The International Journal Of Engineering And Science (IJES)
|| Volume || 4 || Issue || 6 || Pages || PP.47-51 || June - 2015 ||
ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

A Probabilistic Approach Using Poisson Process for Detecting the Existence of Unknown Computer Virus in Real Time

Aswin Kumar Rauta & Yerra.Sankar Rao, Dr T. C. Panda & Dr. Hemraj Saini
Lecturer, Department of Mathematics, S.K.C.G.College, Paralakhemundi, Odisha, India
E-mail:aswinmath2003@gmail.com
Ph.D schlor Shikya O Anusandhan university , Bhubaneswar, Odisha, India
E-mail: sankar.math1@gmail.com
Former Principal, Orissa Engineering College, Bhubaneswar,Odisha, India – 752 050
E-mail:- tc_panda@yahoo.com
Faculty Department of Computer Science & Engineering/IT&CT Jaypee University of Information Technology, Waknagahat-173234.
hemraj1977@yahoo.co.in

ABSTRACT

In this paper a mathematical approach has been studied to ascertain the cyber crime which is a worsening problem that can lead to loss of financial and personal information. However, e-crime is particularly hard to detect since internet is boundless that make evidence hard to collect. We have developed a probabilistic model using Poisson Process to determine the number of infected computers for a period of times when an individual becomes infected by the virus. A qualitatively distinct potential scenario is predicted to fight the cyber-crime. It is believed that the outcomes obtained from the present investigation will provide useful information for application and also serve as a complement to the previous investigations.

Keywords: *Cyber crime, E-crime, Internet, Poisson's Process, Probability, Virus.*

Date of Submission: 25-May-2015 Date of Accepted: 08-June-2015

I. Introduction:

In the age of globalization; the Information Technology (IT) plays a significant and realistic role in the way of bright and flourishing world with the rapid growth of computer and network systems in recent years which provides greater opportunities and options. The automation of companies, banks, educational institution, and railway reservation are reflections displayed everywhere that manifest dependence of human society on these tiny computers. But rapid industrialization and urbanization has brought new forms of crimes involving wider concerns of social order, safety, and security. Today, IT and globalization together gave world the terms like, Universal Access, Digital Divide, e-content, e-Democracy, Hacking, Cyber Crimes and Cyber War, the newly added concepts and problems. Cybercrime takes many forms and has garnered much attention in the media, making information security a more urgent and important priority. Incidences are there, where initially, computer is learned either out of curiosity, pleasure or compulsion (may be official or educational) or latter on learning knowledge turned into delinquency. The user-friendly software has added fuel to the fire making Cyber delinquency much pleasant and easy. This is true, particularly with regard to pornography, vulgar chatting, and piracy. Today, anybody with minimum computer literacy is sufficient to have access to Cyber-criminality and the chances are very less of being trapped by the preventive agencies. These features make Cybercrimes more dangerous and alarming. The most commonly seen crimes involve hacking into computer systems and computer viruses. Now these nuisances also threaten and cause damage via the Internet. Conventional crimes, such as fraud and money laundering, have been creatively changed into new forms by the rising popularity of the Internet. This is the world of internet services and internet users are increasing exponentially. Computer systems now contain millions of records relating to commerce, healthcare, banking, defense and personal information. All this information is at risk of either being misused for fraudulent purposes or modified for malicious reasons. Malicious software, or malware, on the Internet can cause serious problems, not only for services like email and the web, but for electricity, transport and healthcare services due to their increasing Internet dependence. One of the serious threats to the Internet and Computer network is malware attack. A computer virus attaches itself to a program or file enabling it to spread from one computer to another, leaving infections as it travels.

A computer virus can damage hardware, software or files of the systems. Computer viruses have been around from the days of DOS and even earlier, but after the 1990s, they became a potent threat due to the popularity of the internet and removable media. Some reported Viruses are I Love You, Logic Bomb and Melissa. One of the various ways in which computer systems can be compromised is by deploying computer virus/worms .There have been instances in the past where virus/worms have virtually brought the Internet to a grinding. Currently, e-mail is one of the main sources for transmission of virus, worms and Trojans. I Love You (2000) - "I Love You" virus is a computer virus that successfully attacked tens of millions of computers in 2000 when it was sent as an attachment to a user with the text "ILOVEYOU" in the subject line. A computer worm is a code that infects computer system and is able to spread functional copies of it without depending on other codes. Worms spread from computer to computer, but unlike a virus, it has the capability to transmit without any human intervention. Due to the copying nature of a worm and its capability to travel across network the end result in most cases is that the worm consumes too much system memory, causing web servers, network servers and individual computers to stop responding. Some reported worms are Code Red, Slammer. Code Red (2001) - Code Red was a computer worm observed on the Internet on July 13, 2001. It attacked computers running Microsoft's access and web server. Although the worm had been released on July 13, the largest group of infected computers was seen on July 19, 2001. On this day, the number of infected hosts reached 359,000. Slammer (2003) - Slammer worm caused a denial of service on some Internet hosts and dramatically slowed down general Internet traffic. This fast-moving worm managed to temporarily bring much of the Internet to its knees in January 2003. It spread rapidly; infecting most of its 75,000 victims within ten minutes. A Trojan horse is a program that secretly performs its operation under the guise of a legitimate program. The Trojan horse at first glance will appear to be useful software but will actually damage once installed or run on the computer. When a Trojan is activated on the computer the results can vary. Some Trojans are designed to be more annoying than virus and worms as they can cause serious damage by deleting files and destroying information on system Trojans are also known to create a backdoor on computer that gives malicious users access to system, possibly allowing confidential or personal information to be compromised. Unlike viruses and worms, Trojans do not reproduce by infecting other files nor do they self-replicate. Some reported Trojan horses are remote access Trojans (RATs), backdoor Trojans (backdoors) Distributed Denial of Service Attack Trojan horse.

Today, people rely on computers to create, store and manage critical information. Information transmitted over networks has a higher degree of security risk. Thus, it is crucial to protect the computers and data from loss, damage and misuse. Antivirus programs are an effective way to protect a computer against virus, worms and Trojans. An antivirus program protects a computer against malicious codes (virus, worms and Trojans) by identifying and removing them when found in memory, storage media or in any incoming files. One technique that antivirus programs use to identify a virus is to look for virus signatures or virus definitions, which are known specific patterns of virus codes. Most commercial antivirus products make use of a black listing strategy. They rely on databases of virus signatures that are consulted when a new program arrives. Anti-virus tools scan disks and sometimes e-mail looking for known viruses. Updating the antivirus program’s signature files regularly is important as it will download any new virus definitions that have been added since the last update. In this paper we have critically analyzed the effect of antivirus in our model that is; an analysis has been made to study the dynamic behavior of the system with and without antivirus installed in the nodes. If we choose to ignore cyber-crime, the attacker ambition and greed will be encouraged and we will face more serious criminal behaviors. What we must do is fight these cyber-crimes by tracking the instigators through the proper procedures. This is the best way to protect our networks and keep cyberspace secure. Hence, when seeking secure data and website maintenance, we must discuss how legal methods and procedures can be applied to investigate those who engage in cyber and network crime. Crime opportunity theory suggests that humans intend to committee a crime, but the crime intention can be restricted by willpower.

In this paper, we offer solutions to guard against cyber-crime through the implementation of Poisson process. Once the individual computer has been identified as infected, then the possible of all infected computers in the network can be identified over a period of time.

II. Literature Review:

Bimal Kumar Mishra et.al [2] have studied the different epidemic models in computer network which involves the ordinary differential equations only. B.Zaka [1] has studied theory and application of similarity detection techniques in cyber crime. C.Kruegel [4] et.al.have investigated the anomaly detection of web-based attack. Chunhua Feng [5] et.al have studied the oscillatory behavior of a network epidemic SIS model with nonlinear infectivity. Dwarka Shukla [8] et.al have analyzed the user’s web browsing behavior using Markov chain model. J.Lopez Gondar [10] et.al have studied a mathematical model for virus infection in a system of interacting computers. Qingyi Zhu, Xiaofan [17] et.al have presented modeling and analysis of the spread of computer virus. Y.D Shin [24] has developed a new model for cyber crime investigation procedure. Xiaofan Yang [22] et al have investigated the epidemiological modeling of computer viruses. Shelly Xiaonan [20] et.al have studied the use of computational intelligence in intrusion detection systems. J.M.Wong [11] has modeled the spread of a computer virus.

A central drawback common to each of the approaches discussed above is that they fail to address the importance of individual incidents located in a specific time and space. Instead, findings are concerned with general, aggregate patterns; this makes it difficult to draw conclusions regarding how the individual behavior of victims or offenders may be affecting the occurrence and rate of infection. From the literature survey, it makes clear that a little investigation has been done on the cyber crime using Poisson process through probabilistic approach .Many of the authors have modeled the epidemic models through differential equations and used statistical methods. So an attempt has been made by present authors to detect the number of infected computers over a period of time using poisons process through probabilistic model. To the best of our knowledge, this problem has not been considered before, so that the reported results are new.

III. Probabilistic Model:

There is a relatively short incubation period from time when an individual computer becomes susceptible with virus, which causes the infection to the all the computer connecting it until the symptom appears. As a result, it is difficult for the user to be certain of the number of the computers that are infected at any given time. We will now present a first approximation probabilistic model for this phenomenon, which can be used to obtain a rough estimate of the number of infected computers.

Suppose that individual computer contract the virus in accordance with a Poisson process whose rate λ is unknown.

Let us suppose that the time from when an individual become infected until symptoms of the infection appear is a random variable having a known distribution F. Suppose also that the incubation times of different infected individuals are independent.

Let $N_1(t)$ denote the number of individual computers that have shown symptoms of the infection by time t.

Let $N_2(t)$ denote the number that are infected but not shown any symptoms of the infection by time t.

Now, since an individual that contract the virus at time s will have symptoms by time t with probability F(t-s) and will not with probability $\bar{F}$ (t-s).

By a proposition, $N_1(t)$ and $N_2(t)$ are independent Poisson random variable with respective means

$$\mathrm{E}\,[N_1(t)] = \lambda\int_0^t F(t-s)ds = \lambda\int_0^t F(y)dy$$

and

$$\mathrm{E}\,[N_2(t)] = \lambda\int_0^t \bar{F}\,(\mathrm{t}-\mathrm{s})ds = \lambda\int_0^t \bar{F}\,(\mathrm{y})dy$$

Now if we know λ, then we could use it to estimate $N_2(t)$, the number of individual infected but without any outward symptoms at time t, by its mean value E $[N_2(t)]$.

However, since λ is unknown, we must first estimate it.We will presumable know the value of$N_1(t)$, and so we can use its known value as an estimate of its mean E $[N_1(t)$].That is if the number of individual that exhibited symptom by time t is n_1.then we can estimate that

$$n_1 \approx \mathrm{E}\,[N_1(t)\ ] = \lambda\int_0^t F(y)dy$$

Therefore we can estimate λ by the quantity $\hat{\lambda}$ given by

$$\hat{\lambda} = n_1/\int_0^t F(y)dy$$

Using this estimate of λ, we can estimate the number of infected but symptomless individuals at time t by

$$\text{Estimate of } N_2(t) = \hat{\lambda}\int_0^t \bar{F}\,(\mathrm{y})dy$$

$$= \frac{n_1\int_0^t \bar{F}\,(\mathrm{y})dy}{\int_0^t F(y)dy}$$

$$= \frac{n_1\int_0^t \bar{F}\,(\mathrm{y})dy}{\int_0^t\{1-\bar{F}\,(\mathrm{y})\}dy}$$

For F is exponential with mean μ. Then $\bar{F}\,(\mathrm{y}) = \mathrm{e}^{-\mathrm{y}/\mu}$, we get

$$N_2(t) = \frac{n_1\int_0^t \mathrm{e}^{-\mathrm{y}/\mu}dy}{\int_0^t\left\{1-\mathrm{e}^{-\mathrm{y}/\mu}\right\}dy}$$

$$N_2(t) = \frac{n_1\mu(1 - e^{-t/\mu})}{t - \mu(1 - e^{-t/\mu})}$$

Example: Suppose $t = 16$ days, $\mu = 10$ days and $n_1 = 220$,
Then the number of infected but symptomless individuals at time 16 is
Estimate$= \frac{2200(1-e^{-1.6})}{16-10(1-e^{-1.6})} = 218.96$
i.e., if we suppose that forgoing model is approximately correct should be aware that assumption of constant infection rate λ that is unchanging over time is almost certainly a weak point of the model, then if incubation period is exponential with mean 10 days and if total number of individuals that have exhibited by symptoms during of epidemic is 220, then we can expect that approximately 219 individuals are infected though symptomless at time 16 days.

IV. Discussion:

Once the individual computer has been identified as susceptible, then the preferences of each individual computer connected to it through internet can be identified. For example, the the individual computer described above exhibits infection affluently the neighborhood, while another individual computer may exhibits infection affluently the neighborhood. In this way all the computers connected through the internet become infected. The model will be run for different time period, different meantime, different infection rate and the difference between the estimated parameters will be recorded. We propose to look at two aspects of estimates. The first of these is it is a well-known approach to visualizing hot spots. It also represents a prediction model that says the best prediction for the next time period is the surface generated in the current time period. This will allow us to formally determine the amount of improvement we can achieve by explicitly considering change in the environment. By taking the actual locations of crime incidents at a time we can compare the percentile scores for these locations as produced by the various methods.

V. Conclusion:

In this paper we have described a probabilistic model of crime analysis using Poisson's process that can be used for the analysis of both computer and cyber crimes. This method consists of a data collection method to determine the number of infected computer when an individual is known. Because this proposed method is automatable, it can be used in situation where there is a vast wealth of data. This makes it particularly useful analysis in "real world" situations, where the data collection process is relatively easy, but the data analysis is more difficult. This model can be used to predict areas in which future criminal incidents are likely. This methodology will be of interest to members of both the computer and police communities.

References:

[1]. B.Zaka (2009), "Theory and application of similarity detection techniques, IICM, Graz University of Technology: Graz, Austria, p.171.

[2]. Bimal Kumar Mishra and Aditya Kumar Sing (2012), "SI_jRS E-Epidemic model with multiple groups of infection in computer network". International journal of nonlinear science, vol.13, No.3, pp.357-362.

[3]. Bimal Kumar Mishra and Dinesh Kumar Saini (2007), "SEIRS epidemic model with delay for transmission of malicious objects in computer network". Applied mathematics and computation, 188, 1476-1482.

[4]. C.Kruegel and G.Vigna (2003), "Anomaly detection of web-based attack". In Proc.ACM Conference computer and communication security, ACM, pp.251-261.

[5]. Chunhua Feng and Carl S.Pettis(2014), "Oscillatory behavior of a network epidemic SIS model with nonlinear infectivity". Applied Mathematics, 5,203-211.

[6]. D.Biswas (2012), "Probability and statistics", Vol-1, New central Book Agency (p) Ltd.

[7]. Douglas C.Montgomery (2007), "Design and analysis of experiments" Wiley-India Edition.

[8]. Dwarka Shukla and Rahul Singhai (2011), "Analysis of user's web browsing behavior using markov chain model". Int.J.Advanced networking and applications, vol.02, issue: 05, page824-830.

[9]. Hu Xu and Xiang GU (2012), "Extensive design for attack's recognition and resistance of survivable network". Journal of networks, Vol.7, no.2, 222-228.

[10]. J.Lopez Gondar and R.Cipolatti (2003), "A mathematical model for virus infection in a system of interacting computers". Computational and applied mathematics, vol.22, N.2, PP.209-231.

[11]. J.M.Wong, N.Pino, B.Hallahan (2014), "On modeling the spread of a computer virus". Mathematical modeling", spring 2014.

[12]. J.N.Kapoor (2005), "Mathematical modeling, 1st edition, new age international pvt.Ltd.

[13]. K.Wang and S.J.Stolfo (2004), "Anomalous payload-based network instruction detection" in proc.7th int. symposium on recent advances in intrusion detection (RAID).

[14]. M.S.S.Khan (2014), "A computer virus propagation model using delay differential equations with probabilistic contagion and immunity". IJCNC, Vol.6, No.5, sept-2014.

[15]. Mohammad Rasmi and Aman Jantan(2013), "A few algorithm to estimate the similarity between the intentions of the cyber crimes for network forensics".Procedia technology 11,540-547

[16]. Murray R.Spiegel and Ray Meddis (1980), "Probability and statistics".Schaum's outline seriese, McGraw-Hill Book Company.

[17]. Qingyi Zhu, Xiaofan Yang, Jianguo Ren (2012), "Modeling and analysis of the spread of computer virus". Communications in nonlinear science and numerical simulation.17 (12).pp.5117-5124.

[18]. Sheldon M.Ross (2005), "Introduction to probability and statistics for engineers and scientists", third edition, academic press.

[19]. Sheldon M.Ross (2007), "Introduction to probability models", 9th edition, Academic press.

[20]. Shelly Xiaonan Wu and Wolfgang Banzhaf (2010), "The use of computational intelligence in intrusion detection systems: A review". Applied soft computing 10,1-35,doi:10.1016/j.asoc.2009.06.019.

[21]. V.Paxson(1999), "Bro:a system for detecting network intruders in real time". Computer network,vol.31, pp23-24.

[22]. Xiaofan Yang and Lu-Xing Yang (2012), "Towards the epidemiological modeling of computer viruses". Hindawi Publishing Corporation, vol-2012, article ID259671,DOI:10.1155/2012/259671.

[23]. Xie Han and Qiulin Tan (2010), "Dynamical behavior of computer virus on internet". Applied mathematics and computation, vol.217, no.6, pp.2520-2526.

[24]. Y.D Shin (2011), "New Model for cyber crime investigation procedure".JNIT,2(2):P.1-7.

[25]. Y.Srinivasulu, Md.Mastan and M.Kishore Kumar (2012), "Discrete mathematical model of computer worm propagation". ijarcsse.com, vol-2, issue-7, issn-2277128X.

Biographies of authors:

Aswin Kumar Rauta was born in village Khallingi of district Ganjam; Odisha, India in 1981.He obtained his M.Sc. and M.Phil. degree in Mathematics from Berhampur University, Berhampur, Odisha, India. He has qualified NET in 2009 conducted by CSIR-UGC, government of India. He joined as a lecturer in Mathematics in the Department of Mathematics, S.K.C.G.College, Paralakhemundi, Odisha, India in 2011 and is continuing his research work since 2009 and work till now.

Yerra Shankar Rao was born in village Kurula of district Ganjam; Odisha, India in 1982. He obtained his M.Sc. degree in Mathematics in 2004 from Berhampur University, Berhampur, Odisha, India. He has been working as a lecturer in Mathematics in the Department of Mathematics Gandhi Instuete Of Excellent Technocrats Eng.college, Bhubaneswar, and Odisha, India since 2012 and is continuing his research work since 2011 and work till now**.**

Dr T. C. Panda is a Retired Professor in the Department of Mathematics,Berhampur University, India. He has published in journals of high repute and supervised several Ph. D. scholars. His research areas include Fluid Mechanics, and Nonlinear Analysis**.**

Dr Hemraj saini is a Associate professor in the Department of Computer Science & Engineering/IT&CT Jaypee University of Information Technology, Waknagahat. He has published in the journals of high repute and supervise several M.Tech scholars. His area of research in Cyber Security, Enterprise Integration Applications, Software Testing, Cell Planning, Simulation & Modeling.