ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.3 , No.5, Pages : 184-188 (2014)
Special Issue of ICACSSE 2014 - Held on October 10, 2014 in St.Ann's College of Engineering & Technology, Chirala, Andhra Pradesh

Defending Computer Networks from Covert Operations using Chaos Theory

Meruga Vishal Bharadwaj [1] **AVS Sudhakara Rao [2]**

[1]M.Tech Student, Dept of CSE, St. An n's College of Engineering & Technology, Chirala, Prakasam Dist, A.P, India

[2] Associate Professor, Dept of CSE, St. Ann's College of Engineering & Technology, Chirala, Prakasam Dist, A.P, India

***Abstract*— Defending computer networks from covert channels is a challenging task in TCP/IP networks. It is necessary for system engineers to know covert channels in order that they are going to be able to pre-emptively overcome sure security issues. In this paper we propose a new method to find the information hidden in TCP initial sequence numbers (ISNs) through this we can detect difficult covert channels. In our method we create reconstructed phase space by processing space using phase space reconstruction, thus a variety model for detecting covert channels in TCP ISNs. By using this model, a variety algorithm is developed to identify the existence of information hidden in ISNs. Results have demonstrated that our proposed detection method performs high detection accuracy and good reduced computational complexity. We can use this not only in offline but also in online detection.**

***Index Terms*—Covert channel, network, phase space reconstruction, TCP/IP protocols.**

1. INTRODUCTION

Networking becomes one of the most common ways of passing messages in our modern times, on the one hand the applications of covert channels can be used by government agencies and military departments to keep their communications secret. For instance, network administrators can use covert channels to secure network management related communication by hiding it from hackers. On the other hand however, many applications of covert channels are of a malicious or unwanted nature, and therefore pose a serious threat to network security. To send data more securely variety of methods are made one of them are STEGANOGRAPHY, which is art of hiding data in transmission medium so that no one should understand the message transmitted. Most widely available applications of steganography are dedicated to multimedia applications in which hidden data are distributed via files of sound, images and videos .Hiding data at protocols network level is new but it is very important issue for network security.

Different network steganography methods or covert channels based on network can manipulate certain properties of the communications in order to transfer secrete information through medium but without knowing anyone except the covert channels which are in operation. Generally, covert channels are classified into storage and timing channels. Network based covert channels which use protocols as carriers to hide secret information are considered covert storage channels. Thus, developing effective counter measures against covert channels becomes necessary. There are several schemes developed to hide information in TCP/IP headers, among them, information hidden in TCP ISN (Initial Sequence Number) field and IP ID (Identifier) field are the most difficult ones to be detected, since these fields can have arbitrary values within the requirements of the standard

In this paper, we analyze the possible covert channels in TCP/IP protocols and propose a new efficient scheme called Phase Reconstruction Method (PRM) to identify covert channels in TCP ISN and IP Identification fields. First, we use phase space reconstruction method to create a processing space, the first-order difference for the input data—ISNs is used to obtain the reconstructed phase space; then the statistical feature model and classifier are proposed based on the reconstructed phase space. The training dataset used in this scheme is around 350 normal ISNs. Differently from the SVM method, only normal ISNs are used in the training. Simulation results have demonstrated that the proposed method could achieve 100% accuracy in detecting abnormal cases of TCP ISN. The computational complexity is much lower than the SVM. The rest of this paper is organized as follows. The potential covert channels in TCP/IP headers and possible counter measures are described in Section 2; a phase space reconstruction method is briefly introduced to analyze the chaotic nature of TCP ISNs. Then, the statistical feature model and classifier are proposed in Section 3; Section 4 presents our simulation results and performance analysis and finally conclusion.

2. COVERT CHANNELS IN TCP/IP PROTOCOLS

The important protocols for internet community, TCP/IP can be used to communicate across any set of inter-connected networks. TCP/IP covert channels (or TCP/IP steganography) use the fact that some header fields can be changed to move information without impacting normal communications. Normal TCP/IP header structure allows for a number of covert channel options within packet header fields which are normally unused, optional or required to hold random numbers. Here, we group those covert channels into using unused header bit fields, modification of some header fields and using header fields which require taking random numbers.

Some Covert Channel options in TCP/IP header are

1) Covert Channels by Using Unused Header Bit Fields

2) Covert Channels by Modifying Some Header Fields of TCP/IP

3) Covert Channels by Using Some Header Fields Which Require Random Numbers

To understand how to embed secret data in this field let us take a look at the method used to establish and tear down network connection, which is called three-way hand-shake of TCP operation

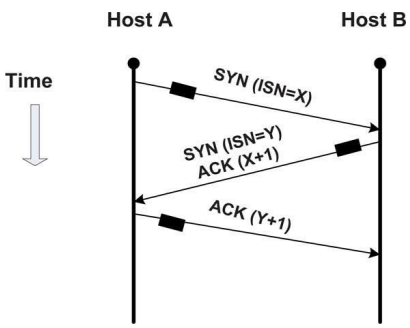

Fig. 1.Three-way handshake in TCP

1) First host A sends a packet with SYN-bit set requesting a new connection. This packet contains the initiating host's sequence number ISN as X for the connection.

2)After receiving packet from A, the received host B sends the packet with an Acknowledgment(ACK) number X+1 with its own initial sequence number Y with SYN bit set

3)Now host A sends a packet with an ACK number Y+1 and the connection is established.

TCP *Sequence Number* is a 32-bit number and is required to hold random number for the first packet. When a new connection is to be established, TCP marks the SYN bit as 1 and chooses a random number as the sequence number, which is called ISN. ISN is randomly initialized by the operating system and are chosen by hosts, hidden messages could be carried through ISNs such as X and Y. In covert Communication sender will sends data through *sequence number field* and the receiver then extracts dataIP identification field's job is to identify the packet IDs and make sure that the fragments of different packets should not reassemble as single packet at the receivers end. Rowland proposed to encode information in TCP ISN and IP ID fields. For the IP ID, each byte of the covert information is multiplied by 256 and directly used as IP ID. For the TCP ISN, each byte of covert information is multiplied by 256 x 65536 and then used as TCP ISN. The secret messages get passed in this way. Rutkowska developed a TCP ISN based covert channel called NUSHU for Linux operating system. Murdoch also dis- cussed two schemes for encoding data with ISNs generated by OpenBSD and Linux. In this paper, we focus on detecting TCP ISN covert channels developed for Windows such as Covert TCP developed by Rowland. The proposed detection method will be compared with SVM method proposed by Sohn *et al.* Sohn *et al* proposed an offline detection scheme that uses SVM to detect covert channels in TCP ISN and IP ID. In this SVM based method , the authors evaluated a three-feature dataset which includes *ISN*, TCP control *Flag*, and TCP header *Checksum*, and used 5000 normal and 5000 abnormal packets to train the SVM. The total number of packets is 10000. The authors claimed that the detection rate by the SVM learning method is high. But this method is complicated and forgot that ISN should contain some structure to avoid receiving remaining packets from previous connection. And this SVM model needs to have knowledge on steganography or we have to collect those remaining packets which is difficult as most of the covert communications are not known.so it is important to operating system to have some sort of identification number correctly. With the current implementation in many operating systems, the kernel increments IP ID field by a fixed amount, from one packet to the next. For example, the kernel in Windows 7 Enterprise is increasing the IP ID field by 1 from one packet to another. This makes the detection of covert channels in IP ID much easier. we focus on detecting hidden messages in TCP ISN field. An novel detection scheme is proposed for identifying covert channels in TCP ISNs. First, phase space reconstruction is used to reveal the dynamic feature of ISNs. The statistical feature model is proposed based on the reconstructed phase space dataset. The high order statistic analysis is conducted to construct the classifier which could separate steganograhic ISNs from normal ISNs.The detection accuracy of PRM is 100%. The computational complexity is greatly reduced and the detection speed is much faster than the one using SVM . Thus, it could be used online for detecting steganographic ISN ISNs.

3. Proposed Detection Model

The use of pseudorandom number generators (PRNGs) has been widely spread when producing ISNs. PRNGs generate a sequence of numbers that approximate the properties of random numbers. Thus, the sequence is not truly random. The random- ness of ISNs makes attackers hard to predict these numbers; the idea not to use truly random numbers for ISNs lies in that if a connection arrives, the randomness of ISNs would make it uncertain that the coming sequence number would be different from a previous incarnation. For IP IDs, as the unique- ness within a given time window ensures that fragments of different packets are not reassembled into one packet on the receiving host, the truly random number should not be used in IP IDs. The PRNG used by the Windows operating system is the most commonly used PRNG. The pseudo-randomness of the output of this generator is crucial for the security of almost any application running in Windows. The PRNG is modeled as a function whose input is a short random seed, and whose output is indistinguishable from truly random bits. Implementations of pseudorandom number generators often use a state whose initial value is a random seed. The state is updated by an algorithm which changes the state and outputs pseudorandom bits, and implements a deterministic function of the state of the generator. Herring observed that pseudorandom number generators are derived from deterministic chaotic dynamic system and made connection between chaos and pseudorandom number generators. To analyze this chaotic/nonlinear behavior, we turn to phase space reconstruction method, which is a useful chaotic/nonlinear signal processing technique. This method to build a spoofing set in predicting ISNs generated by Windows 2000. Some weaknesses of the Windows PRNG were revealed. With this we are also adding encryption before transmitting by plain cipher encryption and Transmission time will be calculated & plotted in Bar Graph.

Phase Space Reconstruction

Chaos can be defined as a random and non-uniform phenomenon in the deterministic nonlinear system and hidden discipline in a complex system can be revealed by chaos theory. Chaos theory makes people aware that often there are certain laws behind the seemly random phenomena. With conventional tools such as Fourier transform, chaos looks like "noise", but chaos has structure seen in the phase space. Phase space reconstruction is the first step in nonlinear time series analysis of data from chaotic systems. It is a useful nonlinear/chaotic signal processing technique to characterize dynamic system, whether low-dimensional or high-dimensional. Reconstructed phase spaces have been proven to be topologically equivalent to the original system and therefore are capable of recovering the nonlinear dynamics of the generating system.

This implies that the fully dynamics of the system are accessible in this space, and for this reason, a phase space reconstruction and features extracted from it can contain more and/or different information than a spectral representation. Phase space reconstruction consists of viewing a time series $X_k = X(kt), K = 1,2,\dots,N$ in a Euclidean space , where m is the embedding dimension and τ is the sampling time. By doing this we expect that the points in R^m form an attractor that preserves the topological properties of the original unknown attractor. Here, an attractor is a set towards which a dynamical system evolves over time. According to this concept, a dynamic system can be described by a phase space diagram, which is essentially a coordinate system, whose coordinates are all the variables that are necessary to completely describe the state of the system at any moment. Among a variety of methods available for phase space reconstruction, the method called "delayed coordinates" is well known and widely used. This method is based on the concept that we can reconstruct missing dimension using its previous and delayed function values as additional coordinates. A given time series, $X^i, i = 1,2,\dots,N,$ can be reconstructed in a multidimensional phase space to represent the underlying Dynamics According to

$$Y_j = (X_j, X_{j-\tau}, \dots, X_{j-2\tau}, \dots, X_{j-(m-1)\tau}), \quad (1)$$

Where j=1,2,…N-(m-1),and m is the dimension of the vector Y_j, also called as embedding dimension, and τ is the delay time. Further expanding we have:

$$Y = [Y_1, Y_2, \dots, Y_j, \dots, Y_M], \quad (2)$$

Where $Y_j$the vector of m dimension an M is is the number of vectors in this multidi-mensional phase space. M Can be given by $M = N - (m-1)$ .Based on the chaos theory the vector fully represents the nonlinear dynamics when $m$is large enough. A correct phase space construction in a dimension m facilitates an interpretation of the underlying dynamics. The physics behind such a reconstruction is that a nonlinear system is characterized by self-interaction, so that a time series of a single variable can carry the information about the dynamics of the entire multiple-variable system.To reveal the hidden structure of ISNs the phase space can be constructed by using "delayed coordinates". For a given sequence of *ISN*(i) numbers, the phase space is constructed as follows:

$$Y_i = (ISN(i), ISN(i-1), \dots, ISN(i-(m-1))),$$

Where i □1,2,..N □m□1,N is the number of ISNs, and m is the

Dimension. Vectors Y_i in the new phase space are formed from time delayed values of the scalar measurements. Used phase space reconstruction to build a spoofing set in predicting ISNs. Instead of using delayed coordinates, the

first-order difference for the input data is used in the phase space construction.

This method shows patterns of the correlation within a set of 32-bit ISNs generated by several operating systems' PRNG. By using "first-order difference" as the coordinates the phase space is constructed

$$x(n) = \mathrm{ISN}(n) - \mathrm{ISN}(n-1)$$
$$y(n) = \mathrm{ISN}(n-1) - \mathrm{ISN}(n-2)$$
$$z(n) = \mathrm{ISN}(n-2) - \mathrm{ISN}(n-3) \quad (5)$$

This is a three-dimensional representation of one-dimensional input data. Here $x(n), y(n)$ and $z(n)$ are called points coordinates.

The PRM Model:

A phase space is created by establishing vectors in R^m . According to the chaos theory, phase vectors can fully represent the nonlinear dynamics if the embedding dimension m is large enough. There are various methods to estimate the m including empirical methods. Different values of such as 2, 3, 4, 5 and 6 were tested in creating the reconstructed phase space as shown below, and m=4,5 and 6 gave us 100% detection accuracy rate. The larger the, the higher the computational complexity is. So m=4 is selected. The coordinates of four-dimensional vector are calculated as follows:

$$x(n) = \mathrm{ISN}(n) - \mathrm{ISN}(n-1)$$
$$y(n) = \mathrm{ISN}(n-1) - \mathrm{ISN}(n-2)$$
$$z(n) = \mathrm{ISN}(n-2) - \mathrm{ISN}(n-3)$$
$$w(n) = \mathrm{ISN}(n-3) - \mathrm{ISN}(n-4)$$

Where and *N* is the number of ISNs. The four–dimensional phase vector r_i is constructed as:

$$r_i = [x(i), y(i), z(i), w(i)], i = 1,2,..M, M = N - 5 \quad (6)$$

Let us use *R* to represent the phase space or dataset formed by the (5) and (6). If the number of ISNs is *N*, the number of phase vectors or elements in the phase space *R* is *N-5*, each is a four-dimensional vector.

$$R = [r1, r2, ... r_M] \quad (7)$$

As shown in fig,2 these vectors in phase space have some level of relations. In order to extract features from the dataset *R*,

we define distance between any two vectors r_i and r_j in the phase space as *R*:

$$d_{i,j} = \sqrt{(x(i) - x(j))^2 + (y(i) - y(j))^2} \times \sqrt{+(z(i) - z(j))^2 + (w(i) - w(j))^2},$$

Proposed Classification Algorithm:

The statistical model is developed by legally generated ISNs. In our experiment we used a dataset of 745 ISNs which are collected by using Win Dump for Windows XP SP3 operating system. Half of these dump used to construct the statistical model in the four-dimensional phase space, and then to obtain the third-order feature of the proposed statistical model.

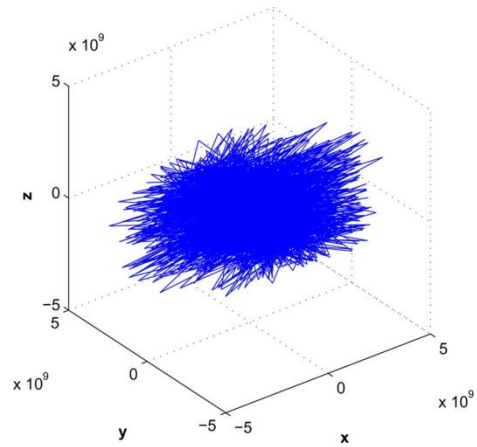

Fig. 2.Three-dimensional differential model.

The other half is used for testing. The stego-ISNs are generated by the algorithm Covert-TCP, in which ISN field is replaced with actual ASCII character to be encoded. The encoding of ASCII code word of letter H is performed by the 72 256 65536. Here, the ASCII code for character H is 72 (Hex). This enables sequence number looks realistic. Using this, packet is sent to the destination host. The destination host, expecting to receive information from client, simply grabs the ISN field of each coming packet to re-construct the encoded data. This way of en-coding secret message in ISNs has been tested in application and is considered a practical breakthrough to hide information in ISNs.

4. Experiments and Results

In Table 1, Data Set1 consists of 745 normal ISNs in which half of them used in training and the other half used in testing and 2000 randomly generated abnormal ISNs generated by Covert_TCP. Data Set2 and Data Set3 have 612 and 1000 normal ISNs, respectively. The 2000 abnormal ISNs are applied
to each testing.

Case	TN	TP	AR
Dataset1	100%	100%	100%
Dataset2	100%	100%	100%
Dataset3	100%	100%	100%

Table 1 (TN STANDS FOR TRUE NEGATIVE RATIO, TP STANDS FOR TRUE POSITIVE RATIO, AND AR STANDS FOR ACCURACY)

By using the third-order feature, we can clearly identify normal and stego-ISNs with accuracy rate 100%.It is known

that network steganalysis is not like image steganalysis, which could be carried out offline. The covert channels need to be detected online because header information (such as TCP header) will be taken off as the packet reached its final destination. It is possible that the secret message has been passed before collecting these stego-packets for training is finished.

5. CONCLUSION

In this paper, we have given an overview of covert channels in computer network proto-cols.We consider the fact that ISNs are produced by pseudorandom number generators, which are derived from deterministic chaotic dynamical systems. Chaos theory is therefore applied in analysing ISNs. Phase space reconstruction method is used to convert one dimensional ISN sequence to a group of four-dimensional vectors to reveal the inherent structure of ISNs The computational complexity involved in this proposed method is$O(3 \times n - samples)$, which is much lower than SVM with $(n\_sample^2 \times m\_features)$. Consequently it can be used for online detection. We also take care of data privacy in case any intruder crack the sequence number so an encryption technique is used before transmitting data using plain cipher encryption. And we also calculate the transmission time and plotted in bar graph by doing this the security of data shared in network is more increased even the hacker takes the packet he cannot read the data

REFERENCES

[1] B. Dunbar, A Detailed Look at Steganographic Techniques and Their Use in an Open-Systems Enviroment, SANS (SysAdmin, Audit, Network,Security) Institute, 2002.

[2] M. Owens, A Discussion of Covert Channels and Steganography SANS (SysAdmin, Audit, Network, Security) Institute, 2002.

[3] K. Szczypiorski, Steganography in TCP/IP Networks. State of the Art and a Proposal of a New System HICCUPS Institute of Telecommunications Seminar [Online]: http://www.tele.pw.edu.pl/krzysiek/pdf/steg-seminar-2003.pdf, Retrieved Jun. 2010

[4] R. J. Anderson and F. A. P. Petitcolas, "On the limits of steganography," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 4, pp. 474–481, May 1998.

[5] S. Attallah, Trusted Computer System Evaluation Criteria,Tech.Rep.DOD5200.28-STD,1985[Online]. Available:http://csrc.nist.gov/publistications/history/dod85.pdf.

[6] C. G. Girling, "Covert channels in LAN's," *IEEE Trans. Software Eng.*,vol. SE-13, no. 2, pp. 292–296, Feb. 1987.

[7] M. Wolf, "Covert channels in LAN protocols," *LNCS*, vol. 396, pp. 91–101, 1989.

[8] D. V. Forte, C.Maruti, M. R. Vetturi, and M. Zambelli, "SecSyslog: An approach to secure logging based on covert channels," in *Proc. First Int. Wksp. Systematic Approaches to Digital Forensic Engineering*, Nov. 2005, pp. 248–263.

AUTHORS

1)MERUGA VISHAL BHARADWAJ received the **B.Tech(CSE)** from JNTU Kakinada, in 2012& pursuing his M.Tech in Computer Science & Engineering from JNTU Kakinada.

2)AVS SUDHAKARA RAO presently working as **Associate professor**, Dept of Computer Science and Engineering, in St.Ann's College of Engineering and Technology. He guided many UG an PG students. He has more than 8 years of teaching experience. He published various international journals and conferences.