第37卷第7期 中国科学技术大学学报 Vol. 37, No. 7
2007年7月 JOURNAL OF UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA Jul. 2007

文章编号:0253-2778(2007)07-0785-06

基于权值加密的隐私安全分布式约束满足问题求解*

王秦辉,陈恩红,王煦法

(中国科学技术大学计算机科学技术系,安徽合肥 230027)

摘要:隐私安全的分布式约束满足问题(distributed constraint satisfaction problem, DisCSP)求解算法可以很好地满足信息敏感的分布式组合求解问题的需要,为了获得更好的求解效率,提出了一种基于权值加密的隐私安全DisCSP的求解算法,对DisCSP问题中的约束基于不同的隐私权值进行加密求解;不需要增加额外的agent进行隐私约束的一致性检查,实现分布式的安全求解策略;对于可能出现的推理信息,用随机选择策略来避免信息泄漏.试验表明,该算法可以减少信息的传递量和计算的复杂性,因而具有更好的求解效率.

关键词:分布式约束满足;隐私安全;加密;异步回退

中图分类号:TP301 **文献标识码**:A

Solving secure distributed constraint satisfaction problems based on encrypted weighted-privacy

WANG Qin-hui, CHEN En-hong, WANG Xu-fa

(*Department of Computer Science and Technology, University of Science and Technology of China, Hefei* 230027, *China*)

Abstract: Algorithms for solving secure distributed constraint satisfaction problems can meet the requirements of information security of distributed combination problems. To improve the solving efficiency, an algorithm for solving secure DisCSP based on encrypted weighted-privacy was presented. The search procedure for solution was encrypted based on weight of different constraints; to realize a real distributed secure DisCSP protocol, additional agents were not introduced to check the consistency of private constraints. Finally, random selection strategy was used to prevent reasoning information leakage. Experimental results show that the algorithm obtains better solving efficiency through reducing the amount of information communicated and the complexity of computation.

Key words: distributed constraint satisfaction; private/secure; encrypt; asynchronously backtrack

研究快报

随着硬件和网络技术的发展,分布式计算环境快速广泛地在各个领域中得到应用,很多人工智能中利用约束满足来求解的问题也越来越多地处于分布式计算环境下,使得DisCSP成为一个十分重要的研究领域.同时对信息敏感的约束关系在求解过程中如何保证其隐私安全成为近来研究的热点.例如在电子商务等情况中,各实体之间的约束通常是不能泄露给竞争者的战略信息.

虽然与传统的集中式约束满足求解算法相比,分布式求解方法能够更好地保护隐私敏感信息的安

* **收稿日期**:2007-02-09;**修回日期**:2007-06-10
基金项目:国家自然科学基金(60573077),新世纪优秀人才支持计划(NCET-05-0549)资助.
作者简介:王秦辉,男,1977年生,博士生. 研究方向:约束满足问题. E-mail: wqh@mail.ustc.edu.cn
通讯作者:陈恩红,博士/教授. E-mail: cheneh@ustc.edu.cn; Tel:0551-3602824

全性[1],但这在数据隐私非常敏感的情况下是很不充分的. 因为在 DisCSP 的求解算法中,当两个 agent 之间控制的变量有约束关系时,这两个 agent 必须互相知道各自对变量的赋值,才能在算法的搜索过程中达到约束一致. 这样,需要由 agent 控制的隐私信息就会泄漏给其他 agent.

为了更好地全面解决隐私泄漏问题,文献[2~5]提出了一系列 MPC-DisCSPx 算法来求解 CSP,这些算法用安全多方计算[6~8]来模拟运算电路[8]作为求解过程,并使用文献[8]中的策略作为求解 CSP 的安全策略. 因为电路的规模总是取决于输入规模,使得该算法的通讯和计算代价总是处于最大情况,另外这种普适的多方计算策略不能利用启发式获得更好的效率,故这种方法只能有限地求解很小的问题. 进而[9]提出了一个更高效率的专门求解 DisCSP 的算法,但是引入了额外的 agent 来求解 DisCSP,是一种"集中式"的求解策略,具有很高的计算复杂性. 文献[10]在此基础上利用安全多方计算来避免增加额外的 agent,但是每次迭代开销都要比文献[9]高得多.

我们在文献[9]的基础上提出了一个基于权值加密的求解 Secure DisCSP 的算法,首先对 DisCSP 问题中不同的约束给予不同的权值,对有隐私安全要求的约束仍然使用不能识别的、同形的和随机的公钥加密方案;此外对约束的加密和置换都在包含约束的 agent 中进行,这个过程可以结合任意的 DisCSP 求解算法,不需要增加额外的 agent,从而实现完全的分布式的安全求解策略;最后对于可能出现的推理信息,我们用随机选择的策略来避免泄漏. 因为加密求解是基于权值的,可以大大减少信息的传递量和计算的复杂性,试验表明,此算法具有更好的求解效率.

DisCSP 问题是变量和约束都分布在不同自治 agent 中的 CSP 问题. 在 CSP 的基础上,可如下定义 DisCSP 问题[11].

定义 1 DisCSP 问题表示为四元组(A,V,D,C),其中:

A 是 n 个 agent 的集合$\{A_1, A_2,\cdots,A_n\}$;

V 是 m 个变量的集合$\{v_1,\cdots,v_m\}$;

D 是所有变量的值域的集合,$D=\{D_1,\cdots,D_m\}$, D_i是变量 v_i的所有可能取值的有限域;

C 是变量之间约束关系的集合 $C=\{C_1,\cdots,C_t\}$,其中每个约束包含一个 V 的子集$\{v_i,\cdots,v_j\}$和一个约束关系 $R\subseteq D_i\times\cdots\times D_j$.

每个 agent 有一个或多个变量,每个变量 v_j属于一个 A_i表示为 belongs(v_j, A_i);变量间的约束关系分布在 agent 内或 agent 之间,当 A_l知道约束关系 C_k时表示为 known(C_k, A_l). 对变量 v_i的赋值记为 $\varepsilon|v_i$, 对变量的子集 $V_i\subseteq V$ 的赋值记为$\varepsilon|V_i$. 当对变量子集 V_i 的赋值满足约束关系 C_i 时定义 $C_i(\varepsilon|V_i)=1$,不满足时有 $C_i(\varepsilon|V_i)=0$.

每个 agent 负责一些变量并决定它们的值,赋值必须满足所有的约束. DisCSP 问题的解的定义为:

定义 2 当且仅当满足下述条件时,DisCSP 问题找到了解:$\forall A_i$, $\forall v_j$存在关系 belongs(v_j, A_i),当 v_j的赋值是 $d_j\in D_j$ 时, $\forall C_k$, $\forall A_l$, known(C_k, A_l)都有 C_k被满足. 也即问题的解为对所有变量的赋值,并且该赋值满足 agent 间及 agent 内的所有约束,记此赋值为 ε^*,有

$$\prod_{i=1}^{t} C_i(\varepsilon^* \mid V_i) = 1. \tag{1}$$

对有的 DisCSP 问题而言,只找到问题的解是不够的,这类问题还要求在求解过程中保证敏感信息或者隐私不会泄漏给其他 agent. 我们对 DisCSP 的定义进行扩充,得到隐私安全 DisCSP 的定义,而且对问题中可能存在的隐私关系进行分类,对不同的隐私安全要求赋予不同的权值.

定义 3 隐私安全 DisCSP 问题是求解过程中保证所要求的隐私或敏感信息不被泄漏的 DisCSP 问题. 可以表示为五元组(A,V,D,C,W),其中 A, V,D,C 同定义 1,W 是对约束有无隐私安全要求进行描述的权值集合 $W=\{w_1,\cdots,w_t\}$. 记求解过程为 Γ,约束 C_k的隐私安全要求定义为

$$\{A_i \mid \text{known}(C_k,A_i)\} \overset{\Gamma}{=} \{A_i \mid \text{known}(C_k,A_i)\},$$

即求解过程中,该约束的信息不会泄漏给其他的 agent. 如下定义权值:

$$w_s = \begin{cases} 1 & \text{要求} \{A_i \mid \text{known}(C_s,A_i)\} \overset{\Gamma}{=} \{A_i \mid \text{known}(C_s,A_i)\}, \\ 0 & \text{其他}. \end{cases}$$

若求解过程满足约束 C_k 的隐私安全要求,记 $S(\Gamma|C_k)=1$,否则 $S(\Gamma|C_k)=0$.

DisCSP 问题中的约束关系存在于 agent 内和 agent 之间,根据约束关系中的变量个数,约束关系可以分为一元约束、二元约束与多元约束. 因为约束

研究快报

关系的传递性，多元约束可以划分成多个二元约束.

一元约束 C^1：即 C_k 满足 $C_k=\langle v_i, R\in D_i\rangle$.

二元约束 C^2：即 C_k 满足 $C_k=\langle\{v_i, v_j\}, R\subseteq D_i\times D_j\rangle$.

在约束关系分类的基础上，我们可以得到如下的隐私安全类别：

一元约束隐私 P^1：当 $C_k\in C^1$ 且 $w_k=1$ 时，C_k 的隐私要求为一元约束隐私.

二元约束隐私 P^2：当 $C_k\in C^2$ 且 $w_k=1$ 时，C_k 的隐私要求为二元约束隐私.

除此之外，还有一类有安全性保护要求的隐私是 agent 通过最后的求解结果由推理可能得到的，称为推理隐私 P^c，这种隐私泄漏的发生不可预知，在有多组解的情况下较易出现. 例如在会议安排问题中，当最后的求解结果有多组解，而且这些解以很大的概率表现出某个参会者在某个时间上有冲突，那么此信息就有可能作为隐私让其他参会者获知.

定义 4　基于权值的隐私安全 DisCSP 问题的解是通过一个满足隐私安全要求的求解过程获得问题的解，包括 ε^* 和 Γ^* 两部分. ε^* 同定义 3. Γ^* 满足

$$\sum_{i=1}^{t} S(\Gamma^* \mid C_i)=\sum_{i=1}^{t} w_i, \tag{2}$$

即求解过程 Γ^* 满足了所有隐私安全要求.

例 1　如图 1 所示，3 个 agent A_1，A_2，A_3 分别表示 3 个会议参加者，每个 agent 控制表示时间和地点的两个变量. 变量 v_1，v_3，v_5 的可取值为 $D_1=D_3=D_5=\{5$ 天工作日$\}$；变量 v_2，v_4，v_6 的可取值为 $D_2=D_4=D_6=\{$省会城市$\}$. 该问题的约束关系有一元约束 $C_1=\langle v_1,\{$M(周一)，W(周三)，F(周五)$\}\rangle$，$C_2=\langle v_2,\{$B(北京)，S(上海)，H(合肥)$\}\rangle$，$C_3=\langle v_3,\{$M，W$\}\rangle$，$C_4=\langle v_4,\{$B，S$\}\rangle$，$C_5=\langle v_5,\{$W，F$\}\rangle$，$C_6=\langle v_6,\{$S，H$\}\rangle$；二元约束 $C_7=\langle(v_1, v_2)\mid\{$(M，B)，(W，B)，(F，S)，(F，H)$\}\rangle$，$C_8=\langle(v_3, v_4)\mid\{$(M，B)，(W，B)，(F，B)$\}\rangle$，$C_9=\langle(v_5, v_6)\mid\{$(W，S)，(W，H)，(F，H)$\}\rangle$表示参加者可以接受的会议时间地点的组合，另外约束 $C_{10}=\langle(v_1, v_3)\mid\{$(M，M)，(W，W)$\}\rangle$表示参加者 1 和 2 有在周一或周三会面的要求，而 $C_{11}=\langle(v_2, v_6)\mid\{$(S，S)，(H，H)$\}\rangle$表示参加者 1 和 3 有在上海或合肥会面的要求. 假设有一全局约束是由于某客观原因会议不能在周三的北京举行，即 $C_{12}=\langle\{(v_1, v_2)\neq$(W，B)，$(v_3, v_4)\neq$(W，B)，$(v_5, v_6)\neq$(W，B)$\}\rangle$. 对 A_1 有 known(C_k, A_1)，$k=1,2,10,11,12$. 对 A_2 有 known(C_k, A_2)，$k=3,4,8,10,12$. 对 A_3 有 known(C_k, A_3)，$k=5,6,9,11,12$. 对些约束可以按照各参加者的要求赋予相应的隐私安全权值，此例中假设参加者对自己提出的约束都有隐私安全要求，则有 $w_i=1, i=1,2,\cdots,11$. 因为 C_{12} 是所有 agent 都可以知道的约束，没有隐私性，所以有 $w_{12}=0$.

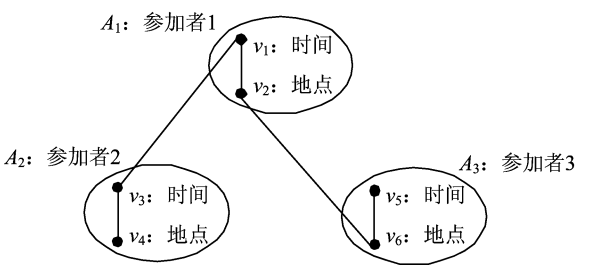

图 1　分布式会议安排的约束网示例

Fig. 1　Constraint network of distributed meeting scheduling

求解 DisCSP 的基本算法有异步回溯算法，异步 weak-commitment 搜索算法和分布式逃逸算法等[11]. 已有的安全隐私 DisCSP 问题的求解过程都是在这些算法的基础上结合不同的安全策略，如文献[9]中利用了 ElGamal 公钥加密，而文献[10]中采用了安全多方计算中的安全共享策略.

我们的求解算法在搜索过程中仍然利用公钥加密体制，因为密文要比明文长，为了减少消息的传递，提出了对约束隐私进行基于权值加密的策略.

基于权值的 ECC 加密策略：对于 $w_i=1$ 的约束 C_i 需要保证其隐私安全而进行加密，但是由于分布式求解的特性，agent 内部的约束在求解时只在内部传递消息，而不用向其他的 agent 发送，因此不可能泄漏内部约束信息. 故对 agent 内的权值 $w_i=1$ 的约束 C_i 不需要进行加密处理，此类约束的隐私安全性可由问题的分布式特性保证. 对 agent 间的权值 $w_i=1$ 的约束 C_i 我们采用椭圆曲线密码体制 ECC[12] 来保证信息安全，这种公钥加密体制具有密钥短，运算快的特点.

隐私安全的约束一致性检查策略：和 DisCSP 求解算法不同的是，这里 agent 之间的约束关系加了密，对变量赋值的一致性检查也要做相应处理，搜索过程仍然以异步回退的方式进行. 以两个有约束关系的 agent 为例，隐私安全一致性检查过程如图 2.

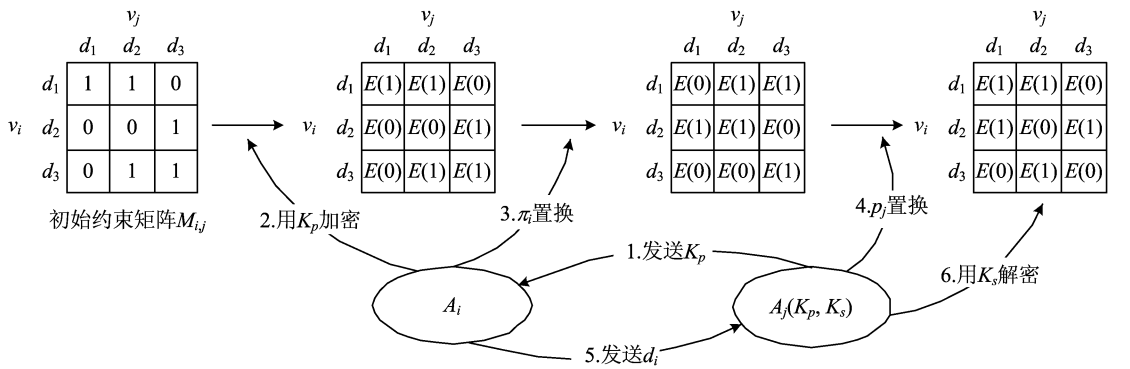

图 2 隐私安全的约束一致性检查

Fig. 2 Check consistency of constraints in security

图 2 中 agent A_i和 A_j所包含的变量分别为 v_i，v_j，它们之间的约束关系表示为约束矩阵 $\boldsymbol{M}_{i,j}$. 矩阵中$(d_i,d_j)=1$ 表示满足约束的赋值. 假设 $i<j$，首先由 A_j生成公钥 K_p和私钥 K_s，然后由 A_i对 $\boldsymbol{M}_{i,j}$进行 ECC 加密，再对加密后的矩阵执行行置换 π_i，图中 $\pi_i=(123)$. A_i将置换后的 $\boldsymbol{M}_{i,j}$发送给 A_j进行列置换 π_j，图中 $\pi_j=(132)$. 约束一致性检查时，A_i选择一个赋值 d_i发送给 A_j，A_j检查矩阵中该赋值对(d_i,d_j)的值 $\boldsymbol{M}_{i,j}(d_i,d_j)$，若解密后 $K_s(\boldsymbol{M}_{i,j}(d_i,d_j))=1$，则说明此对赋值是满足约束的，若 $K_s(\boldsymbol{M}_{i,j}(d_i,d_j))=0$，则 A_i改变赋值，重新开始一个搜索检查过程.

解的随机选择策略：对于有多组解的情况，假设解的个数为 N，则产生一个 N 以内的随机数 random(N)，作为返回解的索引号.

基于权值的隐私安全 DisCSP 问题求解的算法描述如下.

```
Algorithm WSDisCSP
input 基于权值的隐私安全 DisCSP 问题(A,V,D,C,W)
output 问题的解 ε*
  Num_Solution←0;
  PROCEDURE:
    Partial_Solution ←{}, num_var←0;
    Agents Ai(i=1 to n) asynchronously perform:
      if num_var>m
      then Num_Solution←Num_Solution+1,
        record the current value assignment,
        goto PROCEDURE;
      CHECK WEIGHT:
        for each Ck that only known(Ck, Ai) do
          if wk=1 then refine Dk and related Cr;
        for each other Ck∈C² that known(Ck, Ai) do
          if wk=1 then CHECK CONSISTENCY;
      CHECK CONSISTENCY:
        Ai: receive Kp from related Aj;
            Mi,j← Encrypt(Mi,j);
            Mi,j← πi(Mi,j);
        Aj: Mi,j← πj(Mi,j);
            receive di from Ai;
            decrypt Mi,j(di,dj) to Ai;
        Ai: if (decrypt Mi,j(di,dj))=1
            then di←πi⁻¹(πj⁻¹(di));
              add (vi,di) to Partial_Solution;
              num_var←num_var+1;
            else if other di exist
              then select di and send to Aj;
              else goto BACKTRACK;
      BACKTRACK:
        if Partial_Solution is empty
        then announce that there exists no solution;
          terminate the algorithm;
        else num_var←num_var-1;
          remove (vj,dj) from Partial_Solution;
          asynchronous backtracking
    if found no new solution
    then break PROCEDURE
return ε* ←Partial_Solution(random(Num_Solution))
```

在求解算法中没有对所有的约束都进行加密，而是对不同的约束类别，根据隐私安全性权值来进行处理. 这并没有影响求解的隐私安全性，我们在算法中的每个步骤都保证了求解的安全.

(Ⅰ) 对于 agent 内的一元约束和二元约束不需要进行加密处理. 因为这类约束只在 agent 内进行一致性检查，不会泄漏给其他 agent，这也是由分布式约束满足问题的分布式特性所决定的.

研究快报

(Ⅱ) 对于 agent 间的二元约束，因为约束矩阵被加密，首先能够保证该约束信息不会在消息的传递中泄漏给其他的 agent，此外约束矩阵经过两次置换，这样对约束相关的 agent 而言所有的赋值都有相同的概率成为问题的解，而且变量的赋值通过置换随机地重命名，agent 不会知道另一个 agent 所选择的实际原始值，agent 自身也需要通过逆置换才能知道满足约束的赋值是哪个实际原始值. 所以通过对约束矩阵的加密和置换能够保证约束自身的隐私安全，约束相关 agent 的赋值隐私也得到了安全保证.

(Ⅲ) 对于可能出现的推理信息，用解的随机选择策略，可以让问题在有多组解的情况下等概率地返回某个解，降低进行推理的可能.

(Ⅳ) 因为我们的求解算法没用引入额外的 agent 来进行搜索控制或者解码，所以也没有泄漏信息给第三方的风险.

以上分析可以知道算法的求解过程是安全的，满足(2)式.

我们同文献[9]中的算法进行对比试验，4 组试验中设定的 agent 数目分别为 6，8，10，12，变量的值域大小均为 5，每组试验随机生成 15 个隐私安全 DisCSP 问题，最后取平均结果进行比较. 因为文献[9]中的算法是集中式的，图示中用 CSDisCSP 表示，基于权值的求解算法用 WSDisCSP 表示.

图 3 是算法求解过程中发生的信息传递数目的比较. 两种算法的搜索过程与标准的按序回退是相同的. 在最坏的情况下，信息的交换次数是 $O(|D|^m)$. 但是因为 CSDisCSP 在求解过程中引入了搜索控制 agent 和几个解码 agent，所以除了 agent 间的信息传递外，还要同额外的 agent 以及额外 agent 之间都要发生信息传递. 而我们的算法没有引入额外的 agent，信息交换只发生在问题固有的 agent 之间，因而如图 3 所示，我们的算法在求解过程中的信息交换次数是少于 CSDisCSP 的.

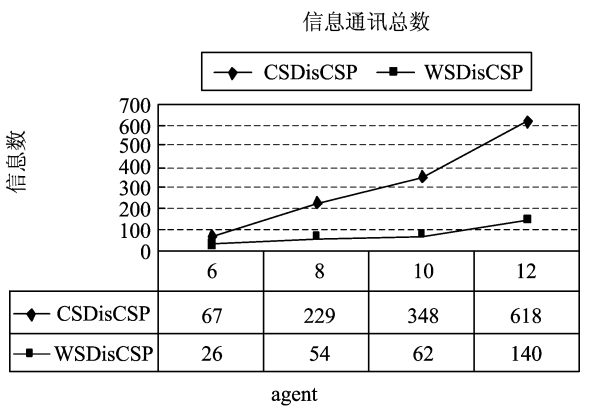

图 3　算法的信息通讯数目比较

Fig. 3　Amount of information communicated comparison between two algorithms

在求解过程中，称 agent 的一次动作为一个运算单元，包括选择赋值、信息发送和接收、约束一致性检查等. 我们的算法是基于权值进行加密处理，相对于 CSDisCSP 的统一加密处理而言，会减少运算单元，而且分布式的求解也会减少集中式求解中的额外运算单元. 这两个方面决定了我们的算法求解中所需的运算单元数大大减少. 从图 4 两个算法求解过程中运算单元总数的比较可以得出这一结果.

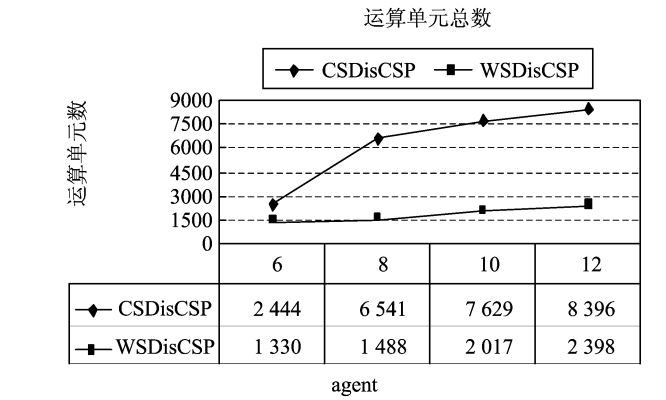

图 4　算法的运算单元数目比较

Fig. 4　Amount of computation units comparison between two algorithms

信息传递和运算单元数目的减少，必然会缩减算法的求解时间，另外 ECC 加密体系易于计算和密文较短的特性也减少了计算的复杂性，进一步提高了 WSDisCSP 的求解效率，结果如图 5 所示.

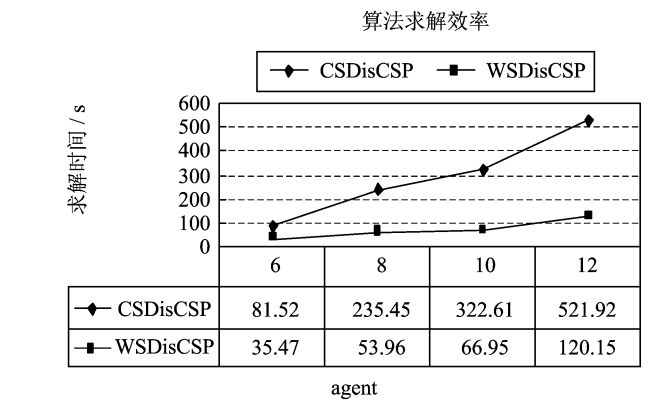

图 5　算法求解效率比较

Fig. 5　Performance comparison between two algorithms

研究快报

实验结果表明,我们的算法通过对不同隐私安全要求,在加密求解过程中进行不同处理可以获得更高的求解效率.与以前的工作进行对比表明,我们的算法进一步减少了可能信息的泄漏,不再需要额外的控制器来参与求解,因而可以在确保隐私安全的前提下引入更好的启发式搜索策略.

参考文献(References)

[1] Faltings B, Yokoo M. Introduction: Special issue on distributed constraint satisfaction [J]. Artificial Intelligence, 2005,161(1-2):1-5.

[2] Silaghi M C. Solving a distributed csp with cryptographic multi-party computations, without revealing constraints and without involving trusted servers [C]//Proc. Workshop on DCR-03 IJCAI, Acapulco, Mexico,2003.

[3] Silaghi M. C, Mitra D. Distributed constraint satisfaction and optimization with privacy enforcement [C]//Zhong N Proc of ICIAT. Los Alamitos: IEEE Press,2004:531-535.

[4] Silaghi M C. Meeting scheduling system guaranteeing n/2-privacy and resistant to statistical analysis (applicable to any DisCSP) [C]//Zhong N Proc of ICWI. Los Alamitos:IEEE Press,2004:711-715.

[5] Silaghi M C. Hiding absence of solution for a distributed constraint satisfaction problem[C]//Russell I Proc of the 18th International FLAIRS Conference. Florida USA: AAAI Press,2005: 854-855.

[6] Goldreich O, Micali S, Wigderson A. How to play any mental game or a completeness theorem for protocols with honest majority[C]//Proc of the 19th Annual ACM STOC. New York: ACM Press,1987: 218-229.

[7] Chaum D, Crépeau C, Damgård I. Multiparty unconditionally secure protocols (extended abstract) [C]// Proc of the 20th Annual ACM STOC. New York: ACM Press, 1988:11-19.

[8] Ben-Or M, Goldwasser S, Wigderson A. Completeness theorems for non-cryptographic fault-tolerant distributed computation (extended abstract) [C]// Proc of the 20th Annual ACM STOC. New York: ACM Press, 1988: 1-10.

[9] Yokoo M, Suzuki K, Hirayama K. Secure distributed constraint satisfaction: reaching agreement without revealing private information [J]. Artificial Intelligence, 2005,161(1-2):229-245.

[10] Kobbi Nissim and Roie Zivan. Secure DisCSP Protocols-From Centralized Towards Distributed Solutions [C]//Proc Workshop on DCR-05 IJCAI, Edinburgh, 2005.

[11] 王秦辉, 陈恩红, 王煦法. 分布式约束满足问题研究及其进展[J]. 软件学报,2006,17(10): 2 029-2 039.

[12] Hankerson D, Menezes A J, Vanstone S. Guide to Elliptic Curve Cryptography [M]. New York: Springer-Verlag, 2004: 153-196.

研究快报