为(0,1]的正实数 $S(v)$与之对应;
- *W* 称为 PDIDG 的依赖因子函数,即对于每条有向边 $e\in E$,均有取值范围为(0,1]的正实数 $W(e)$与之对应.

结点对应的隐私数据项分为直接隐私数据项和间接隐私数据项两类,其中,

(1) 出度为 0 的结点为直接隐私数据项结点;

(2) 出度不为 0 的结点为间接隐私数据项结点.

图 1(c)是一个隐私数据项依赖图的例子,其中,*d*,*f*,*g*,*h* 是直接隐私数据项,*a*,*b*,*c*,*e* 是间接隐私数据项.

- 存在如下直接依赖关系:

$$e\xrightarrow{0.7}\{g,h\},c\xrightarrow{0.6}\{e,f\},b\xrightarrow{0.5}\{d,g\},a\xrightarrow{0.8}\{b,c\};$$

- 存在隐私数据项依赖链,如:

$$chainD_1=(a\xrightarrow{0.8}\{b,c\},c\xrightarrow{0.6}\{e,f\},e\xrightarrow{0.7}\{g,h\}),chainD_2=(a\xrightarrow{0.8}\{b,c\},b\xrightarrow{0.5}\{d,g\});$$

- 存在如下间接依赖关系:

$$c\rightarrow\{g,h\},a\rightarrow\{e,f\},a\rightarrow\{g,h\},a\rightarrow\{d,g\}.$$

直接隐私数据项的敏感度由用户在隐私需求中指定,间接隐私数据项的敏感度根据 PDIDG 中的依赖关系计算获得,只需从直接隐私数据项开始,依次计算间接隐私数据项的敏感度即可.

数据项集合 *S* 中可能仅包含间接隐私数据项或直接隐私数据项,也可能两类都包含,间接隐私数据项 *d* 的敏感度 *sense_d* 的计算公式如下:

$$sense\_d=\max(\max(\{indirect\_data.s|indirect\_data\in S\}),\max(\{dStatement.s|dStatement.dGroup\subseteq S\}))\times\lambda \quad (2)$$

max({*indirect_data.s*|*indirect_data*∈*S*})表示数据项集合 *S* 中间接隐私数据项的最大敏感度,其中,*indirect_data* 表示 *S* 中的间接隐私数据项;max({*dStatement.s*|*dStatement.dGroup*⊆*S*})表示集合 *S* 中直接隐私数据项的最大敏感度,其中,*dStatement* 是用户隐私需求中的隐私数据敏感度声明,*dStatement.dGroup* 表示声明的隐私数据项集合.间接隐私数据项 *d* 的敏感度 *sense_d* 等于两者的最大值乘以依赖因子λ.

例如,考虑直接依赖关系 $a \xrightarrow{0.8} \{b,c,d,e,f\}$,其中,*b*,*c*,*d* 为直接隐私数据项,敏感度分别为 0.3,0.4,0.5,且用户需求中存在敏感度声明 $dstatement_1=(\{b,c\},0.7)$,$dstatement_2=(\{c,d\},0.8)$,假设间接隐私数据项 *e*,*f* 的敏感度分别为 0.6,0.4,则 *a* 的敏感度为 max(max(0.3,0.4,0.5,0.7,0.8),max(0.6,0.4))×0.8=0.64.

为了支持用户隐私需求验证,服务组合提供者一般可由组合流程设计者与隐私分析人员[24]组成,隐私数据的依赖涉及组合中数据输入、输出依赖关系及具体应用的语义分析.为了更精确地表达数据项的隐私依赖关系,本文采用手动方式创建隐私数据项依赖图,由隐私分析人员完成这项工作.隐私分析人员根据服务组合与成员服务的交互情况评估隐私数据项之间的依赖程度,设置依赖因子,由于不知道具体用户的特定隐私需求,需要针对用户请求消息中提供的数据,将所有可能被认定为隐私数据的数据项建立完全依赖图,且不应出现环路.当验证具体用户的隐私需求时,只需要根据用户提供的直接隐私数据项对已建好的完全依赖图进行约简,形成特定用户的依赖图即可.约简方法仅需沿着非用户认定的直接隐私数据项删除相应的结点和边即可.假设图 1(c)是隐私分析人员建立的完全依赖图,若某用户隐私需求中仅认定数据项 *h*,*f* 为直接隐私数据项,约简时需要依次删除结点 *g*,*d*,*b* 以及边⟨*b*,*d*⟩,⟨*b*,*g*⟩,⟨*e*,*g*⟩,⟨*a*,*b*⟩,得到约简后的依赖图如图 1(d).约简算法较为简单,本文不再详述.此外,隐私分析人员建立依赖图时间接隐私数据项的敏感度均为 0,在进行隐私需求验证时,根据用户隐私需求获取直接数据项的敏感度,然后根据公式(2)计算间接数据项的敏感度.

3.2 隐私敏感的服务组合模型

BPEL 已经成为事实上的 Web 服务组合标准语言,本文针对 BPEL 构建隐私敏感的服务组合模型,以支持用户隐私需求的验证.为分析验证组合的性质,研究者提出了大量针对 BPEL 的建模方法.由于 Petri 网既有严格的数学定义,又有直观的图形表示,并且具有丰富的系统描述手段和系统行为分析技术,因此,不少研究者通过将 BPEL 转换为 Petri 网模型来验证服务组合的各种性质.开放工作流网 OWFN 是一种特殊的 Petri 网,OWFN 在传统工作流网的基础上引入与外界交互的消息库所,因此便于分析系统与外界消息的交互行为.Lohmann 在文献[25−27]中详细给出了使用 OWFN 对 BPEL 进行建模的方法,将 BPEL 中的活动转换为变迁,内部控制逻辑转换为内部库所,与外界交互的消息转换为消息库所.服务组合通过消息与成员服务进行交互,用户的隐私数据最终通过消息暴露给成员服务,因此非常适合采用 OWFN 构建服务组合模型.为了能够分析服务组合 BPEL 的隐私数据使用情况,通过扩展 OWFN,提出支持隐私数据暴露分析的隐私开放工作流网 POWFN.

定义 16(**隐私开放工作流网**). $POWFN=(P,T,F,M_0,P_I,P_O,F_{IO},G)$,其中,

(1) *P* 是库所的集合;

(2) *T* 是变迁的集合;

(3) *F* 是流关系的集合,且 $F\subseteq(P\times T)\cup(T\times P)$;

(4) M_0 是初始标识;

(5) P_I 是输入消息库所的集合,$\forall x\in P_I$,前集 ${}^{\bullet}x=\varnothing$;

(6) P_O 是输出消息库所的集合,$\forall x\in P_O$,后集 $x^{\bullet}=\varnothing$;

(7) $P_I\subseteq P\wedge P_O\subseteq P\wedge P_I\cap P_O=\varnothing$;

(8) $F_{IO}\subseteq(P_I\times T)\cup(T\times P_O)\subseteq F$,即$\forall p\in P_I:(t,p)\notin F_{IO}\wedge\forall p\in P_O:(p,t)\notin F_{IO}$;

(9) 记 $T_O=\{t|(t,p)\in(T\times P_O)\}$为输出变迁集合,*G* 是守护函数,$G:T_O\to Expression$,其中,*Expression* 是一个结果为 Bool 型的函数表达式:$f(srv.t,msg.s,srv.p,srv.r)$.*srv.t* 表示服务的信誉度,*msg.s* 表示服务接收消息的敏感度,*srv.p* 表示服务使用隐私数据的目的,*srv.r* 表示服务针对隐私数据的保留期限,*msg*,*srv.p*,*srv.r* 代表了成员服务对用户隐私数据的一次使用行为,函数 *f* 验证成员服务对隐私数据的使用是否满足用户隐私需求.

其中,POWFN 有一个起始库所 $i\in P$,使得 ${}^{\bullet}i=\varnothing$,有一个终止库所 $o\in P$,使得 $o^{\bullet}=\varnothing$,且除消息库所外的其他节点都属于从 i 到 o 的一条路径上.与输入消息库所具有流关系的变迁称为输入变迁,其集合记为 T_I;与输出消息库所具有流关系的变迁称为输出变迁,其集合记为 T_O,T_{IO} 代表输入、输出变迁集合.

为了验证服务组合是否按照用户需求向成员服务暴露隐私数据,需要在输出变迁上增加守护函数,变迁在满足其他触发条件的基础上,只有输出变迁对应成员服务的信誉度与输出消息的使用综合敏感度之间的关系满足用户需求,变迁才能触发.

一个简单的 POWFN 例子如图 2 所示,p_1 为起始库所,p_5 为终止库所,p_2 是输入消息库所,p_4 是输出消息库所,t_1 为输入变迁,t_2 为输出变迁,同时在输出变迁 t_2 上添加守护函数.

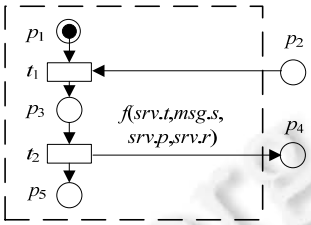


Fig.2 An example of POWFN

图 2 一个 POWFN 例子

BPEL2oWFN[28]是 GNU 开源组织发布的一个开源工具,该项目由 Lohmann 领导,最新版本为 2.0.3,支持 BPEL2.0.该工具将用 BPEL 编写的业务流程转换为 OWFN,转换前首先对 BPEL 的语法及一些逻辑错误进行分析,并分类给出分析结果,通过工具Fiona[29]可以分析 OWFN 的可控性,或者通过其他模型检测工具验证 OWFN 的死锁等 petri 网属性及系统的时序逻辑.因此,采用 OWFN 对 BPEL 进行建模具有现有工具的良好支持,且很容易对生成的开放工作流网模型文件进行解析和扩展,生成 BPEL 的 POWFN 文件.

构建 BPEL 的 POWFN 模型的步骤如下:

(1) 采用工具 BPEL2oWFN 生成 BPEL 的 OWFN 文件.以图 2 为例,OWFN 文件中输出变迁 t_2 对应的代码如下:

TRANSITION t_2 {output}

CONSUME p_3;

PRODUCE p_4, p_5;

(2) 解析 OWFN 文件、BPEL 及 WSDL 文件,最终在 OWFN 文件的输出变迁上添加守护函数,生成 POWFN 文件.同样以图 2 为例,POWFN 文件中输出变迁 t_2 对应的代码如下:

TRANSITION t_2 {output}

CONSUME p_3;

GUARD $f(srv.t,msg.s,srv.p,srv.r)$;

PRODUCE p_4, p_5;

其中,*srv* 为变迁 t_2 对应操作所属成员服务,*msg* 为输出消息,*msg.s* 为消息敏感度,*srv.t*,*srv.p* 和 *srv.r* 分别表示服务信誉度、使用隐私数据的目的和保留期限.值得注意的是,并不是所有输出消息变迁都对应成员服务操作的调用,例如向用户返回结果的操作对应的变迁,这类输出变迁上不添加守护函数.在针对 BPEL 构建隐私敏感的服务组合模型后,若验证 BPEL 的流程相关属性,可以采用相应工具对 OWFN 文件进行分析,若要验证 BPEL 流程的隐私属性,则需要提出新的分析方法.

3.3 隐私需求验证算法

用户隐私需求验证步骤如下:

(1) 服务组合提供者手动构建完全隐私数据项依赖图;

(2) 服务组合提供者针对 BPEL 构建隐私敏感的服务组合 POWFN 模型;

(3) 利用隐私需求验证算法验证组合是否满足用户需求.

在基于本文所提出的方法搭建隐私保护框架时,步骤(1)和步骤(2)在部署服务组合前需要完成,以支持用户隐私需求的验证.用户使用服务组合前,首先提交隐私需求,服务组合通过步骤(3)验证是否满足隐私需求.

为了验证 Web 服务组合是否满足用户的隐私需求,只需验证成员服务的信誉度与隐私数据使用综合敏感度之间的关系是否满足用户的隐私需求.针对服务组合的 POWFN 模型,验证对应的输出消息变迁上的守护函数的执行结果是否为真:若全部为真,则表明服务组合满足用户的隐私需求,即不存在非法的直接隐私暴露和间接隐私暴露.

隐私需求验证算法的具体步骤如下:

(1) 针对用户隐私需求约简完全隐私数据项依赖图,得到具体用户的隐私数据项依赖图.

(2) 根据用户隐私需求中的隐私数据敏感度声明、用户的隐私数据项依赖图计算间接隐私数据项的敏感度,计算方法采用公式(2).

(3) 依次处理 POWFN 模型中具有守护函数的输出变迁,并将验证结果保存到 *PResult* 中.具体步骤如下:

(a) 计算服务组合输出消息的敏感度;

(b) 计算成员服务的隐私数据使用综合敏感度,计算方法采用公式(1);

(c) 判断成员服务隐私数据使用是否满足隐私需求中的敏感度-信誉度函数.

用户隐私需求验证算法伪码如算法 1.

算法 1. 用户隐私需求验证算法.

输入:*UPR*=(*DS*,*fRetention*,*fPurpose*,*fSense_Repute*); //用户隐私需求

CPDIDG; //服务组合的完全隐私数据项依赖图

POWFN. //服务组合的隐私开放工作流网模型

输出:*pResult*. //详细验证结果,每个元素包含服务名称、信誉度、用户要求的信誉度阈值等信息

```
1.  pResult=∅
2.  bResult=true
3.  UPDIDG=getUserDGraph(UPR.DS,CPDIDG)  //得到用户隐私数据项依赖图 UPDIDG
4.  for each dataItem in UPDIDG.indSET do  //indSET 为 UPDIDG 中的间接隐私结点集合
5.    indirectSen=getMaxIndirectSen(dataItem)  //计算依赖的间接隐私结点的最大敏感度
6.    directSen=getMaxDirectSen(dataItem)  //计算依赖的直接隐私数据的最大敏感度
7.    dataItem.s=max(indirectSen,directSen)*getDenpendecyFactor(dataItem)  //得到间接隐私结点敏感度
8.  end for
9.  for each tg in POWFN.TG do  //tg.msg,tg.srv 分别表示变迁对应的输出消息和成员服务
10.   dirDS=getDirectPrivacyItem(tg.msg,UPDIDG)  //得到消息中直接隐私数据项集合 dirDS
11.   indDS=getIndirectPrivacyItem(tg.msg,UPDIDG)  //得到消息中间接隐私数据项集合 indDS
12.   if ((dirDS=∅) and (indDS=∅)) then continue end if  //若消息不包含隐私数据则不执行后续语句
13.   dirMaxS=0
14.   for each UPR.DS.dStatement.dGroup⊆dirDS do  //计算直接隐私数据的最大敏感度
15.     if (UPR.DS.dStatement.s≥dirMaxS) then dirMaxS=UPR.DS.dStatement.s end if
16.   end for
17.   indMaxS=0
18.   for each dataItem in indDS do  //计算间接隐私数据最大敏感度
19.     if (dataItem.s≥indMaxS) then indMaxS=dataItem.s end if
```

```
20.    end for
21.    if (dirMaxS≥indMaxS) then    //tg.msg.s 表示消息敏感度
22.      tg.msg.s=dirMaxS
23.    else
24.      tg.msg.s=indMaxS
25.    end if
26.    tg.msg.fixs=getMsgFixSens(UPR,tg.msg.s,tg.srv.p,tg.srv.r)    //得到隐私数据使用综合敏感度
27.    tRequired=getRequriedSensFromUPR(UPR,tg.msg,fixs)    //获取用户要求的信誉度阈值
28.    if tg.srv.t<tRequired then
29.      bResult=false
30.    end if
31.    pResult.add(tg.srv,tg.srv.t,tg.msg,tg.msg.s,tg.srv.p,tg.srv.r,tg.msg.fixs,tRequired,bResult)
32.    bResult=true
33. end for
34. return pResult
```

算法第(1)步对应第 3 行,由于预先对完全隐私数据项依赖图 CPDIDG 进行处理,将保留的直接隐私结点与需要约简掉的结点和边建立对应关系,可根据用户隐私需求中包含的直接隐私数据项,直接删除需要约简掉的结点和对应的边.设 CPDIDG 的结点数量|*CPDIDG.V*|=*n*,故第 3 行的算法时间复杂度可控制为 $O(n^2)$.

算法第(2)步对应第 4 行~第 8 行,即计算用户隐私数据项依赖图中间接隐私数据项的敏感度.*UPDIDG.indSET* 表示图中间接隐私数据项结点的集合,设|*UPDIDG.indSET*|=*h*,第 5 行的算法时间复杂度为 $O(h)$;第 6 行计算依赖的直接隐私数据的最大敏感度,设隐私数据敏感度声明集合的大小|*UPR.DS*|=*s*,声明中平均包含的数据项数为 *m*,依赖的直接隐私结点数为 *k*,根据公式(2)及集合包含关系的判断方法,第 6 行的算法时间复杂度为 $O(skm)$,故第 4 行~第 8 行的算法时间复杂度为 $O(h(h+skm))$.考虑 *h*,*s*,*k*,*m* 与 *n* 具有线性关系,第 4 行~第 8 行的算法时间复杂度也可以表示为 $O(n^4)$.

算法第(3)步对应第 9 行~第 33 行,即依次处理具有守护函数的变迁.*POWFN.TG* 表示 POWFN 中具有守护函数的变迁集合,设|*POWFN.TG*|=*t*,对其中的每个变迁 *tg* 验证守护函数是否为真.类似对算法第(2)步的分析,第 10 行、第 11 行分别计算消息中的直接隐私数据项集合和间接隐私数据项集合,算法复杂度均为 $O(n^2)$;第 13 行~第 16 行计算直接隐私数据的最大敏感度,根据公式(2)及集合包含关系的判断方法,算法复杂度为 $O(n^3)$;第 17 行~第 20 行计算间接隐私数据的最大敏感度,算法复杂度为 $O(n)$,故第(3)步的算法复杂度为 $O(tn^3)$.

综上所述,算法 1 的总体算法复杂度为 $O(tn^3+n^4)$.

4 实例分析与实验

通过旅行代理(travel agent,简称 TA)对本文提出的方法进行实例分析.TA 采用 Eclipse BPEL Designer 插件开发,由 Apache 服务组合引擎(orchestra director engine,简称 ODE)执行.利用服务组合隐私需求验证原型工具(privacy web service composition checker,简称 PWSCC)验证组合是否满足用户隐私需求,并对隐私需求验证算法的性能进行实验分析.

4.1 实例分析

TA 根据用户旅行计划提供机票预订、酒店预订、在线支付一站式服务,组合了机票预订(flight)、酒店预订(hotel)及在线支付(pay)这 3 个服务.当前,很多服务商都提供这 3 类服务,例如,携程 API 2.0 平台[30]以 Web 服务的方式提供基于 SOAP 协议的数据访问,当前提供酒店预订、机票预订、支付等服务.本文在参考携程 API 2.0 平台的基础上,设计实现旅行代理 BPEL 程序,模拟旅行代理服务,用户成功完成预订的交互情景如图 3 所示.

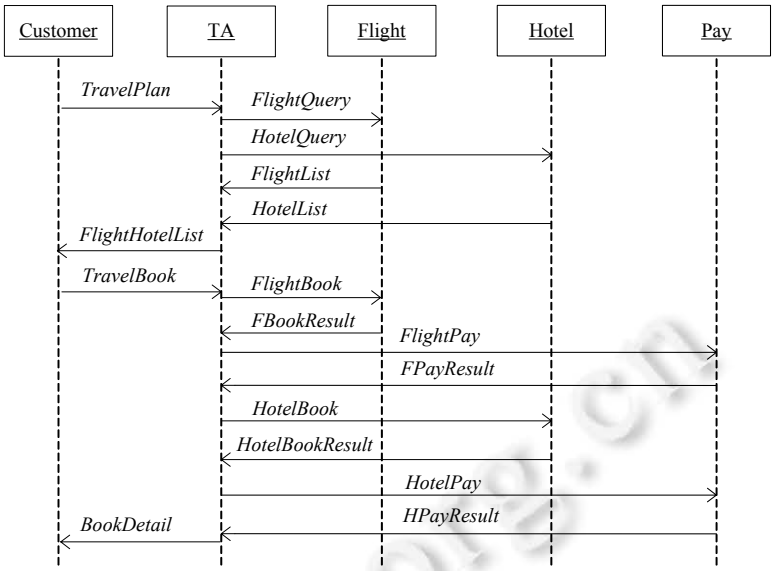


Fig.3 Interaction of successful reservation in TA

图 3 旅行代理成功完成预订的交互情况

用户首先向 TA 发送消息 TravelPlan 提交自己的旅行计划,TravelPlan 中包含了出发地、目的地、出发时间等内容,TA 根据计划分别向 Flight 和 Hotel 服务发送消息 FlightQuery 和 HotelQuery 查询航班和酒店信息. Flight 服务通过 FlightList 消息向 TA 返回满足条件的航班列表,Hotel 服务通过 HotelList 消息返回满足条件的酒店列表,TA 将航班列表和酒店列表通过 FlightHotelList 消息返回给用户.用户通过 TravelBook 消息向 TA 预订选中的机票和酒店,TA 分别向 Flight 和 Hotel 服务发送 FlightBook 和 HotelBook 消息进行预订,并得到对应的返回消息 FbookResult,HBookResult,若预订成功,则分别向 Pay 服务发送 FlightPay 和 HotelPay 消息进行支付. TA 最后将预订信息通过消息 BookDetail 返回给用户,最终完成一站式预订服务.

假设某用户 Bob 的隐私需求定义如图 4 所示.

1. 数据敏感度声明
$statement_1$=({*id_number*},0.6)
$statement_2$=({*name*},0.5)
$statement_3$=({*credit_card_info*},0.8)
$statement_4$=({*email*},0.4)
$statement_5$=({*phone*},0.5)
$statement_6$=({*name*,*phone*},0.6)
$statement_7$=({*id_number*,*name*},0.7)
$statement_8$=({*id_number*,*name*,*phone*},0.8)

2. 期限敏感度函数(*fRetention*,单位:天)

期限	期限敏感度系数
[0,0]	0
(0,10]	0.5
(10,100]	0.7
(100,+∞)	0.9

3. 目的敏感度函数(*fPurpose*)

目的	目的敏感度系数
book_hotel	0.8
book_flight	0.6
pay	0.8
research	0.5
telmarketing	0.7
individual_analysis	0.8
individual_decision	0.8

...

4. 敏感度-信誉度函数(*fSense_Repute*)

fSense_Repute(*s*)=*s*

Fig.4 Example of user Bob's privacy requirement

图 4 用户 Bob 的隐私需求示例

其中,敏感度-信誉度函数为 *fSense_Repute*(*s*)=*s*,表示对成员服务的隐私数据使用行为而言,服务的信誉度应至少等于隐私数据使用综合敏感度,使用行为才被允许.

为支持用户隐私需求的可满足性验证,服务组合 TA 的提供者需要构建隐私敏感的服务组合模型.为便于

说明问题,假设所有服务调用均采用异步调用,且略掉组合中的 assign 等活动,针对 TA 的 BPEL 流程构建的 POWFN 模型如图 5 所示.

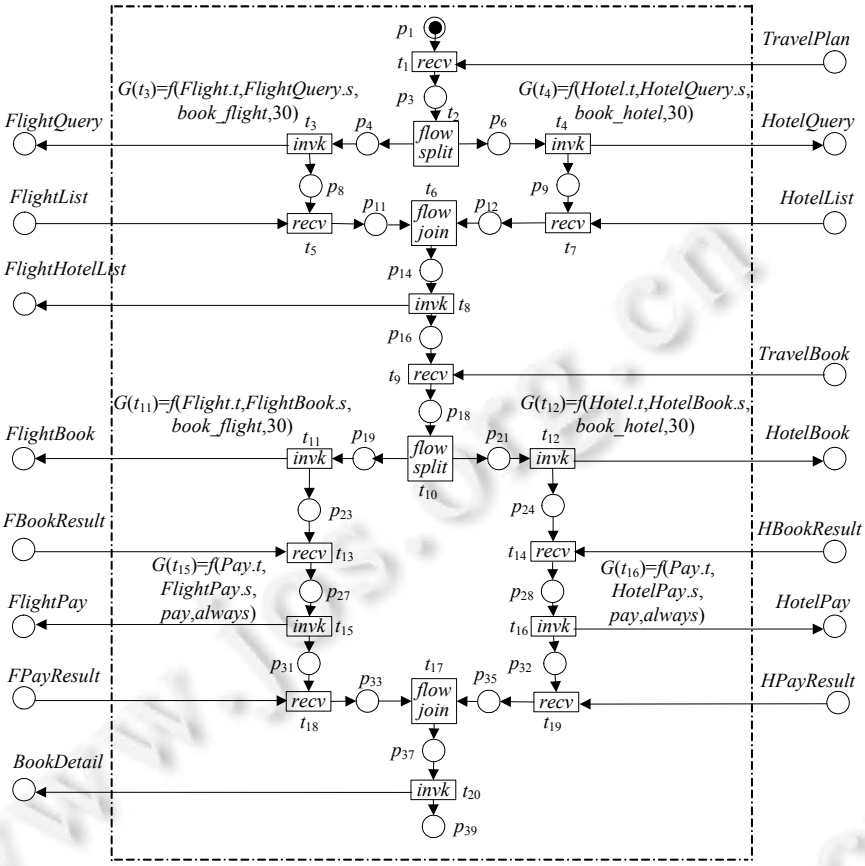


Fig.5 POWFN model of TA

图 5 TA 的 POWFN 模型

隐私分析人员需根据组合的业务逻辑构建完全隐私数据项依赖图,TA 的完全数据项依赖图如图 6(a)所示.

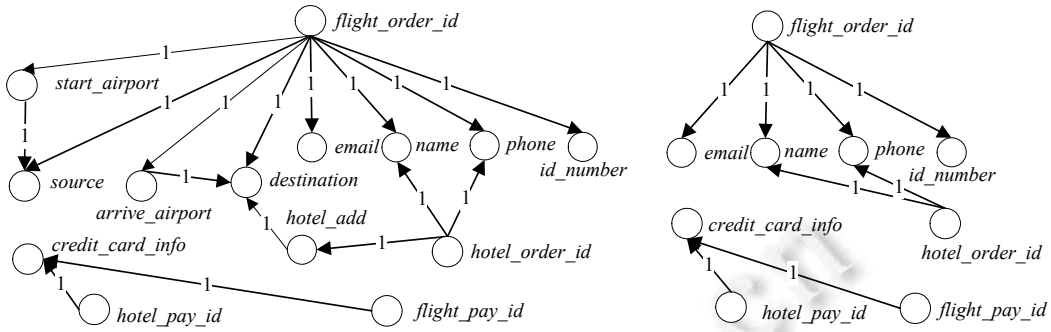
TA 的隐私分析人员需要考虑尽可能多的用户的隐私需求,因此在图 6(a)中将出发城市(*source*)和目的城市(*destination*)也列为直接隐私数据项.组合执行过程中产生的间接隐私数据项出发机场(*start_airport*)依赖于出发城市,达到机场(*arrive_airport*)和酒店地址(*hotel_add*)依赖于目的城市.组合执行产生的订单号依赖于用户预订时提交的直接隐私数据,例如,机票订单号(*flight_order_id*)是机票订单的标识,依赖于 *id_number*,*name* 等直接隐私数据项.

用户 Bob 隐私需求的直接隐私数据项为 *id_number*,*name*,*credit_card_info*,*phone*,*email*,因此,验证 Bob 隐私需求时对图 6(a)进行约简后得到图 6(b).根据 Bob 的隐私需求、TA 的 POWFN 模型及约简得到的用户隐私数据项依赖图,TA 中涉及 Bob 隐私数据的消息见表 1.

假设服务 *Hotel*,*Flight*,*Pay* 使用隐私数据的目的分别为 *book_hotel*,*book_flight*,*pay*,对应隐私保护信誉度分别为 0.6,0.7,0.8,*Hotel*,*Flight* 要求保留期限为 30 天,*Pay* 要求永久保存.根据用户隐私需求验证算法,利用原型工具进行验证,验证结果如图 7 所示.

以机票支付为例,TA 通过 *FlightPay* 消息完成机票支付,消息中包含直接隐私数据项 *credit_card_info* 与间接隐私数据项 *flight_order_id*,隐私数据使用综合敏感度为 0.8×(1+0.9+0.8)/3=0.72.根据用户需求中定义的敏感

度-信誉度函数,要求成员服务的信誉度应至少为 0.72.由于服务 *Pay* 的信誉度为 0.8,故机票支付操作满足用户隐私需求.



(a) 完全隐私数据项依赖图 (b) 用户Bob的隐私数据项依赖图

Fig.6 Privacy data item dependency graph of TA and user Bob

图 6 TA 及用户 Bob 的隐私数据项依赖图

Table 1 Messages involving Bob's private data in TA

表 1 TA 中涉及 Bob 隐私数据的消息

消息	发送者	接收者	直接隐私数据项	间接隐私数据项
FlightBook	TA	*Flight*	*name*, *phone*, *email*, *id_number*	–
HotelBook	TA	*Hotel*	*name*, *phone*	–
FlightPay	TA	*Pay*	*credit_card_info*	*flight_order_id*
HotelPay	TA	*Pay*	*credit_card_info*	*hotel_order_id*

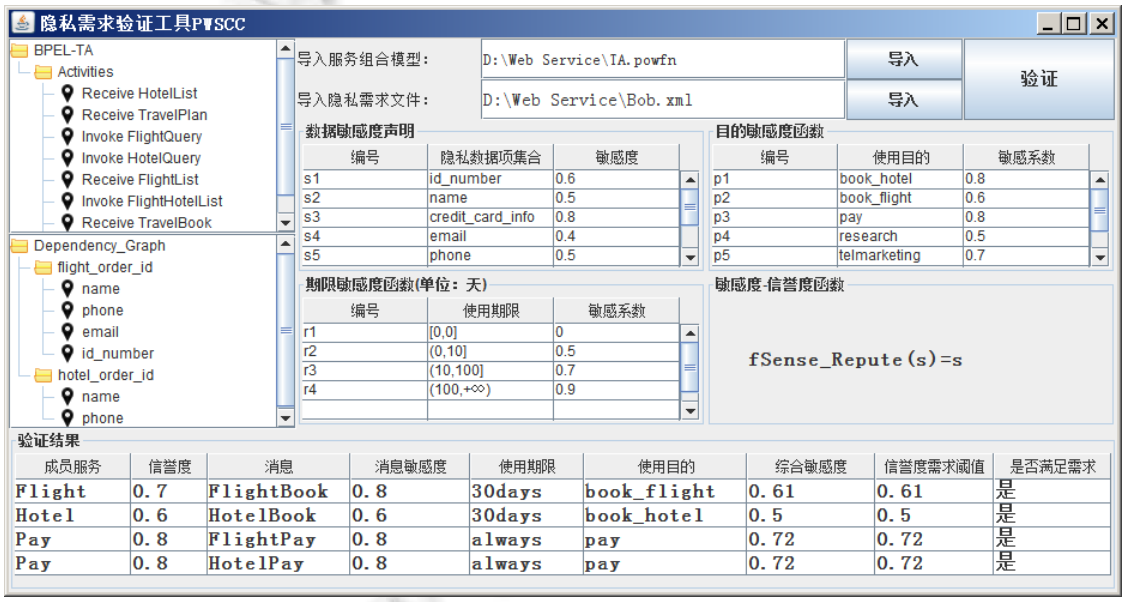


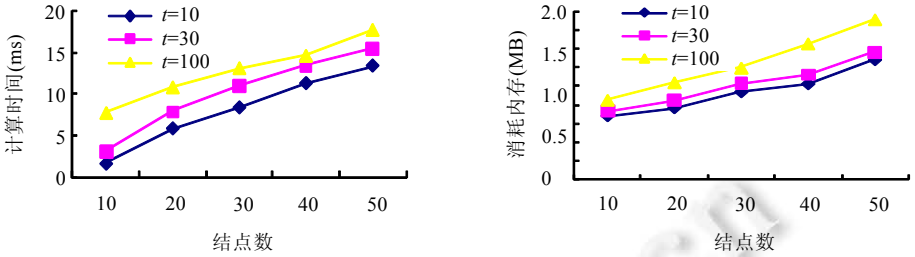
Fig.7 Verification result of user Bob's privacy requirement

图 7 用户 Bob 的隐私需求验证结果

4.2 实 验

通过仿真实验,对隐私需求验证算法的性能进行评估.实验环境为:Intel Core i3 双核 2.4GHz,4G 内存,32 位

Windows 7 系统,编程环境为:Eclipse 4.4.2+JDK1.7.设完全隐私数据项依赖图结点数为 n,取直接隐私数据项结点和间接隐私数据项结点数分别为 $n/2$,随机生成 $2n$ 条依赖边及 $2n$ 条隐私数据敏感度声明,每个消息随机设定 $n/2$ 个隐私数据项,需要验证的变迁数 t 分别取 10,30,100,评估 n 取不同值时算法的性能,实验结果如图 8 所示.



(a) 计算时间随结点数和变迁数的变化 (b) 消耗内存随结点数和变迁数的变化

Fig.8 Performance of algorithm about privacy requirement verification

图 8 隐私需求验证算法的性能

隐私需求验证算法的最坏时间复杂度为 $O(tn^3+n^4)$,主要计算量体现在间接隐私结点敏感度计算及变迁守护函数的验证方面,即算法的第(2)步和第(3)步.由于在实际应用中,隐私数据项依赖图不可能是完全图,且不可能所有消息中包含全部隐私数据项,故在仿真实验中,依赖图边的数量取值为 $2n$,消息包含隐私数据项数取值为 $n/2$.图 8(a)中的实验结果表明:算法在实验设置条件下,执行时间会随着 n 的增大而显著变长.例如,当 n=10,t=10 时,执行时间为 1.7ms;当 n=50,t=10 时,执行时间为 13.4ms.同时,在 n 确定的情况下,t 值的增大对执行时间的影响相对较小.由于实际应用中隐私数据项规模和需要验证的变迁数均较小,因此实验结果表明,隐私需求的验证对用户与服务组合的交互时间影响非常小.

算法的内存消耗如图 8(b)所示,当 n=10,t=10 时,算法内存消耗为 0.68MB;当 n=50,t=100 时,算法内存消耗为 1.71MB.表明随着 n 和 t 的增大,算法消耗内存并不会急剧增长.这是因为依赖图中边的数量、隐私数据敏感度声明的条数、消息中包含隐私数据项的数量均与 n 是线性关系,而 t 的增大仅会增加 POWFN 对内存的需求.

5 结束语

随着服务计算的普及,越来越多的应用通过组合现有服务实现业务功能.但服务组合的业务逻辑往往对用户而言是透明的,且用户与成员服务之间缺乏隐私信息使用的相关协议.如何保证组合按照用户需求暴露隐私信息,成为研究热点.本文在隐私数据及使用情境的敏感度与服务隐私保护信誉度之间建立映射关系,以此构建用户隐私需求.为支持服务组合对用户隐私需求的验证,且能够将需求中对直接隐私数据的使用约束延伸到间接隐私数据,通过隐私数据项依赖图和隐私开放工作流网构建隐私敏感的服务组合模型,提出相应的验证算法,并实现原型工具.在基于本文方法构建隐私保护框架时,用户可以首先通过隐私代理向组合发送隐私需求,只有组合满足需求才进行业务功能的交互,从而有效防止用户隐私信息的非法直接暴露和间接暴露.当前,越来越多的研究关注网络环境中用户对隐私数据及使用情境敏感度的实证分析工作,且已有组织开展对服务的隐私保护认证工作,这为本文的方法提供了有力支撑.同时,本文提出的通过隐私数据及使用情境敏感度与服务隐私保护信誉度的映射关系暴露隐私数据的方法,对服务计算与云计算中的隐私保护问题研究具有一定的探索意义.

下一步的研究工作包括:在本文工作的基础上扩展协商及运行监控机制.当服务组合不满足用户隐私暴露需求时,对用户隐私需求与组合隐私策略进行协商.在组合执行过程中,通过监控机制监控组合是否按照用户需求使用隐私数据.

References:

[1] Papazoglou MP, Van Den Heuvel WJ. Service oriented architectures: Approaches, technologies and research issues. The Int'l Journal on Very Large Data Bases, 2007,16(3):389−415. [doi: 10.1007/s00778-007-0044-3]

[2] Ke CB. Research on privacy analysis and protection method of service composition in cloud computing [Ph.D. Thesis]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2014 (in Chinese with English abstract).

[3] Fan GS, Yu HQ, Chen LQ, Liu DM. Fault diagnosis and handling for service composition based on petri nets. Ruan Jian Xue Bao/ Journal of Software, 2010,21(2):231−247 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/3790.htm [doi: 10.3724/SP.J.1001.2010.03790]

[4] Meziane H, Benbernou S. A dynamic privacy model for Web services. Computer Standards & Interfaces, 2010,32(5):288−304. [doi: 10.1016/j.csi.2010.02.001]

[5] Ke CB, Huang ZQ, Tang M. Supporting negotiation mechanism privacy authority method in cloud computing. Knowledge-Based Systems, 2013,51:48−59. [doi: 10.1016/j.knosys.2013.07.001]

[6] Cranor L, Langheinrich M, Marchiori M, Presler-Marshall M, Reagle J. The platform for privacy preferences 1.0 (P3P1.0) specification. W3C Recommendation. 2002.

[7] Cranor L, Langheinrich M, Marchiori M. A P3P preference exchange language 1.0 (APPEL1.0). W3C Working Draft, 2002.

[8] Parducci B, Lockhart H. eXtensible Access Control Markup Language (XACML) Version 3.0. OASIS, 2013.

[9] Allison DS, EL Yamany HF, Capretz M. Meta model for privacy policies within SOA. In: Proc. of the 2009 Int'l Conf. on Software Engineering (ICSE) Workshop on Software Engineering for Secure Systems. New York: IEEE Press, 2009. 40−46. [doi: 10.1109/IWSESS.2009.5068457]

[10] Hewett R, Kijsanayothin P. On securing privacy in composite Web service transactions. In: Proc. of the 2009 Int'l Conf. for Internet Technology and Secured Transactions. New York: IEEE Press, 2009. 1−6. [doi: 10.1109/ICITST.2009.5402545]

[11] Li YH, Paik HY, Benatallah B. Formal consistency verification between BPEL process and privacy policy. In: Proc. of the 2006 Int'l Conf. on Privacy, Security and Trust (PST): Bridge the Gap Between PST Technologies and Business Services. New York: ACM Press, 2006. 1−10. [doi: 10.1145/1501434.1501466]

[12] Ma Z, Mangler J, Wagner C, Bleier T. Enhance data privacy in service compositions through a privacy proxy. In: Proc. of 2011 the 6th Int'l Conf. on Availability, Reliability and Security. New York: IEEE Press, 2011. 615−620. [doi: 10.1109/ARES.2011.94]

[13] Nakajima S. Model-Checking of safety and security aspects in Web service flows. In: Koch N, Fraternali P, Wirsing M, eds. Proc. of the Web Engineering. Berlin: Springer-Verlag, 2004. 488−501. [doi: 10.1007/978-3-540-27834-4_60]

[14] Hutter D, Volkamer M. Information flow control to secure dynamic Web service composition. In: Clark J, Paige RF, Polack F, Brooke PJ, eds. Proc. of the Security in Pervasive Computing. Berlin: Springer-Verlag, 2006. 196−210. [doi: 10.1007/11734666_15]

[15] Wei J, Singaravelu L, Pu C. A secure information flow architecture for Web service platforms. IEEE Trans. on Services Computing, 2008,1(2):75−87. [doi: 10.1109/TSC.2008.10]

[16] Demongeot T, Totel E, Traon YL. Preventing data leakage in service orchestration. In: Proc. of the 2011 Int'l Conf. on Information Assurance and Security. New York: IEEE Press, 2011. 122−127. [doi: 10.1109/ISIAS.2011.6122806]

[17] Wang H, Qin KF. Research on internet users' personal information sensitivity. Journal of Intelligence, 2013,31(12):171−175 (in Chinese with English abstract).

[18] Hayes CM, Kesan JP, Bashir M, Hoff K, Jeon G. Knowledge, behavior, and opinions regarding online privacy. In: Proc. of the 2014 Research Conf. on Communications, Information and Internet Policy. Washington: SSRN, 2014. 1−34.

[19] Zheng J, Huang Z, Hu J, Wei O, Liu L. Trust-Based privacy authorization model for Web service composition. In: Wu YW, ed. Proc. of the Software Engineering and Knowledge Engineering: Theory and Practice. Berlin: Springer-Verlag, 2012. 307−313. [doi: 10.1007/978-3-642-25349-2_41]

[20] Liu LY, Li Q, Zhu Y, Zhou H, Xiao FX, Huang ZQ. Specification and verification of privacy requirements in Web service compositions. Journal of PLA University of Science and Technology (Natural Science Edition), 2012,13(1):27−33 (in Chinese with English abstract). [doi: 10.3969/j.issn.1009-3443.2012.01.006]

[21] Xu W, Venkatakrishnan VN, Sekar R, Ramakrishnan IV. A framework for building privacy-conscious composite Web services. In: Proc. of the 2006 Int'l Conf. on Web Services. New York: IEEE Press, 2006. 655−662. [doi: 10.1109/ICWS.2006.4]

[22] Yee GOM. Estimating the privacy protection capability of a Web service provider. Int'l Journal of Web Services Research, 2009, 6(2):20−41. [doi: 10.4018/jwsr.2009092202]

[23] Weible RJ. Privacy and data: An empirical study of the influence of types of data and situational context upon privacy perceptions [Ph.D. Thesis]. Mississippi State: Mississippi State University, 1993.

[24] Li YH. A framework to enforce privacy in business processes [Ph.D. Thesis]. Sydney: University of New South Wales Sydney, 2008.

[25] Lohmann N, Massuthe P, Stahl C, Weinberg D. Analyzing interacting BPEL processes. In: Dustdar S, Fiadeiro JL, Sheth AP, eds. Proc. of the Business Process Management. Berlin: Springer-Verlag, 2006. 17−32. [doi: 10.1007/11841760_3]

[26] Lohmann N, Massuthe P, Stahl C, Weinberg D. Analyzing interacting WS-BPEL processes using flexible model generation. Data & Knowledge Engineering, 2008,64(1):38−54. [doi: 10.1016/j.datak.2007.06.006]

[27] Lohmann N. A feature-complete petri net semantics for WS-BPEL 2.0. In: Dumas M, Heckel R, eds. Proc. of the Web Services and Formal Methods. Berlin: Springer-Verlag, 2008. 77−91. [doi: 10.1007/978-3-540-79230-7_6]

[28] bpel2owfn. 2007. http://www.gnu.org/software/bpel2owfn/index.html

[29] fiona. 2006. http://www2.informatik.hu-berlin.de/top/tools4bpel/fiona

[30] ctrip open API platform. 2013 (in Chinese). http://u.ctrip.com/union/Default.aspx

附中文参考文献:

[2] 柯昌博.云服务组合隐私分析与保护方法研究[博士学位论文].南京:南京航空航天大学,2014.

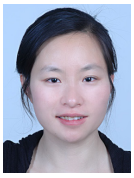
[3] 范贵生,虞慧群,陈丽琼,刘冬梅.基于 Petri 网的服务组合故障诊断与处理.软件学报,2010,21(2):231−247. http://www.jos.org.cn/1000-9825/3790.htm [doi: 10.3724/SP.J.1001.2010.03790]

[17] 王晗,秦克飞.网络用户个人信息的敏感度研究.情报杂志,2013,31(12):171−175.

[20] 刘林源,李清,祝义,周航,肖芳雄,黄志球.Web 服务组合中的隐私需求规约与验证.解放军理工大学学报(自然科学版),2012,13(1):27−33. [doi: 10.3969/j.issn.1009-3443.2012.01.006]

[30] 携程开放平台.2013. http://u.ctrip.com/union/Default.aspx

彭焕峰(1978－),男,山东临沂人,博士生,讲师,CCF 专业会员,主要研究领域为云计算与服务计算,隐私保护,软件形式化验证.



范大娟(1982－),女,博士,讲师,CCF 专业会员,主要研究领域为云计算与服务计算,软件形式化验证.

黄志球(1965－),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为云计算与服务计算,模型检测,嵌入式软件安全性,软件形式化验证.

章永龙(1976－),男,博士生,讲师,主要研究领域为博弈论,拍卖机制设计,云计算.