软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2018,29(6):1739−1755 [doi: 10.13328/j.cnki.jos.005276]

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

服务组合安全隐私信息流静态分析方法*

彭焕峰[1,2], 黄志球[1], 刘林源[3], 李 勇[1], 柯昌博[4]

[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](南京审计大学 电子商务系,江苏 南京 211815)

[4](南京邮电大学 计算机学院,江苏 南京 210023)

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

摘 要: 用户为使用服务组合提供的功能,需要提供必要的个人隐私数据.由于组合的业务逻辑对用户是透明的,且用户与成员服务之间缺乏隐私数据使用的相关协议,如何保证组合执行过程中不发生用户隐私信息的非法泄露,成为当前服务计算领域的研究热点之一.针对隐私保护特征,提出一种服务组合安全隐私信息流静态分析方法.首先,从服务信誉度、隐私数据使用目的及保留期限这 3 个维度提出一种面向服务组合的隐私信息流安全模型;其次,采用支持隐私信息流分析的隐私工作流网(privacy workflow net,简称 PWF-net)构建服务组合模型,并通过静态分析算法分析组合执行路径,检测组合的执行是否会发生用户隐私信息的非法泄露;最后,通过实例分析说明了方法的有效性,并对方法性能进行了实验分析.与现有的相关工作相比,针对隐私保护特征提出了隐私信息流安全模型,且分析方法考虑了隐私数据项聚合问题,从而能够更为有效地防止用户隐私信息非法泄露.

关键词: 服务组合;隐私保护;信息流安全;安全模型;静态分析;工作流网

中图法分类号: TP311

中文引用格式: 彭焕峰,黄志球,刘林源,李勇,柯昌博.服务组合安全隐私信息流静态分析方法.软件学报,2018,29(6):1739−1755. http://www.jos.org.cn/1000-9825/5276.htm

英文引用格式: Peng HF, Huang ZQ, Liu LY, Li Y, Ke CB. Static analysis method of secure privacy information flow for service composition. Ruan Jian Xue Bao/Journal of Software, 2018,29(6):1739−1755 (in Chinese). http://www.jos.org.cn/1000-9825/5276.htm

Static Analysis Method of Secure Privacy Information Flow for Service Composition

PENG Huan-Feng[1,2], HUANG Zhi-Qiu[1], LIU Lin-Yuan[3], LI Yong[1], KE Chang-Bo[4]

[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 211167, China)

[3](Department of E-Commerce, Nanjing Audit University, Nanjing 211815, China)

[4](College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

Abstract: Many service composition scenarios involve the sharing of user's privacy data. Due to the transparency of composition's business logic and lack of privacy protocol between user and member service, how to prevent the leakage of user privacy information has

* 基金项目: 国家自然科学基金(61772270, 61602262, 61562087); 国家高技术研究发展计划(863)(2015AA015303); 江苏省自然科学基金(BK20150865, BK20130735); 江苏省高校自然科学基金(15KJD520001, 13KJB520011)

Foundation item: National Natural Science Foundation of China (61772270, 61602262, 61562087); National High-Tech R&D Program of China (863) (2015AA015303); Natural Science Foundation of Jiangsu Province, China (BK20150865, BK20130735), Jiangsu University Natural Science Foundation (15KJD520001, 13KJB520011)

收稿时间: 2016-10-09; 修改时间: 2016-12-08; 采用时间: 2017-02-15; jos 在线出版时间: 2017-03-31

CNKI 网络优先出版: 2017-03-31 21:54:53, http://kns.cnki.net/kcms/detail/11.2560.TP.20170331.2154.009.html

become a hot research topic in the field of service-oriented computing. A static analysis method of secure privacy information flow for service composition is proposed in this article according to the characteristics of privacy protection. Firstly, a security model is developed to formalize the security policy of privacy information flow on three aspects: service reputation, retention and purpose. Then, the composition is modeled with privacy workflow net, which gives support to the analysis of privacy information flow, and the detection of privacy information leakage is performed by analyzing execution paths of composition. Finally, a case study is included to demonstrate the effectiveness of the proposed method, and the performance experiment is also presented. Compared with the existing relevant works, the security model proposed reflects the characteristics of privacy protection, and the analysis method is able to deal with issues caused by the aggregation of privacy data items. Therefore, the application of this method can prevent the information leakage more efficiently.

Key words: service composition; privacy protection; information flow security; security model; static analysis; workflow net

服务计算是一种新型的分布式计算模式,以服务作为基本元素,支持分布式应用的快速、低成本的组合式开发,受到国内外学术界和工业界的广泛关注.由于 Web 服务开放、动态和自治的特点,隐私信息一旦被收集,用户难以控制服务如何使用和暴露这些信息.特别是在 Web 服务组合应用场景中,由于组合的业务逻辑对用户是透明的,且用户与成员服务之间缺乏相应的隐私信息使用协议,因此难以保证组合执行过程中按照用户意愿使用和暴露用户隐私信息[1].随着隐私侵犯案例的增加,隐私保护问题受到用户越来越多的关注.如何保证组合执行过程中不发生用户隐私信息的非法泄露,成为当前服务计算领域的研究热点之一.

随着社会与科学技术的发展,人们对隐私问题的认识也逐步清晰.1890 年,Warren 与 Brandeis 将隐私定义为“个人独处的权利”[2].1967 年,Westin 将隐私定义为“个体、机构或团体组织决定将它们自身的什么信息、在什么时候、以何种方式与他人进行通信”[3].1997 年,Goldberg 等人强调了个体对其隐私信息的控制能力,将隐私定义为“个体控制其隐私信息被他人收集、保留及暴露的能力”[4].在服务计算领域,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;后者则主要关注隐私数据使用行为的约束与分析,包括用户隐私需求规约方法、服务对隐私需求的可满足性检测、服务隐私策略与用户需求协商及演化、隐私暴露风险分析等研究内容[5].

用户在其隐私需求中定义了相应的访问策略来授权服务对隐私数据的使用行为.面向使用行为的隐私保护要求服务对用户隐私数据的访问都是授权的,即,组合对隐私数据的使用行为应满足用户隐私需求.传统的访问控制机制(如 ACL,RBAC 等)只能控制直接访问的合法性,因未对多个直接访问行为所造成的间接信息传递进行分析和安全性检测,无法保证信息传递过程中的安全.由于信息流控制(information flow control,简称 IFC)技术根据信息流安全策略保证信息的合法流动,越来越多的研究者将 IFC 技术应用于面向使用行为的隐私保护,但当前研究仍然面临如下两个关键问题.

(1) 如何建立面向隐私保护特征的隐私信息流安全模型

用户隐私需求中,对其隐私数据的释放约束条件是多维度的.例如,文献[6]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[7],从隐私数据的使用者、使用目的、保留期限等方面定义了用户隐私需求的元模型,服务只有在满足需求中声明的使用目的和保留期限的前期下,才能够使用相应的隐私数据.因此,需要从隐私数据的使用目的、保留期限等多个维度建立隐私信息流安全模型.

(2) 安全策略实施时,应考虑隐私数据项聚合问题

用户同时释放多个隐私数据项时,往往比释放单个隐私数据项更担心隐私泄露.为此,用户在隐私需求中可能会对隐私数据项组合定义更为严格的释放约束条件.当多条信息流流向同一实体引发隐私数据项聚合时,虽然单条信息流满足安全策略,但多条信息流发生后,可能并不满足用户需求中对隐私数据项组合定义的约束条件,从而发生隐私信息的非法泄露.

虽然已有研究者采用 IFC 技术防止用户隐私信息的非法泄露,但当前研究仅从隐私数据的机密等级单一维度建立信息流的安全模型,并不足以满足用户的隐私保护需求.此外,在安全策略实施时未考虑隐私数据项聚合带来的隐私信息泄露问题.针对上述问题,本文提出一种服务组合安全隐私信息流静态分析方法:首先,针对隐私保护特征,提出隐私信息流安全模型;然后,采用支持隐私信息流分析的隐私工作流网构建服务组合模型,

并通过静态分析算法分析组合执行路径,检测组合的执行是否发生用户隐私信息非法泄露;最后,通过实例分析说明了方法的有效性,并对方法的性能进行了实验分析.

本文主要创新点如下:

(1) 针对隐私保护的特征,从服务信誉度、隐私数据使用目的及保留期限这 3 个维度提出一种面向服务组合的隐私信息流安全模型;

(2) 为对组合中的隐私信息流进行分析,提出了支持隐私信息流分析的服务组合建模方法;

(3) 本文的静态分析方法考虑了隐私数据项聚合带来的隐私信息泄露问题.

本文第 1 节介绍相关研究工作.第 2 节介绍隐私信息流安全模型.第 3 节介绍支持隐私信息流静态分析的服务组合建模方法.第 4 节介绍静态分析方法.第 5 节介绍实例研究,并对方法性能进行实验分析.第 6 节总结全文并展望未来的工作.

1 相关工作分析

不同用户有自身的隐私保护需求,面向使用行为的隐私保护本质上是保证用户需求在服务运行期间不被违背的一种机制.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,从不同维度提出相应的用户隐私需求规约方法.例如:文献[6]根据经济合作与发展组织提出的个人隐私数据保护原则[7],从隐私数据的使用者、使用目的、保留期限等维度定义了用户隐私需求的元模型;文献[8]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,通过比较策略矩阵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.下面从隐私需求维度、需求实施机制、是否考虑隐私数据项聚合问题、是否支持隐私信息间接泄露检测等方面对相关工作进行比较分析,具体见表 1.

Table 1 Comparison of related works

表 1 相关工作之间的比较

参考文献	实施机制	隐私需求维度	是否支持隐私信息间接泄露检测	是否考虑隐私数据项聚合问题
Ref.[8]	访问控制	服务信誉度	否	否
Ref.[9]	访问控制	使用者、使用目的	否	否
Ref.[10]	访问控制	使用目的、保留期限	否	否
Ref.[11]	访问控制	数据敏感度-服务信誉度	是	否
Ref.[13,15,16,18]	信息流控制	数据机密等级-服务安全等级	是	否

用户向组合提供的隐私数据称为直接隐私数据.组合执行过程中会产生新的数据,且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.采用传统的访问控制机制可以保证服务对用户需求中直接隐私数据的授权访问,但无法保证间接隐私数据的合法访问,即,无法支持隐私信息的间接泄露检测.文献[8]采用隐私策略矩阵规约服务的隐私权限,并利用带隐私语义的接口自动机对服务的隐私数据使用行为进行建模,形式化地检验了服务组合行为是否满足隐私授权约束.文献[9]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.文献[10]提出了 WSC-PRBAC 模型,通过全局角色实现对隐私数据的访问控制.为保证间接隐私数据的合法访问,在团队前期工作[11]中,采用敏感度-信誉度函数规约用户需求,并通过隐私数据项依赖图描述组合中间接隐私数据项与直接隐私数据项之间的依赖关系,从而将需求中对直接隐私数据的访问约束扩展到间接隐私数据.该方法本质上仍采用传统的访问控制机制,且对间接隐私信息的保护受限于隐私分析人员定义的隐私数据项依赖图.

信息流控制技术可以有效防止隐私信息的间接泄露,实施隐私信息流安全控制机制将极大提高用户使用服务的信心[12].为此,越来越多的研究者采用信息流控制技术防止组合执行过程中用户隐私信息的非法泄露.机密性、完整性及可用性是信息安全的 3 个基本属性,在面向使用行为的隐私保护研究中,主要考虑用户隐私信息的机密性.为防止组合执行过程中隐私信息非法泄露,文献[13]采用 Denning 的信息流格模型[14]规约安全策

略,将组合 BPEL 转换为 Promela 程序,采用工具 Spin 验证组合是否存在信息的非法泄露.文献[15]关注动态服务组合时用户隐私信息流的安全,提出一种基于安全类型的动态服务组合方法,保证动态服务组合过程中信息流的安全性.为防止工作流并发执行的实例之间存在信息非法泄露,文献[16]通过对变迁标注安全等级的方法构建支持信息流分析的 Petri 网模型,在此基础上提出基于网结构的隐蔽信息流分析方法.然而,此方法并不适用于服务组合场景中面向使用行为的隐私保护.为验证工作流是否满足 Bell-LaPadula 模型[17]提出的"不向下写,不向上读"安全信息流原则,文献[18]通过 Petri 网对工作流进行建模,在可达标识图中为数据库标注所有可能的安全等级,从而枚举出所有可能的状态,在此基础上分析执行路径中信息流的安全性.但随着工作流规模和安全等级数量的增大,不可避免地会带来空间爆炸问题.

当前研究在经典信息流安全模型的基础上构建安全策略,并通过静态分析或者运行时控制的方法防止服务组合中敏感信息的非法泄露.然而,现有研究并没有充分考虑隐私保护的特征,主要存在两方面不足:一方面,并未针对隐私保护的特征提出隐私信息流安全模型;另一方面,安全策略实施时并未考虑隐私数据项聚合问题.

2 隐私信息流安全模型

信息流是指信息间影响交互的一种关系.Denning 于 1976 年在文献[14]中提出信息流的格模型,通过格模型形式化规约了信息流安全策略.为实施用户隐私信息流安全控制机制,首先需要针对隐私保护特征建立隐私信息流安全模型.本节基于信息流格模型,从服务信誉度、隐私数据的使用目的及保留期限这 3 个维度提出隐私信息流安全模型,通过该模型形式化规约服务组合隐私信息流的安全策略.

2.1 敏感度/信誉度格模型

用户对不同隐私数据有不同的敏感度,且在选择功能相同的服务时更倾向于选择信誉度高的服务.因此,在隐私数据的敏感度与服务的信誉度之间建立约束关系,约束用户隐私信息在服务组合中的流动,该策略通过敏感度/信誉度格来表示:

定义 2.1(敏感度/信誉度格). $C=\langle RS,\xrightarrow{c},\Delta\rangle$,其中,

(1) RS 是安全等级的有限集合(即敏感度/信誉度等级);

(2) $\xrightarrow{c}$ 是定义在 RS 上的流关系;

(3) 安全等级运算符Δ满足结合律和交换律,也称为 RS 上的最小上界运算符.

为便于讨论,假设 $RS=\{N,L,M,H,TH\}$.当 RS 表示数据的敏感度集合时,L,M,H,TH 分别表示低、中等、高、最高敏感度等级,N 表示无敏感度.当 RS 表示服务信誉度时,L,M,H,TH 分别表示低、中等、高、最高信誉度等级,N 表示无信誉度.$\forall rs_i,rs_j\in RS, rs_i\xrightarrow{c}rs_j$ 当且仅当类型 rs_i 的信息允许流入类型 rs_j 的实体,例如 $L\xrightarrow{c}M$,$H\xrightarrow{c}TH$.$\langle RS,\xrightarrow{c}\rangle$ 是一个具有下界 N 和上界 TH 的偏序集,且 $\forall rs\in RS, N\xrightarrow{c}rs\wedge rs\xrightarrow{c}TH$.

通过定义易知,C 是一个有限格,格 C 可以用图 1 表示.RS 上存在最大下界运算符Λ,N 为 RS 的最大下界.格 C 从服务信誉度的角度规约了隐私信息的流动策略,即,要求服务只能"向下读,向上写"隐私数据.例如,若服务的信誉度等级为 H,则可读取敏感度等级为 H,M,L 或 N 的数据,只能写敏感度等级为 H 或 TH 的数据.

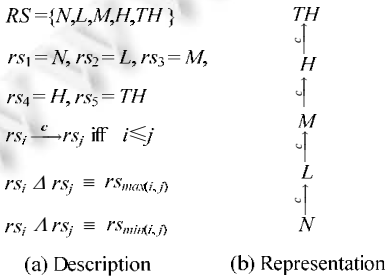


Fig.1 Reputation/Sensitivity lattice

图 1 敏感度/信誉度格

2.2 保留期限格模型

用户对其隐私数据的保留期限有着自身需求,且服务在使用用户隐私数据时也会声明保留期限.显然,只有当服务声明的保留期限等于或短于用户规定的期限时,对应的隐私信息才可以流向服务.因此,需要从保留期限的角度约束隐私信息的流动,该策略可以用保留期限格来表示:

定义 2.2(保留期限格). $R=\langle RT,\xrightarrow{r},\Theta\rangle$,其中,

(1) RT 是安全等级的有限集合 (即保留期限);

(2) $\xrightarrow{r}$ 是定义在 RT 上的流关系;

(3) 安全等级运算符Θ满足结合律和交换律,也称为 RT 上的最小上界运算符.

为便于讨论,假设 RT={*top-retention*,9*days*,5*days*,1*day*,0*day*},其中:0*day* 表示在线交互一旦完成,隐私数据不再保存;*top-retention* 表示永久保存隐私数据.根据具体的安全需求,RT 中元素的数量可以调整,且时间单位可以是小时、月等.

$\forall rt_i,rt_j\in RT, rt_i\xrightarrow{r}rt_j$ 当且仅当类型 rt_i 的信息允许流入类型 rt_j 的实体,例如 *top-retention* $\xrightarrow{r}$ 5*days*, 5*days* $\xrightarrow{r}$ 1*day*.显然,$\langle RT,\xrightarrow{r}\rangle$ 是一个具有下界 *top-retention* 和上界 0*day* 的偏序集,且:

$$\forall rt\in RT, \textit{top-retention}\xrightarrow{r}rt\wedge rt\xrightarrow{r}0day.$$

通过定义易知,R 是一个有限格,格 R 可以用图 2 表示.RT 上存在最大下界运算符⊙,*top-retention* 为 RT 的最大下界.服务和隐私数据均需绑定 RT 中的安全等级,格 R 从保留期限的角度规约了隐私信息的流动策略,即,要求服务只能"向下读,向上写"隐私数据.例如,若服务声明保留期限为 5*days*,则可以使用保留期限为 *top-retention* 或 9*days* 的数据,只能写保留期限为 5*days*,1*day* 或 0*day* 的数据.

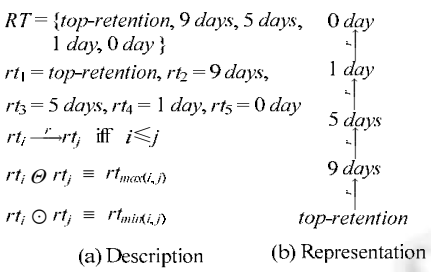


Fig.2 Retention lattice

图 2 保留期限格

2.3 使用目的格模型

用户规定其隐私数据可允许使用目的的有限集合记为 UP,服务声明使用隐私数据目的的有限集合记为 SP,服务使用隐私数据的目的必须匹配用户的隐私需求,即满足 $SP\subseteq UP$.例如,用户规定电子邮箱可以用于当前任务和后续商品推送,则 UP={*current*,*contact*},服务声明的使用目的集合 SP={*current*},满足 $SP\subseteq UP$,从而服务可以使用电子邮箱这一隐私数据.因此,需要从使用目的的角度约束隐私信息的流动,该策略可以用使用目的格来表示:

定义 2.3(使用目的格). $P=\langle PC,\xrightarrow{p},\bullet\rangle$,其中,

(1) PC 是 PS 的幂集,PS 是使用目的的有限集合;

(2) $\xrightarrow{p}$ 是定义在 PC 上的流关系;

(3) 安全等级运算符•满足结合律和交换律,也称为 PC 上的最小上界运算符,$\forall pc_i,pc_j\in PC, pc_i\bullet pc_j\equiv pc_i\cap pc_j$.

为便于讨论,假设 PS={*current*,*admin*,*develop*,*tailoring*,*pseudo-analysis*,*pseudo-decision*,*contact*,*individual-analysis*,*individual-decision*,*telemarketing*,*historical*,*other-purpose*},PS 中的元素来源于 P3P[19]中定义的 12 种使用目的. $\forall pc_i,pc_j\in PC, pc_j\subseteq pc_i\Leftrightarrow pc_i\xrightarrow{p}pc_j$,且 $pc_i\xrightarrow{p}pc_j$ 当且仅当类型 pc_i 的信息允许流入类型 pc_j 的实

体.∅是 *PC* 的上界,表示隐私数据不能以任何目的使用;*PS* 是 *PC* 的下界,表示隐私数据能够以任何目的使用,且 $\forall pc \in PC, PS \xrightarrow{p} pc \wedge pc \xrightarrow{p} \varnothing$.

通过定义易知,*P* 是一个有限格.*PC* 上存在最大下界运算符∘,且 $pc_i \circ pc_j \equiv pc_i \cup pc_j$.为简化表述,设 *PS*={*current*, *contact*,*admin*},此时格 *P* 可以用图 3 表示.服务和隐私数据均需绑定 *PC* 中的安全等级,格 *P* 从使用目的的角度规约了隐私信息的流动策略,即,要求服务只能"向下读,向上写"隐私数据.例如,若服务声明使用目的为{*current*, *contact*},则只能使用目的为{*current*,*contact*}或{*current*,*admin*,*contact*}的隐私数据,只能写使用目的为{*current*, *contact*},{*contact*},{*current*}或∅的数据.

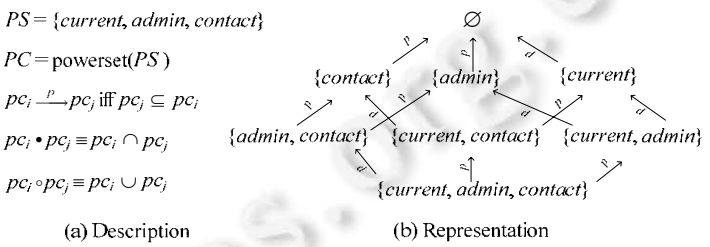


Fig.3 Purpose lattice

图 3 使用目的格

2.4 服务组合隐私信息流安全模型

有限格 *C*,*R*,*P* 分别从服务信誉度、保留期限、使用目的这 3 个维度形式化规约了隐私信息流的安全策略.为从这 3 个维度同时保证用户隐私信息的安全,提出服务组合 BPEL 的隐私信息流安全模型:

定义 2.4(服务组合隐私信息流安全模型). $PFM=\langle V,S,SC,\oplus,\rightarrow\rangle$,其中,

(1) *V* 是变量 *v* 的有限集合,变量 *v* 是隐私数据项的有限集合;

(2) *S* 是成员服务的有限集合;

(3) $SC=RS\times RT\times PC$,是安全等级的有限集合;

(4) 安全等级运算符⊕满足结合律和交换律,也称为 *SC* 上的最小上界运算符;

(5) →是定义在 *SC* 上的流关系.

其中,$\forall sc_i=(rs_i,rt_i,pc_i), sc_j=(rs_j,rt_j,pc_j)\in SC, sc_i\oplus sc_j=(rs_i \Delta rs_j, rt_i \Theta rt_j, pc_i \bullet pc_j)$.$sc_i \rightarrow sc_j$ 当且仅当类型 sc_i 的信息允许流入类型 sc_j 的实体,且 $sc_i \rightarrow sc_j \Leftrightarrow rs_i \xrightarrow{c} rs_j \wedge rt_i \xrightarrow{r} rt_j \wedge pc_i \xrightarrow{p} pc_j$.

因 *C*,*R* 和 *P* 均为有限格,易知$\langle SC,\rightarrow,\oplus\rangle$构成一个有限格.此外,*SC* 上存在最大下界运算符⊗,且 L_{sc}=(*N*,*top-retention*,*PS*)是 *SC* 的最大下界.隐私信息流是安全的,当且仅当 BPEL 程序的活动执行序列不产生违反流关系→的信息流.

3 支持隐私信息流静态分析的服务组合建模

3.1 隐私工作流网

安全隐私信息流的静态分析方法需要验证路径中的每个活动是否满足安全策略,为此,需要对组合的控制流和数据流进行建模.为对工作流进行建模分析,Aalst 以传统的 Petri 网为基础,提出了工作流网(workflow net,简称 WF-net)[20].WF-net 中包含起始库所和终止库所两个特殊库所,且所有节点都属于从起始库所到终止库所的路径上.袁崇义教授对 Petri 网进行扩展,提出了 C_net[21],通过增加数据库所和读写关系,以扩展 Petri 网对数据处理的建模能力.因此,在 WF-net 与 C_net 模型的基础上,提出了支持安全隐私信息流静态分析的隐私工作流网模型:

定义 3.1(隐私工作流网). PWF-net $PN=(P_c,T,F,P_d,R,W,T_S,T_A)$,其中,

(1) (P_c,T,F)是 WF-net;

(2) (P_c,T,F,P_d,R,W)是 C_net;

(3) $P_c\cup P_d\cup T\neq\varnothing\wedge P_c\cap P_d=\varnothing\wedge P_c\cap T=\varnothing\wedge P_d\cap T=\varnothing$;

(4) $F\subseteq P_c\times T\cup T\times P_c$;

(5) $R,W\subseteq P_d\times T$;

(6) $dom(R\cup W)=P_d\wedge dom(F)\cup cod(F)=P_c\cup T\wedge cod(R\cup W)\subseteq T$;

(7) $T_S:T\rightarrow SUBJECT,SUBJECT$ 表示用户、组合及成员服务构成的有限集合;

(8) $T_A:T\rightarrow ACTION,ACTION=\{RECV,SND,ASGN,STRC\}$表示变迁对应的活动类型,前 3 个元素分别表示变迁对应接收消息、发送消息、变量赋值活动.STRC 表示用于建模控制流程结构的辅助变迁,包括 flow split,flow join,enter loop,leave loop,case 等子类型.

其中,P_c,T,F 分别为控制库所集、变迁集、流关系,它们构成一个工作流网;P_d 是数据库所集合;R 表示从 P_d 到 T 的读关系,W 表示写关系,两者图形符号均用—∘表示,区别是读关系小圈端指向变迁,写关系小圈端指向数据库所.对于二元关系 $r\subseteq D_1\times D_2$,其定义域 $dom(r)$和值域 $cod(r)$分别定义为:

- $dom(r)=\{d_1|\exists d_2\in D_2:(d_1,d_2)\in r\}$;
- $cod(r)=\{d_2|\exists d_1\in D_1:(d_1,d_2)\in r\}$.

此外,$r(t)=\{x|(x,t)\in R\}$称为 t 的读集,$w(t)=\{x|(x,t)\in W\}$称为 t 的写集.

定义 3.2(控制视图). 给定一个 PWF-net PN,其控制视图定义为 $PN_{cv}=(P_c,T|_{P_c},F)$,其中,$T|_{P_c}$ 是连接控制库所的变迁集合.

定义 3.3(数据视图). 给定一个 PWF-net PN,其数据视图定义为 $PN_{dv}=(P_d,T|_{P_d},R,W)$,其中,$T|_{P_d}$ 是连接数据库所的变迁集合.

根据图论和集合原理,显然有 $PN=PN_{cv}\cup PN_{dv}$[21].构建 BPEL 的 PWF-net 后,可以通过 PN_{cv} 分析控制流,从而得到执行路径集合,而 PN_{dv} 用于分析数据操作.

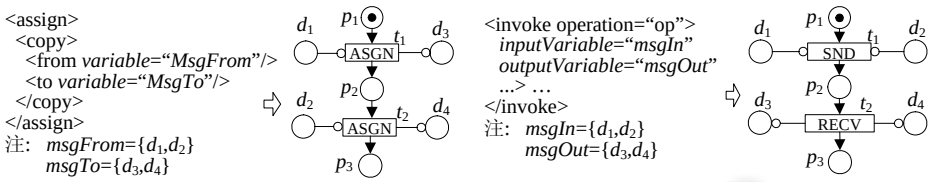
3.2 服务组合的PWF-net建模

BPEL 的活动分为基本活动和结构活动,其中,基本活动包括 assign,receive,reply,invoke 等,结构活动对一组基本活动的执行次序进行了约束,并且可以嵌套使用,如 sequence,flow,if,pick,while 等.BPEL 至 PWF-net 在控制流方面的建模方法同工具 BPEL2oWFN,该工具支持 BPEL 流程至工作流网的完整映射,具体细节可参考文献[22],本文仅在对 assign 活动的建模方法上稍有区别.为细粒度的实施安全策略,在控制流视图的基础上,需要针对隐私数据项构建数据视图.涉及数据操作的基本活动可以映射为 3 类变迁:接收(RECV)、发送(SND)、赋值(ASGN).assign 活动对应 ASGN 类型变迁,receive 活动对应 RECV 类型变迁,reply 与单向 invoke 活动均对应 SND 类型变迁,而双向 invoke 活动对应顺序执行的 SND 类型和 RECV 类型变迁.

下面以 assign 与双向 invoke 活动为例,说明基本活动的建模方法.为细粒度分析隐私信息流,需要根据更新数据项的数量将 assign 活动映射为多个变迁.以图 4(a)为例,assign 活动将变量 *msgFrom* 中数据项拷贝到变量 *msgTo* 中对应的数据项,所以我们映射为 t_1,t_2 这 2 个 ASGN 类型变迁,分别表示对 *msgTo* 中数据项 d_3,d_4 的更新操作.请求响应 invoke 活动建模方法如图 4(b)所示,该活动映射为 SND 类型变迁 t_1 和 RECV 类型变迁 t_2,其中,读取隐私数据项集合为 $r(t_1)=\{d_1,d_2\}$,写隐私数据项集合为 $w(t_2)=\{d_3,d_4\}$.

为分析组合的执行路径,通过增加辅助变迁(对应 STRC 类型)对结构活动进行建模.值得说明的是:本文并不关注流程结构引入的隐蔽通道,因此不需要针对 STRC 类型变迁建模数据操作.以图 5 说明 flow,if,while 结构活动的建模方法,其他结构活动的建模方法类似.为简化表述,图中省略对基本活动的数据操作建模.

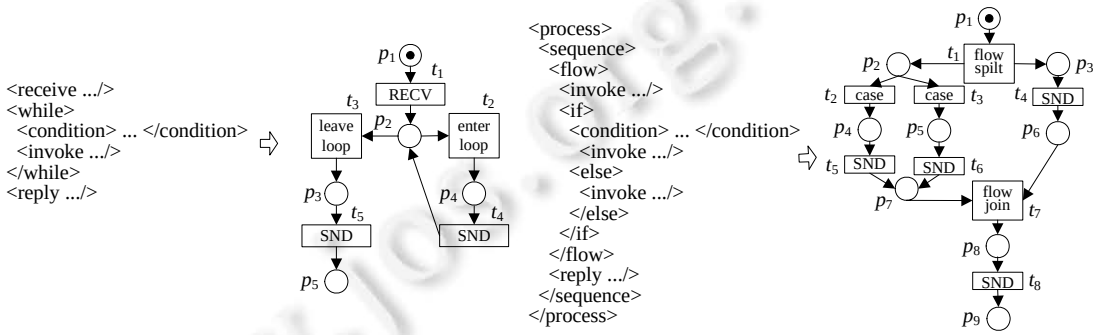
图 5(a)中,变迁 t_2,t_3 是辅助变迁,用于建模结构活动 while.当 t_1 触发后,t_4 对应的 invoke 活动可多次触发,也可能一次也不触发.图 5(b)中,对并发流程中嵌套了选择结构的例子进行建模,变迁 t_1,t_2,t_3,t_7 为辅助变迁,用于表示结构活动,t_4,t_5,t_6,t_8 对应基本活动.



(a) assign活动到PWF-net转换 (b) 请求响应invoke活动到PWF-net转换

Fig.4 Example of transformation from basic activities to PWF-net

图 4 基本活动转换至 PWF-net 示例



(a) while活动到PWF-net转换 (b) flow,if活动到PWF-net转换

Fig.5 Example of transformation from structured activities to PWF-net

图 5 结构活动转换至 PWF-net 示例

4 静态分析方法

静态分析方法的具体步骤如下.

(1) 构建 BPEL 的 PWF-net 模型,且该过程支持对 BPEL 的并发(包括并发活动之间的同步)、选择、循环等结构活动的建模;

(2) 创建 PWF-net 的可达标识图,获得组合的待验证路径集;

(3) 利用静态分析算法检测待验证路径集,以确定是否存在隐私信息非法泄露.

其中,步骤(1)与步骤(2)在部署服务组合前需要完成,用户使用组合前通过步骤(3)检测是否存在隐私信息非法泄露,并将检测结果反馈给用户.

4.1 独立路径集

通过可达标识图获取组合执行路径时,并发活动的存在会导致路径爆炸问题.以图 5(b)为例,其可达标识图如图 6 所示.由于 t_2,t_5 与 t_3,t_6 分别属于两个不同的分支,且每个分支与 t_4 可以并发执行,因此,图 6 中具有 6 条执行路径:$p_1=t_1,t_3,t_4,t_6,t_7,t_8$,$p_2=t_1,t_3,t_6,t_4,t_7,t_8$,$p_3=t_1,t_4,t_3,t_6,t_7,t_8$,$p_4=t_1,t_4,t_2,t_5,t_7,t_8$,$p_5=t_1,t_2,t_5,t_4,t_7,t_8$,$p_6=t_1,t_2,t_4,t_5,t_7,t_8$.

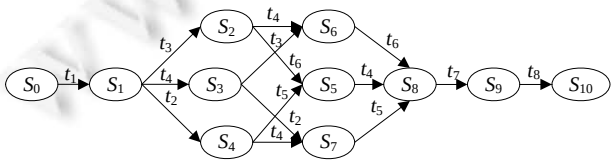


Fig.6 Example of reachability graph

图 6 可达标识图例子

记路径 p 中包含的变迁集合为 $p.T$,具有相同变迁集合的路径集为 PE,则有:$\forall p_i,p_j\in PE,p_i.T=p_j.T$.可达标识图路径集合 $P=PE_1\cup PE_2\cup\ldots\cup PE_n$,其中,$\forall i\neq j,PE_i\cap PE_j=\varnothing$.以图 6 为例,$PE_1=\{p_1,p_2,p_3\},PE_2=\{p_4,p_5,p_6\},P=PE_1\cup PE_2$.在任意两个并发执行的变迁不对应相同成员服务的前提下,PE 中的路径具有如下性质:

性质 1(路径隐私信息流安全验证等价性). PE 中的路径对隐私信息流安全性验证是等价的.

证明:在对路径进行隐私信息流安全性验证时,需要根据安全策略分析隐私数据之间的依赖关系,从而确定数据的安全等级,然后验证成员服务的隐私数据使用行为的安全性.由于服务对隐私数据的使用行为是确定的,故只需证明 PE 的不同路径中数据之间的依赖关系是一致的,即可证明 PE 中路径的隐私信息流安全验证等价性.设 p_i,p_j 为 PE 中的任意两条路径,包含并发变迁集合记为 CT,非并发变迁集合记为 NT.显然,NT 中的变迁执行顺序是一致的,不会导致两条路径中的数据依赖关系不一致.p_i,p_j 的区别在于 CT 中的变迁执行顺序不同.设 t_i,t_j 为 CT 中任意两个并发执行的变迁,且不对应相同的成员服务,对隐私数据的操作具有如下几种情况.

(1) t_i,t_j 读或写不同的隐私数据;

(2) t_i,t_j 均读隐私数据 d;

(3) t_i 写隐私数据 d,t_j 读隐私数据 d;

(4) t_i,t_j 均写隐私数据 d.

显然,情形(1)、情形(2)两种情况不会导致 p_i,p_j 中隐私数据依赖关系的不一致.因 t_i,t_j 的并发执行,需要通过在 t_i,t_j 之间建立同步关系避免情形(3)、情形(4)两种情况,则不会发生因变迁的并发引入多路径的问题.因此,p_i,p_j 对隐私信息流安全性验证而言是等价的. □

定义 4.1(独立路径). 为进行隐私信息流安全性验证,从具有相同变迁集合的路径集 PE 中选取的任意一条路径称为 PE 的独立路径.

定义 4.2(独立路径集). 独立路径集定义为 $PI=\{p_i|p_i\in PE_i,1\leqslant i\leqslant n\}$,其中,$p_i$ 为 PE_i 的独立路径,$P=PE_1\cup PE_2\cup\ldots\cup PE_n$,$P$ 为可达标识图中路径集合.

因此,在任意两个并发执行的变迁不对应相同成员服务的前提下,根据性质 1 可知,可以将独立路径集 PI 作为待验证路径集,不需验证组合的所有执行路径,从而有效减少了验证路径的数量.独立路径集获取算法较为简单,依次遍历从初始标识至终止标识的所有路径,如果当前路径所包含的变迁集合与已遍历路径包含的变迁集合不同,则将该路径作为独立路径加入独立路径集.

值得说明的是:根据路径遍历算法的不同,独立路径集也不同.

4.2 隐私数据项聚合问题

依据文献[6]提出的用户隐私需求元模型,我们从隐私数据、敏感度、使用目的、保留期限这 4 个方面定义用户隐私需求.

定义 4.3(隐私规则). 隐私规则定义为 $r=(ds,sc)$.

其中,ds 是隐私数据项的有限集合,且其安全等级 $\overline{ds}=sc,sc\in SC,sc$ 规定了 ds 的敏感度、保留期限及使用目的.例如,当 *name* 和 *phone* 组合使用时,用户定义其敏感度为 H,且只能用于当前操作和商品推送,服务可以永久保存该数据,则隐私规则可以定义为:$r=(\{name,phone\},(H,top\text{-}retention,\{current,contact\}))$.

一条隐私规则定义了某个隐私数据项集合的敏感度、使用目的、保留期限,即定义了相应隐私数据的安全等级.一般而言,用户对 ds 的使用约束越严格,为 ds 设置的安全等级越高.

定义 4.4(隐私需求). 用户隐私需求定义为 $PR=\{r_1,\ldots,r_n\}$,即隐私规则的有限集合.

成员服务需要声明隐私策略,下面给出成员服务的形式化定义:

定义 4.5(成员服务). 成员服务定义为 $s=(name,sc)$.

其中,*name* 唯一标识了成员服务,$sc=(rs,rt,pc),sc\in SC$,表示服务的隐私策略,rs 表示服务的信誉度,rt 和 pc 分别表示服务声明的隐私数据保留期限和使用目的.

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称为直接隐私数据项;在服务组合执行过程中,新产生且能够间接暴露用户隐私信息的数据项称为间接隐私数据项.直接隐私数据的安全等级来源于用户

需求,间接隐私数据项的安全等级根据信息流的安全策略决定.由于用户需求中规定了隐私数据项组合使用的安全等级,当多个信息流流向同一实体时,安全策略实施时需要考虑隐私数据项聚合问题,主要表现在两方面.

(1) BPEL 中的变量是临时存放数据的容器,由若干隐私数据项组成,变量安全等级的计算需要考虑用户隐私需求中对数据项组合的使用约束.例如,假设用户部分隐私需求为 r_1=({*email*},(*M*,*top-retention*,{*current*,*contact*})),r_2=({*name*},(*M*,1*day*,{*current*})),r_3=({*email*,*name*},(*H*,1*day*,{*current*})).计算变量 *v*={*email*,*name*}的安全等级时,因为用户在需求中定义了 *email*,*name* 组合使用时的安全等级,变量的安全等级不应为 $\overline{v}=\overline{name}\oplus\overline{email}$ =(*M*,*top-retention*,{*current*,*contact*})⊕(*M*,1*day*,{*current*})=(*M*,1*day*,{*current*}).根据隐私规则 r_3, $\overline{v}=\overline{\{email,name\}}$ =(*H*,1*day*,{*current*}).

(2) 在组合执行过程中,成员服务可能多次接收用户隐私数据,虽然单次接收满足安全策略,但多次接收后可能不满足用户隐私需求.仍以情形(1)中的隐私需求为例,假设成员服务 *s* 的安全等级为 $\overline{s}$ =(*M*,1*day*,{*current*})),设 v_1={*email*},v_2={*name*},虽然 $\overline{v}_1\to\overline{s}\wedge\overline{v}_2\to\overline{s}$ 成立,但这两个信息流分别发生后,*s* 接收了数据项集合{*email*,*name*},违反了隐私规则 r_3.

4.3 安全等级绑定

4.3.1 成员服务安全等级绑定

成员服务 *s* 的安全等级可以静态绑定为 *s.sc*.假设 *s* 未接收过任何隐私数据,当其更新或新产生间接隐私数据项时,为满足安全策略的“不向下写”原则,*s* 输出数据的安全等级也应为 *s.sc*,这显然是不合理的.为此,针对成员服务 *s* 提出如下降密策略:由于 *s* 接收到的隐私数据是动态变化的,为防止用户隐私信息泄露,服务 *s* 的安全等级取决于其历史接收到的隐私数据,即 $\overline{s}=\overline{s.hds}$,其中,*s.hds* 表示服务 *s* 曾经接收到的隐私数据项集合.如果 *s* 未曾接收过任何隐私数据项,则 $\overline{s}=L_{sc}$.只要服务的操作满足安全策略, $\overline{s.hds}$ 总等于或低于 *s.sc*.

值得说明的是:检查组合向服务 *s* 释放隐私数据的操作时,*s* 的安全等级仍然绑定为 *s.sc*.

4.3.2 变量安全等级绑定

变量的安全等级绑定有两种方法:静态绑定和动态绑定.BPEL 中的变量是临时存放数据的容器,其安全等级取决于保存的内容,适合采用动态绑定方法.

变量 *v* 中包含的直接和间接隐私数据项的集合分别记为 v_{dir} 和 v_{ind},*v* 依赖的直接隐私数据项集合记为 *DDep*(*v*).由于安全策略的“不向下写,不向上读”原则以及成员服务的降密策略,所以有 $\overline{v}=\overline{DDep(v)}$.*DDep*(*v*)的计算公式如下:

$$DDep(v)=v_{dir}\cup DDep(d_{ind\_1})\cup\ldots\cup DDep(d_{ind\_m}) \tag{1}$$

其中,$d_{ind\_i}\in v_{ind}$,$1\leqslant i\leqslant m$,表示变量 *v* 中的间接隐私数据项.由于用户隐私需求中定义了隐私数据项组合的安全等级,从而引入隐私数据项聚合问题,因此,*v* 的安全等级的计算公式如下:

$$\overline{v}=\overline{ds_1}\oplus\ldots\oplus\overline{ds_i}\oplus\ldots\oplus\overline{ds_n} \tag{2}$$

其中,$ds_i\in 2^{DDep(v)}\wedge ds_i\in PR\_DSet$,$1\leqslant i\leqslant n$,$2^{DDep(v)}$表示 *DDep*(*v*)的幂集,*PR_DSet*={$r_k.ds|r_k\in PR$,$1\leqslant k\leqslant|PR|$},*PR* 表示隐私需求,且 $\overline{ds_i}=r_k.sc, ds_i=r_k.ds$.

4.3.3 间接隐私数据项安全等级绑定

传统的数据流分析技术关注系统实体之间的数据依赖关系,例如:服务 s_1 定义了变量 *v*,服务 s_2 使用变量 *v*,服务 s_2 数据依赖于服务 s_1.这类研究主要应用于数据流相关属性验证、系统演化分析等工作[23−25].信息流控制中的数据流分析则关注数据之间的依赖关系,例如:若存在赋值操作 *a*=*b*,则变量 *a* 依赖于变量 *b*.虽然文献[26]为实施信息流控制机制提出数据依赖分析规则,且这些规则可以分析因变量赋值引入的显式依赖以及流程结构引入的隐式依赖,但在本文的隐私保护场景中,由于考虑了面向成员服务的降密策略及隐私数据项聚合问题,需要提出特定的依赖分析方法.

组合执行过程中更新或新产生的间接隐私数据项记为 d_{new}.由于安全策略的“不向下写,不向上读”原则以及成员服务的降密策略,所以有 $\overline{d_{new}}=\overline{DDep(d_{new})}$,即其安全等级取决于依赖的直接隐私数据项集合

$DDep(d_{new}).d_{new}$ 依赖的直接和间接隐私数据项集合记为 $Dep(d_{new})$,只要得到 $Dep(d_{new})$,根据依赖的传递性,很容易得到 $DDep(d_{new})$.

针对组合的隐私工作流网模型,提出隐私数据项依赖关系分析规则如下.

- **DAR**$_1$. $T_A(t)=ASGN\Rightarrow Dep(d_{new})=r(t),\forall d_{new}\in w(t)$;
- **DAR**$_2$. $T_A(t)=RECV\Rightarrow Dep(d_{new})=T_S(t).hds,\forall d_{new}\in w(t)$;
- **DAR**$_3$. $T_A(t)=SND\Rightarrow T_S(t).hds=T_S(t).hds\cup r(t)$;
- **DAR**$_4$. $\forall d_{new\_i},d_{new\_j},\exists d_{new\_k},d_{new\_k}\in Dep(d_{new\_i})\wedge d_{new\_j}\in Dep(d_{new\_k})\Rightarrow Dep(d_{new\_i})=Dep(d_{new\_i})\cup\{d_{new\_j}\}$.

设 $assign(d_{new},v)$表示 assign 活动,显然 d_{new} 依赖于变量 v,该活动建模为 ASGN 类型变迁.对应 **DAR**$_1$ 有 $Dep(d_{new})=r(t),\forall d_{new}\in w(t)$.其中,$r(t)$表示变迁 t 的读集,$w(t)$表示变迁 t 的写集.设 $receive(s,v)$表示 receive 活动,v 中任意 d_{new} 依赖于 s 曾经接收到的隐私数据项集合,该活动建模为 RECV 类型变迁.对应 **DAR**$_2$ 有 $Dep(d_{new})=T_S(t).hds,\forall d_{new}\in w(t)$.其中,$T_S(t)$表示变迁 t 对应的成员服务,$T_S(t).hds$ 表示该服务曾经接收到的隐私数据项集合.在本文的隐私保护场景中,用户是可信实体,且我们假设用户不会更新或新产生间接隐私数据项,因此,若组合通过 receive 活动从用户接收数据,即 $T_S(t)$=USER,则分析该变迁时不需要应用规则 **DAR**$_2$.设 $reply(s,v)$表示 reply 活动,$invoke(s,v)$表示单向 invoke 活动,本质上均为向成员服务发送变量 v 的内容,均建模为 SND 类型变迁.两者需更新服务 s 接收到的隐私数据项集合,因此对应规则 **DAR**$_3$,即 $T_S(t).hds=T_S(t).hds\cup r(t)$.值得说明的是:由于用户是可信实体,若 $T_S(t)$=USER,该变迁不需要应用规则 **DAR**$_3$.设 $invoke(s,v_1,v_2)$表示请求响应 invoke 活动,建模为顺序执行的 SND 类型和 RECV 类型变迁,需依次应用规则 **DAR**$_3$ 和 **DAR**$_2$.**DAR**$_4$ 用于处理依赖的传递关系,即:如果 d_{new_i} 依赖于 d_{new_k},d_{new_k} 依赖于 d_{new_j},则 d_{new_i} 也依赖于 d_{new_j}.由于方法不考虑流程结构引入的隐蔽通道,所以不需要分析 STRC 类型变迁.

通过规则 **DAR**$_4$,可以从 $Dep(d_{new})$推导出 $DDep(d_{new})$.显然,通过分析隐私数据项的依赖关系,最终可以得到隐私数据项依赖图,其定义如下:

定义 4.6(隐私数据项依赖图). 隐私数据项依赖图 $PDIDG=(V,E)$是有向无环图,其中:V 是隐私数据项的有限集合;E 是边的集合,表示隐私数据项节点之间的依赖关系.节点对应的隐私数据项分为直接和间接隐私数据项两类,其中:出度为 0 的节点为直接隐私数据项节点;出度不为 0 的节点为间接隐私数据项节点.

4.4 信息流控制规则

为了验证路径的隐私信息流安全性,给出隐私信息流控制规则,其中,$secure(t)$=true 表示变迁 t 的触发是安全的.

- **IFCR**$_1$. $T_A(t)=ASGN\Rightarrow secure(t)$=true iff $\overline{r(t)}\rightarrow\overline{d_{new}}$,$\forall d_{new}\in w(t)$;
- **IFCR**$_2$. $T_A(t)=RECV\Rightarrow secure(t)$=true iff $\overline{T_S(t).hds}\rightarrow\overline{d_{new}}$,$\forall d_{new}\in w(t)$;
- **IFCR**$_3$. $T_A(t)$ = $SND\Rightarrow secure(t)$ = true iff $\overline{r(t)\cup T_S(t).hds}\rightarrow\overline{T_S(t)}$.

其中,**IFCR**$_1$ 表明 assign 活动是安全的,当且仅当 $\overline{r(t)}\rightarrow\overline{d_{new}}$.由于 d_{new} 的安全等级采用动态绑定,由 **DAR**$_1$ 可知 $\overline{d_{new}}=\overline{r(t)}$,因此 $\overline{r(t)}\rightarrow\overline{d_{new}}$ 总是满足的.**IFCR**$_2$ 表明 receive 活动是安全的,当且仅当 $\overline{T_S(t).hds}\rightarrow\overline{d_{new}}$,$\forall d_{new}\in w(t)$.同理,由于 d_{new} 的安全等级采用动态绑定,由 **DAR**$_2$ 可知 $\overline{d_{new}}=\overline{T_S(t).hds}$,因此规则 **IFCR**$_2$ 总是满足的.由于用户隐私需求中定义了隐私数据项组合的安全等级,从而引入隐私数据项聚合问题.**IFCR**$_3$ 表明 reply 活动或单向 invoke 活动是安全的,当且仅当 $\overline{r(t)\cup T_S(t).hds}\rightarrow\overline{T_S(t)}$.值得说明的是:若 $T_S(t)$=USER,该变迁不需要应用规则 **IFCR**$_2$ 和 **IFCR**$_3$.对于请求响应 invoke 活动,建模为顺序执行的 SND 类型和 RECV 类型变迁,分别应用规则 **IFCR**$_3$ 和 **IFCR**$_2$ 即可.

4.5 静态分析算法

通过静态分析算法检测待验证的每条路径是否会产生非法隐私信息流,该算法伪码见算法 1.

算法 1. 静态分析算法.

输入:*PR*/*用户隐私需求*/;
PW/*待验证路径集合*/
PWF-net/*组合隐私工作流网模型*/
输出:*bResult*/*布尔型,标识是否存在泄露*/
ILLEGAL_FLOWS/*非法信息流集合*/.

```
1.  bResult=true;
2.  ILLEGAL_FLOWS=∅;
3.  for each p in PW do
4.    init_pdidg(PR,PDIDG)
5.    for each t in p do
6.      if (t.subject==USER) continue; end if
7.      if (t.type==STRC) continue; end if
8.      if (t.type==ASGN) update(PDIDG,t); end if
9.      if (t.type==RECV) update(PDIDG,t); end if
10.     if (t.type==SND)
11.       if (chk_IFCR3(PR,t.subject)==false)
12.         bResult=false;
13.         ILLEGAL_FLOWS.add(p,t);
14.         break;
15.       else
16.         update_hds(t.rs,t.subject);
17.       end if
18.     end if
19.   end for
20. end for
```

算法需要分析 *PW* 中的每条路径 *p*,并依次处理 *p* 中每个变迁 *t*.设 *PW* 中路径数为 *n*,*p* 中变迁数为 *m*.在进行路径分析之前,首先通过第 4 行初始化隐私数据项依赖图 *PDIDG*,根据隐私需求初始化后的 *PDIDG* 中只包含直接隐私数据项节点.第 6 行的含义为:由于用户为可信实体,若变迁对应主体为 USER,不管变迁类型是 RECV 还是 SND,都不需要进行检测.第 7 行的含义为:由于本文分析方法不考虑控制流程引入的隐式泄露,所以不需要检测类型为 STRC 的变迁.第 8 行、第 9 行的含义为:由于间接隐私数据项的安全等级采用动态绑定,根据规则 **IFCR**$_1$,**IFCR**$_2$ 可知:ASGN,RECV 类型的变迁是安全的,只需分别通过规则 **DAR**$_1$,**DAR**$_2$ 更新隐私数据项依赖图 *PDIDG* 即可.若间接隐私数据项不在当前 *PDIDG* 中,则创建相应节点及依赖关系;否则,仅需更新相应的依赖关系.设组合中的隐私数据项数量为 *k*,变迁操作的数据项集合的大小、用户需求 *PR* 中规则的数量等均可以认为与 *k* 具有线性关系.由于每个 ASGN 类型变迁只更新一个间接隐私数据项,因此第 8 行的时间复杂度为 $O(k^2)$.RECV 类型变迁更新写集中间接隐私数据项,所以第 9 行的时间复杂度为 $O(k^3)$.第 10 行~第 18 行处理 SND 类型变迁,其中,第 11 行的 *chk_IFCR*3 函数用于检测变迁的触发是否满足信息流控制规则 **IFCR**$_3$.

为处理隐私数据项聚合问题,该函数首先根据公式(1)得到 $DDep(t.rs\cup t.subject.hds)$,其中,*t.rs* 表示变迁的读集,*t.subject.hds* 表示变迁对应服务 *t.subject* 接收到的隐私数据项集合,该步骤时间复杂度为 $O(k^3)$;然后,根据公式(2)计算 $\overline{DDep(t.rs\cup t.subject.hds)}$,该步骤的时间复杂度为 $O(k^3)$;最后,判断是否满足规则 **IFCR**$_3$ 中的条件 $\overline{DDep(t.rs\cup t.subject.hds)}\rightarrow\overline{t.subject}$,该步骤计算量主要体现在判断使用目的集合的包含关系上,设格 *P* 中使用目的的有限集合 *PS* 中元素数量为 *h*,则该步骤的时间复杂度为 $O(h^2)$.若不满足规则 **IFCR**$_3$,则在第 13 行记录非法信息流信息,并检测下一条路径;若满足,则在第 16 行更新对应成员服务接收到的隐私数据项集合,时间复杂

度为 $O(k^2)$.

综上所述,算法 1 的总体算法复杂度为 $O(nmk^3)$.

5 实例分析与实验

5.1 实例分析

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 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,假设完成机票预订和酒店预订后一并进行支付.针对其隐私数据,假设用户 Bob 定义的隐私需求如下.

- r_1=({*name*},(*M*,1*day*,{*current*,*contact*}));
- r_2=({*phone*},(*M*,1*day*,{*current*,*contact*}));
- r_3=({*id_number*},(*H*,1*day*,{*current*,*contact*}));
- r_4=({*credit_card_info*},(*H*,0*day*,{*current*}));
- r_5=({*name*,*id_number*,*credit_card_info*},(*TH*,0*day*,{*current*})).

其中,用户针对隐私数据项组合{*name*,*id_number*,*credit_card_info*}定义了更为严格的使用约束条件.此外,假设服务 flight,hotel,pay 对应的安全等级分别为(*H*,1*day*,{*current*,*contact*}),(*M*,1*day*,{*current*,*contact*}),(*H*,0*day*,{*current*}).为简化表述,略掉用户查询航班和酒店信息的过程及组合中的 assign 活动,TA 的 PWF_net 模型的控制流视图如图 7 所示.

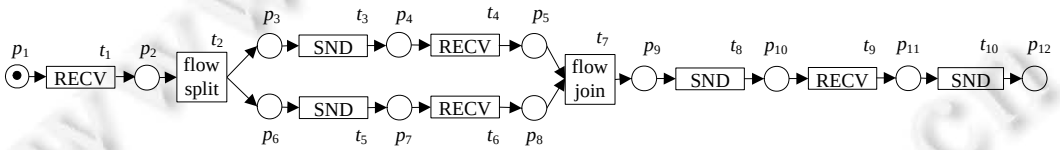


Fig.7 Control view of TA

图 7 TA 控制流视图

{t_3,t_4}与{t_5,t_6}这两组变迁分别对应酒店预订和机票预订操作,每组中的变迁会与另一组中的变迁并发执行,且任意两个并发执行的变迁不对应相同的成员服务,因此根据性质 1,只需验证 TA 的独立路径集,该路径集中只包含一条独立路径.假设选取 p=t_1,t_2,t_3,t_4,t_5,t_6,t_7,t_8,t_9,t_{10} 作为独立路径,通过算法 1 依次验证 p 中的每个变迁,验证过程信息及结果见表 2.

验证过程如下:

(1) t_1 接收用户预订信息,对应用户实体,t_2 属于 STRC 类型变迁,两者不需要验证;

(2) t_3 将隐私数据{*name*,*phone*}发送给酒店预订服务 hotel.因 $\overline{\{name, phone\}}$=(*M*,1*day*,{*current*,*contact*}), $\overline{hotel}$=(*M*,1*day*,{*current*,*contact*}),所以根据规则 **IFCR**$_3$, $\overline{\{name, phone\}} \rightarrow \overline{hotel}$ 成立,t_3 是信息流是安全的.同时,需要根据规则 **DAR**$_3$ 更新服务 hotel 使用的隐私数据项集合 *hotel.hds*={*name*,*phone*};

(3) t_4 接收酒店预订结果,根据规则 **DAR**$_2$,*Dep*(*hotel_order_id*)=*hotel.hds*={*name*,*phone*}.由于间接隐私数据项的安全等级采用动态绑定方法,t_4 是信息流是安全的.同时,需要根据规则 **DAR**$_2$ 更新路径对应的隐私数据项依赖图;

(4) t_5 将隐私数据{*name*,*id_number*}发送给机票预订服务 flight.因 $\overline{\{name,id\_number\}}$=(*H*,1*day*,{*current*,*contact*}), $\overline{flight}$=(*H*,1*day*,{*current*,*contact*}),所以根据规则 **IFCR**$_3$, $\overline{\{name,id\_number\}} \rightarrow \overline{flight}$ 成立,t_5 是信息流是安全的.同时,需要根据规则 **DAR**$_3$ 更新服务 flight 使用的隐私数据项集合:

flight.hds={*name*,*id_number*};

(5) t_6 接收机票预订结果,根据规则 **DAR**$_2$,*Dep*(*flight_order_id*)=*flight.hds*={*name*,*id_number*}.由于间接隐

私数据项的安全等级采用动态绑定方法,t_6是信息流是安全的.同时,需要根据规则 $\mathbf{DAR}_2$ 更新路径对应的隐私数据项依赖图;

(6) t_7属于 STRC 类型变迁,不需要验证;

(7) t_8将隐私数据集合 D={*hotel_order_id*,*flight_order_id*,*credit_card_info*}发送给服务 pay 以完成支付操作.根据之前验证操作建立的依赖关系易知,$DDep(D)$={*id_number*,*name*,*phone*,*credit_card_info*}.由于用户隐私规则 r_5针对隐私数据项组合{*name*,*id_number*,*credit_card_info*}定义了更为严格的使用约束条件,根据公式(2), $\overline{D}$ =(*TH*,0*day*,{*current*}).因 $\overline{pay}$ =(*H*,0*day*,{*current*}),所以 $\overline{D} \rightarrow \overline{pay}$ 不成立,变迁 t_8对应活动会引入用户隐私信息的非法泄露.

值得说明的是:一旦路径中的某变迁引入非法信息流,针对该路径的验证过程即可终止.

Table 2 Verification process and result

表 2 验证过程及结果

变迁	类型	含义	活动输入	活动输出	实体	实体访问的隐私数据集*hds*	是否安全
t_1	RECV	TravelBook	none	*name, id_number, credit_card_info, phone*	user	N/A	Yes
t_2	flow split	flow	none	none	TA	N/A	Yes
t_3	SND	HotelBook	*name, phone*	none	hotel	*name, phone*	Yes
t_4	RECV	HotelBookResp	none	*hotel_order_id*	hotel	*name, phone*	Yes
t_5	SND	FlightBook	*name, id_number*	none	flight	*name, id_number*	Yes
t_6	RECV	FlightBookResp	none	*flight_order_id*	flight	*name, id_number*	Yes
t_7	flow join	flow	none	none	TA	N/A	Yes
t_8	SND	PayRequest	*hotel_order_id, flight_order_id, credit_card_info*	none	pay	null	No
t_9	RECV	PayResponse	none	*pay_result*	pay	null	–
t_{10}	SND	BookResultResp	*hotel_order_id, flight_order_id, pay_result*	none	user	N/A	–

5.2 实 验

首先对算法 1 进行仿真实验,用以评估隐私数据项数量及路径中变迁数量对算法性能的影响.实验的系统环境为 Intel Pentium CPU 3.2GHz,4G 内存,32 位 Windows7 操作系统.编程环境为 Eclipse 4.4.2+JDK1.7.

算法 1 的时间复杂度为 $O(nmk^3)$,其中,n 是路径数量,m 是路径中变迁数量,k 是隐私数据项数量.我们针对单条路径对算法 1 的性能进行实验分析.实验相关参数设置为:隐私数据项数量为 k,其中,直接和间接隐私数据项的数量均为 k/2;用户隐私需求中隐私规则数量为 k,其中,涉及数据项组合的规则数量为 k/2;隐私数据及成员服务的安全等级随机生成,安全等级的集合采用本文第 2 节中的 *SC*;路径中计算量最大的 RECV 和 SND 类型变迁数量均为 m/2,读集和写集中包含的隐私数据项随机生成,且数量均为 k/2.m 分别取 10,50,90,评估 k 不同取值时算法的性能.实验结果如图 8 和图 9 所示.

算法 1 的最坏时间复杂度为 $O(nmk^3)$,主要计算量体现在处理 SND 和 RECV 类型变迁.图 8 中的实验结果表明:在实验设置条件下,当变迁数 m 值确定时,执行时间随着隐私数据项数 k 的增大而显著增大.例如:当 m=10,k=10 时,执行时间为 0.79ms;当 m=10,k=90 时,执行时间为 12.77ms.而当 k 的值确定时,m 值的增大对执行时间的影响则相对较小.

从图 9 可知:在实验设置条件下,当 m=10,k=10 时,算法 1 内存消耗为 0.69MB;当 m=90,k=90 时,内存消耗为 2.68MB.实验结果表明,算法 1 的内存消耗不会随着 m 和 k 的增大而急剧增长.这是因为依赖图中数据项数及边的数量、用户需求中的规则数、消息中的隐私数据项数均与 k 呈线性关系.

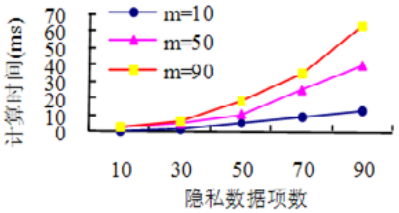


Fig.8 Variation of time cost with the number of privacy data items and transitions

图 8 计算时间随数据项数和变迁数的变化

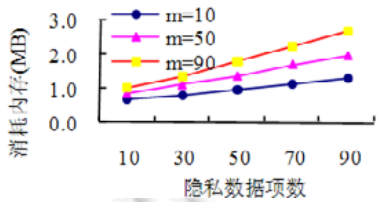


Fig.9 Variation of memory cost with the number of privacy data items and transitions

图 9 消耗内存随数据项数和变迁数的变化

由于组合的PWF-net建模及路径信息获取在部署组合前就已完成,验证开销取决于算法1.上述实验结果表明,方法的计算时间和内存消耗并不会随着变迁数量和隐私数据项数的增大而快速提升.在实际应用中,由于组合的路径、变迁、隐私数据项数均较小,因此本文方法并不会给用户与组合的交互引入过多的负载.

本文方法采用的技术路线与文献[18]类似,为便于比较,以第5.1节中的旅行代理为例,从需要分析的可达标识图中状态数量和需验证的路径数量两个方面进行实验分析,实验结果如图10和图11所示.

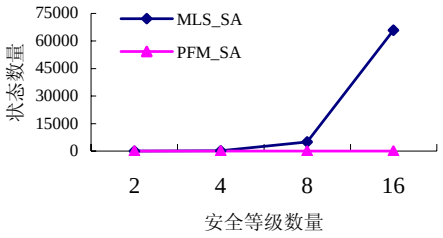


Fig.10 Variation of state number with the number of security classes

图 10 状态数随安全等级数量的变化

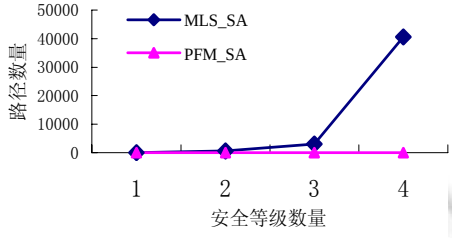


Fig.11 Variation of path number with the number of security classes

图 11 路径数随安全等级数量的变化

MLS_SA表示文献[18]中的方法,PFM_SA表示本文方法.随着安全等级数量的增加,方法MLS_SA需要分析的状态数和路径数会急剧增长.这是由于方法MLS_SA处理RECV类型变迁时,为数据库所标注所有可能的安全等级得到扩展可达标识图,穷举出系统所有可能的状态.而本文方法分析的可达标识图中状态数与路径数并不依赖于安全等级数量.本文方法中,可达图对应的状态数为14,路径数为6.由于TA中并发执行的变迁对应不同的成员服务,故只需对1条独立路径进行分析验证.而且,涉及安全等级的计算量主要体现在判断使用目的集合的包含关系上,时间复杂度为 $O(|PS|^2)$,其中,PS 是使用目的的有限集合.安全模型中安全等级的数量为$|SC|=|RS|\times|RT|\times|PC|$.由于 PC 是使用目的集合 PS 的幂集,且在实际应用中 PS 的元素较多,例如P3P中规定了12种使用目的,故$|SC|$往往较大.因此在实际应用中,本文方法的性能明显优于文献[18]中的方法.

6 总结与未来工作

隐私数据一旦提交给服务组合,用户难以控制组合如何使用和暴露用户隐私信息.如何保证组合执行过程中不发生用户隐私信息非法泄露,成为当前服务计算领域的研究热点之一.本文针对隐私保护特征,从服务信誉度、隐私数据使用目的及保留期限这3个维度提出一种面向服务组合的隐私信息流安全模型,从而形式化规约了隐私信息流的安全策略.为静态实施安全策略,提出了支持隐私信息流分析的隐私工作流网模型PWF-net,在组合的PWF-net模型基础上,通过静态分析算法检测组合执行是否会发生用户隐私信息的非法泄露.最后,通过实例分析说明了方法的有效性,并对方法性能进行了实验分析.

在基于本文方法构建隐私保护框架时,服务提供者在部署组合前创建对应的PWF-net模型及路径信息.用户可以通过隐私代理向组合发送隐私需求,组合通过静态分析算法检测是否存在隐私信息流非法泄露.若不存

在非法泄露,则继续进行使用服务组合.若存在非法泄露,则可以采用如下解决方法:(1) 用户降低隐私数据的释放约束限制;(2) 组合选择其他功能相同的成员服务以满足安全策略.

本文方法仍然存在一些限制,这也是下一步研究的重点.首先,将独立路径集作为待验证路径集的前提条件过于严格,为更好地解决并发变迁引入的路径爆炸问题,后续拟在深入研究并发变迁对隐私数据依赖关系影响的基础上,提出能适用更多情况的路径约简方法.此外,分析方法并未考虑流程结构引入的隐蔽通道,后续研究拟对方法进行扩展,以能够对流程结构引入的隐蔽通道进行检测.

References:

[1] Pearson S. Taking account of privacy when designing cloud computing services. In: Proc. of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing. New York: IEEE Press, 2009. 44−52. [doi: 10.1109/CLOUD.2009.5071532]

[2] Warren SD, Brandeis LD. The right to privacy. Harvard Law Review, 1890,4(5):193−220. [doi: 10.2307/1321160]

[3] Westin A. Privacy and Freedom. New York: Atheneum, 1967.

[4] Goldberg I, Wagner D, Brewer E. Privacy-Enhancing technologies for the Internet. In: Proc. of the 42nd IEEE Int'l Computer Conf. New York: IEEE Press, 1997. 103−109. [doi: 10.1109/CMPCON.1997.584680]

[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.csi.2010.09.001]

[6] 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]

[7] Organization for Economic Co-operation and Development. OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. Organization for Economic Co-operation and Development, 2013.

[8] 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]

[9] 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]

[10] Yan D, Tian Y, Huang J, Yang F. Privacy-Aware RBAC model for Web services composition. The Journal of China Universities of Posts and Telecommunications, 2013,20(1):30−34. [doi: 10.1016/S1005-8885(13)60253-8]

[11] 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 with English abstract). http://www.jos.org.cn/1000-9825/4945.htm [doi: 10.13328/j.cnki.jos.004945]

[12] Bacon J, Eyers D, Pasquier TFJM, Singh J, Papagiannis I, Pietzuch P. Information flow control for secure cloud computing. IEEE Trans. on Network and Service Management, 2014,11(1):76−89. [doi: 10.1109/TNSM.2013.122313.130423]

[13] Nakajima S. Model-Checking of safety and security aspects in Web service flows. In: Proc. of the 4th Int'l Conf. on Web Engineering. Berlin: Springer-Verlag, 2004. 488−501. [doi: 10.1007/978-3-540-27834-4_60]

[14] Denning DE. A lattice model of secure information flow. Communications of the ACM, 1976,19(5):236−243. [doi: 10.1145/360051.360056]

[15] Hutter D, Volkamer M. Information flow control to secure dynamic Web service composition. In: Proc. of the 3rd Int'l Conf. on Security in Pervasive Computing. Berlin: Springer-Verlag, 2006. 196−210. [doi: 10.1007/11734666_15]

[16] Accorsi R, Lehmann A, Lohmann N. Information leak detection in business process models: Theory, application, and tool support. Information Systems, 2015,47:244−257. [doi: 10.1016/j.is.2013.12.006]

[17] Bell DE, Lapadula LJ. Secure computer systems: Mathematical foundations. MITRE Technical Report, 2547, Bedford: MITRE Corporation, 1996.

[18] Knorr K. Multilevel security and information flow in Petri net workflows. In: Proc. of the 9th Int'l Conf. on Telecommunication Systems—Modeling and Analysis. Las Vegas: WASET, 2001. 1−16.

[19] Cranor L, Dobbs B, Egelman S, Hogben G, Humphrey J, Langheinrich M, Marchiori M, Presler-Marshall M, Reagle J, Schunter M, Stampley DA, Wenning R. The Platform for Privacy Preferences 1.1 (P3P1.1) Specification. W3C, 2006.

[20] Van der Aalst WMP. Verification of workflow nets. In: Proc. of the 18th Int'l Conf. on Application and Theory of Petri Nets. Berlin: Springer-Verlag, 1997. 407−426. [doi: 10.1007/3-540-63139-9_48]

[21] Zhou GF, Du ZM. Petri nets model of implicit data and control in program code. Ruan Jian Xue Bao/ Journal of Software, 2011, 22(12):2905−2918 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/3956.htm [doi: 10.3724/SP.J.1001.2011.03956]

[22] Lohmann N. A feature-complete petri net semantics for WS-BPEL 2.0. In: Proc. of the 4th Int'l Workshop on Web Services and Formal Methods. Berlin: Springer-Verlag, 2007. 77−91. [doi: 10.1007/978-3-540-79230-7_6]

[23] Kazhamiakin R, Pistore M. Static verification of control and data in Web service compositions. In: Proc. of the 4th IEEE Int'l Conf. on Web Services. New York: IEEE Press, 2006. 83−90. [doi: 10.1109/ICWS.2006.124]

[24] Song M, Wei ZX, Yin GS. Evolution analysis of data flow oriented internetware service. Ruan Jian Xue Bao/Journal of Software, 2013,24(12):2797−2813 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4396.htm [doi: 10.3724/SP.J.1001.2013.04396]

[25] Song W, Ma XX, Lu J. Instance migration in dynamic evolution of Web service compositions. Chinese Journal of Computers, 2009, 32(9):1816−1831 (in Chinese with English abstract). [doi: 10. 3724/SP.J.1016.2009.01816]

[26] She W, Yen IL, Thuraisingham B, Huang SY. Rule-Based run-time information flow control in service cloud. In: Proc. of the 9th IEEE Int'l Conf. on Web Services. New York: IEEE Press, 2011. 524−531. [doi: 10.1109/ICWS.2011.35]

附中文参考文献:

[8] 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]

[11] 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 [doi: 10.13328/j.cnki.jos.004945]

[21] 周国富,杜卓敏.程序代码中隐含数据与控制的 Petri 网建模技术.软件学报,2011,22(12):2905−2918. http://www.jos.org.cn/1000-9825/3956.htm [doi: 10.3724/SP.J.1001.2011.03956]

[24] 宋敏,韦正现,印桂生.面向数据流的网构软件服务动态演化分析.软件学报,2013,24(12):2797−2813. http://www.jos.org.cn/1000-9825/4396.htm [doi: 10.3724/SP.J.1001.2013.04396]

[25] 宋巍,马晓星,吕建.Web 服务组合动态演化的实例可迁移性.计算机学报,2009,32(9):1816−1831. [doi: 10.3724/SP.J.1016.2009.01816]

彭焕峰(1978－),男,山东临沂人,博士生,副教授,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.

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.

刘林源(1981－),男,博士,讲师,主要研究领域为云计算与服务计算,系统可靠性与安全.

李勇(1983－),男,博士生,讲师,主要研究领域为实证软件工程,机器学习.

柯昌博(1984－),男,博士,讲师,CCF 专业会员,主要研究领域为基于本体的软件工程,SaaS 服务中的隐私增强技术.