Potential Issues in Novel Computational Research: Artificial Immune Systems

Anjum Iqbal[1] and Mohd Aizaini Maarof[2]

Group on Artificial Immune Networks and Security (GAINS), Faculty of Computer Science and Information Systems
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
Emails: [1]anjum@siswa.utm.my, [2]maarofma@fsksm.utm.my

Abstract- *Recent trends in computational research show the significance of computations based biological research, for example, bioinformatics and system biology, and biologically inspired computational research, which includes genetic algorithms, artificial neural networks, and evolutionary programming. Artificial Immune Systems (AIS) is a novel computing paradigm inspired from Human Immune System (HIS). The dramatic features of HIS like distributability, uniqueness, adaptability, mobility, imperfect-detection, and anomaly-detection, have grabbed the attention of computational researchers for adopting HIS principles to design AISs that are robust, scalable, highly flexible, resilient to sub-versioning, and degrade gracefully. Designers of AIS make use of their creative abilities for designing computational systems by extracting corresponding metaphors from HIS. AIS is growing as a core paradigm for wide application area including: intrusion detection, data analysis, pattern recognition, optimization, robotics, distributed control, and bioinformatics etc. The field of AIS, being in its infancy, bears enormous research potentials. This paper aims to throw light on some prospective research issues in the field. The effort is based on; 1) latest research references and 2) our experience for pioneering AIS based research infrastructure in the Universiti Teknologi Malaysia (UTM). It may help new researchers to jump start AIS research and save their precious time.*

Keywords: *Artificial Immune Systems, biologically inspired computing, novel research, potential issues*

1. INTRODUCTION

Artificial Immune Systems (AIS) is a novel computing paradigm inspired from Human Immune System (HIS) [1, 2]. The dramatic features of HIS, like distributability, uniqueness, adaptability, mobility, imperfect-detection, and anomaly-detection, have grabbed the attention of computational researchers for adopting HIS principles [3] to design AISs, which are robust, scalable, highly flexible, resilient to sub-versioning, and degrade gracefully [4]. Designers of AIS make use of their creative abilities for designing computational systems by extracting corresponding metaphors from HIS. The success of all analogies between computing and living systems ultimately rests on our ability to identify the correct level of abstraction [5]. AIS is growing as a core paradigm for wide application area including: intrusion detection, data analysis, pattern recognition, optimization, robotics, distributed control, and scheduling etc [1, 2]. The field of AIS, being in its infancy, bears enormous research potentials. The influential work that initiated AIS research was a paper by Farmer and Perelson [6] in 1986. The first AIS model ARTIS [7], claimed as general model, was developed in 2000.

The AIS research is *interdisciplinary* in nature, that is, involving both biological and computational sciences. This type of research is considered complex for a number of reasons, especially at its early stages [9]. The researchers have to understand two distinct fields and then, making use of their creative abilities, apply their learning to various problem areas. Therefore, exploration of potential research issues in interdisciplinary research areas is a relatively tough task. The main contribution of this paper might be the presentation of some general but potential research issues without going into complex details of interdisciplinary research. The included issues are; scaling of detectors, matching rules, costimulation, main goal of immune system, and distributed testing. The effort is based on; 1) latest research references and 2) our experience for pioneering AIS based research infrastructure in the Universiti Teknologi Malaysia (UTM).

Fortunately, a lot of latest research literature, about AIS, is available on the web and can be accessed freely. This includes Masters and PhD thesis, bibliographies, conference proceedings [11], technical reports and open-source software [8]. Also researchers in the field are encouraging and responsive. This witnesses the commitment of AIS research community for promoting this novel research area.

We, the GAINS members, are trying to avail the opportunity of joining AIS research community and capturing pace of the research at this earlier stage. Our significant effort till now is the reverse engineering of LYSIS [8]; an AIS based *intrusion detection system* (IDS). This has opened the door for our AIS based research. We hope to put further efforts in various application areas in near future. The current main focus of our research is AIS based computer security, especially IDS.

The following section 2 gives the overview of AIS without going into complex biological details, so to absorb readers' attention and motivate them for the novel computational research. Where required, the reader can refer the literature for further details. Section 3 describes

the basic AIS algorithms and their integration to achieve the main AIS task. For simplicity, we have described only those details, which are required to elaborate the potential research issues given in section 4. There are a number of researchable problems in the field, but the issues described in section 4 might be considered among the most general issues. Section 5 concludes the paper.

2. ARTIFICAL IMMUNE SYSTEM OVERVIEW

One vital system, which is involved in constant activity, never shirking its duty, is the defense system (immune system). This system protects the body from all kinds of invaders day and night and works with great assiduity, just like a fully equipped army for the host body, which it serves [10]. The immune system is a complex of cells, molecules and organs which has proven to be capable of performing several tasks, like pattern recognition, learning, memory acquisition, generation of diversity, noise tolerance, generalization, distributed detection, secheduling and optimization. Based on immunological principles, new computational techniques are being developed, aiming not only at a better understanding of the system, but also at solving engineering problems. The *immune engineering* makes use of immunological concepts in order to create tools for solving demanding machine-learning problems using information extracted from the problems themselves [1, 2].

Artificial Immune System research is connected to immunological research. The connection of interdisciplinary research, that is computations intensive biological research and biologically inspired computational research is shown in figure 1. The story starts with the research in a biological laboratory. This research provides the principles of biological systems. Following these principles, computational scientists design computational models of biological systems. These models are either used to design simulators, which help biological researchers in unveiling the truth and refining previously explored biological principles, or they help computational scientists in designing biologically inspired computational techniques like artificial neural networks, genetic algorithms, evolutionary programming and the novel field of artificial immune systems. The simulation of biological systems is relatively straightforward task because of the exact mimicking of the biological system. On the other hand, applying biological principles to computational systems requires deep involvement of the human creativity. The success of all analogies between computing and living systems ultimately rests on our ability to identify the correct level of abstraction [5]. It means that the designer of an AIS has to thoroughly understand the principles of HIS, as explored by an immunologist (for details readers are referred to [10] and [8]). Then he will have to analyze the computational system for which he is going to design an AIS. The success of the system depends on the success of abstractions he explores for the required AIS.

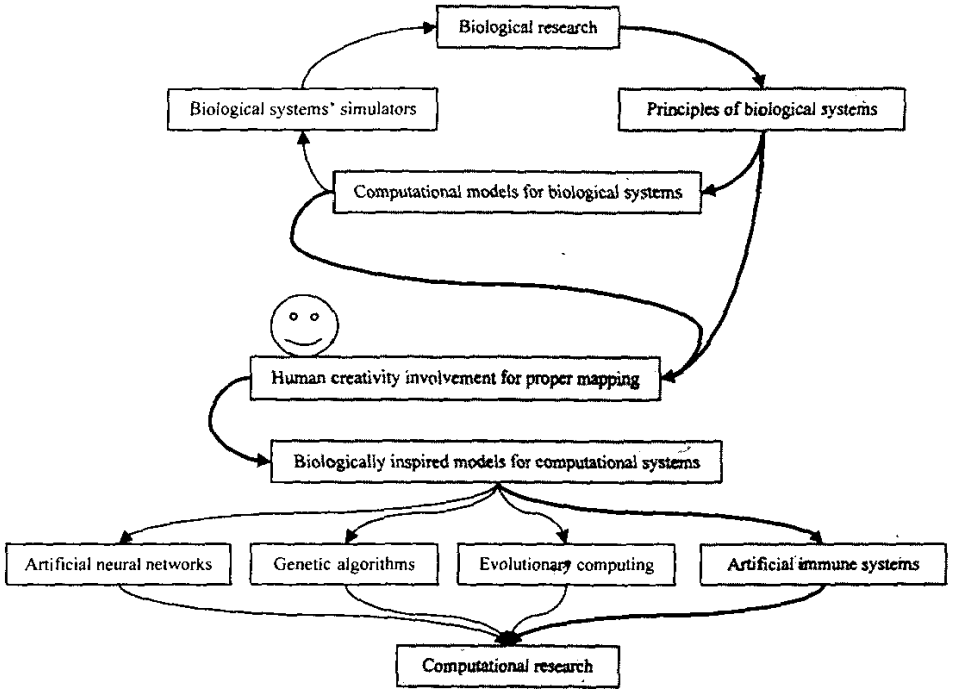

Figure 1. the connection of two distinct fields in interdisciplinary research

The influential work that initiated AIS research was a paper by Farmer and Perelson [6] in 1986. This paper introduced a dynamic model of the immune system based on Jerne's network hypothesis that was simple enough to simulate on a computer. The ARTIS [7] is the first model aiming to be the general AIS model. S.A Hofmeyr and S. Forrest [7] believe that it is fruitless to design intelligent systems in complete isolation from the environments in which they exist. The most natural domain in which to begin applying immune system mechanisms is computer security [12]. The LYSIS [7] is an application of ARTIS to the field of computer security, a network based IDS. It is open-source and downloadable from [8]. LYSIS is not suitable for industrial level applications due to numerous limitations, but provides strong base for designing and developing industrial level AIS based IDSs. The AIS is not only limited to computer security, it is also delivering benefits to a variety of application area including robotics, data analysis, scheduling, anomaly detection, pattern recognition [1, 2].

3. ARTIFICIAL IMMUNE SYSTEM ALGORITHMS

As discussed in the previous section, AIS may not be the exact replica of HIS but it is designed for an application extracting suitable abstractions from natural immune system. Different cells, molecules and mechanisms of HIS are mapped to well-formatted strings of AIS according to the application requirements. Following sections describe basic AIS algorithms, negative selection and clonal selection, derived from HIS mechanisms.

3.1 Negative Selection Algorithm

The human immune system contains an organ called *thymus* that is located behind the breastbone. The thymus is responsible for the maturation of T-cells, one of the important cell types in immune system. It is protected by a blood barrier capable of efficiently excluding non-self *antigens*, cells or molecules not belonging to human body, from the thymic environment. Thus, most elements found within the thymus are representative of self instead of non-self. After T-cells are generated, they migrate into the thymus where they mature. During this maturation, all T-cells that recognize self-antigens, the cells and molecules belonging to human body, are excluded from the population of T-cells; a process termed *negative selection*. All T-cells that leave the thymus to circulate throughout the body are said to be *tolerant* to self, i.e., they do not respond to self [1, 2, 4, 10].

A negative selection algorithm [13] has been proposed in the literature with applications focused on the problem of anomaly detection, such as computer and network intrusion detection, time series prediction, image inspection and segmentation, hardware fault tolerance, and pattern recognition.

Given an appropriate problem representation the designer of AIS defines the set of strings to be protected and call it the *self- set*. Based upon the negative selection algorithm, generate a set of strings called *detectors* that will be responsible to identify all elements that do not belong to the self-set, i.e., the non-self elements. The negative selection algorithm runs as shown in figure 2. After generating the set of detectors, the next stage of the

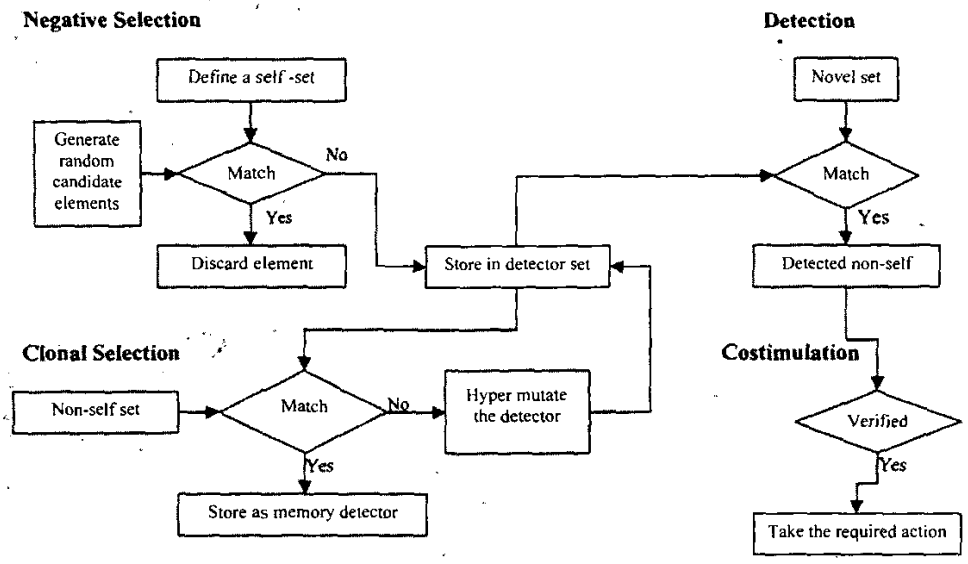

Figure 2. Basic algorithms of Artificial Immune Systems and their integrated function

algorithm consists in *monitoring* the system for the presence of non-self. Assume a set of strings that might be composed of the self set plus other new strings, or it can be a completely novel set.

For all elements of the detector set that corresponds to the non-self strings, check if it recognizes (matches) an element of novel set and, if yes, then a non-self string was recognized and an action has to be taken. The resulting action of detecting non-self varies according to the problem under evaluation.

3.2 Clonal Selection

Complementary to the role of negative selection, *clonal selection* (see figure 2) is the theory used to explain how an immune response is mounted when a non-self antigenic pattern is recognized by a B-cell, another important type of immune system cells (reader is referred to [1, 2, 4, 10] for details). In brief, when a B-cell receptor recognizes a nonself antigen with a certain affinity, it is selected to proliferate and produce antibodies in high volumes. The antibodies are soluble forms of the B-cell receptors that are released from the B-cell surface to cope with the invading nonself antigen. Antibodies bind to antigens leading to their eventual elimination by other immune cells. Proliferation in the case of immune cells is asexual, a mitotic process; the cells divide themselves (there is no crossover). During reproduction, the B-cell progenies (clones) undergo a hyper *mutation*[1] process that together with a strong selective pressure, result in B-cells with antigenic receptors presenting higher affinities with the selective antigen. This whole process of mutation and selection is known as the *maturation of the immune response* and is analogous to the natural selection of species. In addition to differentiating into antibody producing cells, the activated B cells with high antigenic affinities are selected to become memory cells with long life spans. These memory cells are pre-eminent in future responses to this same antigenic pattern, or a similar one [14].

4. POTENTIAL RESEARCH ISSUES

4.1 Scaling Issue of Detectors

When the negative selection algorithm is applied to a broader range of self it requires generation of exceptionally large amounts of detectors and causes an unacceptably long computation time. Also when the self definition widened, the string needed to encode a detector lengthened. As the result of the long length of detectors, the number of detectors required to gain an acceptable *false negative error*, incorrect detection of self, become huge, and thus requires an unacceptably long computation time [15].

[1] Generation of new cells in HIS and elements in AIS having characteristics different from their parents

Considering that the AIS has a relatively short history compared to other artificial intelligent techniques, it may be advantageous to compose AIS with other mature intelligent techniques. In particular, this type of advantage can be notable when the differences between the human immune system and the artificial system cause the artificial immune algorithms to struggle with real world problems. Hybridization with mature algorithms could provide a more powerful solution than the sole adoption of artificial immune algorithms. The scalability of current artificial immune algorithms might be greatly increased when they adopt other data mining algorithms that have been proven to be successful at handling a massive amount of data, see figure 3. One of the main research topics that have been widely studied in the data mining field is scalability. There are several algorithms available that scan a massive amount of data within a reasonable time and produce some summaries of collected data. These algorithms can be used to pre-process raw data and thus reduce the amount of data that has to be handled by the AIS. These algorithms can be also used as a feature constructor that constructs useful features from raw data. Such an algorithm can be a good front-engine, defining a more optimized self set to be passed to the AIS [15].

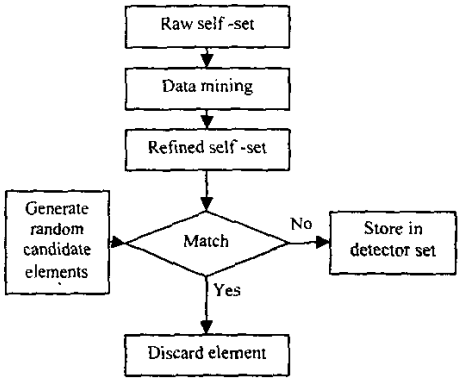

Figure 3. Negative selection with data minning

4.2 Issue of Matching Rules

Another identified drawback of the negative selection algorithm is the adoption of the r-contiguous rule to check the match between a given detector and antigen. The negative selection algorithm requires an appropriate number of detectors in order to produce acceptable error and detection rates. The established formula that approximates the appropriate number of detectors is applicable only when the algorithm uses the r-contiguous matching function. However, the r-contiguous matching rule is too simple to be used for determining the match between complex and high-dimensional patterns. Since the r-contiguous bit matching only measures the contiguous bits of two given strings, it is hard to guarantee that it can detect correlations in complex self and non-self patterns [15].

Justin Balthrop et al. [16] introduced a new matching rule, r-chunks, and showed that it performs better than full length r-contiguous bits matching on one LYSIS [7] based data set. The r-chunk is appealing because it is easier to analyze mathematically and it scales well as the length of detector increases (both in terms of number of detectors that are required for a given level of coverage). An important avenue for future research is to conduct experiments on various applications and to develop mathematical understanding of the properties of this system. The other caveat concerns the simplified version of LISYS used to conduct the experiments. It will be important to confirm how well permutations and r-chunks perform in the context of the complete LISYS system.

4.3 Issue of Costimulation

The costimulation is a process by which preliminary detection of antigen by one type of immune system cells is verified by another type of cells. This is actually a process of antigen verification. It would be harmful to take any action against the suspect without confirming its antigenecity. The process of costimulation requires second highly authentic source of information to generate costimulation signal. Efficient and effective costimulation would certainly help reduce *false positive* rates, incorrect detection of non-self.

For an AIS, the process of costimulation is important to achieve overall benefits from the system. The issue of costimulation might be tackled in two ways; 1) improving preliminary detection hence reducing need of costimulation, and 2) improving the costimulation mechanism.

Current AIS models lack automatic and efficient costimulation methods. The ARTIS [7] generates a costimulation request, which is processed manually by a human operator bringing into play his own intelligence about the system. The costimulation process can be improved with the help of second reliable source of information about the antigen presence.

In order to resolve costimulation issue, DynamiCS [15] employed the use of hypermutation to produce the effect of gene library evolution. This additional extension has been designed to fine-tune generated memory detectors so that the system obtains higher *true positive rates*, correct detected non-self, without increasing the amount of costimulation.

4.4 Issue of Main Goal

As the AISs are designed on the principles of HIS, therefore immunological viewpoints about the main goal of HIS are extremely important for designing good AISs and defining their main goal.

Conventionally it is thought that immune system completes the task of discriminating self from non-self, using various algorithms amongst which *negative selection* algorithm [7, 13] is the most prominent one. This self / non-self discrimination viewpoint has the following questionable issues, which need attention of AIS researchers [17]:

i. Negative selection is bound to be imperfect, not having exact matching, and therefore auto-reactions (false positives) are inevitable.
ii. The self/non-self boundary is blurred since self and non-self antigens often share common regions.
iii. Self changes over time. Therefore, one can expect problems with memory cells, which later turn out to be inaccurate or even auto-reactive.

Over the last decade, a new theory has become popular amongst immunologists. It is called the Danger Theory. It points out that there must be discrimination happening that goes beyond the self/non-self distinction. Future AIS applications might derive considerable benefits, and new insights, from the Danger Theory [17, 18].

4.5 Issue of Distributed Testing

The definition of distributed detection introduced by Hofmeyr and Forrest [7, 12] is restricted. It assumes that LYSIS, an AIS based network intrusion detection system, operates only under a broadcast LAN, local area network, environment. A broadcast LAN environment transfers identical input network packets to all the local hosts in a domain. Although LYSIS has different sets of detectors at local hosts, they are exposed to exactly the same set of input network packets. This kind of environment is a very special case. With a switched Ethernet, for example, each host only can experience network packets transferred to it and thus network packets handled by each detector set are different from each other. Due to this rather special circumstance, LYSIS was able to achieve several novel features such as scalability originated from the absence of communication among different detector sets, and robustness. This implies that no AIS having truly distributed components have been developed for IDS, intrusion detection system. The potential advantages of a distributed IDS have only been discussed theoretically in the literature, including the artificial immune model proposed by Jung Won Kim [15] and some other work [19]. None of these proposed models has been tested in a real distributed environment. Further development of distributed AIS and its study would be an important research avenue to be pursued [15].

5. CONCLUSIONS

The AIS research is a complex and novel interdisciplinary research. Therefore, identification of potential research issues in this field may consume a lot of precious time of researchers, first learning complex biological knowledge and then applying for AIS design

and development. A comprehensive overview of some general issues must save this time and enable to jump start the respective research. We have described some of the most general issues including; scaling of detectors, matching rules, costimulation, main goal of immune system, and distributed testing. The effort is based on; 1) latest research references and 2) our experience for pioneering AIS based research infrastructure in the Universiti Teknologi Malaysia (UTM). Care has been taken to avoid complex biological details to motivate new researchers for this state of the art computational research.

ACKNOWLEDGMENT

The authors are grateful to Ministry of Science, Technology and the Environment (MOSTE) Malaysia grant under vote number 74022 for pioneering intrusion detection and response research in Malaysia. We are also thankful to ICARIS-2003 bursary committee for awarding bursary to the main author.

REFERENCES

[1] L. N. de Castro, F. J. V. Zuben (1999). *Artificial Immune Systems: Part I A Survey of Applications.* Technical Report, DCA – RT 01/99, State University of Campinas, SP, Brazil, December 1999.

[2] L. N. de Castro, F. J. V. Zuben (2000). *Artificial Immune Systems: Part II A Survey of Applications.* Technical Report, DCA – RT 02/00, State University of Campinas, SP, Brazil, February 2000.

[3] A. Somayaji, S. Hofmeyr, and S. Forrest (1998). "Principles of a Computer Immune System." In *1997 New Security Paradigms Workshop*, pp75-82, ACM 1998.

[4] http://www.cs.unm.edu/~immsec/html-imm/immune-system.html, Last accessed on 30 August, 2003.

[5] S. Forrest, J. Balthrop, M. Glickman and D. Ackley (2002). "Computation in the Wild." In *the Internet as a Large-Complex System*, edited by K. Park and W. Willins: Oxford University Press. July 18, 2002.

[6] J. D. Farmer, N. H. Packard, and A. S. Perelson (1986). "The immune system, adaptation and machine learning." *Physica D*, 22:187-204.

[7] S. Hofmeyr and S. Forrest (2000). "Architecture for an Artificial Immune System." *Evolutionary Computation Journal* Vol. 8, No. 4, pp. 443-473. (2000).

[8] http://www.cs.unm.edu/~judd/lisys/ , Last accessed on 30 August, 2003.

[9] Junhyong Kim (2002). "Computers Are from Mars, Organisms Are from Venus." *IEEE Computer*, Volume: 35, Issue: 7, July 2002, Pages: 25-32.

[10] Harun Yahya (2001). *The Miracle of the Immune System.* Goodword Books, 2001.

[11] http://www.aber.ac.uk/icaris-2002/Proceedings/, Last accessed on 30 August, 2003.

[12] S. A. Hofmeyr (1999). "An Immunological Model of Distributed Detection and its Application to Computer Security." *PhD Dissertation*, University of New Mexico, 1999.

[13] S. Forrest, A.S. Perelson, L. Allen, R. and Cherukuri (1994). "Self-nonself discrimination in a computer." In *Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy*, Los Alamitos, CA: IEEE Computer Society Press.

[14] L. N. de Castro and J. Timmis (2002). "Artificial Immune Systems: A Novel Paradigm to Pattern Recognition." L Alonso J Corchado and C Fyfe, editors. In *Artificial Neural Networks in Pattern Recognition*, pages 67-84. University of Paisley, January 2002.

[15] Jung Won Kim (2002). "Integrating Artificial Immune Algorithms for Intrusion Detection." *PhD Thesis*, University College of London, July, 2002.

[16] J. Balthrop, F. Esponda, S. Forrest and M. Glickman (2002). "Coverage and Generalization in an Artificial Immune System." In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2002)*, New York, pp. 3-10.

[17] U. Aickelin and S. Cayzer (2002). "The Danger Theory and Its Application to Artificial Immune Systems." In *Proceedings of the International Conference on Artificial Immune Systems (ICARIS-2002)*, September 2002.

[18] U. Aickelin et al (2003). "Danger Theory: The Link between AIS and IDS." *In Proceedings of International Conference on Artificial Immune Systems (ICARIS-2003)*, September 2003.

[19] Dasgupta, D. (1999). "Immunity-Based Intrusion Detection Systems: A General Framework." In *Proceedings of the 22nd National Information Systems Security Conference (NISSC)*, October 18-21, 1999.