Certain Methods for Investigating Epidemics and Preventing the Spread of Viruses in Self-Organizing Systems

Suren Poghosyan
Institute for Informatics and Automation Problems of NAS RA
Yerevan, Armenia
e-mail: psuren55@yandex.ru

Yeghisabet Alaverdyan
Institute for Informatics and Automation Problems of NAS RA
EKENG CJSC
Yerevan, Armenia
e-mail: ealaverdjan@gmail.com

Vahagn Poghosyan
Institute for Informatics and Automation Problems of NAS RA
Xilinx Armenia LLC
Yerevan, Armenia
e-mail: povahagn@gmail.com

Sergey Abrahamyan
Institute for Informatics and Automation Problems of NAS RA
Yerevan, Armenia
e-mail: serj.abrahamyan@gmail.com

Agit Atashyan
Institute for Informatics and Automation Problems of NAS RA
Synopsys Armenia CJSC
Yerevan, Armenia
e-mail: agitatashyan1@gmail.com

Hrachya Astsatryan
Institute for Informatics and Automation Problems of NAS RA
Yerevan, Armenia
e-mail: hrach@sci.am

Yuri Shoukourian
Institute for Informatics and Automation Problems of NAS RA
Yerevan, Armenia
e-mail: shouk@sci.am

***Abstract* — Investigation and control of epidemic phenomena, as well as introduction of methods of preventing the spread of viruses, are always up-to-date. In this context, the paper presents newer methods of modeling the dynamics of epidemics in self-organizing systems based on cellular automata. Methods and approaches for behavioral analysis of epidemic waves, also for statistical data processing, are promoted and validated in terms of obtained theoretical results. Recovery of causality in multilevel graphs simulating epidemic situations observed over time is given accordingly. Enabling collective observation and operation of dynamic processes along with conducting comprehensive analysis of epidemiological procedures and dynamical predicting of target parameters is achieved by embedding the multi-user software platform designed for the purpose. Cryptographic algorithms and data anonymization techniques of proven security have been suggested to protect personal information at all levels of data acquisition, storage, transfer and processing. A method for synthetic data generation enabling simulation and investigation of the spread of viruses has been introduced and estimated accordingly.**

***Keywords* — self-organizing system, cellular automata, epidemic phenomenon, simulation, cryptographic algorithms.**

I. INTRODUCTION

Computational modeling and simulation of epidemiological spread of viruses is a special subclass in self-organizing complex systems development.

Description of the evolution of epidemics with the involvement of self-organizing criticality models has already been proposed by many researchers. In particular, in [1], the idea of criticality was promoted in describing the occurrence and spread of virus outbreaks, also, the virus aggressiveness and the critical value of the virus density on the population are estimated. It was shown that near critical points, a remarkable increase in the percentage of active carriers is observed, and this fully coincides with the prognosis suggested by the theory of critical phenomena. Significant prolongation of epidemics is detected at values above the critical points of the virus density.

A mathematical model of self-organizing criticality is presented in [2]. Here, the authors proposed new approaches to predict the spread of viruses by presenting the nature of the spread by a mathematical function, which can be successfully used to measure localized infectious energy. In contrast to classical approaches, where the severity of the infection is traditionally calculated depending on the population, the authors obtained the mapping of the places where the infection was confirmed. The authors also interpret the model under investigation as a cellular automaton implementing a two-threshold self-organizing criticality.

In [3], the authors clearly state the applicability of critical phenomena investigation theory to the research and modeling of the dynamics of epidemic spreading among the real population.

The work [4] substantiates the applicability of self-organizing complex systems' properties in the field of epidemiology. In particular, coherency between the non-linear dynamic paradigm of pathological processes (transition from health to disease) and self-organizing critical systems is revealed, that is, the dynamic processes are characterized by: a phase transition; dynamics far from equilibrium; non-linearity; irreversibility, and heterogeneity. The second major achievement of the paper is the verification of the theory of self-organizing critical systems against reliable results of epidemiological investigation.

II. Proposed Solution, Arguments And Premises

The sandpile model [10-11] is the simplest theoretical model for describing self-organizing systems, and is fully applicable to simulate epidemiological outbreaks. In order to make the study of sandpile model based physical phenomena applicable and more effective, a joint research toolkit has been implemented with the possibility of simulation and visualization. The toolkit allows multi-users to calculate various physical and information characteristics of the model and to provide a visualization of joint research in changing, maintaining, and transferring the model states in 2D environment. It also enables simulation of cluster and cloud systems based on the sandpile and rotor-router models aiming at the optimal distribution of users' dynamic tasks [8-9]. Statistical data provided, the toolkit will help in visualizing and controlling epidemic waves on the multilevel graph thus contributing to the discovery of the causality of epidemics while detecting and tracking the epidemiological outbreaks. The set of essential characteristics of the model includes: the risk zone of the infection; disease intensity; speed of propagation of the infection frontal wave, etc.

By interpreting the spread of viruses as the dissemination of information on a multilevel graph in gossip/broadcast model, a number of our theoretical results achieved in this direction have been applied in construction and investigation of fault-tolerant schemes for epidemic spreading. Expansion and adaptation of the results obtained, also development of newer algorithms enables conducting a comprehensive analysis of epidemiological procedures and thereby predicting the models' target parameters dynamically. This also creates premises to make the behavior of the studied model at least partially controllable. Tracking and administration of the epidemiological waves on a multi-user software platform with involvement of gossip/broadcast optimal and fault-tolerant schemes and sandpile / rotor-router models, is a novelty in the field.

Provided accurate (or generated synthetic) population data, the topological space (multilevel graph) of multi-level possible contacts between patients are suggested to construct by modeling the relations, such as: «neighbors», «relatives», «colleagues», «residents», and others. On the derived graph, association of personal data with medical statistics and probabilistic characteristics obtained allows interpreting the epidemic wave as a stochastic process. Algorithms developed are based on the sandpile model and derived theorems. Also, the system dynamics follows basic rules developed for the purpose.

Below, certain specifications of the model formal description are given.

Consider an undirected graph $G=(V,E)$ with the set of vertices $V=\{v_1,v_2,\dots,v_N\}$ and the set of edges E. Each vertex $v_i\in V$ is assigned a variable h_i, which takes integer values and presents the height of the sand at that vertex, and h_i^{max} denotes the maximal allowed height for the vertex v_i in the graph G. For a d-dimensional lattice, $h_i^{max}=2d+1$.

Let C_T denotes the set of heights h_i, which determines the configuration of the system at a given discrete time T. A configuration is called stable, if all heights satisfy $h_i < h_i^{max}$. Given $\deg(v_i)$ indicatng the degree of v_i, the vertex v_i is called closed, if $h_i^{max}=\deg(v_i)$.

Consider a stable configuration C_T at a given time T. We add a grain of sand at a random vertex $v_i\in V$ by setting h_i to h_{i+1} (we assume that the vertex is chosen randomly with a uniform distribution on the set V). This new configuration, if stable, defines C_{T+1}. If $h_i \geq h_i^{max}$, then the v_i becomes unstable and topples by losing h_i^{max} grains of sand, meanwhile all neighbors of v_i receive one grain.

Note that if the vertex is open, then the system loses grains. During the toppling of the closed vertices, the number of grains is conserved. Note also that toppling of a vertex may cause some of its neighbors to become unstable. In this case, the neighboring vertices also topple according to the same toppling rule. Once all unstable vertices are toppled, a new stable configuration C_{T+1} is obtained. If the finite connected graph G has at least one open vertex, then all vertices become stable after finite number of topplings. Moreover, the new stable configuration is independent of the toppling order. Therefore, the dynamics of the system is well defined. Let $\hat{a}_i$ be an operator, which acts on sandpile configurations and adds a grain at vertex i. Obviously, $\hat{a}_i\hat{a}_j=\hat{a}_j\hat{a}_i$. This is the reason why the sandpile model is called Abelian.

III. Sand Game: Modeling Epidemics

Aimed at the implementation of the self-organizing behavior of the spread of diseases/viruses, in the scope of the paper, a pattern of a "serious game" has been developed. The game serves as a suitable platform for conducting visual control and learning of self-organizing criticality models. Required statistical data provided, the game software platform can be expanded and adapted accordingly to enable detecting and controlling real epidemic waves. Deployment of the multi-user platform for studying dynamic processes in the cloud, and then on a supercomputer, allows state and medical authorities to gain authenticated access to epidemiological data about newly infected, recovered, or self-isolated patients. Monitoring and full control of epidemic spread of diseases/viruses will be visualized for preferred surveilled areas.

The SandGame [13] is constructed and developed on the Abelian Sandpile model. The game implements a 2-dimensional connected lattice with $n\times n$ size, where n is the length (nodes count) by any direction of the lattice (torus).

On the 2D lattice, every edge with random direction is assigned an arrow parameter; also contours on the graph are removed. Let D_i denote the number of arrows directed to the node i. The model will become unstabilizable if each node topples at least once during the overall toppling. As each node has 4 neighbors, we put $4-D_i$ grains on every node i. Total grains' count will be $4\times n^2-\sum D_i=2\times n^2$, which is an indispensable number of grains that rends the model infinitely unstable. Statistical methods [10] applied, an infinitely unstable state is achieved for the average count of grains equal to $2.12588... \times n^2$. By toppling $n^2 - 1$ grains on a 2D lattice, users need to topple $0.1258 \times n^2$ grains in average in order to wrap up the game.

Architecture of the game comprises an engine and a user-friendly interface to interact with.

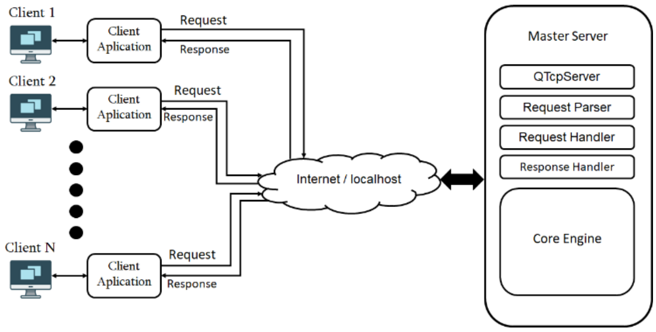

Figure 1. Architecture of the SandGame.

The entire software is written in the C++ language equipped with necessary QT library functions to implement Graphical user interface, also TCP Client-server architecture was involved to best solve communication issues between distributed network end-users.

Initially, when launching the Game, exactly two sands are allocated at each node of the Torus. Then, Players' names and dimensions of the torus are selected. The Game window illustrates a lattice where each node presents a patient.

Simulation of epidemics is conducted as follows: the user selects a node the state of which changes to an infectious. The infected node, in turn, makes its neighbors contactors. If the node/patient was previously indicated as a contractor, the node becomes infectious and thus spreads the infection to its neighbors, recursively.

The Game ends up with all the nodes/patients infection, which transfers the overall system into an infinite loop of recovering/infecting permanently.

Simulation allows to intervene into the process of the spread of the infection by imposing constraints in patients' displacement in order to bypass the infection wave and escape getting infected. Meanwhile, spreading the infection from one infected node to its neighbors in an arbitrary sequence of multiple repetitions is achieved by extending the rotor-router model for the generated multilevel graph. Here, spread of the infection from one node to another is considered to be a displacement of the virus from the current "starting point" selected during quasi-random walk to the current "end point".

For the sandpile model, we introduce a so-called Node Infection Risk Function depending on the number of infected nodes and on the membership of individuals to the corresponding risky age groups. The Risk Function has been developed by involving known theorems and our theoretical results obtained for sandpile and rotor-router models [9-11]. The simulation results obtained in the SRG model make it possible to locate the "non-infectious enclave" areas and enable the bypass or neutralization of the epidemic wave. Algorithms for detecting epidemics, also for discovering enclaves and predicting the spread of epidemics have been developed based on our results obtained for rotor-router models [5-7].

The rotor-router model on a graph describes a discrete-time walk accompanied by deterministic evolution of configurations of rotors randomly placed on vertices of the graph. We prove the following property: if at some moment in time the rotors form a closed clockwise contour on the planar graph, then the clockwise rotations of rotors generate a walk which enters the contour at some vertex v and performs a number of steps inside the contour so that the contour formed by rotors becomes anti-clockwise, and then leaves the contour at the same vertex v. This property generalizes the previously proved theorem for the case when the rotor configuration inside the contour is a cycle-rooted spanning tree, and all rotors inside the contour perform a full rotation. We use this proven property for an analysis of the sub- diffusive behavior of the rotor-router walk.

IV. Approaches For Data Anonymization And Synthetic Data Generation

To answer the needs to safely share and handle personal data, application of data anonymization and pseudonymization techniques are suggested. Data anonymization adds an extra layer of security for collected data, whereas data pseudonymization significantly reduces restrictions in the handling of personal data under the General Data Protection Regulation (GDPR). With pseudonymization, the subject descriptors are replaced by one or several pseudonyms. Data anonymization refers to the process of removing direct and indirect personal identifiers, such as an address, telephone number, image, date of birth, workplace, social status, etc. Data anonymization ensures that individuals are no longer identifiable such that the identity of a person cannot be deduced using anonymized data, and therefore, is not counted to be personal data, and thus, is out of scope of GDPR. With anonymized data, data breach fully eliminates, as there is no any link between the patient and his/her data. Classical techniques for anonymizing and randonymizing data can involve

- cryptographic data scrambling; directory replacement (changing the names of individuals while maintaining consistent relations between the values);
- ID to identify a patient. Information that directly identifies the individual is stored in a separate location. The anonymization is done via deletion of the separately stored information that identifies the patient;
- Information masking by involving speedy operations, such as XORing the data with predefined unique secret tokens.
- Generalization of data by introducing definite intervals instead of definite values makes it difficult (or even impossible) to retrieve the exact values associated with the patient.

Aimed at experimental data generation, algorithms for mirroring the statistical properties of the epidemiological data are suggested, which replace and do not reveal information regarding real patients. The following approaches have been implemented for synthetic data generation: given the knowledge of the disease/virus distribution, fit the synthetic data to the known distribution meanwhile handling possible overfittings. Deep generative models, such as Variational Autoencoder (VAE) and Generative Adversarial Network (GAN) can also be effectively used.

V. Approaches For Personal Data Protection

Generally, availability of information on personal data in databases is risky from the viewpoint of information security, and, in case of legislative restrictions, is sometimes even impossible. For the purpose, internationally recognized algorithms of proven security for storing and processing of collected confidential data are suggested in order to ensure secure collection, storage, transfer and processing of the information.

Thus, the patients' names, surnames, addresses and workplaces are suggested to encrypt via AES-256 symmetric cryptographic algorithm, whereas the patients' ages are concealed via Order-revealing encryption [12] allowing for efficient range queries, sorting, and threshold filtering on encrypted data. A fully homomorphic data encryption is suggested to operate the patients' encrypted data on the cloud.

To identify the users that request access to a system, network, or device, strong authentication techniques are recommended to meet the required level of assertion. Techniques, such as Password-based authentication; Multi-factor authentication; Certificate-based authentication; Biometric authentication; Token-based authentication, can be applied for asserting the legitimate users of the system.

Acknowledgment

This work was supported by the RA Science Committee, in the frames of the research project 20TTAT-RBe016.

References

[1] Y. Contoyiannis, S.G. Stavrinides, M.P. Hanias, M. Kampitakis, P. Papadopoulos1, S. Potirakis, "On the effectiveness of imposing restrictive measures in a graded Self-organized Criticality epidemic spread model: The case of COVID-19", arXiv:2004.00682, 2020.

[2] Stelian Ion and Gabriela Marinoschi, "A self-organizing criticality mathematical model for contamination and epidemic spreading", *Discrete & Continuous Dynamical Systems B*, vol. 22, no. 2, pp. 383-405, 2017.

[3] C.J. Rhodes, H.J. Jensen, and R.M. Anderson, "On the critical behaviour of simple epidemics", *Proc Biol Sci.* vol. 264, pp. 1639-1646, 1997.

[4] P. Philippe, "Epidemiology and Self-Organized Critical Systems: An Analysis in Waiting Times and Disease Heterogeneity", *Nonlinear Dynamics Psychol Life Sci*, vol. 4, pp. 275-295, 2000.

[5] V.S. Poghosyan and V.B. Priezzhev, "Euler tours and unicycles in the rotor-router model", *J. Stat. Mech.* P06003, 2014.

[6] Vl.V. Papoyan, V.S. Poghosyan and V.B. Priezzhev, "A loop reversibility of the rotor-router walk", *J. Phys. A: Math. Theor.* vol. 48, p. 285203, 2015.

[7] Vl.V. Papoyan, V.S. Poghosyan and V.B. Priezzhev, "Spiral Structures in the Rotor-Router Walk", *J. Stat. Mech.* 043207, 2016.

[8] Vl.V. Papoyan, V.S. Poghosyan and V.B. Priezzhev, "Rotor-Router Walk on a Semi-infinite Cylinder", *J. Stat. Mech.* 073209, 2016.

[9] H.E. Nahapetyan, J.-P. Jessel, Su.S. Poghosyan, Yu.H. Shoukourian, "A multi-user and multi-purpose CA simulator", *CSIT, IEEE conference proceedings*, pp. 23-26, 2017.

[10] Su.S. Poghosyan, V.S. Poghosyan, V.B. Priezzhev and P. Ruelle, "Numerical Study of the Correspondence between the Dissipative and Fixed Energy Abelian Sandpile Models", *Phys. Rev. E*, vol. 84, p. 066119, 2011.

[11] J.G. Brankov, V.S. Poghosyan, V.B. Priezzhev and P. Ruelle, "Transfer matrix for spanning trees, webs and colored forests", *J. Stat. Mech.* P09031, 2014.

[12] K. Lewi and D. J. Wu, "Order-revealing encryption: new constructions, applications, and lower bounds", *ACM Conference on Computer and Communications Security*, pp. 1167-1178, 2016.

[13] H.E. Nahapetyan, "An Example of a Multiplayer Serious Game on 3D Sandpile Model", *Mathematical Problems of Computer Science* vol. 48, pp. 5-13, 2017.