Soft Computing Techniques in Cyber Defense

Dinesh Kumar Saini and Vikas Singh
1Faculty of Computing and Information Technology, Sohar University, Oman
1Sr. Research Fellow, University of Queensland, Brisbane, Australia
2 Birla Institute of Technology and Science, Pilani, India

ABSTRACT
Artificial intelligence is a technology to make the machines human compatible. Various techniques like- heuristic search (Generate and test, Hill climbing, BFS, DFS, Problem reduction, constraint satisfaction, means-ends analysis etc.), game playing, understanding, planning, NLP, Learning, commonsense, predictions and actions, and expert systems are there to make the system intelligent. This paper primarily focuses on the problem of malicious objects in the cyber space and discusses the usage of different AI techniques to overcome the mentioned problem. Every day the information stored in the computer and the information in transit faces threats due to malicious objects which further leads to a big loss. This paper discusses the possibilities to incorporate the AI techniques to analyze the data and finding and restricting the malicious objects. To make a proactive cyber defense system how the system log can be explored to find the malicious object and how the AI techniques like heuristic search can help in this process is discussed. This paper also discusses about the uncertainty of attacks and gives some views to implement AI techniques such as probability and Bayes' theorem, certainty factors and rule based system, Bayesian networks, fuzzy logic etc.

General Terms
Computer Science, Information Technology, Security, Cyber Defense, Artificial Intelligence

Keywords
Cyber Security, Compartmentalization, Cyber hardening, Authentication, Access control, Confidentiality, Integrity

1. INTRODUCTION
Due to numerous malicious objects and misuse of existing technologies the information stored at an interconnected computer in Internet and the information in transit is not secured [1]. Any time a cyber-attack can occur and destroy the valuable information which causes a big loss to the society. In today's time various organizations such as power corporations, finance, telecommunications, health care, transportation, water, defense and the Internet, is highly vulnerable to cyber-attack and that such attacks could damage the whole economy so as to permanently and negatively alter the way of life [2,3]. So it is important to protect this valuable information from these malicious attacks by providing some means of cyber defense. The major problem in cyber defense is to predict the time of next attack because the time of attack is totally stochastic [4,5].

Also the information available from the analysis of the data gathered from the surroundings of the system is incomplete and insufficient to predict the future of cyber-attacks Hence to make the information complete and sufficient for the right prediction of the cyber-attacks this is needed to go in the lap of AI[6].

2. ARTIFICIAL INTELLIGENCE
It is a domain of techniques which is used to make the machines human compatible, by these techniques can include the intelligence in the machines so that they can think, interpret the situations and take the decision. Artificial Intelligence techniques can be used to restrict the cyber-attacks [7,8,9]. A general AI based simulation process can be represented by the following figure-1.

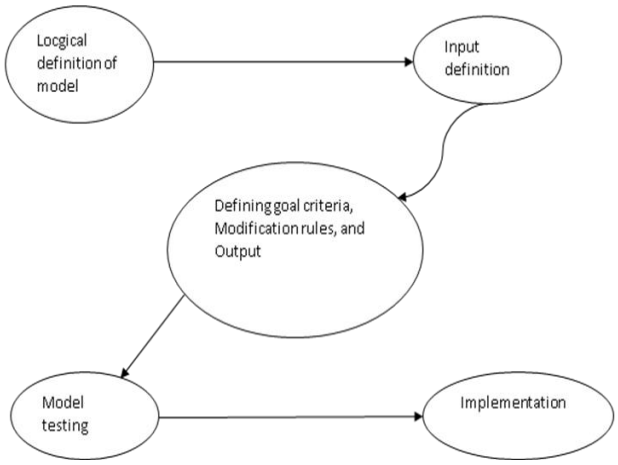

Figure-1: AI based simulation process

3. POSSIBLE USAGE OF AI TECHNIQUES
3.1 Expert System
A computer expert may detect a cyber-attack but it requires a lot of requirement gathering in the form of files, print outs or secondary backup. It is also a time consuming process which is just opposite to our desire i.e. fast or pre detection of attacks. These problems can be solved by using the expert system. It is basically a question answer system which makes decisions after analyzing the situations [15]. Possible queries of such an expert system to detect the cyber-attacks may be as under:

- Has more than one user tried to logon with the same password? If yes, then the password is compromised.
- Did a system file change? If yes, which one? Who changed it? Can it be repaired? If not, flag the file and report the violation.
- Did a user try to logon unsuccessfully 10, 100, 1000 times? If yes, then a hacker is probing the system, close the connection.
- Have specific protected files been requested or altered? If yes, then identify the source and record the details for authorities.
- Is a user making requests of the system that are out of the norm? If yes, then flag the user and restrict access.

- Are all users on the system authorized? If no, shut down unauthorized users and alert administrators.
- Did all users enter the system via the normal logon procedure? If no, then trace their origin and log them off the system.

3.1.1. Certainty Factor and Rule-Based System

A rule-based system is represented by a set of rules. If each rule is associated with a certainty factor then it is known as rule-based system with certainty factor. The certainty factor is the measure of the extent to which the evidence that is described by the antecedent of the rule supports the conclusion that is given in the rule's consequent

3.2 Artificial Neural Network

These are the systems that simulate intelligence by attempting to reproduce the types of physical connections that occur in animal brains [7]. They can learn dynamically by the existing worms, viruses or Trojan horses about their characteristics and produces the new threat possibilities and warnings. But the problem is consistent training implementation so that it can produce exact new cases and complexity of determining the inputs [13].

3.3Fuzzy Logic

These are the systems which represent the degree of truthfulness or falsehood of a statement on the basis of the degree of situation awareness. So it can improve the effectiveness of cyber security system. Problem with this approach is to validate and verify the fuzzy logic in current large sized Internet. These AI techniques can be integrated to design the effective cyber defense system [14].

3.4 Bayesian Network

Bayesian networks offer the AI researcher a convenient way to attack a multitude of problems in which one wants to come to conclusions that are not warranted logically but, rather, probabilistically. Furthermore, it allows attacking these problems without the traditional hurdles of specifying a set of numbers that grows exponentially with the complexity of the model [10].

Probably the major drawback to their use is the time of evaluation (exponential time for the general case). So Bayesian networks will be the wave of future in the field of cyber defense.

4. ELEMENTS OF AN AI-BASED CYBER DEFENSE SYSTEM

There can be following possible elements of an AI based cyber defense system [12].

- Sensors and Exploitation (Eyes of the system to detect an adversary).
- Situation awareness (A process that transforms sensed data into a decision).
- Defensive Mechanism (A technology to counter threat).
- Command and Control (A process of masking and executing decisions).
- Strategies and tactics (It is knowledge of what constitutes a good decision and configuring them according to situations).
- Science and Engineering (It is a foundation of design, composition, building, and maintenance of cyber defense system).

Figure-1 shows the integrated AI based cyber defense system architecture for the distributed nature of Internet. In this the sensor collects the information from the distributed ports and sent the analyzed data to the situation awareness element which makes the decision to further combine to generate the final decision to execute the defensive mechanism.

Here, both the elements Science and Engineering and Strategies and Tactics will be implemented in every element of the system.

The main challenge to this architecture is to prepare the input of the Decision Making Element due to:

- Temporal nature of information
- Incompleteness of information
- Uncertainty of information

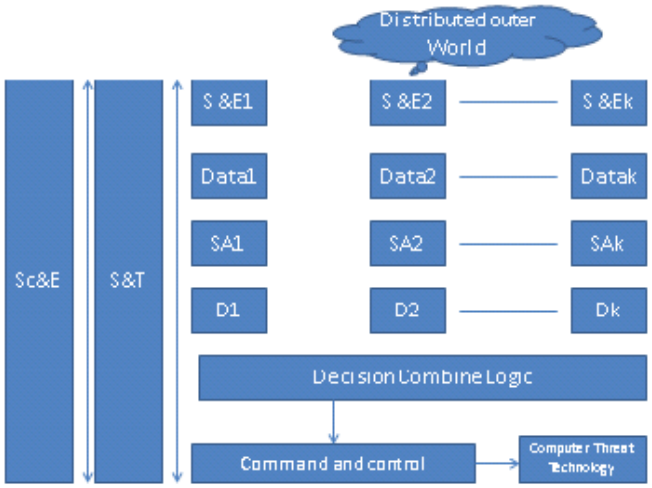

Figure-2: Cyber Defense prototype

S&E-Sensors & Exploitation
SA-Situation Awareness
D-Decision
Sc&E-Science & Engineering
S&T-Strategies & Engineering

4.1 The behavior of the above prototype

- *V* is a finite set of threshold values of a particular type of activity such as repetition of information from a particular IP. The training data can be collected from the incoming information to the Network. After analyzing this information different elements of the set V and their threshold values can be decided.

- *S* is a finite set of system vulnerabilities. The strategy of attack follows some fixed steps such as to choose an IP for attack which is vulnerable, attack on it and make it as new attacker. So to make it effective certain vulnerable nodes are to be found out to represent the set S. These nodes attack first and increase the number of new attacker.

- *U* is a finite set of command and control such as observe, halt, or kill the anomalous process. These elements of set U are the actions which are to be taken in response to attack by the defensive parts of the system. Which action is to be taken can be decided by the level of attack and its cost.
- $S \& E \rightarrow \{v \mid v \in V\}$. On the basis of analysis results of training data, the defensive elements came to know about the threshold values of different suspected events. These events are nothing but the

elements of set *V* after seeing the threshold values. If event value is greater than threshold value then it is included in set S&E.

- $Data \rightarrow \{s \,\&\, e \mid s \,\&\, e \in S \,\&\, E\}$. All the information filtered on the basis of the comparison of event occurrence and threshold value the '*Data'* set has to be created.

- $D \xrightarrow{SA}$ *{data}.* It is not necessary that all the elements of '*Data'* set are malicious objects or prone to attacks due to some misunderstanding of situation or incompleteness of information. So by using data mining or text mining under artificial intelligence it is necessary to justify the completeness and proper understanding to decide the final decision.

- *Decision combine logic* → *{Union of all* D_i | $i=1,2,\ldots,k.\}$. By using different intelligent sensors different decisions has to be created and finally they need to be combined for one single universal decision.

- *Command and control* → *{s | s ε U}.* Once the universal decision has to be come out, the appropriate command and control is decided from the set U.

5. USE OF RULE-BASED SYSTEM TO DETECT THE MALICIOUS MATCH

Let *U* is a finite and closed set of data patterns. *S* is a set of normal data patterns and *N* is a set of anomalous data patterns such that

$$U = S \cup N \text{ and}$$
$$S \cap N = \phi$$

A detection system is has two components, first a classification function *f*, and a data pattern *M*.

Detection system $(D) = (f, M)$

The detection system, *D,* classifies all the data patterns in two categories, normal and anomalous by using classification function.

$$(f, s) = \begin{cases} normal : if s \in M \\ anomalous : otherwise \end{cases}$$

Here we are give the input to the detection system after ensuring the completeness of the data. So, the first part of the detection system is the classification function which is based on the AI rule-based system.

In the network the information comes in the form of TCP SYN packets. So the pattern matching will be of binary strings. Let S_l is a set of all representations of a string then AI rule for matching will be-

Rule: Find a single string a_r that matches '*a'* but does not match any self string *s*, i.e. $Match(a_r, a)$ and $\neg Match(a_r, s), \forall s \in S_I$. If no such template exists then '*a'* is an anomalous string.

The matching can be done on the bases of Hamming distance. According to Hamming match, two strings '*a'* and '*b'* match under the Hamming match rule if they have the same bits in at least '*r'* (Hamming distance) positions. So if *a* = 011001 and *b* = 001111, then only if $r \leq 3$ is there a match, that is, $Match(a, b)$.

6. AI TECHNIQUES TO PREPARE THE COMPLETE AND ACCURATE DATA

6.1 Heuristic Techniques

A computer can be best represented in two states- Normal or Abnormal (under attack). So the whole thing can be represented in the form of state space problem, where the factors affecting the state may be system file corruption, trying the password more than 100 times, more than on user tried to logon in a system, etc. Now after getting the representable form of the problem we can uses heuristic techniques to solve it. By these type of techniques can help to find the solutions which are good but not best because these techniques are highly dependent on domain-specific knowledge. Some of Heuristic technologies are such as Depth First Search, Breadth First Search, Generate and Test, Best first search, Problem Reduction, Constraint satisfaction, and Means-Ends-Analysis [6]. .

6.2 Reasoning with Uncertainty

Number of Cyber-attacks can occur at any time, we cannot predict the time of the next attack, it is totally stochastic. So the problem of uncertainty arises. To solve this problem we can use the AI technique "Reasoning with uncertainty". In this type of situation the information is not consistent and completes [22]. To complete the information to find out the weaker points and then neglect them for further cases. When all the possibilities are found, choose the best as the solution. The information of attacks from one source is considered and analyzes it on the basis of time or some specific event to predict the next attack. The uncertain knowledge must be represented in the form of rules, predicate logics or in some well-formed formulas. To represent uncertainty there may be several key issues like-

- How can the knowledge base be extended to allow inferences to be made on the basis of lack of knowledge as well as on the presence of it?
- How can the knowledge base be updated properly when a new fact is added to the system?
- How can the knowledge be used to help resolve conflicts when these several inconsistent no monotonic inferences that could be drawn?

These issues can only be solved by reasoning which needs to detect the weakest point and ignoring it for finding the next reason till the solution.

6.3 Use of Game Playing

If the situation is structured like a chess game then it is easy to measure the success or failure. Some one can say that it is easy due to less amount of knowledge in a simple game. It is not true in many of the games. In chess, branching factor is around 35, each player might make 50 moves and in order to examine the complete game tree, one would have to examine 35100 positions [12]. This technique is also a matter of interest to solve the cyber-attack problems. In cyber space also there is a large number of nodes and it is more similar to a complex game which can be solved. Some advance game playing techniques are such as minmax search procedure, adding Alpha-Beta cutoffs with additional refinements such as waiting for quiescence, secondary search, using book moves, alternatives to minmax, and iterative deepening.

7. DISCRIMINATIVE MODELS FOR PATTERN IN MALICOUS CODE

Discriminative models are also very useful in cyber defense systems. Instead of modeling joint probability distribution over observation and label sequences, discriminative models define a conditional distribution p(y|x) over observation and label sequences. This means that when identifying the most likely label sequence for a given observation sequence, discriminative models use the conditional distribution directly, without bothering to make any dependence assumption on observations or enumerate all the possible observation sequences to calculate the marginal probability p(x).

7.1. Maximum Entropy Markov Models (MEMMs)

For finding the patterns in the signature we can use MEMMs, which are a form of discriminative models for labeling sequential data. MEMMs consider observation sequences to be conditioned upon rather than generated by the label sequence. Therefore, instead of defining two types of distribution, a MEMM has only a single set of separately trained distributions of the form, which represent the probability of moving from state y to y' on observation x. The fact the each of these functions is specific to a given state means that the choice of possible states at any given instant in time t+1 depends only on the state of the model at time t. Figure 10 show the graphic structure of MEMMs [25]. This concept is also used in cyber defense system design.

$$p(y'|x) = p(y'|y,x)$$

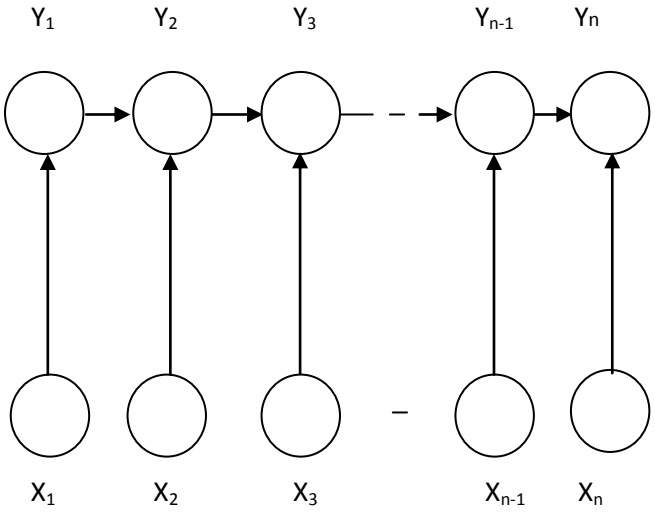

Figure 3: Graphic structure of first-order MEMMs in the signature

Given an observation sequence x, the conditional probability over label sequence y is given by

$$p(y|x) = p(y_1|x_1)\prod_{t=2}^{n} p(y_t|y_{t-1}, x_{t-1}$$

Treating observations as events to be conditioned upon rather than generated means that the probability of each transition may depend on non-independent, interacting features of the observation sequence. Making use of maximum entropy frame work and defining each state-observation transition function to be a log-linear model, equation (8) can be calculated as:
(10)

$$p(y'|x) = \frac{1}{Z(y,x)} \exp\left(\sum_k \lambda_k f_k(y',x)\right)$$

where Z(y, x) is a normalization factor; λk are parameters to be estimated and fk are feature functions. The parameters can be estimated using Generalized Iterative Scaling (GIS) [25] Each feature function can be represented as a binary feature. For example:

$$f(y',x) = \begin{cases} 1 \text{ if } b(x) \text{is true and } y = y' \\ 0 \text{ otherwise} \end{cases}$$

Despite the differences between MEMMs and HMMs, there is still an efficient dynamic programming solution to the classic problem of identifying the most likely label sequence given an observation sequence. A variant Viterbi algorithm is given by [25].

7.2. Sequential Labeling-Based Extraction Methods

By casting information extraction as sequential labeling, a set of labels need to be denned first according to the extraction task. For example, in metadata extraction from research papers [25], labels such as TITLE, AUTHOR, E-MAIL, and ABSTRACT are defined. A document is viewed as an observation sequence x. The observation unit can be a word, a text line, or any other unit. Then the task is to find a label sequence y that maximize the conditional probability p(y|x) using the models described above.

In generative models, there are no other features that can be utilized except the observation itself. Due to the conditional nature, discriminative models provide the flexibility of incorporating non-independent, arbitrary features as input to improve the performance. For example, in the task of metadata extraction from research papers, with CRFs we can use as features not only text content, but also layout and external lexicon. Empirical experiments show that the ability to incorporate non-independent, arbitrary features can significantly improve the performance.

On the other hand, the ability to incorporate non-independent, arbitrary features of discriminative models may sometimes lead to too many features and some of the features are of little contributions to the model. A feature induction can be performed when training the model to obtain the features that are most useful for the model [25].

7.3 Nonlinear Conditional Random Fields

Conditional random fields (CRFs) are the state-of-the-art approaches in information extraction taking advantage of the dependencies to do better extraction, compared with HMMs However, the previous linear-chain CRFs only model the linear-dependencies in a sequence of information, and is not able to model the other kinds of dependencies (e.g., nonlinear dependencies [28] In this section, we will discuss several nonlinear conditional random field models.

7.4. Condition Random Fields for Relational Learning

This concept of conditional random fields can also be used in the relational learning in the patterns in the malicious codes. HMMs, MEMMs and linear-chain CRFs can only model dependencies between neighboring labels. But sometimes it is

important to model certain kinds of long-range dependencies between entities. One important kind of dependency within information extraction occurs on repeated mentions of the same field. For example, when the same entity is mentioned more than once in a document, such as a person name Robert Booth, in many cases, all mentions have the same label, such as SEMINAR-SPEAKER. An IE system can take advantage of this fact by favoring labeling that treat repeated words identically, and by combining feature from all occurrences so that the extraction decision can be made based on global information. Furthermore, identifying all mentions of an entity can be useful in itself, because each mention might contain different useful information. The skip-chain CRF is proposed to address this.

The skip-chain CRF which is a linear-chain CRF with additional long-distance edges between similar words can also be used in finding the patterns. . These additional edges are called skip edges. The features on skip edges can incorporate information from the context of both endpoints, so that strong evidence at one endpoint can influence the label at the other endpoint.

Formally, the skip-chain CRF is defined as a general CRF with two clique templates: one for the linear-chain portion, and one for the skip edges. For an input x, let it be the set of all pairs of sequence positions for which there are skip edges. The probability of a label sequence y given on x is modeled as:

$$p_\lambda(y|x) = \frac{1}{Z(x)} \exp\left(\sum_{t=1}^{T} \sum_{k} \lambda_k \cdot f_k(y_{t-1}, y_t, x, t)\right.$$

$$\left. + \sum_{(u,v)\in G} \sum_{l} \lambda_l f_l(y_u, y_v, x, u, v)\right)$$

where Z(x) is the normalization factor, fk is the feature function similar to that in equation (12) and f1 is the feature function of the skip edges. λk and λ1 are weights of the two kinds of feature functions.

Because the loops in a skip-chain CRF can be long and overlapping, exact inference is intractable for the data considered. The running time required by exact inference is exponential in the size of the largest clique in the graph's junction tree. Instead, approximate inference using loopy belief propagation is performed, such as TRP.

7.5. 2D CRFs for Web Information Extraction & Dynamic CRFs

[28] Propose 2D conditional random fields (2D CRFs). 2D CRFs are also a particular case of CRFs. They are aimed at extracting object information from two-dimensionally laid-out Web pages. The graphic structure of a 2D CRF is a 2D grid, and it is natural to model the 2D laid-out information. If viewing the state sequence on diagonal as a single state, a 2D CRF can be mapped to a linear-chain CRF, and thus the conditional distribution has the same form as a linear-chain CRF. As a particular case, a factorial CRF (FCRF) was used to jointly solve two NLP tasks (noun phrase chunking and part-of-speech tagging) on the same observation sequence. Improved accuracy was obtained by modeling the dependencies between the two tasks.

8. CONCLUSION

In the field of cyber defense, artificial intelligence techniques can be used to handle the different problems efficiently such as

- Incompleteness of the information
- Uncertainty of the information
- Situation awareness and
- Good decision making.

For fetching meaningful data from the training-data data-mining or text-mining is used with the good combination of artificial intelligence techniques to remove the uncertainty of information, incompleteness of information and to decide the proper situation awareness.

The malicious attacks are poorly predicted in the real world. For the good prediction of the malicious attacks different artificial intelligence techniques are used. These artificial intelligence techniques are such as

- Rule based system with certainty factor
- Expert system
- Artificial neural network
- Fuzzy logic
- Bayesian networks
- Advance game playing
- Reasoning with uncertainty
- Heuristic techniques etc.
- Maximum Entropy Markov Models (MEMMs)
- Nonlinear Conditional Random Fields
- 2D CRFs for Web Information Extraction & Dynamic CRFs

The framework for cyber defense system given in the paper is the good example of the use of artificial intelligence techniques to deal with the incompleteness of information, situation awareness, uncertainty of information and decision making. Rule-based system can be used to detect the malicious match; it also shows the use of AI.

9. ACKNOWLEDGEMENTS

Author would like to thank Faculty of Computing and Information Technology, Sohar University and Faculty of Engineering and Information Technology University of Queensland Brisbane, Australia, for supporting and providing research opportunity. The author would also like to thank the reviewers for the review. The authors would like to thank BITS Pilani India also for the research support.

10. REFERENCES

[1] Dinesh Kumar Saini “Sense the Future” Campus Volume 1- Issue 11, Page No14-17, February 2011.

[2] Antonatos S., Akritidis P., Markatos E. P., Anagnostakis K. G. Defending against hit-list worms using network address space randomization. Proceedings of the 2005 ACM workshop on Rapid malcode. ACM Press New York NY, USA. pp. 30-40; 2005.

[3] Dinesh Kumar Saini “A Mathematical Model for the Effect of Malicious Object on Computer Network Immune System” Applied Mathematical Modeling, 35(2011) Page No. 3777-3787 USA, doi:10.1016/.2011.02.025.

[4] Bimal Kumar Mishra and Dinesh Kumar Saini “Mathematical Models on Computer viruses” Elsevier International Journal of Applied Mathematics and Computation, 187(2), 929-936. Volume 187, Issue 2, 15 April 2007, Pages 929-936. USA

[5] Bimal Kumar Mishra and Dinesh Kumar Saini "SEIRS epidemic model of transmission of malicious objects in computer network" Elsevier International Journal of Applied Mathematics and Computation, Volume 188, Issue 2, 15 May 2007, Pages 1476-1482. USA

[6] Gorodetsky V., Karsaev O., Samoilov V. Multi-agent and Data Mining Technologies for Situation Assessment in Security-related Applications. Monitoring, Security, and Rescue Techniques in Multi-agent Systems, Series of books, Advances in Soft Computing, Springer, pp. 411-422, 2005.

[7] Gaines B.R., Fuzzy reasoning and logics of uncertainty. Proceedings of the sixth international symposium on Multiple-valued logic. IEEE Computer Society Press Los Alamitos, CA, USA. pp. 179-188, 1976.

[8] Dinesh Kumar Saini and Hemraj Saini "VAIN: A Stochastic Model for Dynamics of Malicious Objects", the ICFAI Journal of Systems Management, Vol.6, No1, pp. 14- 28, February 2008. INDIA

[9] Hemraj Saini and Dinesh Kumar Saini "Malicious Object dynamics in the presence of Anti Malicious Software" European Journal of Scientific Research ISSN 1450-216X Vol.18 No.3 (2007), pp.491-499 © Euro Journals Publishing, Inc. 2007 http://www.eurojournals.com/ejsr.htm EUROPE

[10] Jan-Erik Lane, The logic of means-end analysis , Quality and Quantity Springer Verilog, Volume 20, Number 4, Netherland, pp. 339-356, 1986.

[11] Kienzle D.M., Elder M.C. Recent worms: a survey and trends. Proceedings of the 2003 ACM workshop on Rapid malcode. ACM Press New York, NY, USA. pp. 1-10; 2003.

[12] Sayadjari O.S. Cyber Defense: art to science. Communication of the ACM. 2004, Volume-47, Issue-3, ACM Press New York, NY, USA, pp. 52-57, 2004.

[13] Dinesh Kumar Saini and Hemraj Saini "Proactive Cyber Defense and Reconfigurable Framework for Cyber Security" International Review on computer and Software (IRCOS) Vol.2. No.2. March 2007, Pages 89-98. ITALY

[14] Rich E., Knight K. Artificial Intelligence. Second Edition, Chapter-12, pp- 307-326, TMH: New York, 1991.

[15] Regina Barzilayy, Daryl McCullough, Owen Rambow,Jonathan DeCristofaroz, Tanya Korelsky, Benoit Lavoie, A New Approach to Expert System Explanations, [A data base from Internet]. Available at:http://www.cogentex.com/papers/explanation-iwnlg98.pdf

[16] Dinesh Kumar Saini and Nirmal Gupta "Fault Detection Effectiveness in GUI Components of Java Environment through Smoke Test", Journal of Information Technology, ISSN 0973-2896 Vol.3, issue3, 7-17 September 2007.

[17] Dinesh Kumar Saini and Nirmal Gupta "Class Level Test Case Generation in Object Oriented Software Testing, International Journal of Information Technology and Web Engineering, (IJITWE) Vol. 3, Issue 2, pp. 19-26 pages, march 2008. USA

[18] Nordlander T.E. AI Surveying: Artificial Intelligence in Business. A thesis from De Montfort University, 2001. Available at: http://www.csd.abdn.ac.uk/~tnordlan/Publications/Artificial_Intelligence

[19] Dinesh Kumar Saini, Lingaraj A. Hadimani and Nirmal Gupta "Software Testing Approach for Detection and Correction of Design Defects in Object Oriented Software" Journal of Computing, Volume 3, Issue 4, April 2011, ISSN 2151-9617, Page No. 44-50.

[20] Jabbar Yousif and Dinesh Kumar Saini "Hindi Part-of-Speech Tagger Based Neural Networks" Journal of Computing, Volume 3, Issue 2, March, 2011, ISSN 2151-9617, Page No. 59-66.

[21] Dinesh Kumar Saini and Bimal Kumar Mishra "Design Patterns and their effect on Software Quality" Vol.5, No.1, January 2007, Page356-365 ACCST Research Journal, INDIA

[22] Hemraj Saini, Dinesh Kumar Saini, Nirmal Gupta, "Cyber Defense Architecture in Campus Wide Network System" International Journal of Theoretical and Applied Information Technology (JATIT)-(E-ISSN 1817-3195, ISSN 1992-8645), April 2008. Europe.

[23] Dinesh Kumar Saini "Testing Polymorphism in Object Oriented Systems for improving software Quality" ACM SIGSOFT Volume 34 Number 2 March 2009, ISSN: 0163-5948, USA

[24] Lakshmi Sunil Prakash, Dinesh Kumar Saini and Kutti N.S. "Integrating EduLearn Learning Content Management System (LCMS) with Cooperating Learning Object Repositories (LORs) in a Peer to Peer (P2P) architectural Framework" ACM SIGSOFT Volume 34 Number 3 May 2009, ISSN: 0163-5948, USA.

[25] J Lafferty, A McCallum "Conditional random fields: Probabilistic models for segmenting and labeling sequence data" Proceedings of the 18th International Conference on Machine Learning 2001 (ICML 2001), pages 282-289.

[26] Dinesh Kumar Saini, Jabar H. Yousif, and Wail M. Omar "Enhanced Inquiry Method for Malicious Object Identification" ACM SIGSOFT Volume 34 Number 3 May 2009, ISSN: 0163-5948, USA.

[27] H Saini, D.K.Saini and N.Gupta "E-Business System Development: Review on Methods, Design Factors, Techniques and Tools with an Extensive Case Study for Secure Online Retail Selling Industry" Journal of science and Technology (ISSN 0974-6846), Vol.2. No.5, May 2009, India.

[28] J Zhu, Z Nie, JR Wen, B Zhang Proceedings of the 22nd 2d conditional random fields for web information extraction", extracted from microsoft.com, 2005dl.acm.org

[29] Wail M.Omar, Dinesh K. Saini and Mustafa Hassan "Credibility Of Digital Content in a Healthcare Collaborative Community" Software Tools and Algorithms for Biological Systems in book series "Advances in Experimental Medicine and Biology, AEMB" Springer, Volume 696, Part 8, Page No. 717-724, DOI: 10.1007/978-1-4419-7046-6_73,