Grid Computing Services for Parallel Algorithms in Medicine and Biology

DRAGOS AROTARITEI[1], MARIUS TURNEA[1], MARIUS CRISTIAN TOMA [2], RADU CIORAP[1], MIHAI ILEA[1]
[1,2]University of Medicine and Pharmacy "Gr.T.Popa"
[1]Faculty of Biomedical Engineering
Str. Universitatii Nr 16, Iasi 700115
ROMANIA
dragos_aro@yahoo.com

Abstract: - The compartmental models using differential equations are basic models in epidemiology. The temporal evolution of spatial models for epidemic spreading is suitable for parallelization and GRID services are solutions for speeding the algorithms used in these models. We investigate several computational aspects of parallel algorithms used in cellular automata model and small world networks model. The four compartmental small world network model of disease propagation (SEIR) is parallelized. In the second application, we have found an asymptotic solution of zero degree for a nonlinear differential parabolic equations system with a unique, small parameter, in a cancerous disease model. We got a cancerous cells density using simulation in three stages according with some system parameter value. Grid service has been constructed for these numeric simulations.

Key-Words: - Epidimiological models, cancerous cells density, parallel algorithms, high performance computing

1 Introduction

The distributed computing is an efficient solution for applications that requires high computational effort, information retrieval from high resources geographically distributed or both. Such applications use interconnected networks of computers or supercomputers, very large databases, software instrument for storage and retrieval, advanced devices and scientific instruments ([3]-[5]).

The GRID computing solution is used for implementation of few epidemiological models, but the applications are mainely focused on sharing very large data bases. Statistic models used in these applications are based on serial approach with no parallelization. Few applications used the GRID computational advantages ([3]-[5]). eMicrob [3] build a GRID platform (eMicrob miniGrid) provide secure access to heterogeneous data and expensive resources in different locations. A system for distributed cohort characterization is proposed in [4]. The system is applied to study of the first episode psychosis [4]. GISE is a flexible service built on Globus 4 grid infrastructure and it has been tested in an epidemic monitoring and surveillance system [5].

In our knowledge, the GRID services related to epidemiological models usually refer to data management across the GRID (breast cancer, mammography, etc.) and only few applications refer to code paralellization the implement epidemiological models.

The epidemiology is one of the standard methods used for identification of population health ([6]-[10]). The most epidimiological phenomena have a mathematical model which permits the simulation and a prediction of disease evolution. A common mathematical model is made by differential equations [8]. An important part of these models use multi-stages (compartmental) approaches. These compartmental approaches that use ordinary differential equations (ODE) are suitable to be implemented on computer systems in order to simulate the temporal and spatial evolution of phenomena [8].

The mathematical models that use ODE can be solved sequentially by iterative methods, numerical methods, or parallelization of the solver algorithm that is based on Euler or Runge-Kutta methods. The efficiency of these algorithms depends of how much overhead is given by communication among processors and the load balancing of the tasks.

The maladies spectrum for a finite population can be sporadic [10], endemic (regular, with continue apparition), epidemic (continue increase of number of affected persons) or pandemic (many countries are affected). Many of these models are affected by seasonal variations (e.g. influenza that are more frequent in the winter) or have seasonal variation that are known as cycle of burst after a number of years. These templates are identified by seasonal patterns [11]. The most common model of seasonal variation is the periodic function based that use sinus or cosinus formulas.

The interest in mathematical models for epidemiology has grown exponentially in the last years. Some models involves aspects such as passive immunity, gradual loss of vaccine and disease-acquired immunity, stages of infection, vertical transmission, disease vectors, macroparasitic loads, age structure, social and sexual mixing groups, spatial spread, vaccination, quarantine, and chemotherapy [10]. Special models have been proposed for diseases such as measles, rubella, chickenpox, whooping cough, diphtheria, smallpox, malaria, onchocerciasis, filariasis, rabies, gonorrhea, herpes, syphilis, avian flu, and HIV/AIDS. The most common models are presented as a set of ordinary differential equation (ODE) or partially differential equations (PDE). In the last years we remark an increased interest related to paralellization of ODE that are used to describe epidimiological models ([11]-[17]).

In the context of disease transmission, some of the studies focused on several forms of computer-generated networks that are defined in terms of how individuals are distributed in space (which may be geographical or social) and how connections are formed ([17]-[24]). This complex process simulates the spatial spread of disease that happens within real populations. We can mention the epidimiological models that fall in this case: random networks, small-world networks (SWN), spatial networks, scale-free networks, exponential random graph models, lattices, and cellular automata (CA) [21]. In what follow we made experiments using CA and SWN.

Cellular automata (CA) are characterized by their discretization of space and time. The epidemiological model using cellular automata is a model that focuses on spatial spreading of a disease. Cellular automata consist of spatial grid placed by cells that are characterized by discrete time and state. At each discrete time, we perform iteration in which cells are updated using certain rules. Corridors of spread in cellular automata can be considered to improve model with real situation when infected individuals can move toward other locations (via train, bus or car) and construct a new infection node.

The term small-world network refers to networks where over a regular lattice small number of shortcuts is introduced. A small-world structure is similar to situation when clusters of connected individuals (social groups) have contact with “nearby” groups and “far-off” groups via the sparse long-range links.

Lattices display high clustering but long path lengths take many steps to move between two randomly selected individuals. Small-world networks offer a means of moving between the rigid arrangement of lattices and the unstructured connections of network models. The high level of clustering means that most infection occurs locally, but short path lengths mean that epidemic spread through the network is rapid and the disease is unlikely to be contained within small regions of the population.

In the last years, a number of application related biomedical application including epidimiological models and cancerous disease have been proposed and implemented on computational Grid ([16]-[17]).

For the model of cancer disease that we presented there, a differential equation in which the singular parameter that multiplies maximum order derivatives for which the boundary function phenomenon shows up ([25]-[28]). The main propose is building an asymptotic approximation of the perturbed problem solution both on the exterior of boundary function vicinity and inside the vicinity. This approximation has an asymptotic character. The asymptotic solution construction generally reduces to the solving of a problem less difficult than the perturbed one. The practical value of this method is determined by the possibility of finding effectively that asymptotic solution with the help of a simple problem.

In the section two we present the epidemic SEIR and the cancerous disease model. In the section three we present our proposal for parallelization of algorithms used in the two models. In the next section the experimental results are presented followed by a section where grid services are described. In the last section we summarize the conclusions of our work.

2 Epidemic Models and Cancerous Disease Model

To find out of an adequate model that fit for one epidemic is a difficult operation. The mathematic

model is a tradeoff between simplicity, accuracy and generality. A model should approximate what happens in the real world. A complex model could have a greater accuracy but it could be too difficult to be parameterized and understood. The most common models are compartmental models [8, 10].

2.1 The SEIR Model

The four compartmental small world network model of disease propagation has four categories of populations: *S* - Susceptible (the fraction of susceptible individuals, those individuals able to contact the disease), *E* – Exposed (the fraction of exposed individuals, those individuals that have been infected but are not yet infectious), *I* – Infective (the fraction of individuals that are able to transmitting the disease), and *R* – Recovered (the fraction of individuals who become immune).

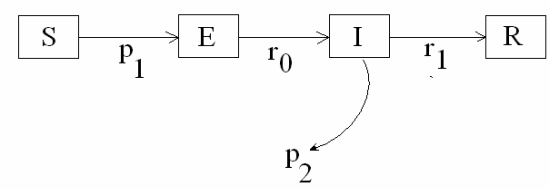

Fig. 1. The four compartmental SWN model of disease propagation

The compartmental model and transitions are showed in Fig. 1. Suppose the birth and death rate μ is constant. The equations of basic SEIR model are [22]:

$$dS/dt = \mu - \beta(t)SI - \mu S \quad (1)$$

$$dE/dt = \beta(t)SI - (\mu + \alpha)E \quad (2)$$

$$dI/dt = \alpha E - (\mu + \gamma)I \quad (3)$$

$$S + E + I + R = N \quad (4)$$

In the equations above $1/\alpha$ is the mean latent period for disease and $1/\gamma$ is the mean infection period [20]. The parameter $\beta(t)$ represents the force of infection (infection rate) and can be constant $\beta = \beta_0$ = constant or can be seasonal that is time-varying with time *t* in years:

$$\beta(t) = \beta_0(1 + \beta_1 \cos(2\pi t)) \quad (5)$$

The paths of transmission (state transition graph) are depicted in Fig. 2. The infected individuals can create susceptible individuals to who are linked with some probability. The immediate neighbors will become infected with probability p_1 meanwhile the long range links will become infected with probability p_2.

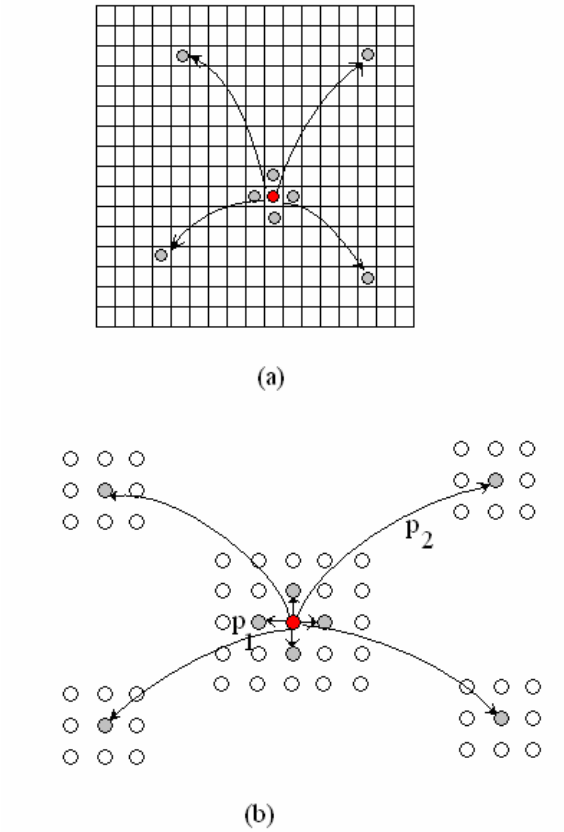

Fig. 2. The short-range and the long-range network links (a) grid locations (b) the local arrangement of nodes in small network/

Exposed individuals will become infected with probability r_0 and finally, infected individuals will become immune (recovered) with probability r_1. The degree distribution can be truncated power-law form (6) or discrete exponentially decaying distribution (7). Our experiments use the form presented in (7):

$$f_X(x) = \frac{1}{C}e^{-x/\mu}, \quad C = \frac{1}{1 - e^{-1/\mu}} \quad (6)$$

$$p(n_2^i = e^k) = \frac{1}{k}e^{-\frac{k}{\mu}} \quad (7)$$

Cellular automata (CA) are characterized by their discretization of space and time. Epidemiology model using cellular automata is a model that focuses on spatial spreading of one disease. Cellular automata consist of spatial grids placed by cells that are characterized by discrete time and state. At each discrete time, we perform iteration in which cells update on certain rules. Corridors of spread in

cellular automata can be considered to improve model with real situation when infected individuals can move toward other locations (via train, bus or car) and these individuals can construct a new infection node.

The corridors are considered only in four directions (*North*, *South*, *East* and *West*). These corridors permit the spread of disease without the links cell-by-cell. The extension with the other four intermediary points complicates the problem and these extensions will be considered for further research.

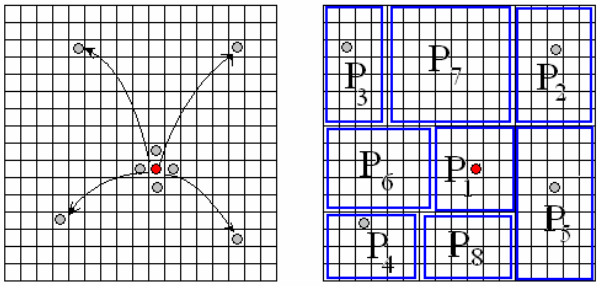

Fig. 3. The allocation of processors

For each simulation we seed the model with one initial infection. We denote some variables inspired by [22]. The population of N is assumed that there are no other births or a death during simulation that is the population N is constant. The population is disposed on a grid with $L \times H$ rectangular area, so in this case $L=H$, $N= L^2$.

Let's be *np* the number of available processors. The parallelization in our case is simple but it is clear that the difficulties that arise are mainly due to map in the case of random contour. In the case of rectangular grid, the partition of the map among *np* processors is simple. Each processor has a rectangle of dimension *M= (int) N/np* for the first *np-1* processors meanwhile the processor *np* take the rest from rectangle. An arbitrary partition based on heuristics can be taken also into account. The partition could have a different number of squares.

The urban population from cities has an increased probability to grow the infected people if a single infected is present in the city area. The heuristic of partition is based on trial to allocate the equal distribution of population of each processor and to allocate an entire city to one processor without splitting the location among processors (Fig. 3).

Partition of N cells in M rectangles (M << N)
Start the seed of disease in point $S(i_s, j_s)$
Allocate the processor P_1 to rectangle R_1 where $S(i_s, j_s) \in R_1$
Allocate the rectangles M_i to processor P_i for $1<i<np+1$
Repeat until no area to be allocated
Allocate the rectangles M_k to processor P_i
for $1<i<np+1$, $np<k<M$
Allocate the boundaries of each area to corresponding processor
Start SEIR model

For step=1 to step = max_steps
SEIR model and find the new E and I points
(short range and long range)

Identify the area and processor allocated to new
E and I points

Verify the boundaries and correct the processor allocation
End for

Collect all the data to Processor P_1

Fig. 4. The proposed algorithm

Our proposal for parallel version of serial algorithm is depicted in Fig. 4. These heuristics are based on population distribution on the map where the big cities are more susceptible to infection long range and short range so the processor must cover the entire area of one city.

2.2 The Cancerous Disease Model

The second model refers to *n* asymptotic solution of zero degree for a nonlinear differential parabolic equations system with a unique, with small parameter, in a cancerous disease application ([27]-[28]). The model is given by differential equation with small parameter with perturbation.

The perturbation presented in this work is a singular one, which causes major changes of the solutions. For the model that we are going to present in this work is being representative a differential equation in which the singular parameter that multiplies maximum order derivatives for which the boundary function phenomenon shows up. The main propose is building an asymptotic approximation of the perturbed problem solution both on the exterior of boundary function vicinity and inside the vicinity. This approximation has an asymptotic character. The asymptotic solution construction generally reduces to the solving of a problem less difficult than the perturbed one. The practical value of this method is determined by the possibility of finding effectively that asymptotic solution with the help of a simple problem.

$$\begin{cases} \varepsilon \dfrac{\partial U}{\partial t} = \varepsilon^2 . \dfrac{\partial^2 U}{\partial x^2} + U(1-U-aV) \\ \dfrac{\partial V}{\partial t} = D\dfrac{\partial^2 V}{\partial x^2} + bV(1-V) + aUV - cVW \\ \dfrac{\partial W}{\partial t} = \dfrac{\partial^2 W}{\partial x^2} - W + dVW \end{cases} \quad (8)$$

$$\begin{cases} U(t,x) = U_0(x,t) + \pi_0 U(x,\delta) + P_0 U(\rho,t) \\ V(t,x) = V_0(x,t) + \pi_0 V(x,\delta) + P_0 V(\rho,t) \\ W(t,x) = W_0(x,t) + \pi_0 W(x,\delta) + P_0 W(\rho,t) \end{cases} \quad (9)$$

The perturbed problem (P_ε) is in the form given by equation (8). The initial conditions for this perturbed problem are U(x,0,ε)=A; V(x,0,ε) =B;W(x,0,ε)=C. We shall consider the zero order problem (P_ε) asymptotic development in this shape (9).

The conditions (10) in medical terms, are stating that the cancerous cells situated in the malign phase are not being able to move along the omega border mass. Consider Ω as a cancerous cells habitat.

$$\frac{\partial V}{\partial n} = \frac{\partial W}{\partial n} = 0, x \in \partial\Omega \quad (10)$$

$$\frac{\partial U}{\partial n} = 0, t > 0 \quad (11)$$

The normal vector at Ω border U,V,W represents the three stages cancerous cells density. The W last stage is considered the pre-malignant stage. For a three stage mutation model which could lead to malignant cells, the evolution of the stages is of form presented in Fig.5.

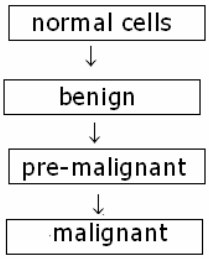

Fig. 5. The three stage mutation model

For instance in esophagus cancer, the cells go through the next stages: the initial one, rapid cells growth and their invasion, angiogenesis and metastasis (malign stage). The cancerous stages with the W density are those cells that had suffered mutations in the previous stages and have adjusted to the new conditions. The dVW term represents the interaction of the cells that had suffered mutations with the surrounding cells in their struggle for surviving and new resources.

The D constant represents the ability to move of the cancerous cells in report to the healthy cells. If D = 0, the cancerous cells metastasized. The fact that D is singular shows us he moving capacity of the cells that had suffered mutations along their three stages of resource struggling. In the medical literature, D is also called migratory factor. The ε parameter from the first equation of (P_E) problem measures the cell mobility of the pre-malignant stage cells compared to the healthy cells one.

Considering ε small, the migratory pre-malignant stage cells capacity grows. We can say that the e parameter measures the pre-malignant cells aggressiveness towards healthy cells in their competition for surviving. The a,b,c,d coefficients are called interaction coefficients.

2.3 Parallelization of Cancerous Disease Model

The parallel algorithm is shortly described in Fig. 6. The allocation of processors the proposed algorithm in depicted in Fig. 7. Let's be *m* users that want to rule the model with different parameters simultaneously (the lower bound is *m*=1). We denote by $P_i(j, k)$ the process *i*, that belong to user *k* in the step *j*.

P_0: *Set the parameters of model;*
Set t_{total}, t_0 ;
Initialise $U(t_0)$, $V(t_0)$, $W(t_0)$
P_1: $r' = rand_{Poisson}$;
$r = f(U(t_i), V(t_i), W(t_i))$;
P_2: *compute* $U(t_{i+1})$ *depend on r and r'* ;
P_3: *compute* $V(t_{i+1})$ *depend on r and r'*;
P_4: *compute* $W(t_{i+1})$ *depend on r and r'* ;
P_5: $i = i+1$;

Fig. 6. The parallel proposed algorithm for cancerous disease application

The pipeline for a fixed number of iterations *q* for all users is $LP = \{P_0(1,1), P_0(1,2),..., P_0(1,m), P_1(1,1), P_1(1,2),...., P_1(1,m), P_2(1,1), P_2(1,2),...., P_2(1,m), ..., P_4(1,1), P_4(1,2),...., P_4(1,m), P_5(1,1), P_5(1,2),...., P_5(1,m), ..., P_1(2,1), P_1(2,2),...., P_1(2,m), ..., P_5(q,1), P_5(q,2),...., P_5(q,m)$. If we have *p* processors, each processor have o block of *p* processes, from ordered list LP, in order block after

block. The graph of scheduling processes in parallel is shown in Fig. 7.

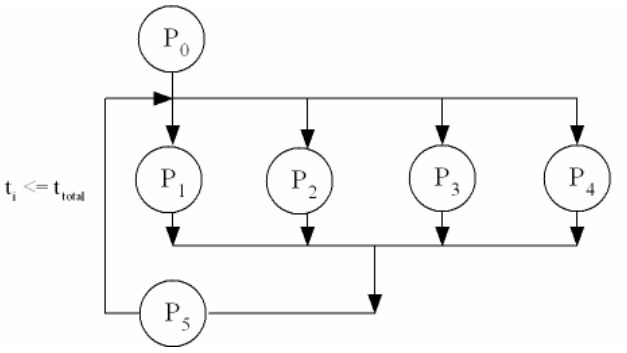

Fig. 7. The allocation of processors cancerous disease application

3 Experimental Results

The model has the parameters set to L= 1500, $N=L^2$ = 2250000, $n_1 = 4$, r_0 follow a geometric distribution $f_X(x) = (1-p)^{x-1}p$, $p_1 = 1/n1(0.27-\mu p_2)$, $r_0 = 0.135$, r_1 = 0.25, μ = 7 values partially inspired from [13] with SEIR model for SARS. We used p_2 in the range [0, 0.07].

We tested the simulation of spread disease in parallel implementation for 2-9 processors. The tests have been made under several restrictions: (1) the simulation is made for *T* days meanwhile the spread of disease doesn't cross the border of rectangles allocated to processors (2) the scalability of the algorithm is tested only for maximum 8 processors (3) we compared the parallel algorithm with serial version that is the run on a single processor.

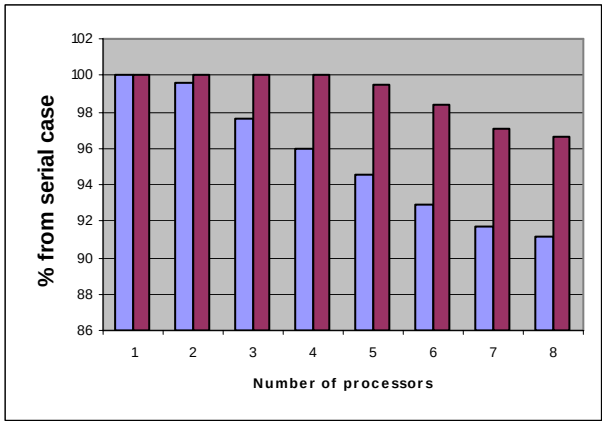

Fig. 8. The result of simulation (the best case and the unfavorable case).

The evolution after 122 days is presented (in percent time of the serial algorithm) as follow: 2 processors – 99.6%, 3 processors – 97.6%, 4 processors - 96%, 5 processors 94.6 %, 6 processors – 92.9% 7 processors – 91.7%, 8 processors – 91.2%. We must remark that the performance of parallel algorithms increase with the evolution of disease but it is limited by the number of processors. Each new area gets a processor allocation and this processor will compute all the operation from cells in the allocated area.

The cellular model proved to have problems. The main problem is that long range seeds for new cluster of infection, despite the corridor improvement have difficulties. The corridors are difficult to be modeled according to principle of cellular automaton and the realistic case. By these reasons, only preliminary results are reported in this paper regarding cellular automata model. The experiments uses a city location, a medium size city (450000 inhabitants and only vertical corridors has been taken into account. The results after 257 days are not very satisfactory so we not continued in this direction.

The experiments with cancerous disease model started with different values in order to find the numerical solutions to the system of singularly perturbed equations. The Simulation of the solution for unperturbed problem over the parameters: a=0.3, b=0.1, c=0.7, d=0.4 is presented in Fig. 9-10.

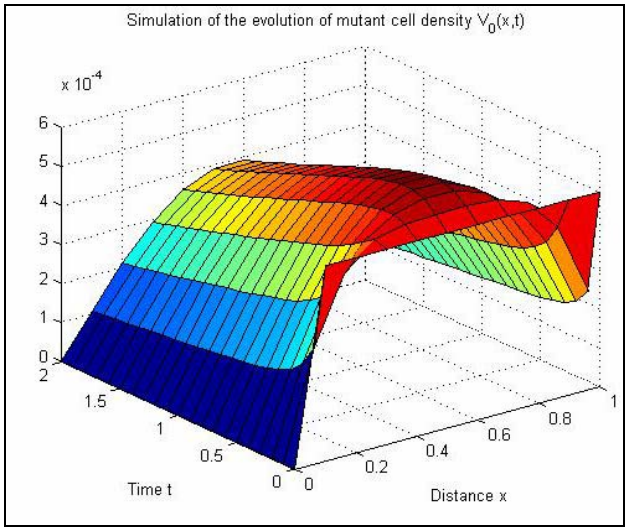

Fig. 9. The Simulation of the solution for unperturbed problem over the parameters: a=0.3, b=0.1, c=0.7, d=0.4 (V_0).

The survival of the mutant cell under selection pressure may lead to another mutation. The total number of cancer cells in a body depends on the rates at which they divide and are destroyed by the immune system.

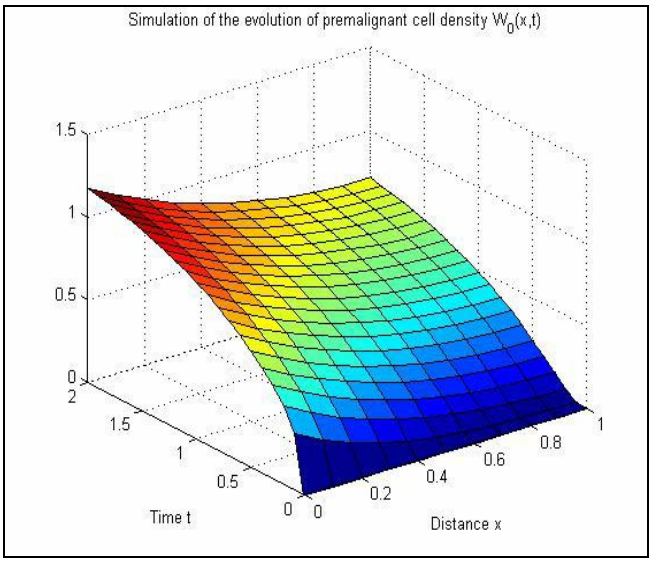

Fig. 10. The Simulation of the solution for unperturbed problem over the parameters: a=0.3, b=0.1, c=0.7, d=0.4 (V_0).

Simulation of the solutions in perturbed problem can be interpreted as the tumour feeds on oxygen and nutrients that diffusive to its surface. But these supplies cannot penetrate deep into the tumour, so cells in the case become dormant or die of starvation. Simulation of the solutions in perturbed problem can be interpreted as asymptotic solution involving partial differential equations often encounters the difficulty connected with non-smoothness of some of the terms of the asymptotic expansions.

On the other hand, the zeroth and the first order approximations to the solution, which are usually sufficient for practical proposes can be obtained in a somewhat simpler way, using a boundary function method. We will consider the case where the nonsmoothness of the terms in the asymptotic series has a different nature. We assumed all constant parameters (a,b,c,d) are independent from reactions including the effect of continuous changes of environment on DNA loading to the mutation. These changes are inevitable and may cause a sequence of mutations. The more aggressive mutant cells are able to exploit the environment and the resources of gells of previous stage and have a better chance to survive.

4 Grid Services

The services of data analysis used in our Grid present two distinct components: a component of data analysis developed in C++ using C++ Standard Template Library and MPI 2.0 models of communication. The end-user service grid component that makes connection between the two modules is developed in Java (Fig. 7).

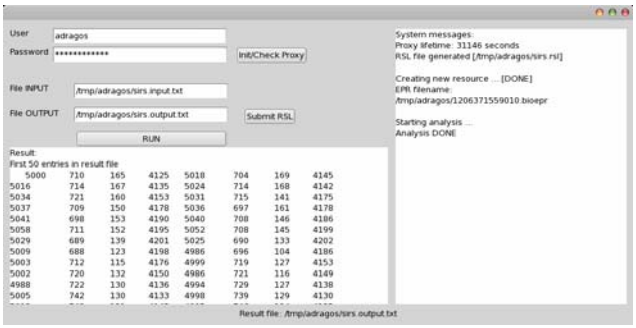

Fig. 11. The GUI for epidimiological service developed in JAVA

The scripts for BioGridEpidemiologyService, the Grid Epidemiology service that run the models are presented shortly below. We extracted sequences from the script because the script is too long to be entirely presented.

```
<?xml version="1.0" encoding="UTF-8"?>

<definitions name="BioGridEpidemiologyService"
   targetNamespace="http://127.0.0.1/
                    namespaces/BioGridEpidemiologyService_instance"
   xmlns="http://schemas.xmlsoap.org/wsdl/"
   xmlns:tns="http://127.0.0.1/
                    namespaces/BioGridEpidemiologyService_instance"
   xmlns:wsdl="http://schemas.xmlsoap.org/wsdl/"
   xmlns:wsrp="http://docs.oasis-open.org/wsrf/2004/06/
                    wsrf-WS-ResourceProperties-1.2-draft-01.xsd"
   xmlns:wsrpw="http://docs.oasis-open.org/wsrf/2004/06/
                    wsrf-WS-ResourceProperties-1.2-draft-01.wsdl"
   xmlns:wsdlpp="http://www.globus.org/namespaces/2004/10/
                    WSDLPreprocessor"
   xmlns:xsd="http://www.w3.org/2001/XMLSchema">

   <wsdl:import
         namespace="http://docs.oasis-open.org/wsrf/2004/06/
                    wsrf-WS-ResourceProperties-1.2-draft-01.wsdl"
         location="../../schema/wsrf/properties/
                    WS-ResourceProperties.wsdl" />


   <!-- TYPEZ -->
   <types>
      <xsd:schema targetNamespace="http://127.0.0.1/
                    namespaces/BioGridEpidemiologyService_instance"
         xmlns:tns="http://127.0.0.1/
                    namespaces/BioGridEpidemiologyService_instance"
         xmlns:xsd="http://www.w3.org/2001/XMLSchema">
<!--
         <xsd:element name="JobProperties">
           <xsd:complexType>
           <xsd:sequence>
               <xsd:element name="jobRSL" minOccurs="1"
                    maxOccurs="1" type="xsd:string"/>
               <xsd:element name="poxyPath" minOccurs="1"
                    maxOccurs="1" type="xsd:string"/>
               <xsd:element name="keyFile" minOccurs="1"
                    maxOccurs="1" type="xsd:string"/>
               <xsd:element name="certFile" minOccurs="1"
                    maxOccurs="1" type="xsd:string"/>
           </xsd:sequence>
           </xsd:complexType>
         </xsd:element>
-->
....
```

```
<!-- PORTYPEZ -->
<portType name="BioGridEpidemiologyPortType"
    wsdlpp:extends="wsrpw:GetResourceProperty"
    wsrp:ResourceProperties=
                    "tns:BioGridEpidemiologyResourceProperties">

    <operation name="methodWriteSettings">
      <input message="tns:MethodWriteSettingsInputMessage"/>
      <output message="tns:MethodWriteSettingsOutputMessage"/>
    </operation>

    <operation name="methodWriteLogSettings">
      <input message="tns:MethodWriteLogSettingsInputMessage"/>
      <output message="tns:MethodWriteLogSettingsOutputMessage"/>
    </operation>

    <operation name="methodClientStartJob">
      <input message="tns:MethodClientStartJobInputMessage"/>
      <output message="tns:MethodClientStartJobOutputMessage"/>
    </operation>

    <operation name="methodReadJobStatus">
      <input message="tns:MethodReadJobStatusInputMessage"/>
      <output message="tns:MethodReadJobStatusOutputMessage"/>
    </operation>

  </portType>
</definitions>
```

The Grid services require the following resources: *jobSettingsRP* (file name for configuration of desired analysis), *JobLogSettingsRP* (prefix path), and *ClientJobStatusRP* (job status). The methods are: *methodWriteSettings, methodWriteLogSettings, methodClientStartJob* (Java COG call are used to start the Grid Services – the *RSL* script will be sent to GRAM for run). The JAVA source files used by Grid service are presented in Fig. 12.

- BioGridEFactoryService.java
- BioGridEpidemiologyQNames.java
- BioGridEpidemiologyResourceHome.java
- BioGridEpidemiologyResource.java
- BioGridEpidemiologyService.java

Fig. 12. The JAVA source files

The configuration file and the RSL WS-GRAM used in services are presented below.

deploy-jndi-config.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<jndiConfig xmlns="http://wsrf.globus.org/jndi/config">

<!-- Instance service -->
<service name="BioGridEpi/BioGridEpidemiologyService">
      <resource name="home" type=
       "biogride.factory.impl.BioGridEpidemiologyResourceHome">
      <resourceParams>

            <parameter>
                  <name>resourceClass</name>
                  <value>biogride.factory.impl.
                              BioGridEpidemiologyResource</value>
            </parameter>

            <parameter>
                  <name>factory</name>
                  <value>org.globus.wsrf.jndi.BeanFactory</value>
            </parameter>

            <parameter>
                  <name>resourceKeyType</name>
                  <value>java.lang.Integer</value>
            </parameter>

            <parameter>
                  <name>resourceKeyName</name>
      <value>{http://127.0.0.1/namespaces/
                        BioGridEpidemiologyService_instance}
                        BioGridEpidemiologyResourceKey</value>
            </parameter>

      </resourceParams>
      </resource>
</service>
```

BioGridEpidemiologyService

```
<?xml version="1.0" encoding="UTF-8"?>
<job>
      <executable>/export/home/adragos/ws_gram/sir.bin</executable>
      <directory>/export/home/adragos/ws_gram</directory>
      <argument>/tmp/adragos/fis1.txt</argument>
      <argument>/tmp/adragos/fis2.txt</argument>
      <stdout>/tmp/sirs.stdout</stdout>
      <stderr>/tmp/sirs.stderr</stderr>
      <count>1</count>
      <jobType>single</jobType>
</job>
```

We used the mechanisms of Globus Toolkit 4 in order to identify the user. Any user will be identified by grid proxy. The compilation of sources for grid service is made by ./globus-build-service.sh bgeFactoryTest.

5 Conclusion

For each simulation we seed the model with only one initial infection. Two or more seeding points could be a realistic situation but this aspect will be extended in the further results. The results of simulation are based on synthetic data. It is clear that in practice, some coefficients can have particular values that depend on estimation results in the first days of epidemics. We can operate with the limits of the parameters in the sense of pessimistic and optimistic case (lower and upper limits) but the intervals are usually very large. In some cases, some values could have as result the extinction of epidemic in very few days, which is an unrealistic case (or very less probable).

The communication among processors for a small number of processors could overcome the benefit of parallelization. The proposed algorithm can be useful for a large area (e.g. pandemic spread of diseases) and a fine granulation of partitions that must be allocated to processors. Also, the partition contour is very important to exploit the benefices of parallelism. Meanwhile, a laced optimal contour

produces in increased of computational effort especially in the case of fine granulation: the processor must verify each cell on the border in order to avoid going in the partition allocated to other processor.

However, the model proved to have a good approximation for evolution of epidemic disease in the sense that by modification of the parameters including the probability of disease we can reach any cell from lattice in a reasonable number of days and the model can cover a very large class of real epidemic spreads.

The proposed solution solves a practical problem that involves ODE used for epidimiological models. In the future research we will extend the spatial epidemiological model to a very large geographic area. This extension needs considerable amount of geographic and demographic data that need time to be collected.

Another direction of research is the implementation of a more general parallel solver for Grid applications. This parallel solver will be focused on various compartmental models with application in epidimiological models.

Acknowledgments. The Grant CEEX II03/31.07.2006 "Academic Grid for Complex Applications – GRAI and CEEX 11 – 005.3/ 14.09.2007 NEOSIM have supported the research for this paper.

References:

[1] G. Fox, A.J.G. Hey (Eds), *The Grid Computing: Making the Global Infrastructure a Reality*, Wiley & Sons, 2003.

[2] J. Guo, et. al., *eMicrob: A Grid-Based Spatial Epidemiology Application*, Lecture Notes in Computer Science, Vol. 3516, 2005, pp. 472-475.

[3] J. Ainsworth, R. Harper, I. Juma, I. Buchan, *PsyGrid: Applying e-Science to Epidemiology*, 19th IEEE International Symposium on Computer-Based Medical Systems CBMS 2006, 2006, pp. 727-732.

[4] A. Tirado-Ramos, P.M.A. Sloot, A. G. Hoekstra, M. Bubak, *An integrative approach to high-performance biomedical problem solving environments on the Grid*, Parallel Computing, Vol. 30, 2004, pp. 1037–1055.

[5] E. Gallo et. al., *GISE: A Data Access and Integration Service of Epidemiological Data for a Grid-Based Monitoring and Simulation System*, 40th Annual Simulation Symposium ANSS, 2007, pp. 267-274.

[6] R.M. Anderson, R.M. May, B. Anderson, *Infectious Diseases of Humans. Dynamics and Control*, Oxford University Press, 1992.

[7] D.J. Daley, J.Gani, *Epidemic Modelling: An Introduction*, Cambridge University Press, 2001.

[8] O. Diekmann, J.A.P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*, John Wiley & Sons, 2000.

[9] J. Giesecke, *Modern Infectious Disease Epidemiology*, A Hodder Arnold Publication, 2001.

[10] H.W. Hethcote, The Mathematics of Infectious Diseases, SIAM REVIEW, Vol. 42-4, 2000, pp. 599–653.

[11] K. Burrage, *Parallel methods for systems of ordinary differential equations*, Oxford University Press Inc., New York, 1995.

[12] M.J. Miguel, et. al., *Parallelization of Implicit-Explicit Runge-Kutta Methods for Cluster of PCs*, Lecture Notes in Computer Science, Vol. 3648, 2005, pp. 815-825.

[13] P.J. van der Howen, B.P. Sommeijer, *Parallel ODE solvers*, Proceedings of the 4th international conference on Supercomputing, Vol. 18-3b, 1990, pp. 71-81.

[14] D. Petcu, M. Dragan, Designing an ODE Solving Environment, LNCSE, Vol. 10, 1999, pp. 245-266.

[15] M. Korch, T. Rauber, *Scalable Parallel RK Solvers for ODEs Derived by the Method of Lines*, Euro-Par 2003. Parallel Processing, LNCS, Vol. 2790, 2003, pp. 830–839.

[16] S. Ichikawa, Y. Fujimura, *Iterative Data Partitioning Scheme of Parallel PDE Solver for Heterogeneous Computing Cluster*, Proc. Applied Informatics AI 2002, 2002, pp. 364-369.

[17] F. Ismail, Z.A. Majid, M. Bin Suleiman, U.K.S. Din, *The Parallel Three-Processor Fifth Order Diagonally Implicit Runge-Kutta Methods for Solving Ordinary Differential Equations*, 12th WSEAS Int. Conf. on APPLIED MATHEMATICS, Cairo, Egypt, 2007, pp. 184-188.

[18] R.A. Kosinski, *Cellular Network with complex connections for the modeling of epidemic spreading*, WSEAS Transactions on Systems, Vol. 3, 2004, pp. 2651-2656.

[19] H. Fukś, R. Duchesne, A.T. Lawniczak, *Spatial correlations in SIR epidemic models*, Proc. of WSEAS MATH 2005, Canun, Mexico, 2005, pp. 108-113.

[20] S. Venkatachalam, A.R. Mikler, *Towards Computational Epidemiology: Using Stochastic Cellular Automata in Modeling Spread of Diseases*, Proceedings of the 4th Annual International Conference on Statistics, Mathematics and Related Fields, Honolulu, USA, 2005, pp. 1-16.

[21] S. H. White, A.M. del Rey, G. R. Sanchez, *Modeling epidemics using cellular automata*, *Applied Mathematics and Computation*, Vol. 186, Issue 1, 2007, pp. 193-202.

[22] M. Small, C.K. Tse, *Clustering model for transmission of the SARS virus: application to epidemic control and risk assessment*, Physica A: Statistical Mechanics and its Applications, Vol. 351, 2005, pp. 499-511.

[23] M.J. Keeling, K.T.D. Eames, *Networks and epidemic models*, J. R. Soc. Interface, Vol. 2, 2005, pp. 295–307.

[24] J.L. Aron I.B. Schwartz, *Seasonality and period-doubling bifurcations in an epidemic model*, Journal of Theoretical Biology, Vol. 110, 1984, pp. 665-67.

[25] M. Höhle, U. Feldmann, *RLadyBug—An R package for stochastic epidemic models*, Computational Statistics & Data Analysis, Vol. 52 , Issue 2, 2007, pp. 680-686.

[26] S. Lloyd, *Integrative Biology - the challenges of developing a collaborative research environment for heart and cancer modelling*, Future Generation Computer Systems, Vol. 23, 2007, pp. 457–465.

[27] R. Ahangar, X.-B. Lin, *Multistage evolutionary model for Carciogenesis Mutations*, 5th Mississippi State Conference On Differential Equations and Computational Simulations, Conference 10, 2003, pp. 33-53.

[28] K.T. Bogen, *Cell proliferation Kinetics and Multistage Cancer Risk Models*, Journal of the National Cancer Institute,Vol.81, No. 4, 1989, pp. 267-277.