Worm Analysis through Computer Simulation (WAtCoS)

Madihah Mohd Saudi, Kamaruzzaman Seman, Emran Mohd Tamil and Mohd Yamani Idna Idris.

Abstract— Computer viruses have received a lot of attention. In fact, the best-known viruses have not been viruses at all, but worms, programs that spread through networks instead of modifying programs. Both viruses and worms reproduce themselves and defensive measures have focused on stopping or slowing their spread. Ultimately, though, there is no defense better than a comprehensive security strategy that embraces user education, crisis-response teams, and technologically sound security measures including, but not limited to, those that relate specifically to the threats posed by viruses and worms. Defense against harm can consist of preventing the harm from occurring, limiting the extent of the harm, or recovering from the harm after it has occurred. This research aims to resolve the confusion in identifying visualization, simulation and games in teaching malware analysis. Computer simulation has greater impact and based on research that had been carried out it is identified as one of the best approach in teaching worm analysis.

Index Terms— Worm analysis, visualization, simulation, game.

I. DEFINITION

A. Visualization

Visualization is any technique for creating images, diagrams, or animations to communicate a message [Herbert and James, 1998]. Visualization through visual imagery has been an effective way to communicate both abstract and concrete ideas since the dawn of man. Examples from history include cave paintings, Egyptian hieroglyphs, Greek geometry, and Leonardo da Vinci's revolutionary methods of technical drawing for engineering and scientific purposes.

Visualization today has ever-expanding applications in science, engineering product visualization, all forms of education, interactive multimedia, medicine etc. Typical of a visualization application is the field of computer graphics. The invention of computer graphics may be the most important development in visualization since the invention of central perspective in the Renaissance period. The development of animation also helped advance visualization.

B. Simulation

A simulation is an imitation of some real thing, state of affairs, or process. The act of simulating something generally entails representing certain key characteristics or behaviors of a selected physical or abstract system [Narayanasamy et al. 2005]. A computer simulation is an attempt to model a real-life or hypothetical situation on a computer so that it can be studied to see how the system works.

C. Computer Game

The history of the computer game is, in parts, a history of technology. The computer game requires technology capable of handling large amounts of data and of representing this data. The first computer game is generally assumed to be the game *Spacewar!*, developed in 1962 at MIT (Stephen Russell). The players can shoot each other, turn their ships, and accelerate. Naturally, the goal is to hit the other player before being hit yourself.

Formally, a game is best defined as a goal-directed and competitive activity that involves some form of conflict [Sauvé et al. 2005], conducted within a framework of agreed rules [Lindley 2003]. The operator and/or user of a game is referred to as the player or gamer. A game is a structured or semi-structured activity, usually undertaken for enjoyment and sometimes also used as an educational tool. The term "game" is also used to describe simulation of various activities e.g., for the purposes of training, analysis or prediction.

Simulation Games are just one genre of computer games. Simulation games are mixtures of games of skill, chance, and strategy that result in the simulation of a complex structure. Most of the simulation games are general games for educational purposes, but more and more company specific games, tailored for specific organizational aims can be seen.

II. PREVIOUS WORKS

A. Visualization

[Donna Gresh et al. 2001] describe a visualization system designed for interactive study of proteins in the field of computational biology. Their system incorporates multiple, custom, three-dimensional and two-dimensional linked views of the proteins. The visualization environment that they have developed is intended to facilitate the study of proteins for researchers in the field of computational biology, where the motion and behavior of proteins and other molecules are studied in the computer rather than in the test tube.

Madihah Mohd Saudi is with the Faculty Science and Technology, Islamic Science University of Malaysia (USIM), Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia (email: madihah@usim.edu.my).

Professor Dr. Kamaruzzaman Seman is with the Faculty Science and Technology, Islamic Science University of Malaysia (USIM), Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia (email: drkzaman@usim.edu.my).

Emran Mohd Tamil is with the Faculty of Computer Science and Information Technology, University of Malaya (UM), Kuala Lumpur, Malaysia (email: emran@um.edu.my).

Mohd Yamani Idna Idris is with the Faculty of Computer Science and Information Technology, University of Malaya (UM), Kuala Lumpur, Malaysia (email: yamani@um.edu.my).

Another previous work on visualization is by [Thomas Baxley et al. 2006]. They develop an animated visualization tool to teach the concepts of various attacks on Local Area Networks. Understanding how LAN attacks work and knowing the vulnerabilities in the protocol design of LANs are an important part of education in computer networks and network security. Understanding how attacks on LAN work requires knowledge in both network hardware and protocol software. Students must know how hubs, switches, network interface cards (NICs) and Address Resolution Protocol (ARP) work in great details. They are also required to know the data structures like ARP cache table, switch port mapping table and Ethernet frames and ARP packets. In addition, the actual attack includes multiple phases including scanning, table poisoning and traffic interception. Because of these complexities, many network security class students in the previous semesters experienced difficulties understanding these concepts even after hours of lectures. This tool is targeted to assist instructors who teach college level network security and computer networks. The tool accurately and realistically shows attacks such as ARP Poisoning, Port Stealing and MAC Flooding. They integrated features such as high degree user interaction, play and pause, tooltips and quizzes. This software is intended to be used in undergraduate computer networks and network security courses. However, anyone with interest learning LAN security can benefit from the software.

While [Dino Schweitzer and Wayne Brown, 2007] propose the use of interactive visualizations as an effective means to actively engage students in the classroom. Engaging students in the learning process has been shown to be an effective means for education. Several methods have been proposed to achieve this engagement for computer science and other disciplines. Active learning is one such technique that incorporates interactive classroom activities to reinforce concepts and involve the students. Visualizations of computer science concepts such as algorithm animations can be used for these activities. They have developed and used ICV's (Interactive Classroom Visualizations) in several of computer science courses including algorithms, data structures, computer graphics, security, cryptography, and introductory computer science.

B. Simulation

Previous scanning worms have been significant sources of network congestion and have had catastrophic effects on switches, routers and end systems. [Ihab Hamadeh et al. 2005] focus on the simulation of the secondary resource-exhaustion effects that scanning worms cause on network protocols operating in Internet routers. The SQL Slammer/Sapphire worm and the Ramen worm generated large volumes of scans destined to multicast addresses creating a storm of Source Active (SA) messages that propagated across Multicast Source Discovery Protocol (MSDP) enabled networks. Specifically, they describe a preliminary simulation study on the effect of the spread of the Ramen and SQL Slammer/Sapphire worms on the multicast infrastructure and their ultimate goal is to create a realistic simulation platform to evaluate and tune techniques to mitigate the effects of scanning worms on network protocols.

Another previous work is about on how fast and how far a worm could spread by making use of mobile computers and wireless networks. [Everett Anderson et al. 2005] use existing data of real users working in a campus-wide wireless environment over the course of several months to provide realistic data on mobility and connectivity patterns. They perform simulations based on this data to observe how a worm might propagate using only local wireless connections and human user mobility.

Next previous work is from [Jin Feng 2002]. He shares the experience of using computer simulation technology in an interior lighting design class to improve the teaching and learning environment. The use of simulation technology has revolutionized the teaching and learning environment of lighting design. Through the virtual experience of the complete cycle of design, build and evaluation, the students obtained better understanding of the relationship between lighting plans, specifications, selection of interior materials, and actual lighting effects and technical measurement. The use of simulation technology also opens up new possibilities to support our effort in the paradigm change from illuminance-based design to luminance-based design, and eventually realize the integration of interior design and lighting design.

C. Computer Game

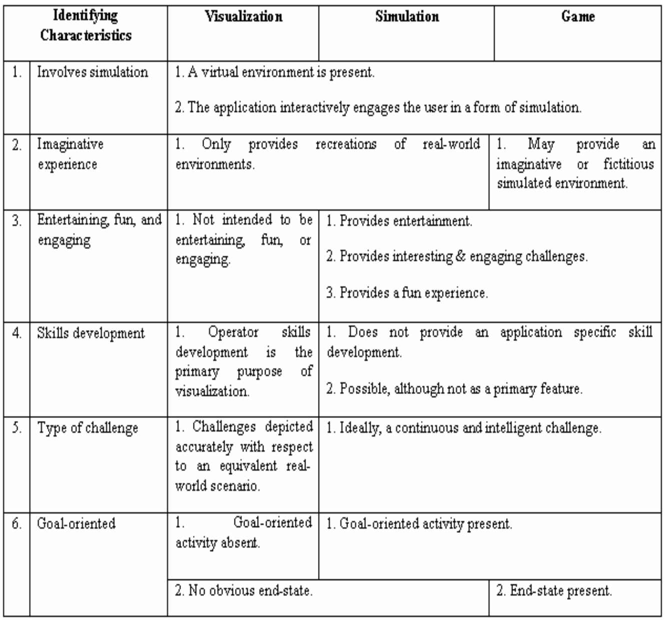
The Security Protocol Game [Dr Leonard G C Hamey, 2002] is a highly visual and interactive game for teaching secure data communication protocols. This game provides a simple representation of public key and secret key cryptographic systems and related algorithms. Students use the game to simulate protocols and explore possible attacks against them. Specifically, the game provides representations for plain text and encrypted messages, message digests, digital signatures and cryptographic keys. Using these representations, students can construct public key certificates and perform multiple encryption, tunneling and encrypted key transmission. They can simulate a wide range of protocols including authentication, key exchange and blind signature protocols. Application protocols such as Transport Layer Security and Pretty Good Privacy can be simulated in detail. The game clearly reveals the key issues of confidentiality, integrity, authentication and non-repudiation in secure data communications. Used as a small group learning activity, students gain a deep understanding of protocol design and operation issues. The game is suitable for use in tertiary and professional education courses for managers and information technology students at all levels.

Second previous work is on simulation game for the course “Simulation Game in Electric Economics”. Their paper [A. Turtiainen et al. 2002] presented the course and how a WWW-application has been used in teaching economics as a game. The basic idea of the simulation game was to teach the students how to operate on the liberalized electricity markets. This was done by simulating management of a fictitious electricity company. In order to run their companies, students needed to handle the determination of electricity sales tariffs, operate on the liberated (e.g. spot market and the financial instruments’ market) and also take care of the company’s image. The simulation game was designed for a web-use only. This electric economics game appeared to be a good way of understanding the basics of electric economics. It does, however, require some basic knowledge but its power lies in the way of doing things. Usually students learn better if they have to use their abilities instead of only reading or writing.

Learning scientists are increasingly turning to computer and video games as tools for learning. [Kurt Squire et al. 2003] examines what learning occurs when an electromagnetism simulation game is used in a school for students. Game mechanics enabled students to confront weaknesses in understandings, and physics representations became tools for understanding problems. The goal of Supercharged! Game is to help learners build stronger intuitions for electromagnetic concepts. With this game, they suggest that simulation computer games can be effective tools in helping students understand complex physics phenomena.

III. COMPARISON

Based on the research and observation that had been carried out, it is need for carried out an abstract distinction among visualization, simulation and games by building and assessing a common taxonomy based on the characteristics (The results are presented in Figure 1).



Identifying Characteristics		Visualization	Simulation	Game
1.	Involves simulation	1. A virtual environment is present. 2. The application interactively engages the user in a form of simulation.		
2.	Imaginative experience	1. Only provides recreations of real-world environments.		1. May provide an imaginative or fictitious simulated environment.
3.	Entertaining, fun, and engaging	1. Not intended to be entertaining, fun, or engaging.	1. Provides entertainment. 2. Provides interesting & engaging challenges. 3. Provides a fun experience.	
4.	Skills development	1. Operator skills development is the primary purpose of visualization.	1. Does not provide an application specific skill development. 2. Possible, although not as a primary feature.	
5.	Type of challenge	1. Challenges depicted accurately with respect to an equivalent real-world scenario.	1. Ideally, a continuous and intelligent challenge.	
6.	Goal-oriented	1. Goal-oriented activity absent.	1. Goal-oriented activity present.	
		2. No obvious end-state.		2. End-state present.

Figure 1. Comparison between Visualization, Simulation and Game

1. Involves simulation. While using visualization, simulation and game, the applications in question have to be identified as containing some simulation elements. In particular, a virtual environment that tries to recreate some form of fictitious or real-world environment is necessary.

2. Imaginative experience. An imaginative virtual experience may include experiences that have elements of fiction or fantasy, or an experience that simply deviates from reality. In the quest to provide interesting and exciting worlds, most games involve "unreal" fictional elements that contribute to an imaginative experience. Unlike visualization and simulation, there are games that simulate the presence of fantasy worlds (e.g., Master of Orion III, MechWarrior 4). But visualization and simulation cannot use imaginative elements as an accurate representation of the real world. It is necessary to train operators to develop their skills in virtual environments in real-world situations. Thus the absence of imaginary experiences can be used to distinguish visualization and simulation from games.

3. Entertaining, fun and engaging. An entertaining experience may be defined as an interesting or amusing one. An interesting experience engages the attention of the player and provides excitement, and might also arouse curiosity or emotion. The intent of games and simulation is to engage players in a fun and entertaining experience, while the intent of visualization is to train and develop the skills of its operators.

4. Skills development. The motivation for developing the visualization is to maximize the rate at which operators develop their skills, while the operators' objective is to maximize their performance in the task being simulated. For games and simulation, however, the entertainment features of an application are the highest priority. For these reasons, visualization support high-fidelity simulations with a greater degree of verisimilitude, while games and simulation games only make a best effort at creating a representation that is consistent and accessible.

5. Type of challenge. Ideally, games and simulation attempt to provide a continuous flow of intelligent challenges to engage the players. Lately, much research was done to introduce emergent challenges in games (i.e., the notion of a "good surprise") to eliminate lackluster challenges due to repetitive predefined content. However, introducing random, varying, unpredictable, or sometimes nondeterministic, content in visualization to create interesting and engaging challenges is undesirable and, in many cases, inappropriate. This is because the challenges in visualization have to be well-designed reproductions of real-world scenarios, so that an operator can develop useful skills, reproducible in real-life, without visualization. The presence of random, unpredictable, varying, and non-deterministic challenges can then be used to identify games and simulation.

6. Goal-oriented. Goal-oriented activities include any activity or set of activities that are conducive to achieving a desirable end-state in a game by a player. Simulation and game are goal-oriented activity but visualization not a goal-oriented activity. The end-state of a game is that associated with the notion "end of the game." It is achieved when an adequate number of victory conditions, as determined by the game, are met. There are a number of possible victory conditions. Visualization and simulation not involved end-state. However, in game, the end-state present.

IV. ADVANTAGES AND DISADVANTAGES

Nowadays the visualization is the most important approach to extract relevant information from the huge of data produced by today's computational and experimental works. Visualizations are now recognized as a powerful approach to get insight on large datasheets produced by scientific experimentation's and simulations and the introduction of these 3D models are a way for a better understanding of this information, and to a better performance of all visualization process. However, visualization design stresses on achieving a higher accuracy in the situation/environment being visualized. Elements to make it interesting or exciting are either not considered or avoided.

As for simulation, traditionally, simulations were used to study the behavior of a system as it evolves over time. This is done by first modeling the system and then developing a simulation model. The model usually takes the form of a set of assumptions concerning the operation of the system. The assumptions can be mathematical, logical, or symbolic relationships between the entities/objects of interest. The disadvantage for simulation is the model has to be validated before it can be used to predict or reproduce the behavior of the system being modeled under varying sets of circumstances. More recently, simulation models are used in a real-time interactive mode to derive pleasure and enjoyment to provide entertainment. Simulations can be applied for the teaching of facts, concepts and principles and to train specific skills. In fact, the participants in a simulation should be able, after a training program, to apply learned principles to new situations such as: make decisions, solve problems and work in small groups.

Compared to game, it is a powerful medium for learning and self expression. Essentially, games are developed for different purposes, but two of them are seen to be more relevant: education and demonstration. Also the purpose can be detailed: to describe which is illustrate or demonstrate an issue, a situation or a process; to demonstrate which is a method or a technique; to practice which is to train and educate; to reflect which is an experiment and obtain response; to prepare which is to increase or direct the attention towards a certain situation. While the advantage, it is important to not formulate the purpose of the game too wide and it must be developed according to exact/clear focus. This is because it will be use by different groups in different situations.

V. INTEGRATION OF TEACHING MALWARE ANALYSIS

Malicious code (or malware) is defined as software that fulfills the deliberately harmful intent of an attacker. Malware analysis is the process of determining the behavior and purpose of a given malware sample (such as a virus, worm, or Trojan horse). This process is a necessary step to be able to develop effective detection techniques and removal tools. Currently, malware analysis is mostly a manual process that is tedious and time-intensive. To mitigate this problem, a number of analysis tools have been proposed that automatically extract the behavior of an unknown program by executing it in a restricted environment and recording the operating system calls that are invoked [Andreas et al. 2006]. The problem of dynamic analysis tools is that only a single program execution is observed. Unfortunately, however, it is possible that certain malicious actions are only triggered under specific circumstances (e.g., on a particular day, when a certain file is present, or when a certain command is received). Figure 2 shows the flow for handling worm attack which is produced based on our research. This procedure had been tested in our lab and it works effectively and efficiently.

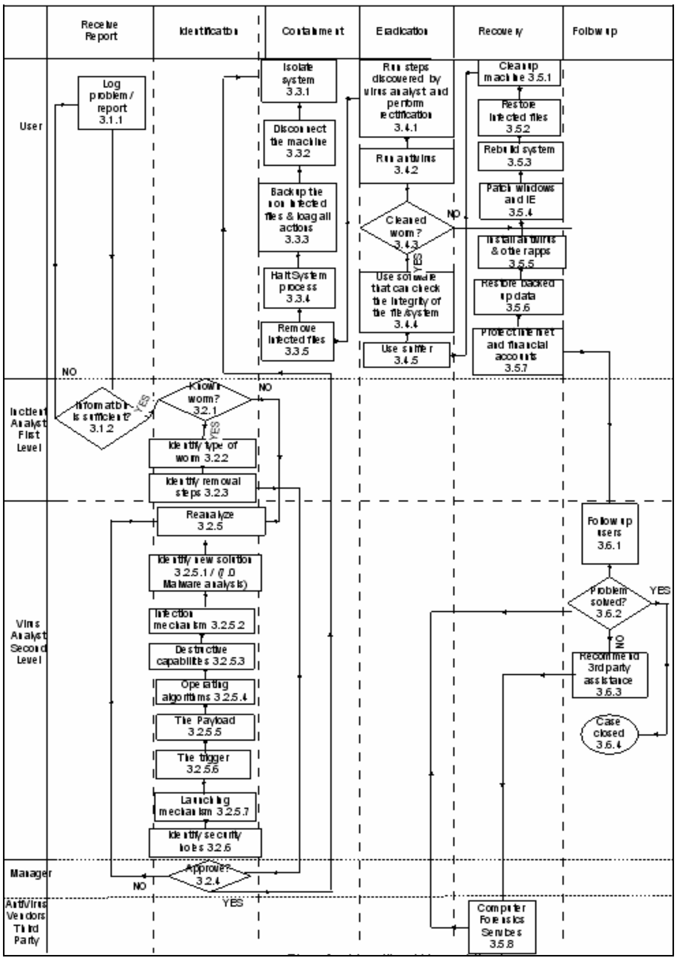

Receive Report
Identification
Containment
Eradication
Recovery
Follow up
User
Log problem / report 3.1.1
Isolate system 3.3.1
Disconnect the machine 3.3.2
Backup the non infected files & log all actions 3.3.3
Halt System process 3.3.4
Remove infected files 3.3.5
Run steps discovered by virus analyst and perform rectification 3.4.1
Run antivirus 3.4.2
Cleaned worm? 3.4.3
NO
YES
Use software that can check the integrity of the file/system 3.4.4
Use sniffer 3.4.5
Cleanup machine 3.5.1
Restore infected files 3.5.2
Rebuild system 3.5.3
Patch windows and IE 3.5.4
Install antivirus & other apps 3.5.5
Restore backed up data 3.5.6
Protect internet and financial accounts 3.5.7
NO
Incident Analyst First Level
Information is sufficient? 3.1.2
YES
Known worm? 3.2.1
NO
YES
Identify type of worm 3.2.2
Identify removal steps 3.2.3
Virus Analyst Second Level
Reanalyze 3.2.5
Identify new solution 3.2.5.1 / (7.0 Malware analysis)
Infection mechanism 3.2.5.2
Destructive capabilities 3.2.5.3
Operating algorithms 3.2.5.4
The Payload 3.2.5.5
The trigger 3.2.5.6
Launching mechanism 3.2.5.7
Identify security holes 3.2.6
Follow up users 3.6.1
Problem solved? 3.6.2
YES
NO
Recommend 3rd party assistance 3.6.3
Case closed 3.6.4
Manager
Approve? 3.2.4
NO
YES
AntiVirus Vendors Third Party
Computer Forensics Services 3.5.8

Fig. 2. Flow for Handling Worm Attack

For this paper, we would like to investigate method of teaching in malware analysis. Helping students to understand complex ideas on malware analysis will always be problematic for teaching professionals. Often, the students can be limited by not only their imagination, but by their experiences. When trying to explain something that is outside of the students' imagination, it is often helpful to have either simple animations or even interactive simulations that the students can explore. The creation of interactive simulations can greatly help educators get their point across and, as a result, help students comprehend the ideas.

Based on research that has been done, simulation technique can be applied to malware analysis, with educational objectives. It is hoped that the learning benefits of the simulation will transfer to the real world - which learners will be able to apply knowledge that they have gained to problems outside of the simulation. Fig. 3, 4, 5 and 6 shows the storyboard on malware analysis simulation.

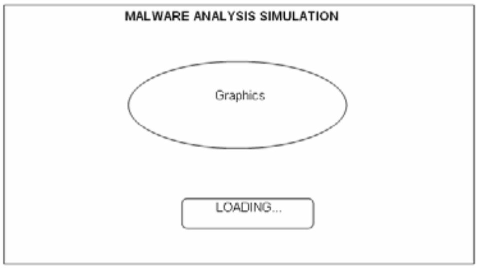


Fig.. 3. Loading Page

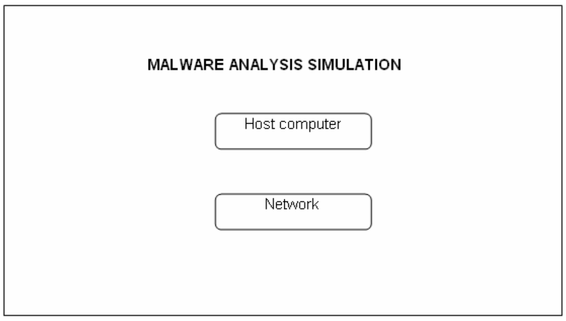


Fig.. 4. Menus

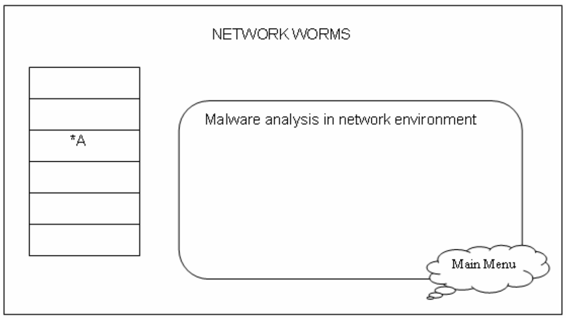


Fig.. 5. Type 1 – Network Worms

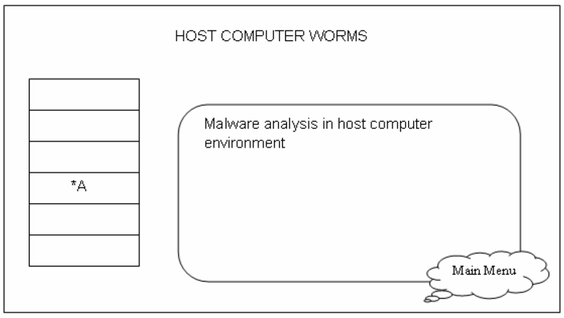


Fig. 6. Type 2 - Host Computer Worms

VI. CONCLUSION

Simulation is a useful tool in many areas of computer science education. A review of examples of simulation indicates that great potential can be realized much more rapidly. The simulation is at least as effective as other methods for teaching knowledge about facts, concepts and application of knowledge. It is believe that simulation in worm analysis have greater impact on participant's attitudes than other instructional techniques.

REFERENCES

[1] Joao Rafael Galvao, Paulo Garcia Martins, Mario Rui Gomes. 2000. "Modeling Reality with Simulation Games for a Cooperative Learning". Proceedings of the 2000 Winter Simulation Conference.

[2] Donna Gresh, Frank Suits, and Yuk Yin Sham. 2001. "Case Study: An Environment for Understanding Protein Simulations Using Game Graphics". IEEE.

[3] Dino Schweitzer and Wayne Brown. 2007. "Interactive Visualization for the Active Learning Classroom". ACM 1-59593-361-1/07/0003.

[4] Thomas Baxley, Jinsheng Xu, Huiming Yu, Jinghua Zhang, Xiaohong Yuanand Joseph Brickhouse. 2006. "LAN Attacker: A Visual Education Tool ". ACM 1-59593-437-5/00/0006.

[5] Ihab Hamadeh, Jason Hart, George Kesidis and Venkat Pothamsetty. 2005. "A Preliminary Simulation of the Effect of Scanning Worm Activity on Multicast". Proceedings of the Workshop on Principles of Advanced and Distributed Simulation (PADS'05), IEEE.

[6] Jin Feng. 2002. "Computer Simulation Technology and Teaching and Learning Interior Lighting Design". ACM.

[7] Everett Anderson, Kevin Eustice, Shane Markstrum, Mark Hansen and Peter Reiher. 2005. "Mobile Contagion: Simulation of Infection & Defense". Proceedings of the Workshop on Principles of Advanced and Distributed Simulation (PADS'05), IEEE.

[8] A. Turtiainen, T. Mannila, S. Kuusiluoma and L. Korpinen. 2002. "Simulation Game in Teaching Electric Economics". IEEE.

[9] Dr Leonard G C Hamey. 2002. "Teaching Secure Communication Protocols Using a Game Representation". Australian Computer Society, Inc.

[10] Kurt Squire, Mike Barnett, Jamillah M. Grant and Thomas Higginbotham. 2003. "Electromagnetism Supercharged! Learning Physics with Digital Simulation Games". ACM.

[11] Herbert L. Dershem and James Vanderhyde. 1998. "Java Class Visualization for Teaching Object-Oriented Concepts". ACM.

[12] Viknashvaran Narayanasamy, Kok Wai Wong, Chun Che Fung and Shri Rai. 2005. "Distinguishing Games and Simulation Games form Simulators". ACM.

[13] Ero Carrera and Gergely Erdélyi. 2004. "Digital Genome Mapping – Advanced Binary Malware Analysis". Virus Bulletin Conference September 2004.