24.02.2014

Universal Computational Cluster

Overall system activities, communication protocol and plugin design

Universal Computational Cluster (UCC) is meant for computing NP-hard problems by distributive exact algorithms. This document presents overall system activities on UML Activity diagrams, specifies the communication protocol by showing sequence diagrams and giving the XML Schema files for all the types of messages used by the system and as an appendix gives the crucial abstract class libraries for the Task Solvers.

1. System activities

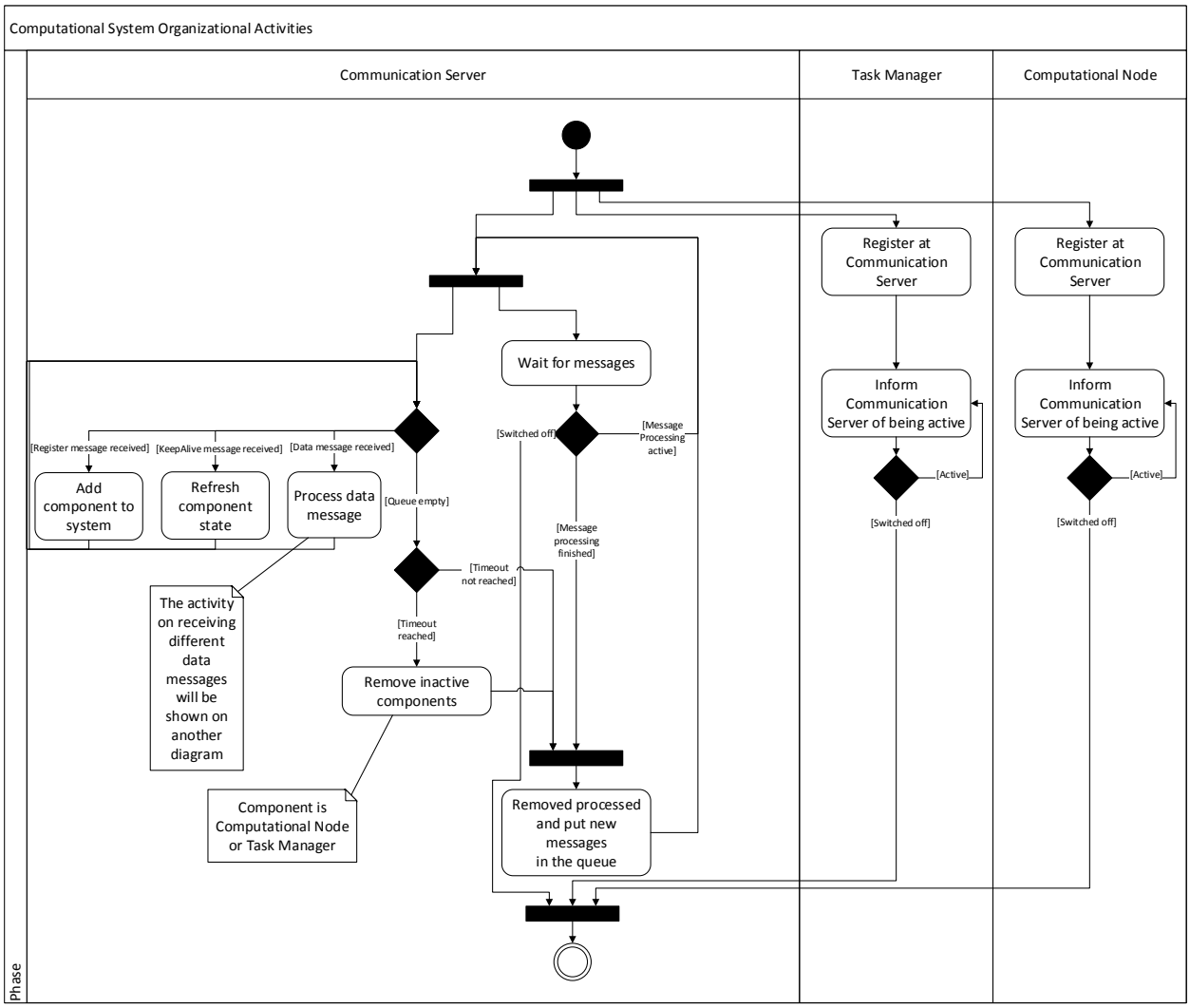

Figure 1 Overall system activities concerning the setup of the cluster

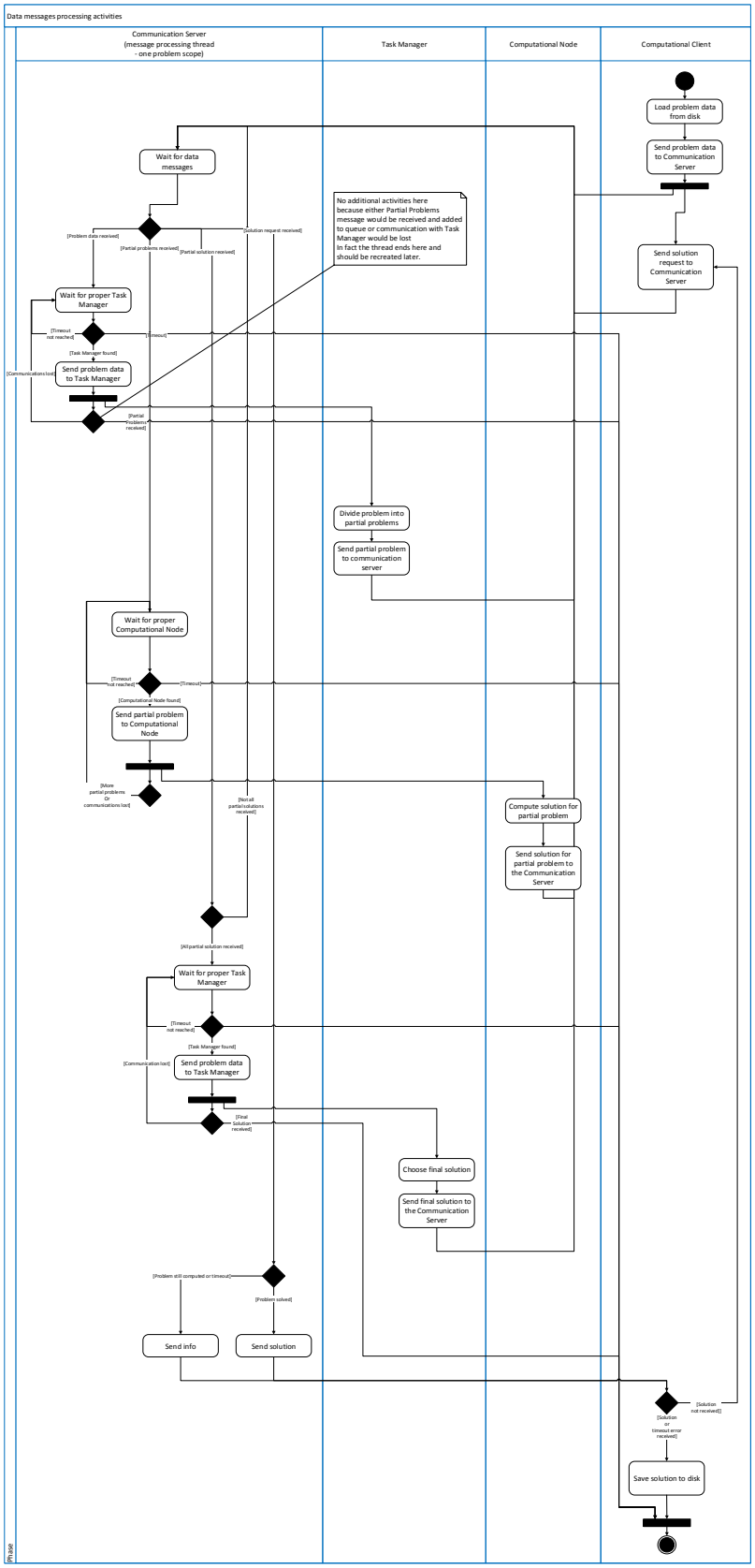

Figure 2 Additional system activities while performing computations

Fig. 1 presents how the components connect with each other and what activities are maintained even in idle state of the cluster. Task Manager (TM) and Computational Node (CN) register to the Communication Server (CS) and constantly inform CS of their state (thus maintaining the TCP/IP connection open).

Fig. 2 shows all the additional activities of the system after Computational Client (CC) connects to it and requests computations for a given problem to be performed. Both figures represent only one instance of each of the CN, TM and CC but possibly there could (and should) be many of them.

The important thing is that CS is always awaiting new messages (which are read almost instantly), regardless of the state of various problems send for computation. The processing of each of the message could be done separately and in parallel by different threads and the current state of problems is stored in the memory and not as a state of the system.

2. Communication protocol design

The system uses the TCP/IP protocol in order to send messages. Only the CS has a listens on a given port and all the other components start with connecting to CS. All the messages are a text XML messages.

Fig. 3 presents a scenario with two TM, two CN and one CC. The scenario shows that each of the TM could be chosen for dividing the problem/merging the solutions and presents what happened when one of the CN crashes.

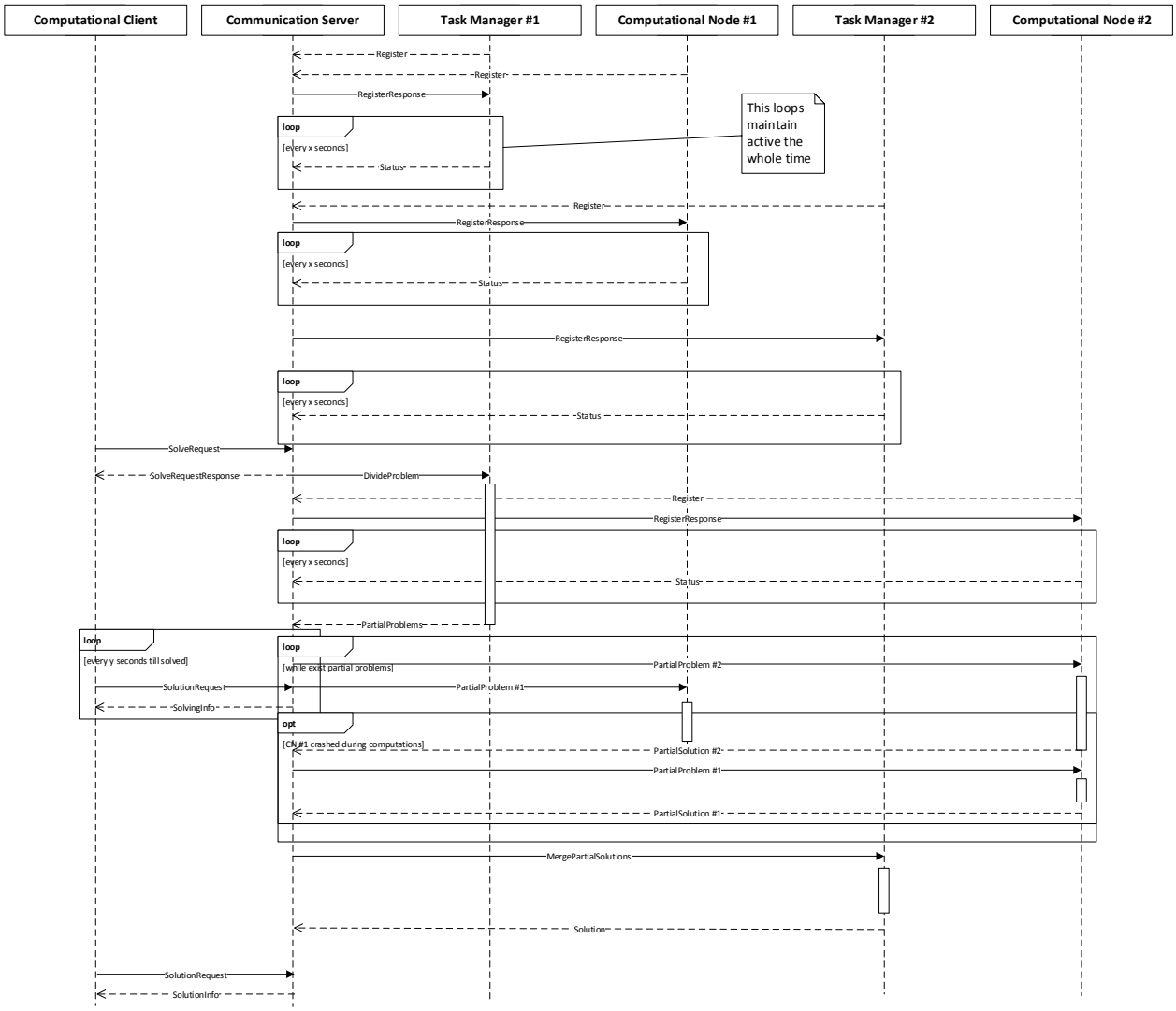

Figure 3 Sequence diagram presenting communication protocol messages

Register message

Register message is sent by TM and CN to the CS after they are activated. In the register message they send their type `TaskManager` or `ComputationalNode`, the type of problems they could solve and the computational power of the component (note that the protocol would support both the registration of many components from the same computer with one thread and registration of one component with many threads).

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified"
targetNamespace="http://www.mini.pw.edu.pl/ucc/"
xmlns:xs="http://www.w3.org/2001/XMLSchema">
<xs:element name="Register">
    <xs:complexType>
      <xs:sequence>
        <!-- defines the type of node (either TM or CN) -->
        <xs:element name="Type">
          <xs:simpleType>
            <xs:restriction base="xs:string">
              <xs:enumeration value="TaskManager" />
              <xs:enumeration value="ComputationalNode" />
            </xs:restriction>
          </xs:simpleType>
        </xs:element>
       <!-- gives the list of names of the problems which could be solved (probably sth
       like DVRP-[group no.]) -->
        <xs:element name="SolvableProblems">
          <xs:complexType>
            <xs:sequence>
              <xs:element maxOccurs="unbounded" name="ProblemName" type="xs:string" />
            </xs:sequence>
          </xs:complexType>
        </xs:element>
       <!-- the number of threads that could be efficiently run in parallel -->
       <xs:element name="ParallelThreads" type="xs:unsignedByte" />
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Register Response message

Register Response message is sent as a response to the Register message giving back the component its unique ID and informing how often it should sent the Status message.

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified"
targetNamespace="http://www.mini.pw.edu.pl/ucc/"
xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="RegisterResponse">
    <xs:complexType>
      <xs:sequence>
       <!-- the ID assigned by the Communication Server -->
       <xs:element name="Id" type="xs:unsignedLong" />
       <!-- the communication timeout configured on Communication Server -->
       <xs:element name="Timeout" type="xs:time" />
      </xs:sequence>
```

```
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Status message

Status message is sent by TM and CN to CS at least as frequent as a timeout given in Register Response. In the Status message the component reports the state of its threads, what they are computing (unique problem instance id, task id within the given problem instance, the type of problem instance) and for how long.

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified"
targetNamespace=http://www.mini.pw.edu.pl/ucc/
xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="Status">
    <xs:complexType>
      <xs:sequence>
       <!-- the ID of node (the one assigned by server) -->
       <xs:element name="Id" type="xs:unsignedLong" />
       <!-- list of statuses for different threads -->
       <xs:element name="Threads">
          <xs:complexType>
            <xs:sequence>
              <xs:element maxOccurs="unbounded" name="Thread">
                <xs:complexType>
                  <xs:sequence>
                     <!-- information if the tread is currently computing something  -->
                    <xs:element name="State">
                      <xs:simpleType>
                        <xs:restriction base="xs:string">
                          <xs:enumeration value="Idle" />
                          <xs:enumeration value="Busy" />
                        </xs:restriction>
                      </xs:simpleType>
                    </xs:element>
                    <!-- how long it is in given state (in ms) -->
                     <xs:element name="HowLong" type="xs:unsignedLong" />
                    <!-- the ID of the problem assigned when client connected -->
                    <xs:element minOccurs="0" name="ProblemInstanceId"
                    type="xs:unsignedLong" />
                    <!-- the ID of the task within given problem instance -->
                    <xs:element minOccurs="0" name="TaskId" type="xs:unsignedLong" />
                    <!-- the name of the type as given by TaskSolver -->
                    <xs:element minOccurs="0" name="ProblemType" type="xs:string" />
                  </xs:sequence>
                </xs:complexType>
              </xs:element>
            </xs:sequence>
          </xs:complexType>
        </xs:element>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Solve Request message

Solve Request message is sent by the CC to CS. It gives the type of the problem instance to be solved, optionally the max time that could be used for computations and the problem data in base64.

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified"
targetNamespace=http://www.mini.pw.edu.pl/ucc/
xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="SolveRequest">
    <xs:complexType>
      <xs:sequence>
       <!-- the name of the type as given by TaskSolver -->
        <xs:element name="ProblemType" type="xs:string" />
       <!-- the optional time restriction for solving the problem (in ms) -->
        <xs:element minOccurs="0" name="SolvingTimeout" type="xs:unsignedLong" />
       <!-- the serialized problem data -->
       <xs:element name="Data" type="xs:base64Binary" />
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Solve Request Response message

Solve Request Response message is sent by CS to CC as an answer for the Solve Request. It provides CC with unique identifier of the problem instance.

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified"
targetNamespace=http://www.mini.pw.edu.pl/ucc/
xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="SolveRequestResponse">
    <xs:complexType>
      <xs:sequence>
       <!-- the ID of the problem instance assigned by the server -->
        <xs:element name="Id" type="xs:unsignedLong" />
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Divide Problem message

Divide Problem is sent to TM to start the action of dividing the problem instance to smaller tasks. TM is provided with information about the computational power of the cluster in terms of total number of available threads.

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified"
targetNamespace=http://www.mini.pw.edu.pl/ucc/
xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="DivideProblem">
    <xs:complexType>
      <xs:sequence>
        <!-- the problem type name as given by TaskSolver and Client -->
        <xs:element name="ProblemType" type="xs:string" />
```

```
        <!-- the ID of the problem instance assigned by the server -->
        <xs:element name="Id" type="xs:unsignedLong" />
        <!-- the problem data -->
        <xs:element name="Data" type="xs:base64Binary" />
        <!—the total number of currently available threads  -->
        <xs:element name="ComputationalNodes" type="xs:unsignedLong" />
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Solution Request message

Solution Request message is sent from the CC in order to check whether the cluster has successfully computed the solution. It allows CC to be shut down and disconnected from server during computations.

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified"
targetNamespace=http://www.mini.pw.edu.pl/ucc/
xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="SolutionRequest">
    <xs:complexType>
      <xs:sequence>
        <!-- the ID of the problem instance assigned by the server -->
        <xs:element name="Id" type="xs:unsignedLong" />
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Partial Problems message

Partial problems message is sent by the TM after dividing the problem into smaller partial problems. The data in it consists of two parts – common for all partial problems and specific for the given task. The same Partial Problems schema is used for the messages sent to be computed by the CN.

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified"
targetNamespace=http://www.mini.pw.edu.pl/ucc/
xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="SolvePartialProblems">
    <xs:complexType>
      <xs:sequence>
        <!-- the problem type name as given by TaskSolver and Client -->
        <xs:element name="ProblemType" type="xs:string" />
        <!-- the ID of the problem instance assigned by the server -->
        <xs:element name="Id" type="xs:unsignedLong" />
        <!-- the data to be sent to all Computational Nodes -->
        <xs:element name="CommonData" type="xs:base64Binary" />
        <!-- optional time limit – set by Client (in ms) -->
       <xs:element minOccurs="0" name="SolvingTimeout" type="xs:unsignedLong" />
       <xs:element name="PartialProblems">
          <xs:complexType>
            <xs:sequence>
              <xs:element maxOccurs="unbounded" name="PartialProblem">
                <xs:complexType>
                  <xs:sequence>
```

```
                    <!-- Id of subproblem given by TaskManager -->
                    <xs:element name="TaskId" type="xs:unsignedLong" />
                    <!-- Data specific for the given subproblem -->
                    <xs:element name="Data" type="xs:base64Binary" />
                  </xs:sequence>
                </xs:complexType>
              </xs:element>
            </xs:sequence>
          </xs:complexType>
        </xs:element>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Solutions message

Solutions message is used for sending info about ongoing computations, partial and final solutions from CN, to TM and to CC. In addition to sending task and solution data it also gives information about the time it took to compute the solution and whether computations were stopped due to timeout.

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified"
targetNamespace="http://www.mini.pw.edu.pl/ucc/"
xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="Solutions">
    <xs:complexType>
      <xs:sequence>
        <!-- the problem type name as given by TaskSolver and Client -->
        <xs:element name="ProblemType" type="xs:string" />
        <!-- the ID of the problem instance assigned by the server -->
        <xs:element name="Id" type="xs:unsignedLong" />
        <!-- common data which was previously sent to all Computational Nodes (possibly
could be stored on server as TaskManagers could have changed) -->
        <xs:element minOccurs="0" name="CommonData" type="xs:base64Binary" />
        <xs:element name="Solutions">
          <xs:complexType>
            <xs:sequence>
              <xs:element maxOccurs="unbounded" name="Solution">
                <xs:complexType>
                  <xs:sequence>
                    <!-- Id of subproblem given by TaskManager – no TaskId fo
              final/merged solution -->
                    <xs:element minOccurs="0" name="TaskId" type="xs:unsignedLong" />
                    <!-- Indicator that the computations ended because of timeout -->
                    <xs:element name="TimeoutOccured" type="xs:boolean" />
                    <!-- Information about the status of result (Partial/Final) or status
              of computations (Ongoing) -->
                    <xs:element name="Type">
                      <xs:simpleType>
                        <xs:restriction base="xs:string">
                          <xs:enumeration value="Ongoing" />
                          <xs:enumeration value="Partial" />
                          <xs:enumeration value="Final" />
                        </xs:restriction>
                      </xs:simpleType>
                    </xs:element>
```

```
                <!-- Total amount of time used by all threads in system for computing
            the solution / during the ongoing computations (in ms) -->
                <xs:element name="ComputationsTime" type="xs:unsignedLong" />
                <!-- Solution data -->
                <xs:element name="Data" minOccurs="0" type="xs:base64Binary" />
              </xs:sequence>
            </xs:complexType>
          </xs:element>
        </xs:sequence>
      </xs:complexType>
    </xs:element>
  </xs:sequence>
 </xs:complexType>
</xs:element>
</xs:schema>
```

3. Configuration

For whole solution to work the following parameters should be configured:

- Communications Server: listening port number
- Communications Server: communications timeout (status refresh interval)
- Other components: address and port of the CS

4. Q&A (in polish)

W dniu 27 lutego 2014 09:43 użytkownik Michal Cwienczek <cwienczekm@student.mini.pw.edu.pl> napisał:

>Czy moglibysmy prosic o bardziej szczegolowy opis pol w wiadomosiach XML?

Prośba uwzględniona.
Do 10.03. dokumentacja zostanie uzupełniona w tym zakresie.

W dniu 27 lutego 2014 09:44 użytkownik Michal Cwienczek <cwienczekm@student.mini.pw.edu.pl> napisał:

>Czy beda z gory zadane TaskSolvery i czy dostaniemy je do "podpiecia" na zywo do projektu czy mamy
>sami implementowac przed zlozeniem projektu?

Implementacja Task Solverów jest jednym z głównych zadań projektów,
natomiast wszystkie Task Solvery będą miały za zadanie znajdować dokładne rozwiązania dla instancji problemu DVRP i jeżeli będą Państwo wytwarzali w technologii .NET to Task Solver ma być zrealizowany w formie wtyczki implementującej dostarczoną klasę abstrakcyjną - w razie potrzeby będziemy się starali dostarczać analogiczne biblioteki do innych technologii.

Więcej (oprócz przesyłanych w poprzednim semestrze) publikacji i materiałów nt. DVRP znajdą Państwo m.in. u mnie na stronie przedmiotu MSI2.

W dniu 27 lutego 2014 09:47 użytkownik Michael Cwienczek <michael.cwienczek@gmail.com> napisał:

>Chcielibysmy zaproponowac zobostrzenie (zawężenie) aby komunikacja odbywala sie po protokole
>HTTP zgodnie z jego standardami. Swoja propozycje usprawiedliwiamy tym, ze jest to juz ogolny
>standard komunikacji na swiecie.

Nie zgadzam się - dodatkowy narzut związany z protokołem HTTP (zarówno po stronie implementacyjnej jak i komunikacyjnej) jest tutaj całkowicie zbędny.

W dniu 27 lutego 2014 09:54 użytkownik Michal Cwienczek <cwienczekm@student.mini.pw.edu.pl> napisał:

>wydaje mi sie ze brak jest w dokumentacji opisu jak CN czy tez CC odkrywa obecnosc CS w systemi e. Czy
>jest wysylany jakis broadcast? Jak serwer informuje inne wezly o tym jaki jest jego IP?

Serwer o niczym nie informuje - jego IP i port są traktowane jako znane wartości i stanowią element konfiguracji systemu (sekcja Configuration) w dokumentacji.

W dniu 1 marca 2014 21:53 użytkownik Konrad Oleksiuk <oleksiukk@student.mini.pw.edu.pl> napisał:

>Na początku podane było przez państwa, że projekt może być napisany w dowolnej technologii. Uznaję
>więc, że to nie tylko C# i Java, ale też Ruby, Python, Erlang czy cokolwiek.

Oczywiście, proszę się nie sugerować dostarczoną DLLką, powstała ona ponieważ znakomita większość studentów PL, będzie pisać w .NET.

>1 .Czy jest jakiś konkretny powód dla którego jako schemat komunikacji został wybrany wspierany
>głównie przez MS xml + schema? Praktycznie żaden język nie ma tak dobrego wsparcia jak C#, co
>teoretycznie obliguje do korzystania z niego. Czy nie moglibyśmy wprowadzić bardziej generycznego
>formatu, jakim jest JSON? Rozmawiałem z częścią osób i również podzielają moje zdanie, że podany
>format komunikacji jest praktycznie nie do zaakceptowania przy innym języku, a implementacja obsługi
>tego formatu mija się z celem, bo nikt nie zamierza poświęcać po 15-30h tygodniowo, żeby to w ogóle
>działało.

XML Schema jest:

a) jednym z powszechnie przyjętych sposobów definiowania komunikatów
b) co najmniej równie wspierany jest przez Javę (JAXB), Pythona (PyXB) jak i pewnie pozostałe technologie...
Zwykle Google: [technologia] code generation from XSD pomaga znaleźć co trzeba.

>2. <member
>name="M:UCCTaskSolver.TaskSolver.Solve(System.Byte[],System.Byte[],System.TimeSpan)">
>To również można podciągnąć pod punkt pierwszy: Wymaganie używania C# jako TaskSolver

Jak we wstępie - ten XML to tylko dokumentacja do DLLki.

>3. Ktoś jest w stanie to [XML] w ogóle wyjaśnić? Bo dokumentacja zupełnie pomija, do czego w ogóle >miałoby to służyć.

Problem był zgłaszany, do 10.03. pola zyskają opisy.

>4. The system uses the TCP/IP protocol in order to send messages. Only the CS has a listens on a given >port and all the other components start with connecting to CS. All the messages are a text XML >messages.

>Nie bardzo rozumiem, dlaczego akurat TCP/IP i konkretny port. Czy nie łatwiejszym rozwiązaniem jest >zastosowanie po prostu bardziej abstrakcyjnego protokołu http, który pozwala na standaryzację >połączenia / wysyłania danych między elementami, niezależnie od języka?

To protokół HTTP jest bardziej konkretny niż TCP/IP, HTTP korzysta z TCP/IP a więc rezygnujemy z narzutów komunikacyjnych (niewielkich) jak i komponentowych (potencjalnie dużych). Przesyłanie XMLi/JSONów jako stringów jest również niezależne od języka.

>5. Czy projekty z grupy anglojęzycznej w ogóle były brane pod uwagę?

Ze względu na to, że państwo mieli nieco dłuższy termin na oddawanie dokumentacji nie byłem w stanie uwzględnić jej w procesie wyboru, scalania i kompilacji do wersji obowiązującej.

W dniu 2 marca 2014 22:25 użytkownik kozickit <kozickit@student.mini.pw.edu.pl> napisał:

>Czy możliwość posiadania wielu wątków dla jednego komponentu powinna dotyczyć też TaskManager-a?

Zgodnie z generalnym celem projektu - wymienionym w preambule do pierwszych wytycznych (zeszły semestr) - celem projektu jest napisanie systemu o maksymalnej wydajności, w ramach posiadanych zasobów.

Wydaje się, że wielowątkowość TM (poza ew. wątkami przyjmującymi i obsługującym wiadomości) może nie być istotna ze względu na możliwość współistnienia TM i CN na jednej maszynie, natomiast wielowątkowy TM ma znaczenie, jeżeli zakładają Państwo wykonywanie intensywnych obliczeń przez TM przy scalaniu zadań - wtedy jego wielowątkowość będzie miała znaczenie.

W dniu 2 marca 2014 22:28 użytkownik kozickit <kozickit@student.mini.pw.edu.pl> napisał:

>Zmiana nazwy wiadomości `PartialProblems` na `SolvePartialProblems`, aby uniknąć używania w >wiadomości takiej samej nazwy elementu w różnym kontekście.

Uwzględnię przy nanoszeniu poprawek.

W dniu 2 marca 2014 22:30 użytkownik kozickit <kozickit@student.mini.pw.edu.pl> napisał:

>Zmiana elementów SolvingTimeout, oraz ComputationsTime z obecnego formatu typu String na czas
>wyrażony w milisekundach typu unsignedLong.

Dobry pomysł.
Zmienimy.

W dniu 2 marca 2014 22:32 użytkownik kozickit <kozickit@student.mini.pw.edu.pl> napisał:

>W wiadomości RegisterResponse element Timeout powinien być ustalony z góry, ponieważ grupy
>mogłyby przyjąć zupełnie różne wartości tego elementu, powodując problemy w komunikacji
>komponentów różnych grup.

Są 2 timeouty:
1. pierwszy dotyczy timeout'ów komunikacji - czyli jak często komponenty powiadamiają przy pomocy statusów, że żyją - ten timeout jest ustawiany tylko na serwerze a pozostałe komponenty otrzymują go jako parametr dopiero po skontaktowaniu się z serwerem
2. dotyczy czasu na obliczenia - ustala go CC - po tym czasie obliczenia mają się zakończyć - niekoniecznie pełnym sukcesem, ale jakimś rozwiązaniem

Także zostaje tak jak jest.

W dniu 2 marca 2014 22:37 użytkownik kozickit <kozickit@student.mini.pw.edu.pl> napisał:

W obecnym kształcie biblioteka *UCCTaskSolver.dll* z abstrakcyjną klasą `TaskSolver` uniemożliwia poprawną implementację metody `GetInstance()`. Obecnie metoda ta w klasie pochodnej nie może być określona jako statyczna (co zakłada wzorzec *singleton*).

Dziękuję za uwagę, zdecydowanie nie przemyślałem tego fragmentu rozwiązania, zwłaszcza singletona - jeżeli już miałby być to przecież trzeba go jakoś sprytniej opakować bo task solverów może być dużo różnych.

W każdym razie ta klasa zostanie poprawiona.