# Cyber Defense: Mathematical Modeling and Simulation

Dinesh Kumar Saini

***Abstract*—This paper gives an introduction to a realistic problem of current cyberspace i.e. cyber attacks and defense. In beginning of the paper we describe the objects used for cyber attacks and tell the means of spreading them such as secondary memory, e-mail attachments, instant messages or malicious bots. After this paper describe the roll of mathematical modeling and simulation to solve the problem with an extensive literature review. Cyber defense architecture and possible elements of cyber defense model are also studied for finding the research gaps. At the last this paper finds some gaps and possible ways to bridge these gaps.**

***Index Terms*—Virus, worms, differential equation, instance messaging, FTP, E-mail, cyber-hardening**

## I. Introduction

The advent of Internet/Network technology in past three decades has led to sea change in the way data is transferred and information exchange takes place. Over the years coupled with technological development and need, Internet technology has grown, offering numerous functionalities and facilities. The growth of Internet technology has thrown severe challenges in form of requirement of a suitable cyber defense system to safeguard the valuable information stored on system. Towards this goal it is proposed to study and understand the various malicious objects and develop a mathematical model to represent their behavior [1],[2]. Initially we did the study of self-replication and self-propagation of malicious objects such as virus, worm, Trojan horse, Bots etc. [1], [2].

## II. Mathematical Modeling

In one sentence modeling is- "Study of a system before actually it is build and implemented." Sometimes it is not feasible to implement a system or real world problem in the actual environment due to the huge cost and large time. So it is better to build a prototype or model and study the behavior of the system. A model is not only a substitute of the actual system, it also the simplification of the system [3], [4].

### A. Some Examples of Different Models

*Physical Model:* These are based on some analogy of Mechanical, Electrical, or Electric and Hydraulic Systems.
*Mathematical Model:* Systems which can be represented in the form of mathematical equations like- Demand and supply system

Manuscript received May 10, 2012, reviewed June 13, 2012.
D. K. Saini is with Computing and Information Technology, Sohar University, Sultanate of Oman and Faculty of Engineering and IT, University of Queensland, Brisbane, Australia (e-mail:dinesh@soharuni.edu.om)

*Physical Static Model:* These physical models do not change their behavior as time changes like- water-tank model.

*Physical Dynamic Model:* These physical models change their behavior as time changes like- spring suspension system or equivalent electric system.

*Mathematical Static Model:* These mathematical models give a mathematical equation when the system is in equilibrium state like- demand supply system.

M*athematical Dynamic Model:* In these mathematical models allow the change of system attributes as the function of time like- oscillatory motion.

*Mathematical Static Analytical Model:* These are small static mathematical model which can be solved by traditional math.

*Mathematical Static Numerical Model:* These are complex static mathematical model which can be solved by simulation.

*Mathematical Dynamic Analytical Model:* These are small dynamic mathematical model which can be solved by traditional math.

*Mathematical Dynamic Numerical Model:* These are complex dynamic mathematical model which can be solved by simulation.

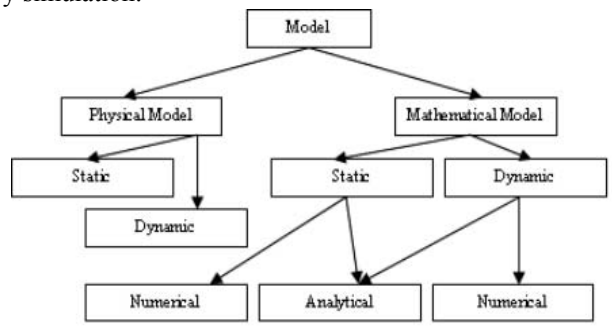


Fig. 1. Mathematical models

## III. Principle of Modeling with Respect to Cyber Defense System

Cyber attacks are the major problem of today's world. To overcome this problem it is necessary to understand the behavior of malicious objects. For this mathematical modeling can play an important role. It can help to fix the possible parameters of the malicious object those are important to tell how the malicious object can propagate in the Internet. Except this mathematical modeling helps to the following scenarios to build the cyber defense system [5,6] :

*Why***?** Cyber defense is due to various malicious objects like virus, worm, Trojan horse, spam etc. and certain technologies like P2P, Instant Messaging, bots, Phishing etc. misused for scams. For providing defense against them we must be aware of the behavior of these malicious objects and

it can only be understood by using modeling.

*Find***?** From the network traffic comes to our system we must know about the existence of the malicious objects or the data which can force our system to behave abnormally. The length of the data, type of the data, source of the data, how it can affect our system, etc are the various things to know.

*Given***?** The things which we know and to be analyzed to predict the future aspects of them. Findings of the relevant data such as TCP dump from a specific port, time stamp, accumulation time of data etc.

*Assume***?** Some necessary assumptions we need to make such as life time of data, time for aggregation of data, aggregation time for connections, and number of connections the data to be collected, etc. Also, the circumstances in which assumptions can be applied need to be identified.

*How***?** The physical principles those govern the model such as data/text mining, intelligent techniques, fuzzy logics, uncertainty principles etc. Also, how should we look for the model- distributed, laired, isolated, etc?

*Predict***?** Model should predict the malicious objects. It can be represented by mathematical equations, the calculations that to be made, and the state of the information.

*Valid***?** Certain tests are to be carried out to judge the validity of model, also the consistency of the model with respect to its principles and assumptions.

*Verified***?** To check whether the model is verified means the predictions are good, tests are to be identified and carried out. It is useful in terms of the initial reasons it was done.

*Improve***?** Simulated results helps to identified parameter values that are not adequately known, variables that should have to include and/or assumptions that could be lifted to improve the model. Implement the iterative loop that we can call "model-validate-verify-improve-predicate."

*Use***?** To make the security model adaptable what further exercises we can do, identify and include.

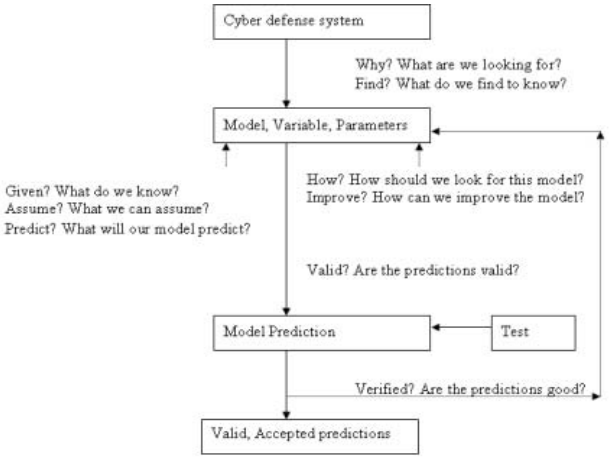


Fig. 2. Mathematical modeling process

## IV. Malicious Objects and Defense

Current Internet system having more than 233 million of computers till Jan-2004 [3], are prone to threat from various Malicious objects, may be of any type like – Worm, Virus, Trojan horse, Spam etc. and they can spread over the Internet through –

- Secondary memory (Floppy, Hard Disk, CD-ROM etc).
- E-mail (Attachments).
- Instant Messaging (FTP, Text Messaging, Chat etc.).
- Malicious Bot Programs [4].

The attacks on the computer are totally stochastic. We do not know the actual time of next attack on the computer. But on the basis of probability concepts in simulation we can find the probability of the attack at an instance of time.

If stochastic variable (Time of attack) can take I different values, $x_i$ (i = 1, 2, …, I), and the probability of the value xi being taken is $P(x_i)$, the set of numbers $P(x_i)$ is said to be a probability mass function. Since the variable must taken one of the values, it follows that

$$\sum_{i=1}^{I} P(X_i) = 1$$

Probability mass function can be defined as

$$P(x_i) = n_i / N$$

where N= total number of attacks and $n_i$ number of attacks from a specified source.

A cumulative distributed function can also be found which gives the probability of stochastic attacks' being less than or equal to a given value. Different measures of probability functions can be used for the study of the stochastic system such as mean, mode, median, Standard deviation, etc. Models characteristic equations can be of two types – Linear and non-Linear. Non-linear system can be represented by Partial Differential Equations (PDE). Consider that malicious object has propagation property P, depends upon various other factors like- A, B, C …etc. It can be represented as

$$P = f(A, B, C...).$$

The velocity can be represented as

$$\partial P/\partial t = \partial f(A, B, C,...)/\partial t.$$

and the acceleration rate can be represented as

$$\partial^2 P/\partial t^2 = \partial^2 f(A,B,C,...)/\partial t^2.$$

Once the simulated results obtained by the use of certain approximation techniques mentioned below can be used for complementing the data generated by simulation as well as validation:-

Taylor series expansion: Any function that has derivatives can be expanded by Taylor's Formula, The value of the independent variable, x, in a region near x = a, a function f(x) can be approximated by the polynomial

$$F(x) = f(a) + f'(a)(x-a) + (f''(a)/2!)*(x-a)^2 + ..... + (f^{(n)}(a)/a!)*(x-a)^n.$$

Finite difference approximation methods: This method transforms a partial deferential equation over small intervals. This is of two types-

Forward difference Approximation: It calculates the function gradient at various points by the formula:

$$f'(x_i) = (f(x_{i+1}) - f(x_i))/\Delta x$$

Backward difference approximation: It also calculates the function gradient at various points by the formula:

$$f'(x_i) = (f(x_i) - f(x_{i-1}))/\Delta x$$

Higher order derivatives: These can be calculated to describe the various important points in the distribution by the following formula-

$$f^{(n)} = (f^{(n-1)})'$$

Some regression tests such as Polynomial regression tests can also be used to validate the model. It finds that the values

can be fitted into a polynomial or not.

Once the characteristic equation is derived then results can be empirically/analytically validated on the basis of available standard mathematical hypothesis. The first thing for mathematical model validation is the dimensional homogeneity, which requires that each term has the same net dimensions [7]. Secondly, the models can be validated by checking qualitative and limit behavior. Except these some other things can also be considered, depending on how large the errors are? What is the accuracy and precision? Are the data fitted into the uniform curve? The data can be prepared by mean, mode, median or standard deviation. These data can be compared easily and help us to understand the behavior of malicious objects [8].

For example some definitions of virus can be modeled as follows:

Def-1: This definition defines the simple virus which infects the existing files and makes them able to behave in the similar way.

Let $P$ is a set of programs, $v \in P$, and $p_i \in P$. If v is a virus then the following equation must be true.

$f(v) = f(p_i)$, where function $f(v)$ and $f(p_i)$ gives the behavior of programs v and pi.

***Def-2:*** This is the definition of virus in respect of a fixed time t.

$$T(v, p_i, e, t, S) = \log \frac{f(v, e, t, S)}{f(p_i, e, t, S)} \tag{1}$$

where function $f(v, e, t, S)$ and $f(p_i, e, t, S)$ gives the behavior of program v and $p_i$ respectively at time t in the system S at the occurrence of event e.

Now if function $T(v, p_i, e, t, S) = 0$ then program $v$ is a virus otherwise not.

***Def-3:*** This is the definition of virus in respect of a continuous time interval $\Delta t$.

$$T(v, p_i, e, t, S) = \log \frac{\int_{t=\tau_0}^{t=\tau_1} f(v, e, t, S)dt}{\int_{t=\tau_0}^{t=\tau_1} f(p_i, e, t, S)dt} \tag{2}$$

where function $\int_{t=\tau 0}^{t=\tau 1} f(v, e, t, S)$ and $\int_{t=\tau 0}^{t=\tau 1} f(p_i, e, t, S)$ gives the behavior of program v and $p_i$ respectively in the time interval $\Delta t = \tau_1 - \tau_0$ in the system S at the occurrence of event e. Now if function $T(v, p_i, e, t, S) = 0$ then program $v$ is a virus otherwise not.

Now, for modeling a cyber defense system one should know about the different components needed. Here is a good division of cyber defense components by O. Sami Saydjari in Fig. 3.

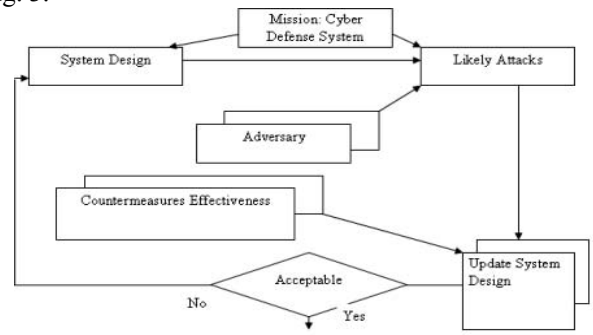


Fig. 3. Cyber defense system design model

These are sensors and exploitation, situation awareness, defensive mechanism, command and control, strategies and tactics, and science and Engineering [6]. He also tried to set an analogy in between an "art of war" to technology and he got a success in doing such thing at some extent

## V. Discussion and Summary

Although various models and solutions are provided, the following problems remain to be solved-

- All the characteristics of the malicious object's propagation have to be represented in the form of mathematical equations so that it gives the right direction to predict the future behavior. Another thing is to implement that characteristic equation in the existing environment (Existing Intranet and Internet), so to make a realistic model such that it can be implemented in real time environment without much change.
- Once a model is provided then its accuracy can be analyzed and improved on the basis of simulated results. This accuracy helps to protect the system by attack of malicious object. Like Code Red worms have been modeled with high accuracy and now it can be controlled easily [10], [11].Most of the models are with certain accuracy limits but still there is scope to improve them. So we will target our goals to improve these models.
- Most of the malicious objects having lot of characteristics, which is to be determined and because of this there is increase in the behavioral complexity [13]. We need to generalize each characteristic and narrow down the characteristic domain.
- Most of the cyber defense systems, when implemented create overheads like- slow down the existing system performance, increase the packet length or take more time for comparison etc. So it is needed to provide such a cyber defense system, which does not create much more extra overheads. Both the defensive systems (after attack recovery, and before attack) should be predictable enough on the basis of mathematical modeling with high accuracy. .
- In cyber defense it is difficult to trace the attacker due to lower education and awareness [15], [16]. Only few

countries have some inadequate policies, and education and awareness system. So, to restrict such type of malicious activities, our society needs to devise a proper set of policies.

## REFERENCE


[1] D. K. Saini, "A Mathematical Model for the Effect of Malicious Object on Computer Network Immune System," *Applied Mathematical Modeling*, vol. 35, pp. 3777-3787 USA, doi:10.1016/.2011.02.025, 2011.

[2] B. K. Mishra and D. K. Saini, "Mathematical Models on Computer viruses," *Elsevier International Journal of Applied Mathematics and Computation*, vol. 187, no. 2, vol. 187, no. 2, pp. 929-936. USA, 2007.

[3] B. K. Mishra and D. K. Saini, "SEIRS epidemic model of transmission of malicious objects in computer network," *Elsevier International Journal of Applied Mathematics and Computation*, vol. 188, no. 2, pp. 1476-1482. USA, 2007.

[4] D. K. Saini and B. K. Mishra, "Design Patterns and their effect on Software Quality," *ACCST Research Journal*, vol. 5, no. 1, pp. 356-365 India, 2007.

[5] D. K. Saini and H. Saini, "Proactive Cyber Defense and Reconfigurable Framework for Cyber Security," *International Review on computer and Software* (IRCOS), vol. 2. no. 2. pp. 89-98, 2007.

[6] D. K. Saini and N. Gupta, "Fault Detection Effectiveness in GUI Components of Java Environment through Smoke Test," *Journal of Information Technology*, ISSN 0973-2896 vol. 3, no. 3, pp. 7-17 2007.

[7] D. K. Saini and H. Saini, "VAIN: A Stochastic Model for Dynamics of Malicious Objects," *the ICFAI Journal of Systems Management*, vol. 6, no. 1, pp. 14- 28, 2008.

[8] H. Saini and D. K. 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. [Online]. Available: http://www.eurojournals.com/ejsr.htm

[9] D. K. Saini and N. Gupta, "Class Level Test Case Generation in Object Oriented Software Testing," *International Journal of Information Technology and Web Engineering*, (IJITWE) vol. 3, no. 2, pp. 19-26 2008.

[10] H. Saini, D. K. Saini, and N. 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.

[11] D. K. Saini, "Testing Polymorphism in Object Oriented Systems for improving software Quality," *ACM SIGSOFT*, vol. 34, no. 2, 2009.

[12] L. S. Prakash, D. K. Saini, and N. S. Kutti, "Integrating EduLearn Learning Content Management System (LCMS) with Cooperating Learning Object Repositories (LORs) in a Peer to Peer (P2P) architectural Framework," *ACM SIGSOFT*, vol. 34, no. 3, 2009.

[13] D. K. Saini, J. H. Yousif, and W. M. Omar, "Enhanced Inquiry Method for Malicious Object Identification," *ACM SIGSOFT*, vol. 34, no. 3, 2009.

[14] D. K. Saini, L. A. Hadimani, and N. Gupta, "Software Testing Approach for Detection and Correction of Design Defects in Object Oriented Software," *Journal of Computing*, vol. 3, no. 4, April 2011.

[15] D. K. Saini, "Security Concerns of Object Oriented Software Architectures" *International Journal of Computer Applications*, vol. 40, no. 11, pp. 41-48, 2012.

[16] D. K. Saini and Y. Sharma, "Soft Computing Particle Swarm Optimization based Approach for Class Responsibility Assignment Problem," *International Journal of Computer Applications*, vol. 40, no. 12, pp.19-24, 2012.