Computational Cluster

Katarzyna Węgiełek Paweł Własiuk Kamil Sienkiewicz Marcin Wardziński

 $19 \ {\rm stycznia} \ 2015$

Spis treści

1	$\mathbf{W}\mathbf{s}$	tęp	3
2	Dia	gramy Przypadków Użycia	5
	2.1	Komunikacja	6
		2.1.1 Komponenty z serwerem	6
		2.1.2 Serwer z Menadżerem Zadań	8
		2.1.3 Serwer z Węzłem obliczeniowym	9
		2.1.4 Serwer z Klientem	10
3	Dia	gramy Aktywności	11
	3.1	Zlecenie rozwiązania problemu	11
	3.2	Odczytanie wyniku	12
	3.3	Węzeł obliczeniowy	13
	3.4	Wybieranie zadania do podzielenia	15
	3.5	Podział zadania na podproblemy	16
	3.6	Zgłaszanie wyniku częściowego do menadżera zadań	17
4	Dia	gramy stanów	18
	4.1	Rodzaj zadania	18
	4.2	Zadanie	18
	4.3	Obliczenie	19
	4.4	Menadżer zadań	19
	4.5	Węzeł obliczeniowy	20
5	Ser	wer komunikacyjny	21
	5.1	Cykl życia zadania	21

6	Kor	nunikacja	24
	6.1	Nawiązywanie połączenia	24
	6.2	Lista rozwiązywalnych problemów	26
	6.3	Rozwiązywania zadania	28
	6.4	Odczytanie rozwiązania	30
	6.5	Schema	31
		6.5.1 Uzyskiwanie połączenia	31
		6.5.2 Wiadomości	32

Rozdział 1

Wstęp

Tematem projektu jest stworzenie dokumentacji dla klastra obliczeniowego, który będzie umożliwiał rozwiązywanie problemów o dużej złożoności. System składa się z serwerów komunikacyjnych, menadżerów zadań, klientów i wezłów obliczeniowych.

Rolą serwera jest zapewnienie komunikacji pomiędzy wszystkimi komponentami. Elementy systemu wymieniają informacje wykorzystując protokół XML. W klastrze może być kilka serwerów. Jeden z nich jest serwerem głównym - zarządza przekazywaniem komunikatów i nadzoruje stan wszystkich elementów systemu. Pozostałe pełnią rolę pomocniczą - na nich znajdują się kopie zapasowe częściowych i ostatecznych rozwiązań problemów, zadań oczekujących na rozwiązanie i informacji o trwających obliczeniach. W przypadku awarii serwera głównego, jeden z serwerów pomocniczych przejmuje jego zadania. Wykonywaniem kopii danych zajmuje się serwer główny. On też informuje menadżerów zadań, klientów oraz węzły obliczeniowe o istnieniu serwerów pomocniczych, aby w przypadku awarii możliwa była ich zamiana. Zadaniem aplikacji klienckiej jest pobranie danych wejściowych od użytkownika i wysłanie do serwera komunikacyjnego zlecenia rozwiązania podanego problemu. Gdy obliczenia zostaną zakończone, klient odbiera od serwera rozwiązanie zadania i prezentuje otrzymane wyniki użytkownikowi.

Menadżer zadań zajmuje się organizacją obliczeń i uzyskaniem ostatecznego wyniku z rozwiązań cząstkowych. Po otrzymaniu problemu od serwera komunikacyjnego, dzieli go na mniejsze podproblemy, przeznaczone do rozwiązania przez pojedyncze węzły obliczeniowe. Następnie wysyła je do serwera i oczekuje na wyniki częściowe. Na ich podstawie oblicza rozwiązanie końcowe i

przesyła je na serwer. Każdy menadżer zadań zawiera jeden lub kilka modułów, odpowiedzialnych za rozwiązanie konkretnego typu problemu. Wysyła on do serwera nazwy typów problemów, jakie potrafi obsłużyć, dzięki czemu serwer wie, które zadania może do niego kierować.

Zadaniem węzła obliczeniowego jest rozwiązywanie podproblemów otrzymanych od serwera komunikacyjnego. Węzeł nie ma dostępu do informacji o całym zadaniu i nie zna wyników uzyskanych przez inne węzły. Otrzymuje jedynie problem częściowy, wyznaczony przez menadżera zadań, rozwiązuje go i wysyła do serwera wyniki swoich obliczeń. Węzeł składa się z jednego lub więcej modułów, zawierających algorytmy do rozwiązywania konkretnych typów podproblemów. Zanim otrzyma zadanie, wysyła do serwera informację dotyczącą rodzajów problemów, które potrafi rozwiązać. Podczas wykonywania algorytmu węzły na bieżąco informują serwer, że obliczenia trwają. W przypadku awarii węzła lub przerwania połączenia, do serwera przestaną dochodzić komunikaty i będzie on mógł skierować dany problem do innego węzła, aby uzyskać brakujący wynik częściowy.

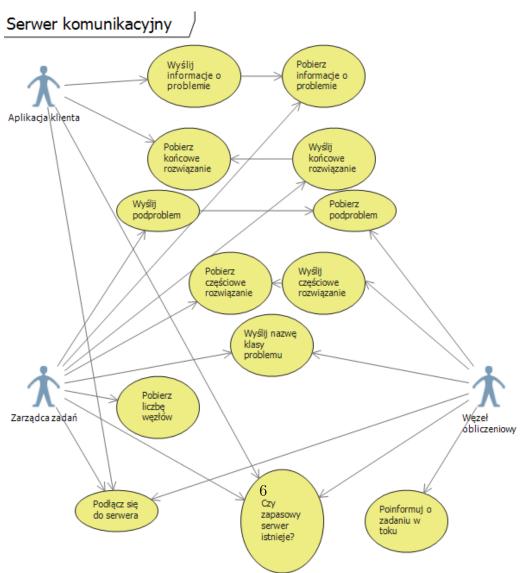
Przykładowym zadaniem, które można rozwiązać korzystając z opisanego klastra obliczeniowego jest problem marszrutyzacji (Dynamic Vehicle Routing Problem).

Rozdział 2

Diagramy Przypadków Użycia

2.1 Komunikacja

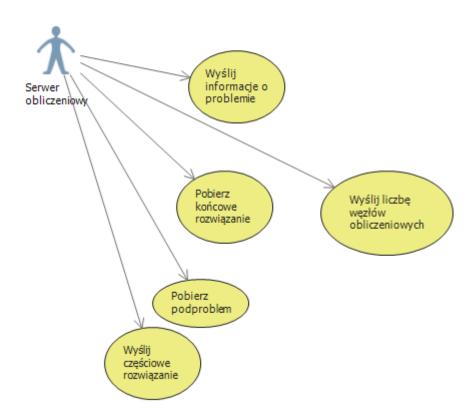
2.1.1 Komponenty z serwerem



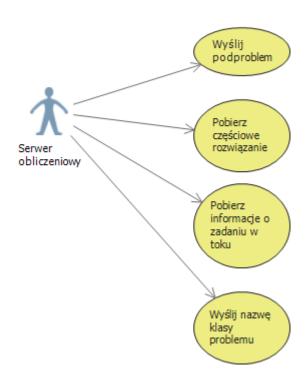
Cała komunikacja musi odbywać się przez serwer, dlatego na początku swojego działania każdy komponent musi się z nim połączyć. Serwer pośredniczy w przesyłaniu informacji o problemie między klientem a Menadżerem Zadań, końcowego rozwiązania od Menadżera Zadań do Klienta, jak również rozwiązań pośrednich między Węzłami Obliczeniowymi, a Menadżerem Zadań. Jego użycie jest również wymagane do przesyłania częściowych podproblemów między Menadżerem Zadań i Węzłami obliczeniowymi.

Do dodatkowych funkcji serwera, nie związanych bezpośrednio z komunikacją należy podanie liczby dostępnych Węzłów obliczeniowych Menadżerowi zadań oraz informowanie o istnieniu zapasowego serwera. Serwer może zażądać informacji o klasie problemu od Menadżera Zadań i Węzłów Obliczeniowych. Węzły obliczeniowe mają obowiązek informować serwer o tym, czy obliczenia nadal się odbywają.

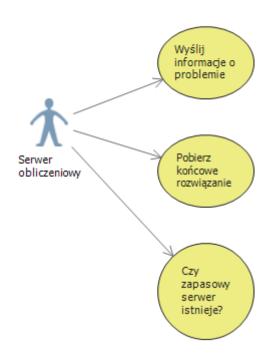
2.1.2 Serwer z Menadżerem Zadań Zarządca zadań $\Big/$



2.1.3 Serwer z Węzłem obliczeniowym Węzeł obliczeniowy



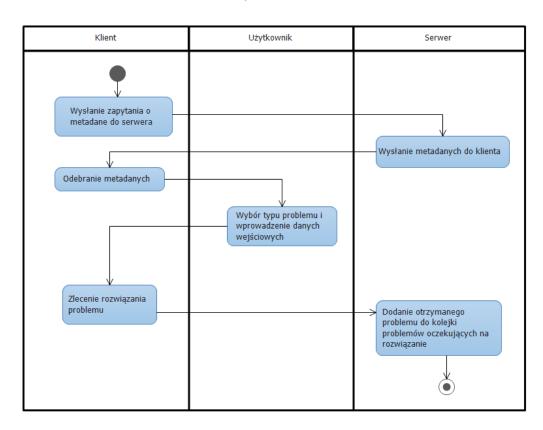
2.1.4 Serwer z Klientem Aplikacja klienta /



Rozdział 3

Diagramy Aktywności

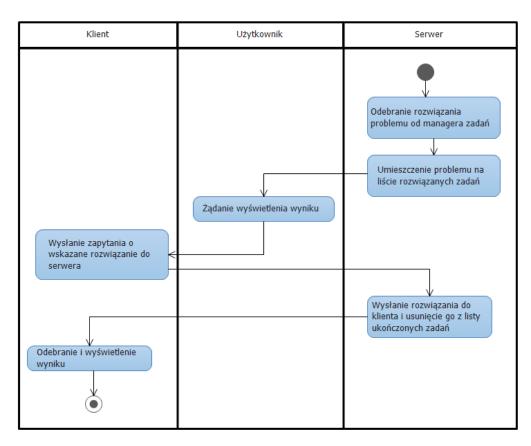
3.1 Zlecenie rozwiązania problemu



Aplikacja kliencka nawiązuje połączenie z serwerem głównym, używając adresu IP serwera zapisanego w pliku konfiguracyjnym. Wysyła zapytanie o metadane - adresy IP serwerów backup'owych oraz nazwy klas problemów,

które mogą być rozwiązane przez ten klaster obliczeniowy (takie, że istnieje przynajmniej jeden menadżer zadań potrafiący obsłużyć dany typ problemu). Serwer przesyła metadane do klienta. Następnie użytkownik wybiera spośród dostępnych nazw klas problemów typ zadania jakie ma zostać rozwiązane i wprowadza do programu wszystkie potrzebne dane wejściowe. Aplikacja kliencka wysyła do serwera zlecenie rozwiązania problemu i podane przez użytkownika dane w formacie XML. Serwer odbiera zlecenie przysłane przez klienta i umieszcza problem w kolejce problemów danego typu oczekujących na rozwiązanie. Zadanie znajduje się w kolejce, dopóki któryś z menadżerów zadań, potrafiących rozwiązać problem tej klasy, nie zakończy obliczeń i nie będzie mógł się nim zająć.

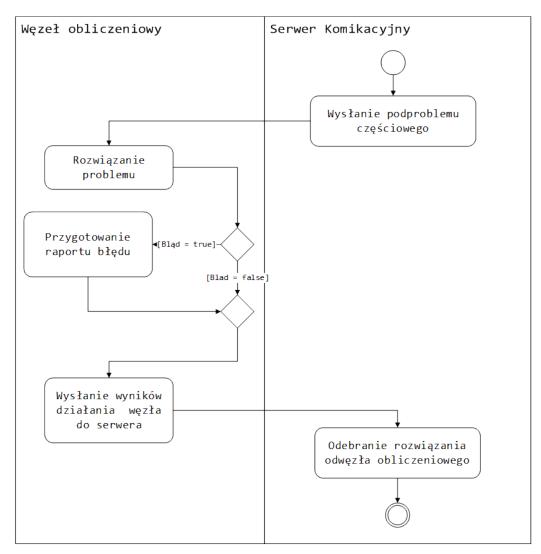
3.2 Odczytanie wyniku



Po otrzymaniu ostatecznego rozwiązania od *menadżera zadań*, serwer umieszcza problem na liście ukończonych, ale jeszcze nieodczytanych rozwiązań.

Następnie użytkownik wybiera z listy problem, którego rozwiązanie chce zobaczyć. Aplikacja kliencka wysyła do serwera żądanie pobrania wskazanego wyniku. Serwer wysyła rozwiązanie odpowiedniego problemu do klienta i usuwa je z listy ukończonych zadań. Klient odbiera wyniki i wyświetla je użytkownikowi.

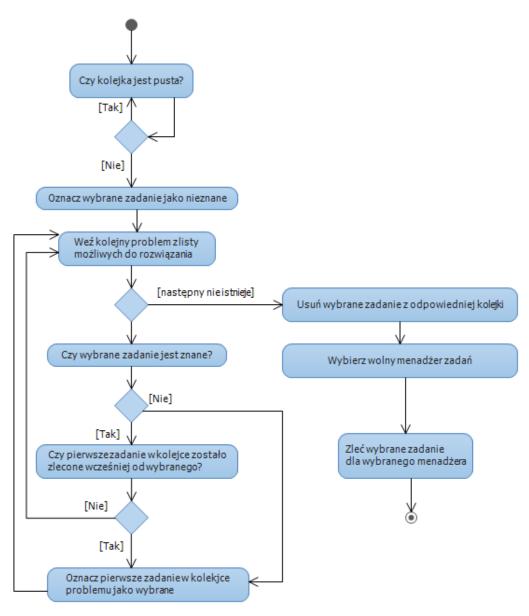
3.3 Węzeł obliczeniowy



Węzeł obliczeniowy przy uruchomieniu zgłasza swoją obecność serwerowi komunikacyjnemu. Informacje temat serwera znajdują się w pliku konfiguracyjnym węzła. Węzeł wysyła do serwera informacje na temat typów proble-

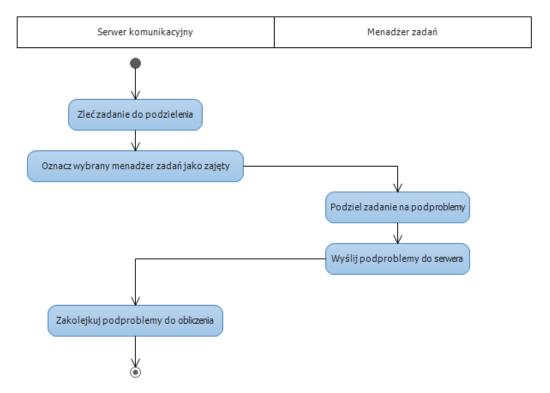
mów które jest w stanie rozwiązać, dzięki temu serwer komunikacyjny może uwzględniać go przy przesyłaniu do **Menadżera Zadań** informacji na temat ilości węzłów potrafiących rozwiązać dany typ problemu. Węzeł obliczeniowy nie wykonujący w danym momencie obliczeń otrzymuje od serwera skolejkowany podproblem, podzielony na części przez **Menadżer Zadań**. Zadaniem węzła obliczeniowego jest rozwiązanie otrzymanego zadania oraz przesłanie rozwiązania do serwera. Jeżeli podczas wykonywania obliczeń wystąpi błąd, węzeł ma za zadanie przygotować raport o błędach a następnie wysłać go do serwera.

3.4 Wybieranie zadania do podzielenia



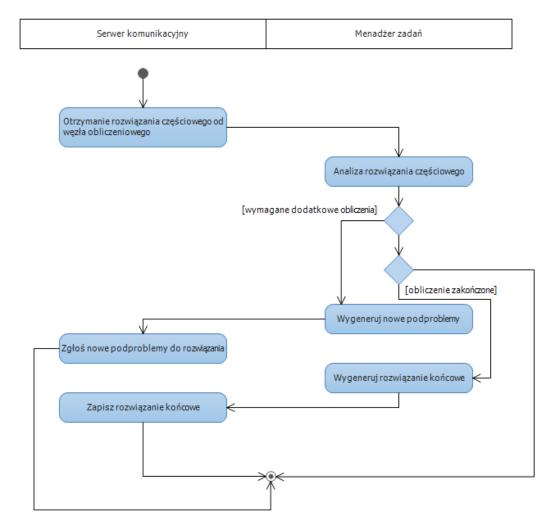
W celu przeprowadzania procesu dzielenia problemów na podproblemy w sposób optymalny z zachowaniem kolejności potrzebne było opracowanie specjalnej struktury danych do ich przechowywania. (Gdzie opisać tą strukturę? Tutaj to się nie nadaje. Najlepiej byłoby to umieścić oddzielnie.)

3.5 Podział zadania na podproblemy



Wybrane zadanie do przetworzenia poprzez **serwer komunikacyjny** trafia do wolnego **menadżera zadań** potrafiącego rozwiązać dany problem. Po podzieleniu go na podproblemy, **menadżer zadań** odsyła definicje podproblemów z powrotem do **serwera komunikacyjnego**, który umieszcza je na innej kolejce do późniejszego obliczenia.

3.6 Zgłaszanie wyniku częściowego do menadżera zadań

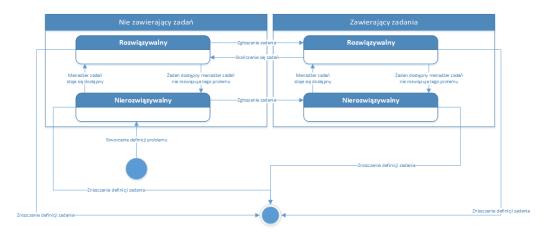


Zaraz po otrzymaniu wyniku częściowego od **węzła obliczeniowego serwer komunikacyjny** odsyła go do **menadżera zadań** który zajmuje się daną instancją problemu. **Menadżer zadań** po analizie rozwiązania częściowego może wygenerować nowe podproblemy, które zostaną zgłoszone analogicznie jak w przypadku początkowego podziału zadania. Może również oznaczyć problem jako rozwiązany i odesłać pełne rozwiązanie do późniejszego pobrania przez użytkownika.

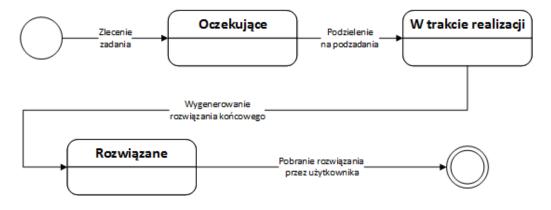
Rozdział 4

Diagramy stanów

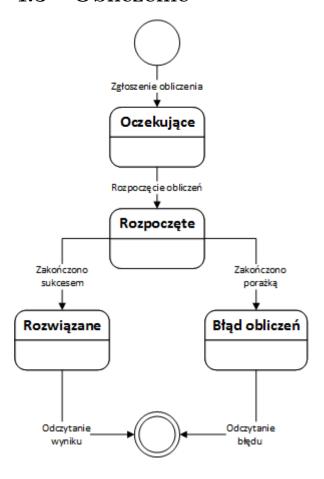
4.1 Rodzaj zadania



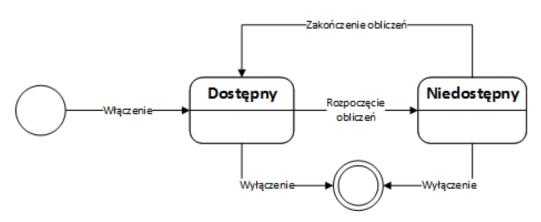
4.2 Zadanie



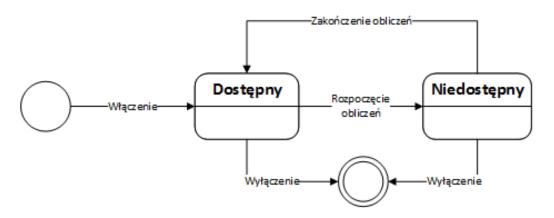
4.3 Obliczenie



4.4 Menadżer zadań



4.5 Węzeł obliczeniowy

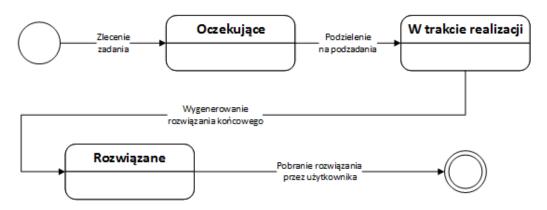


Rozdział 5

Serwer komunikacyjny

5.1 Cykl życia zadania

Dla każdego zadania możemy wyodrębnić kilka jego głównych stanów. Nowo zlecone zadanie trafia nieprzetworzone do odpowiedniej kolejki, oczekując na przesłanie do menadżera zadań i podzielenie na łatwiejsze do rozwiązania podproblemy. Po podzieleniu, zadanie czeka na rozwiązanie wygenerowanych podzadań, a po ich otrzymaniu wygenerowane jest rozwiązanie, które oczekuje potem na pobranie przez użytkownika.



Rysunek 5.1: Diagram stanów zadania

W celu optymalnego przechowywania zleconych przez użytkownika zadań opracowaliśmy strukturę ich przechowywania w której główny podział stanowi rodzaj problemu (TSP, DVRP itp.), a jego w obrębie znajdują się osobne listy FIFO dla nowo zgłoszonych problemów, podproblemów do rozwiązania

oraz podproblemów do scalenia. Rodzaj problemu dla przyspieszenia obliczeń przechowywuje także informacje o menadżerach zadań i węzłach obliczeniowych, które potrafią go obsłużyć.

Serwer komunikacyjny w momencie wyboru zadania filtruje rodzaje problemów w poszukiwaniu takich, które są mają oczekujące zadania oraz są rozwiązywalne, czyli posiadają do dyspozycji wolny menadżer zadań oraz mają zlecone jakieś zadanie. Spośród tych rodzajów wybieramy zadanie zlecone najwcześniej, by zachować oczekiwaną kolejność zajmowania się z zadaniami.

Przy przydzielaniu podzadań dla poszczególnych węzłów komunikacyjnych postępujemy analogicznie.

Rozwiązania podzadań odsyłamy do menadżera zadań w celu scalenia po otrzymaniu ich wszystkich. Menadżer po analizie odpowiedzi może zdecydować o wygenerowaniu rozwiązania końcowego lub zlecić podzadania dodatkowe i odłożyć końcową odpowiedź na termin późniejszy.

Rysunek 5.2: Wybór zadania przez serwer komunikacyjny Oznacz "wybrane zadanie" jako nieznane Czy istnieje następny rodzaj problemu z oczekującymi zadaniami? Weź następny rodzaj problemu, który ma oczekujące zadania Ilość rozwią zujących menadżerów zadań większa od 0. Znajdź zadanie oczekujące zgłoszone najwcześniej lub zadanie jeszczeniewybrane Oznacz to zadanie jako wybrane Czy wybrano zadanie? Obsłuż wybrane zadanie 23

Rozdział 6

Komunikacja

6.1 Nawiązywanie połączenia

Uruchomienie całego systemu rozpoczyna się od uruchomienia serwera komunikacyjnego. Od tej chwili komponenty systemu tj. węzły obliczeniowe i menadżery zadań mogą zgłaszać swoją obecność w klastrze. Każdy z komponentów systemu posiada w swoim pliku konfiguracyjnym adres IP i port serwera komunikacyjnego oraz jego backup'a. Zaraz po uruchomieniu komponent zgłasza swoją obecność do głównego serwera komunikacyjnego. Po trzech nieudanych próbach następuje wysłanie identycznej informacji do serwera zapasowego. Każdy z komponentów w takiej wiadomości informuje o swoim rodzaju podaje IP i port na którym działa. Na tej podstawie serwer będzie komunikował się z tymi elementami systemu. Przykładowa wiadomość powinna wyglądać w następujący sposób:

examples.xml

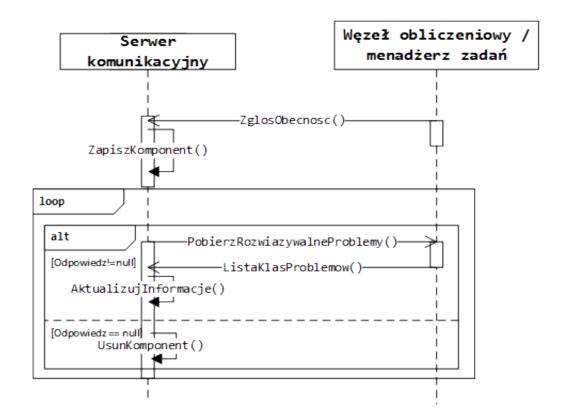
Parametr ComponentType informuje o rodzaju komponentu systemu. Wartości, które może przyjąć ten parametr to:

- TaskManager w przypadku gdy wiadomość pochodzi od menadżera zadań,
- ComputationalNode gdy wiadomość pochodzi od węzła obliczeniowego

Zadaniem serwera komunikacyjnego jest utrzymywanie listy aktywnych komponentów systemu, oraz przechowywania informacji na temat klas problemów, które dane komponenty obsługują. W tym celu serwer regularnie, co pewien określony czas będzie odpytywał wszystkie swoje komponenty o listę klas problemów możliwych do rozwiązania. W przypadku gdy takiej odpowiedzi nie otrzyma uznaje komponent za wyłączony i usuwa jego dane z pamięci. W przypadku gdy komponent odpowie na jego żądanie serwer na podstawie otrzymanych danych uzupełnia/aktualizuje informacje o rozwiązywalnych problemach przez komponenty. Przykładowa odpowiedz komponentu przedstawiona jest poniżej:

examples.xml

Diagram przedstawiający uzyskiwanie połączenia przez komponent systemu:



6.2 Lista rozwiązywalnych problemów

Proces rozwiązywania zadania przez klaster obliczeniowy rozpoczyna się na poziomie aplikacji klienckiej. Przed wysłaniem problemu do rozwiązanie, aplikacja kliencka musi pobrać z serwera informację na temat klas problemów rozwiązywalnych w danej chwili przez klaster obliczeniowy (wynika ona z obecnie dostępnych Menadżerów zadań i węzłów obliczeniowych). Aplikacja wysyła do serwera zapytanie postaci:

gdzie wartość węzła MessageType (definiującego typ wiadomości) ustawiona jest na ProblemsListRequest.

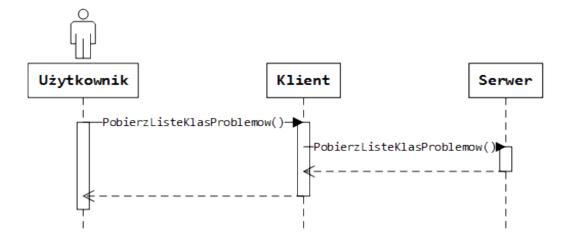
Serwer w odpowiedzi na to żądanie odsyła wiadomość, w której zawarte są informacje o wszystkich typach problemów rozwiązywalnych przez klaster. Informacje zostają przetworzone i przedstawione użytkownikowi.

examples.xml

Przedstawiony powyżej fragment kodu to przykładowa odpowiedź serwera. Dla każdego typu problemu zostają przekazane dwie informacje:

- problemClassName nazwa problemu,
- problemClassId guid problemu, który w jednoznaczny sposób identyfikuje typ problemu.

Ten etap komunikacji przedstawi poniższy diagram:



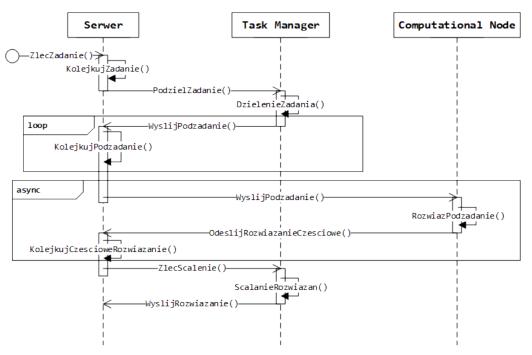
6.3 Rozwiązywania zadania

Po wykonaniu czynności opisanych w poprzednim rozdziale, użytkownik może zlecić zadanie *klastrowi obliczeniowemu* Aplikacja kliencka wysyła do serwera wiadomość z informacją o typie rozwiązywanego problemu, jak też danych wejściowych. Przykładem takiej wiadomości jest poniższy przykład:

examples.xml

Wiadomość zawiera 4 parametry konieczne do stworzenia, rozwiązania i zidentyfikowania zadania przez użytkownika:

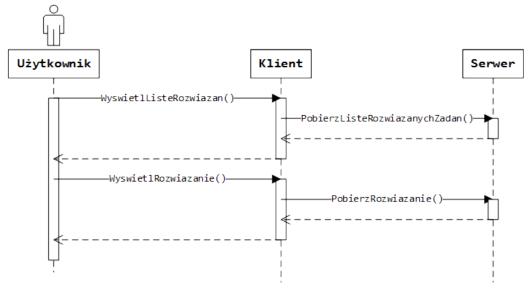
- MessageType parametr ustawiony na TaskOrder informuje serwer o typie wiadomości,
- problemClassId identyfikator problemu, umożliwiający rozwiązanie go przed odpowiednie części systemu,
- Data parametry typu String zawierający dane wejściowe dla danego typu problemu. Dane te przedstawione są w formacie XML specyficznym dla danego typu problemu.
- TaskName nazwa zadania zdefiniowana przez klienta, umożliwi w przyszłości wyszukanie rozwiązania.



Klaster obliczeniowy wykonuje ciąg czynności koniecznych do otrzymania rozwiązania, tzn:

- (a) przesłanie polecenia podziału zadania od serwera do menadżera zadań,
- (b) podział zadania i odesłanie stworzonych podproblemów prowadzących do otrzymania rozwiązania,
- (c) przesłanie każdego z podzadań do węzłów obliczeniowych
- (d) przeprowadzenie obliczeń i odesłanie częściowych rozwiązań do serwera
- (e) przesłanie częsciowych rozwiązań do $menadżera\ zadań$ w celu stworzenia rozwiązania
- (f) odesłanie pełnego rozwiązania zadania do serwera

6.4 Odczytanie rozwiązania



Proces odczytania rozwiązania jest podobny do początkowego etapu obliczania. Tym razem *aplikacja kliencka* musi poprosić *serwer* o listę rozwiązanych zadań, wyświetlić je i umożliwić użytkownikowi wybór typu konkretnego zadania. Razem z lista rozwiązanych problemów serwer przesyła:

- name nazwa zadania (wcześniej zdefiniowana przez użytkownika),
- id identyfikator zadania, wykorzystywany jest przy zapytaniu o rozwiązanie konkretnego zadania,
- problemClassName typ klasy problemu

Następnie zostaje przesłana prośba o pobranie rozwiązania dla wybranego zadania. Rozwiązanie zostaje przesłanie w postaci stringu w formacie XML specyficznym dla danego typu problemu. Poniżej przedstawione zostały dwa inne dwa przykłady komunikatów

• lista rozwiązanych zadań przesyłana do klienta

examples.xml

• rozwiązanie przesłane do klienta

examples.xml

6.5 Schema

6.5.1 Uzyskiwanie połączenia

```
</xs:simpleType>
14
15
    <xs:element name="ConnectionMessage">
      <xs:complexType>
17
        <xs:sequence>
18
          <xs:element type="components" name="ComponentType"/>
19
          <xs:element type="ip" name="IP"/>
          <xs:element type="xs:integer" name="Port"/>
        </xs:sequence>
      </rs:complexType>
    </xs:element>
25 </xs:schema>
```

connectionMessageSchema.xsd

6.5.2 Wiadomości

```
1 <?xml version="1.0" encoding="UTF-8"?>
 qualified" xmlns:xs="http://www.w3.org/2001/XMLSchema">
            <xs:simpleType name="guid">
                  <xs:restriction base="xs:string">
                        <xs:pattern value="([0-9a-fA-F]{8}-[0-9a-fA-F]{4}-[0-9a-fA-F
                                   ] \{4\}-[0-9a-fA-F] \{4\}-[0-9a-fA-F] \{12\}) | ( \{[0-9a-fA-F] \{8\}-[0-9a-fA-F] \{8\}
                                   [4]-[0-9a-fA-F]\{4\}-[0-9a-fA-F]\{4\}-[0-9a-fA-F]\{12\}\}"/>
                  </xs:restriction>
            </xs:simpleType>
            <xs:simpleType name="messageTypes">
 9
                  <xs:restriction base="xs:string">
10
                       <xs:enumeration value="ProblemsListRequest"/>
11
                        <xs:enumeration value="ProblemsList"/>
                       <xs:enumeration value="TaskOrder"/>
                       <xs:enumeration value="TaskSolvedListRequest"/>
                       <xs:enumeration value="TaskSolvedList"/>
                       <xs:enumeration value="TaskSolved"/>
                       <xs:enumeration value="TaskSolvedRequest"/>
                       <xs:enumeration value="SubtaskOrder"/>
                       <xs:enumeration value="SubtaskSolved"/>
19
                       <xs:enumeration value="NodeError"/>
                       <xs:enumeration value="SolvableProblemsList"/>
21
                        <xs:enumeration value="SolvableProblemsRequest"/>
                  </xs:restriction>
23
            </xs:simpleType>
```

```
25
    <xs:element name="Message">
      <xs:complexType>
27
        <xs:sequence >
          <xs:element type="messageTypes" name="MessageType"/>
29
          <xs:element type="guid" name="problemClassId" minOccurs="0"/>
30
          <xs:element type="guid" name="problemId" minOccurs="0"/>
31
          <xs:element type="xs:string" name="Data" minOccurs="0"/>
          <xs:element name="ProblemClassList" minOccurs="0">
            <xs:complexType>
              <xs:sequence>
35
                <xs:element name="ProblemClass" maxOccurs="unbounded"</pre>
                    minOccurs="0">
                  <xs:complexType>
                    <xs:simpleContent>
38
                     <xs:extension base="xs:string">
39
                       <xs:attribute type="xs:string" name="problemClassName"</pre>
40
                            use="required"/>
                       <xs:attribute type="guid" name="problemClassId" use="</pre>
41
                           required"/>
                     </r></xs:extension>
42
                   </xs:simpleContent>
                  </rs:complexType>
                </xs:element>
              </xs:sequence>
46
            </rs:complexType>
          </xs:element>
          <xs:element type="xs:string" name="TaskName" minOccurs="0"/>
          <xs:element name="TasksList" minOccurs="0">
            <xs:complexType>
51
              <xs:sequence>
52
                <xs:element name="Task" maxOccurs="unbounded" minOccurs="0">
                  <xs:complexType>
                   <xs:simpleContent>
                     <xs:extension base="xs:string">
56
                       <xs:attribute type="xs:string" name="name" use="</pre>
57
                           required"/>
                       <xs:attribute type="guid" name="id" use="required"/>
58
                       <xs:attribute type="xs:string" name="problemClassName"</pre>
59
                            use="required"/>
                     </xs:extension>
                    </xs:simpleContent>
                  </rs:complexType>
62
```

 ${\it message Schema.xsd}$