Modelling and Solving QoS Composition Problem Using Fuzzy DisCSP

Xuan Thang Nguyen
Faculty of Information and Communication Technologies
Swinburne University of Technology
Melbourne VIC 3122, Australia
Email:xnguyen@ict.swin.edu.au

Ryszard Kowalczyk
Faculty of Information and Communication Technologies
Swinburne University of Technology
Melbourne VIC 3122, Australia
Email: rkowalczyk@ict.swin.edu.au

Manh Tan Phan
Faculty of Information and Communication Technologies
Swinburne University of Technology
Melbourne VIC 3122, Australia
Email: tphan@ict.swin.edu.au

Abstract—Web service compositions have attracted considerable efforts in the context of supporting enterprise application integrations. For a composite service, in addition to its functional requirements, QoS requirements are important and deserve a special attention. The central question to a QoS composition problem is how to compose a service from different sub-component services so that its overall QoS can satisfy certain requirements. In this paper, we propose an agent-based method using Fuzzy Distributed Constraint Satisfaction Problem (Fuzzy DisCSP) techniques to solve this problem. We show that by using the composition structures, local constraints can be constructed and used with DisCSP. We also present an a new algorithm called the Fuzzy constraint satisfaction Algorithm for Distributed Environment (FADE) to solve the problem and discuss our experiment in building a prototypical system to prove the feasibility of our approach.

I. INTRODUCTION

There have been tremendous developments in the area of Web services recently. Web services’ capability for integrating distributed and heterogeneous applications across organizational boundaries has attracted considerable research in various areas including Web service discovery, composition, and management. Web service composition focuses on the problem of composing a value-added composite Web service from different Web services. There are two key requirements in in the composition process: finding syntactic and semantic matches of the component Web services and satisfying non-functional requirements or Quality of Service (QoS). We call the latter - QoS composition problem for composite Web services as opposed to the former - functionality composition problem. Because component Web services of a composite Web service can also be composite, a composite Web service can have any number of composition nested level. In parallel to the advancement of Web services, the MAS and AI community has shown an increasing interest in the Distributed Constraint Satisfaction Problem (DisCSP) in the last few years. DisCSP has been widely viewed as a powerful paradigm for solving combinatorial problems arising in distributed, multi-agent environments. A DisCSP is a problem with finite number of variables, each of which has a finite and discrete set of possible values and a set of constraints over the variables. These variables and constraints are distributed among a set of autonomous and communicating agents. A solution in DisCSP is an instantiation of all variables such that all the constraints are satisfied. Fuzzy DisCSP is a DisCSP with soft constraints, i.e. different constraints have different important levels. In negotiation context, fuzzy constraints have been used widely to model service clients and providers’ preferences over attribute values of products. Fuzzy constraints are useful to describe the preferences or utility functions of Web service providers. This paper propose a Fuzzy DisCSP-based technique to solve the QoS composition problem. Our algorithm models providers’ preferences as fuzzy constraints and supports the notions of preference priorities, i.e. different providers may have different priorities of being satisfied. The rest of the paper is organized as follows. A literature review on Web service compositions and DisCSP are presented in the next section. In Section 3, we model the QoS composition problem as a DisCSP. Section 4 proposes an algorithm to construct constraints from a Web service composition structure. An enhancement of the Asynchronous Aggregate Search (AAS) algorithm, which can handle multiple variables in each agent with a new heuristic developed to exploit QoS parameters’ characteristics, is presented in Section 5. Section 6 discusses our initial experiment in building a prototypical system as a proof of concept for our proposal. Finally, conclusions and future work are discussed in Section 7.

II. RELATED WORK

There have been several studies on Web service compositions. A survey of various studies on functionality compositions can be found in [17]. QoS compositions are discussed in [9], [1], [7], [5]. In [12], numeric calculations of execution time and cost of a composite service are presented. In [7], the authors discuss a method for QoS aggregation based on Web service composition patterns [19]. More related work on QoS planning can be found in [9], in which the authors propose an approach for selecting optimal sub-providers from a list of service providers. While these research focus on functional and QoS compositions at the same time, they solve the composition problems at each provider independently. Consequently, the real complexity of composing Web services which lead to many nested compositions or have may dependencies is...
overlooked. Also, these works do not make use of possible collaborations between different providers in a composition. This collaboration is proven to be important as the new Web service choreography standard of WS-CDL [21] become more matured.

To overcome this limitation, we propose to use Fuzzy DisCSP techniques in which service providers from different nested levels in a composition can collaborate to solve the QoS composition problem together. To our best knowledge, there has been no application of DisCSP in this area so far. DisCSP is a major technique for coordination and conflict resolutions in a distributed and collaborative environment. A DisCSP is a constraint satisfaction problem in which the variables and constraints on these variables are distributed among independent but communicating agents. One important motivation behind the DisCSP paradigm is that it allows agents to keep their constraints privately while permits the solution to be solved globally. Many studies have been done in proposing algorithms for solving DisCSPs, such as Asynchronous Backtracking (ABT) [23], Asynchronous Weak-Commitment (AWC) search [22], and lately Asynchronous Distributed constraint Optimization with quality guarantees (ADOPT) [13], [2] to name a few. ABT is normally considered as the base algorithm for others.

Fuzzy DisCSP has lately captured the interest of MAS community in the context of negotiation. In [8], utility theory and fuzzy constraint-based are employed for the negotiation process. These works focus on bilateral negotiations and when many agents take part, a central coordinating agent may be required. The work in [10] is interesting as it promotes a framework for multi-lateral negotiation with a rotating central coordinating agent. However it does not take advantage of the current advancement of DisCSP. Also, even though there has been substantial research on different levels of importance of constraints in the Priority Fuzzy Constraint Satisfaction Problem (PFCSP) framework [11], little work has been done on priorities of agents in terms of satisfaction levels. In practice, different Web service providers may have different priorities for their satisfaction levels to be honored. For example, a composite Web service provider (main contractor) would naturally have higher priority than a component Web service providers (sub contractors). In [6], an iterative approach is used to solve a Fuzzy DisCSP through as a series of DisCSP is presented. This algorithm shares some similarity with our work. This algorithm is one among a few available algorithms for solving Fuzzy DisCSP. However the algorithm also does not address the satisfaction priorities. Because the algorithm can be extended to multivariables and n-ary constraints, we will use it to compare against our algorithm.

Util now, two popular problems in the DisCSP application domain are Meeting Scheduling and SensorCSP [3]. In the Meeting Scheduling problem, multiple agents negotiate to find a meeting time that can satisfy all agents’ personal constraints. In the SensorCSP there are a number of sensors modeled as agents and targets where a target is tracked if k sensors are tracking it at the same time. The sensors have to cooperate to track all the targets with their own constraints that a sensor can only track one target at a time. In this paper, we introduce another problem that can be modeled and solved with DisCSP - the problem of QoS composition. This problem is of a great interest to the agent community.

III. MODELING QOS COMPOSITION PROBLEM AS DISCSP

The DisCSP and Fuzzy DisCSP can be formally defined as follows.

Definition 1: A Distributed Constraint Satisfaction Problem $P = \langle X, D, C, N \rangle$ is a problem defined on a set of variables $X = \{X_1, ..., X_n\}$, a discrete finite domains for each of the variables $D = \{D_1, ..., D_m\}$, and a set of constraints $C = \{C_1, ..., C_m\}$ on possible values of variables. These variables and constraints are distributed among a set of agent $A = \{A_1, ..., A_k\}$. If an agent $A_t$ holds a constraint $C_q$, it also must hold all variables contained in $C_q$. A solution for $P$ is an assignment of variables in $X$ from their domains so that all constraints in $C$ are satisfied.

In some DisCSP framework like the one introduced with ABT [23], the definition goes further by assigning owners (agents) to variables. Proposals on a variable can be made only by the owner of that variable. Slightly different frameworks have been used in the literature [16] where agent owns problems and shares variables. In this paper, we use the second approach as it is more flexible to model and solve problems with multi variables.

Definition 2: A Fuzzy Distributed Constraint Satisfaction Problem $P^f$ is an extension of a DisCSP $P = \langle X, D, C, N \rangle$ in which a constraint $C^f_{ij}$, $j = 1, ..., m$ on a subset of the variables $\{X_{j1}, ..., X_{jk}\} \subseteq X$ represents a fuzzy relation defined on the Cartesian product space $D_{j1} \times ... \times D_{jk}$ characterized by a membership function $\mu_{C^f_{ij}}$. That is

$$\mu_{C^f_{ij}}: \prod_{p=1}^{k} D_{jp} \rightarrow [0,1], j=1, ..., m.$$ 

The global satisfaction degree of an assignment $x_S$ of all variable in $X$ is defined as:

$$sat(x_S) = \oplus\{ \mu_{C^f_{ij}}(x_S | C^f_{ij}) | C^f_{ij} \in C^f \}$$

in which $x_S | C^f_{ij}$ is the projection of $x_S$ into variable space of $C^f_{ij}$ or the part of assignment in $x_S$ for variables of $C^f_{ij}$. $\oplus$ is a T-norm operator and commonly used as the min function. In general, $x_S$ is a solution of $P^f$ if it maximizes the global satisfaction degree, or it has the global satisfaction degree greater than or equal to a predefined threshold value. For each assignment, we define the agent satisfaction level for an agent $A_n$ is the satisfaction degree of the assignment on the constraints known by $A_n$, i.e.

$$sat_{A_n}(x_S) = \oplus\{ \mu_{C^f_{ij}}(x_S | C^f_{ij}) | C^f_{ij} \in C^f, C^f_{ij} held by A_n \}$$

In the QoS composition problem, multiple service providers participate in the service composition process. Each provider has its own constraints on QoS levels of some parameters. The QoS levels may be continuous but in most cases the service providers and consumers are only interested in discrete values of the QoS levels; for example, discrete values of cost, disk space and response time. Some constraints of the providers
can be revealed, for example through different advertised classes of service and others must be completely private such as resource limitations, business rules, organizational policies and service composition structures. The goal of the QoS composition problem is to find a solution to all providers so that every providers’ constraints are satisfied. In addition, each provider has a different level of satisfaction on a solution. Therefore, it is desirable to maximize the global satisfaction level for all providers.

Essentially, a QoS service composition problem is a combinatorial problem and can be modeled in DisCSP framework. In this model each service provider can be considered as an agent (an autonomously processing entity) in a constraint network. Each QoS parameter is mapped into a variable in the constraint network; and the set of providers’ constraints is mapped into the network’s constraint set. The problem of maximizing the global satisfaction level is equivalent to the constraint network with fuzzified constraints.

To facilitate the discussion, we start by a simple example without preferences. Figure 1 shows an example of five agents A₁, A₂, A₃, A₄, A₅ together forming an Attraction Service composite service. Each agent has its own constraints on QoS.

In Figure 2, c(S) and t(S) define the total cost and response time of a composite service S while cᵢown and tᵢown define the cost and response time introduced by the agent Aᵢ itself. The c variable can take a price between $1 and $100 (USD) and the t variable can take an integer value between 1 and 100 (ms) for all services. In a more complex example, to negotiate on the values of cost and response time, A₃ and A₄ may need to negotiate on other variables as well. In other words, the set of variables at each agent may not represent the same set of QoS parameters. Also, the example in Figure 1 gives an impression of a tree structure. However, this is not always true as many consumers can use the same session of a Web service.

To show the differences between not applying and applying DisCSP techniques, we consider the two most commonly used approaches for QoS composition problems [9], [1], [7], [5]:

- In the first approach, QoS composition is solved incrementally. Firstly, A₁ negotiates with A₂ and A₃ for attraction finding and booking. A₂ agrees to provide a FindAttraction service with, for example, {1ms, $5} for {response time, cost}. A₃ agrees to provide its BookAttraction service with, for example, {4 ms, $10} because it is confident to find sub-providers to support these values. After that, A₃ contacts and negotiates with A₄ and A₅. If A₄ agrees on {1ms, $5} and A₅ agrees on {3ms, $5} for their services respectively, a solution is found. Otherwise, A₃ has to find a substitute of A₄ and A₅.
- In the second approach, synchronous backtracking is used to solve the QoS composition problem. A₁ proposes some cost and response time values to A₂ and A₃. For example, it proposes {1ms, $1} to A₂ and {4ms, $5} to A₃. A₂ accepts the proposal while A₃ subsequently negotiates with A₄ and A₅ before responding to A₁. In the negotiation, if A₄ and A₅ cannot satisfy with any proposals from A₃; A₃ backtracks to A₁ with a refusal on {4ms, 5 USD} values. A₁ then has to negotiate for another value with A₂ and A₃. The above steps are repeated until A₄ and A₅ can agree on some proposals from A₃.

![Fig. 1. A scenario of a composite Attraction Service which has a nested composition structure with the Book Attraction service offered by A₃.](image)

![Fig. 2. Illustration of different constraints held by each agent](image)

It can be seen that both approaches are synchronous, i.e. each agent has to wait for responses from other agents before it can proceed. Also, each agent is only aware of its local composition problem which is formed whenever a request from a client arrives. For example, A₃ is only aware of a composition problem when it receives a proposal from A₁. To address this limitation, our approach allows every agent to be aware of the global QoS requirement, and encourages them to take part in the solving process asynchronously as soon as possible after a conceptual functional composition is formed.
and before and contracts are signed. In the functionality composition phase, when an agent receives a problem id from its consumers, it generates a new problem id if necessary, and then forwards this to its providers to inform them that there is a global QoS problem which needs to be solved. In particular, a problem id received in combination with the sender address creates a unique context to identify a particular global QoS problem.

IV. CONSTRAINT DEDUCTION FROM COMPOSITION STRUCTURE

As stated in the previous part, constraints can be formed from different items such as business rules, organizational policies, or composition structures. Translations from business rules and organizational policies into QoS constraints may vary from organizations to another since these rules and policies can be represented and interpreted differently. In this paper, we use the term “QoS constraint formulation” to describe the process of finding these constraints/relationships from a composition input. The composition is presented in BPEL4WS (the Business Process Execution Language for Web services) format. BPEL4WS is selected because it is a popular Web service composition language supported by different Web service vendors. Our approach bases on important work in [19], [18], [7]. In [19] and [18], the authors conduct a survey and collect a set of workflow patterns which have been used in workflow languages today, including Web service composition languages. In [7] the authors map these workflow patterns into QoS composition patterns. We combine these results to form a mapping from BPEL4WS into QoS composition patterns as shown in Table 1. It is worth noting that Table 1 only presents composition patterns supported by BPEL4WS. Our QoS constraint formulation process consists of the following three main steps:

1) Construct constraint formulas for QoS composition patterns: Sequence, AND (Split/Join), XOR (Split/Join) and OR (Split/Join).
2) Iteratively decompose the composite structure through BPEL workflow patterns into smaller sub-compositions until this cannot be done any further.
3) Iteratively re-construct the constraints in a composition from constraints of its sub-compositions. The re-construct uses formulas available in step 1 above.

In reference to other efforts on QoS aggregation [7], [9], [5], [12], our constraints can be considered as means to compute the QoS aggregations. The difference is that we construct the formulas to aggregate QoS instead of compute a value. Formulas for simple composition patterns in step 1 with different QoS attributes can be found in [7]. Table 2 presents a partial list of these. It is important to know that our algorithm does not restrict to the formulas presented in [7].

<table>
<thead>
<tr>
<th>No.</th>
<th>Workflow Pattern (BPEL pattern)</th>
<th>QoS Composition Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sequence (Sequence)</td>
<td>Sequence</td>
</tr>
<tr>
<td>2</td>
<td>Parallel Split (Flow/Link)</td>
<td>AND Split</td>
</tr>
<tr>
<td>3</td>
<td>Synchronization (Flow/Link)</td>
<td>AND Join</td>
</tr>
<tr>
<td>4</td>
<td>Exclusive Choice (Switch/Link)</td>
<td>XOR Split</td>
</tr>
<tr>
<td>5</td>
<td>Simple Merge (Switch/Link)</td>
<td>XOR Join</td>
</tr>
<tr>
<td>6</td>
<td>Multi-choice (Link)</td>
<td>OR Split</td>
</tr>
<tr>
<td>7</td>
<td>Implicit Termination (By Default)</td>
<td>Structural Patterns</td>
</tr>
<tr>
<td>8</td>
<td>M.I. Without Synchronization(Flow)</td>
<td>AND Split and AND Join</td>
</tr>
<tr>
<td>9</td>
<td>Deferred Choice(Pick)</td>
<td>XOR Split</td>
</tr>
<tr>
<td>10</td>
<td>Interleaved Parallel (Serializable Scope)</td>
<td>Sequence</td>
</tr>
<tr>
<td>11</td>
<td>Cancel Activity (Terminate)</td>
<td></td>
</tr>
</tbody>
</table>

TABLE I
WORKFLOW PATTERNS TO QoS COMPOSITION PATTERNS

<table>
<thead>
<tr>
<th>No.</th>
<th>Workflow Pattern (BPEL pattern)</th>
<th>QoS Composition Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sequence (Sequence)</td>
<td>Sequence</td>
</tr>
<tr>
<td>2</td>
<td>Parallel Split (Flow/Link)</td>
<td>AND Split</td>
</tr>
<tr>
<td>3</td>
<td>Synchronization (Flow/Link)</td>
<td>AND Join</td>
</tr>
<tr>
<td>4</td>
<td>Exclusive Choice (Switch/Link)</td>
<td>XOR Split</td>
</tr>
<tr>
<td>5</td>
<td>Simple Merge (Switch/Link)</td>
<td>XOR Join</td>
</tr>
<tr>
<td>6</td>
<td>Multi-choice (Link)</td>
<td>OR Split</td>
</tr>
<tr>
<td>7</td>
<td>Implicit Termination (By Default)</td>
<td>Structural Patterns</td>
</tr>
<tr>
<td>8</td>
<td>M.I. Without Synchronization(Flow)</td>
<td>AND Split and AND Join</td>
</tr>
<tr>
<td>9</td>
<td>Deferred Choice(Pick)</td>
<td>XOR Split</td>
</tr>
<tr>
<td>10</td>
<td>Interleaved Parallel (Serializable Scope)</td>
<td>Sequence</td>
</tr>
<tr>
<td>11</td>
<td>Cancel Activity (Terminate)</td>
<td></td>
</tr>
</tbody>
</table>

TABLE II
FORMULAS OF CONSTRAINTS FOR SIMPLE COMPOSITION PATTERNS OF SEQUENCE, AND AND OR

<table>
<thead>
<tr>
<th>No.</th>
<th>Workflow Pattern (BPEL pattern)</th>
<th>QoS Composition Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sequence (Sequence)</td>
<td>Sequence</td>
</tr>
<tr>
<td>2</td>
<td>Parallel Split (Flow/Link)</td>
<td>AND Split</td>
</tr>
<tr>
<td>3</td>
<td>Synchronization (Flow/Link)</td>
<td>AND Join</td>
</tr>
<tr>
<td>4</td>
<td>Exclusive Choice (Switch/Link)</td>
<td>XOR Split</td>
</tr>
<tr>
<td>5</td>
<td>Simple Merge (Switch/Link)</td>
<td>XOR Join</td>
</tr>
<tr>
<td>6</td>
<td>Multi-choice (Link)</td>
<td>OR Split</td>
</tr>
<tr>
<td>7</td>
<td>Implicit Termination (By Default)</td>
<td>Structural Patterns</td>
</tr>
<tr>
<td>8</td>
<td>M.I. Without Synchronization(Flow)</td>
<td>AND Split and AND Join</td>
</tr>
<tr>
<td>9</td>
<td>Deferred Choice(Pick)</td>
<td>XOR Split</td>
</tr>
<tr>
<td>10</td>
<td>Interleaved Parallel (Serializable Scope)</td>
<td>Sequence</td>
</tr>
<tr>
<td>11</td>
<td>Cancel Activity (Terminate)</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3. Constraint formation algorithm
These formulas can be customized to suit particular scenarios and QoS characteristics.

For a more detailed explanation, the pseudo-code for our constraint formulation process is listed in Figure 3. In our algorithm, the BPEL4WS composition is represented as a ProcessDef object which has a base Activity object. The Activity class represents all possible activities available in BPEL4WS processes such as receive”, “reply”, “invoke”, “sequence” and “switch”. Hence, an Activity object is composed of other Activity objects. XPath class represents an XPATH expression. The main idea behind the Construct-Constraint algorithm in Figure 3 is to iteratively decompose a BPEL activity and apply equivalent QoS composition formulas (table 2) at each iteration. New variables are added whenever “invoke” and “receive” activities are encountered. In the line 10, F is a constraint function (constructed from Table 2) for QoS composition patterns. For an example, a composite service with a structure listed in Figure IV after passing through the Construct-Constraint function will produce (t(Shipping)= max( t(ShipmentAir), t(ShipmentWater), t(ShipmentLand))) for the input of qos-name as response time.

V. SOLVING THE QoS COMPOSITION PROBLEM WITH DISCSP ALGORITHMS

A. The ADE algorithm

There have recently been many publications on DisCSP algorithms. Traditionally these algorithms are developed and demonstrated in the context of the Meeting Scheduling and Sensor Network problems as discussed at the beginning of this paper. Often the techniques used in these algorithms combine tree-search with backtracking, look-ahead, and back-jumping. However, there are some characteristics that make the QoS composition problem different from the Meeting Scheduling and Sensor Network problems:

- Each agent holds more than one variable.
- Each agent holds a set of interested QoS parameters and the variables that represent these parameters.
- Local constraints in QoS problem can be very complex.
- Provider agents are not willing to reveal their consumer agents’ addresses to others.

Fig. 4. An example of BPEL activity

```xml
<flow name= "Shipping">
  <sequence>
    <invoke name= "ShipmentAir"/>
  </sequence>
  <sequence>
    <invoke name= "ShipmentWater"/>
  </sequence>
  <sequence>
    <invoke name= "ShipmentLand"/>
  </sequence>
</flow>
```

Fig. 4. An example of BPEL activity

Fig. 5. AAS ok?, nogood and channel processing

- Agents may want to hide private information from others as much as possible.

In searching for a suitable DisCSP algorithm, those above characteristics are the most important criteria for us. Whilst most algorithms such as ADOPT and IDIBT can be extended so that one agent can hold more than one variable, substantial effort is required for this and for handling complex private constraints. In addition, the back-jumping technique in some algorithms require undesirable revelations of agent consumers’ addresses to others. Asynchronous Aggregate Search (AAS) [16] is a good candidate since it allows one agent to maintain a set of variables and these variables can be shared. AAS differs from most of existing methods in that it exchanges aggregated consistent values (in contrast to a single value in ABT) of partial solutions. This reduces the number of backtracks. However, private information is revealed more in AAS than in ABT because of the aggregation. Using trusted servers where critical information of organizations is hosted to prevent privacy loss is not very practical for Web services. In this work, we rate privacy as the most important issue. Also, our goal is to develop a simple but effective algorithm in which agents can exploit special characteristics of the QoS composition problem. As such, we propose an extension of AAS with three following enhancements and modifications:

- Incorporating local CSP solvers into agents to solve complex private constraints.
- Making use of common characteristics of QoS parameters to speed up the solving process.
- Reducing the aggregation in AAS into one value per variable to reduce privacy loss.

Similar to AAS, in our algorithm called constraint satisfaction Algorithm for Distributed Environment (ADE), agents are assigned with priorities so that they can be arranged in order. A composite Web service provider always has higher

```xml
process-ok(<x_j,s_j,h_j>) do
1. if(history(x_j) invalidates h_j) return;
2. add <x_j,s_j,h_j> to agent-view
3. update nogood list store
4. check-agent-view
```

```xml
process-nogood(A_j,~N) do
1. update agent-view for any new assignments in ~N
2. send setup channel request to A_j
3. if ~N is invalid return;
4. insert ~N into the nogood list
5. update nogood list store
6. check-agent-view
```

```xml
process-channel(<A_j, x_j, pa_j>) do
1. setup relay-channel for A_j
```

Authorized licensed use limited to: SWINBURNE UNIV OF TECHNOLOGY. Downloaded on March 15,2010 at 05:46:27 EDT from IEEE Xplore. Restrictions apply.
priority than the providers of its components. Every agent proceeds with a similar process. Information about the outside world learnt by each agent is stored in an agent view and a nogood list. An agent view is a set of values that the agent believes to be assigned to the variables belonging to the higher priority agents. Agents exchange assignments and nogoods. An assignment has the form $\{\bigwedge_{i=1}^{m}(x_{i}=a_{i}), h_{j}, p_{k}\}$ which indicates that the variable $x_{i}$ is assigned a value $a_{i}$, $h_{j}$ is the message history [16], $p_{k}$ is the pseudo-agent address and will be explained later. The assignment is an AAS aggregation of single value. A nogood list holds the assignments of values to the variables during the solving process which cause inconsistency. ‘ok?’ or ‘nogood’ messages are used as in AAS. The ‘ok?’ message is used to inform the lower priority agents of new value assignments. A nogood is used to backtrack the assignment that causes inconsistency between constraints. In Figure 6, we use $A^{+}$ to denote the set of agents that are linked to $A$ and have priorities higher than the priority of $A$. Similarly, $A^{-}$ is denoted as the set of agents that are linked to $A$ and have priorities lower than the priority of $A$. $V^{+}$ is the set of variables the agent share with $A^{+}$, and $V^{-}$ is with $A^{-}$. As illustrated in Figure 6, when an agent receives an ‘ok?’ message, it validates the message and adds the assignment in the message into its local view. It then checks the local view consistency. To handle the complexity of local constraints, we use a local CSP solver inside each agent. In the check-local-view procedure, first the agent updates any new assignment. It detects if there is any unknown variable in the assignment. If there is such a variable then the agent asks the message sender to setup a relay channel so that it can communicate to the agents who hold this variable. In addition, it generates a pseudo-agent address for each of those agents. These pseudo-agent addresses will be mapped to the actual agent addresses at the sender. In essence, the relay channel is equivalent to the add-link mechanism in ABT. However it protects an agent from revealing its address to unknown agents by relaying all messages through a known agent. In the check-local-view, if the agent view is inconsistent then the agent’s local CSP solver is invoked to find new value assignments for this agent’s variables. If a solution is found, the agent sends new ‘ok?’ messages with the assignments to the lower priority agents; otherwise, it sends a nogood message back to the ‘ok?’ message sender. The detailed description of the extended AAS algorithm and performance analysis is a subject of another paper (under review).

In the context of Web service composition, QoS parameters have particular characteristics that normal constraint variables have not. Agents (service providers) often have had their constraint preference as a monotonic function over a QoS value. In other words, there is a preference operator $\preceq$ defined between any two values $a^{(1)}$ and $a^{(2)}$ in the domain of variable $x$. If $a^{(1)} \preceq a^{(2)}$ then we say that the agent prefers $a^{(2)}$ to $a^{(1)}$ for $x$. In addition, this operator can also be defined in the aggregation of assignments as following: $\bigwedge_{i=1}^{m}(x_{i}=a_{i}) \preceq \bigwedge_{i=1}^{m}(x_{i}=a_{i}^{2})$ if $\forall i \in \{1..m\} a_{i}^{1} \preceq a_{i}^{2}$. Note that this relation defines only on a subset of the Cartesians products of aggregations. The preference for A2 in Figure 2, for example, can be that $t(FindAttraction)=1 \land t(FindAttraction)=5 \preceq 3$. While the ADE algorithm follows the principles of AAS, the local CSP solver uses a new heuristic for selecting new values. When a new assignment is generated a heuristic criterion could be to choose the value with not less preference (if it is comparable to the previous rejected assignments) to the lower priority agent. This preference is further exploited in our Fuzzy DisCSP algorithm in the next section.

B. FADE algorithm

Our Fuzzy constraint satisfaction Algorithm for Distributed Environment (FADE) algorithm is an extension of ADE for Web service providers with preferences. As pointed out in [8], a FCSP can be modelled as a collection of crisp CSPs at different levels of constraint satisfaction. Using the well-known principle of the resolution identity [20,21] in fuzzy set theory each fuzzy constraint $C^{f}_{j}$ in definition 2 can be decomposed into a union of non-fuzzy (crisp) constraints that are its $\alpha$-cuts as follows:

$C^{f}_{j}(x) = \sum_{\alpha} C^{f}_{j,\alpha}(x)$

in which $C^{f}_{j,\alpha}(x)$ is an $\alpha$-cut of $C^{f}_{j}$ defined as $C^{\alpha}(x) = \{x \mid c_{j}(x) \geq \alpha, 0 \leq \alpha \leq 1\}$. The solution of the Fuzzy DisCSP in definition 2 is to find a solution at the highest $\alpha$-cut level.

**check-agent-view do**

1. when agent-view and current-aggregate are inconsistent
2. $V =$localCSP.solve(agent-view, nogood-list, local-constraints)
3. if $V$ is null
4. backtrack
5. else
6. reset current-aggregate
7. for $\forall a \in V$
   if a is new for $A^{+}_{k}$
9. append new history and send ok message to $A^{+}_{k}$
10. current-aggregate = current-aggregate $\cap$ a
11. else
12. if a is needed
13. current-aggregate = current-aggregate $\cap$ a

*Fig. 6. AAS check-agent-view*
This hints us to start searching from the highest $\alpha$-cut level and move down rather than solving all DisCSPs. This is also the principle of our algorithm.

FADE models the Fuzzy DisCSP as a set of crisp DisCSPs at different $\alpha$-cut levels. It solves a crisp DisCSP by using the ADE algorithm. The solving process starts by having on agents agree on a set of $\alpha$-cut levels. At each agent, the search begins from the highest level. If a solution in this level can not be found, the search then concedes to the next $\alpha$-cut level. It is important to know that at the beginning every agent start with their highest $\alpha$-cut levels of 1.0. However during the search, these $\alpha$-cut levels can become different. Figure 8 illustrates this. Assuming that the satisfaction priorities of agents in the figure decrease from left to right. The figure depicts a search running on the shaded areas: at the $\alpha$-cut level of 0.75 for the left-hand side agent, 0.5 for the middle agent, and 0.25 for the right-hand side agent. Agents with higher satisfaction priority have higher $\alpha$-cut level at any instance of time if possible. In more details, the solving process of Fuzzy DisCSP happens in three phases: $\alpha$-cut level forwarding, DisCSP solving, and level conceding. The second and third phases can be iterative. In the level forwarding phase, every agent forward their $\alpha$-cut levels of preference to others and together they assemble a global set of $\alpha$-cut levels. In the DisCSP solving phase, the agents use asynchronous backtracking technique (ADE algorithm) to solve the crisp DisCSP at assigned $\alpha$-cut levels. The third phase is triggered by the second phase when no solution can be found. Agents which causes no solution, starting from the ones with lowest satisfaction priority, are asked to move to the next $\alpha$-cut level. The second and third phases are iterated until the time that a solution is found.

In FADE, in addition to the nogood content in a nogood message. The sender also creates the explanation for this nogood and sends along with the nogood message. The explanation is created as follows. An agent during the search for a new solution in check-local-view procedure, it takes the union of violated constraints. If the union set is empty, then all $\alpha$-cut constraints in this set forms the explanation. Otherwise, the set of $\alpha$-cut constraints in every recorded nogoods plus the current $\alpha$-cut constraint make up the explanation. The receiver of the nogood message records this exaption in association with the nogood and remove any nogood with have the $\alpha$-cut level greater than the current searching $\alpha$-cut level. These remove possible duplicated searches at different $\alpha$-cut levels and hence increase the efficiency of the overall search.

VI. IMPLEMENTATION

In this part, we describe our initial agent-based implementation of a framework for DisCSP with our toolkit J4WSM (Jade for Web Services Management) to support our proposal. J4WSM is a re-factor of our previous toolkit called WS2JADE that integrates Jade agents with Web services [14], towards WS management. J4WSM, which consists of a Jade agent platform and other utilities including WS4JADE, runs as a service inside a J2EE container. In this initial version, J4WSM runs under JBoss. J4WSM is similar to Blue-Jade [4] but targets Web Services Management particularly. J4WSM is designed to use services instead of making API calls.

A. FADE algorithm

The main components in J4WSM consists of three main modules: constraint formations, local CSP solver and DisCSP protocols. The constraint formation module is to construct and keep the constraints available at each provider. Constraints in J4WSM are presented in XPath expressions. At this version, constraints are created from composite service structures (in BPEL format) by using the algorithm specified in Section 3. In J4WSM, we do not specify any particular CSP solver engines. Different CSP solver engines can be used as long as there is an adapter which can translate XPath expressions into the CSP solver languages. For our experiment, we use NSolver [15] for the local CSP solver. Whenever a provider needs to solve the local constraints, it invokes NSolver to
find a solution. We have developed an adapter to translate XPath expression into NSolver on the fly. This adapter can be downloaded from [20]. In this version, J4WSM only supports one Fuzzy DisCSP protocol which corresponds to the FADE specified in the previous part. This protocol is developed as an interaction protocol in Jade. Each provider in the DisCSP is represented by a Jade agent. These agents must understand a protocol and agree to use it before the solving process can start. Different components in J4WSM are presented in figure 8. Figure 9 shows the improvement of FADE over the iterative DisCSP algorithm in [6]. The graph shows the average improvement of 10 trials for each value in the horizontal axis. The improvement is measured in term of processing cycles which has been popular used to compare performance between different DisCSP algorithms. For each agent, one cycle consists of reading all incoming messages, invoking its CSP solver to find a solution and sending messages [23]. It can be seen that the reduction of cycle increases with the number of agents and reaches around 90% for 15 agents.

The structure of the composition in this experiment is built incrementally by adding one agent each time from the first agent A1 until a balanced binary tree with depth = 4 (i.e there are 15 agents in total) is formed. The experiment is carried out for two parameters: cost and response time. Each agent A_k has a business rule: c(A_k) \geq \max_{i \in q_k} \tau_i - t(A_k) + C_T \in [1..2]. The constraints deduced from the structure of the own composite service: t(A_k) = t_{own}(A_k) + \sum_{\forall A_i \in A_k} t(A_i), c(A_k) = c_{own}(A_k) + \sum_{\forall A_i \in A_k} c(A_i) in which t(A_k) and c(A_k) are response time and cost of the composite service at A_k, t_{own}(A_k) and c_{own}(A_k) are response time and cost introduced by A_k’s own service respectively. In the experiment, the domain of the cost variable for an agent with depth=1 is [1.25\ldots1] and the domain of this agent’s time response variable is also [1.25\ldots1]. T_0=4ms and C_T=1S. Each provider has a reference function described by a fuzzy membership function with four crisp numbers at the levels of 1.0, 0.75, 0.5 and 0.25. The purpose of this experiment setup is to mimic a real scenario of Web service composition and demonstrate the applicability of DisCSP technique into the QoS composition problem. Figure 10 shows a debug trace for 4 agents with J4WSM in which the agent performative PROPOSE is used for ok? and REJECT-PROPOSAL is for nogood.

VII. CONCLUSIONS

This paper proposes a new approach of Fuzzy DisCSP application into the QoS composition problem. An algorithm to construct constraints from the composition topology is proposed. We also develop an enhanced version of AAS for multiple variables with a new heuristic for distributed agents to exploit QoS parameters’ characteristics. In addition, the iterative algorithm to solve the QoS composition problems with different Web service providers’ preferences - FADE is described. Our future work will concentrate on a framework which allows agents to exchange not only assignments but also possible constraints during the solving process, as a part of J4WSM toolkit. We believe that this is beneficial for the QoS composition problem and useful in developing new Fuzzy DisCSP algorithms for the solving process.

REFERENCES