www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume - 3 Issue -9 September, 2014 Page No. 8312-8316

Detection of TCP/IP Covert Channel based on Naïve-Bayesian Classifier

Vibhor Kumar Vishnoi[1], Sunil Kumar[2]

[1]M.Tech student in Department of Computer Science& Engineering,
ABES Engineering College, Ghaziabad
Under U.P. Technical University, Lucknow
vibhorvishnoiyash@gmail.com

[2]Astt. Professor in Department of Computer Science& Engineering,
ABES Engineering College, Ghaziabad
Under U.P. Technical University, Lucknow
Sunil.kumar@abes.ac.in

Abstract: ***A covert channel is any methodology of communication that`s accustomed illicitly transfer data, so breaking the protection policy of a system. A network covert channel is a covert communication by hiding covert messages in to overt network packets. Any shared resource will probably used as a covert channel. In recent years with the development of various hiding methods, network covert channel has become a new kind of threat for network security. A covert channel is an unintended design within legitimate communication whose motto is to leak information as a part of rudimentary protocols. In fact, most detection systems can detect hidden data in the payload, but struggle to cope with data hidden in the IP and TCP packet headers. The vast number of protocols in internet seems ideal as a high-bandwidth vehicle for covert communication. Due to unwanted and malicious nature of covert channel applications and as it poses a serious security threat to network, it is recommended to detect covert channels efficiently. This paper presents a review of TCP/IP covert channel design and their detection scheme and presents a proposed method based on Naïve-Bayesian classifier to detect covert channels in TCP ISN and IP ID fields of TCP/IP packet.***

Keywords: TCP/IP covert channel, storage channel, Timing channel, TCP, IP, network security.

1. Introduction

Since Lampson proposed the concept of covert channel in 1973, covert channel has been taken as an important issue in the field of information security [1]. According to Lampson "a communication channel is covert if it is neither designed nor intended to transfer information at all". Later work defines a covert channel as "a communication channel that allows a process to transfer information in a manner that violates the system's security policy" [2]. This definition is now more commonly accepted. Initially, covert channels were identified as a security threat on monolithic systems i.e. mainframes but recently focus has shifted towards covert channels in computer network protocols.

A common analogy employed for discussing the dynamics of covert communications is one known as the "prisoner's problem" [3]. It involves two prisoners, called here for Simplicity Alice and Bob, who need to communicate with each other in order to devise an escape plan. It also involves a warden, Wendy, who oversees all interprisoner communications, and can monitor them in one of two ways:

- She can examine all messages, and let them pass or deny them based on what she sees. This is a passive approach.
- She can modify the message slightly to make sure it is not precisely what was sent, without changing the meaning of the message. By modifying the message it is assumed that she might frustrate any attempt of embedding a secret message in the communication. This is an active approach.

Ideally, the prisoners find a way to communicate which doesn't raise suspicion from the warden. But the warden must accept that there is a risk that some covert communication may be attempted, and pose and hypothesis of how it might function. Covert channel is different from cryptography as its main aim is to hide the existence of transmission whereas cryptography does not hide the existence of message but transform it in a form that is only readable by receiver. In cryptography there is no intention to hide the communication. Covert channel in computer network protocols and steganography are closely related but often confused. Steganography involves hiding of information in audio, visual, or textual content. While steganography requires some form of content as a cover, covert channel requires some network protocol as a carrier.

As network covert channels are communication channels that are not designed nor intended to exist, the communication streams must be embedded inside authorized channels. They may be based on existing protocols from OSI low layers (e.g.: IP, TCP, UDP) to OSI high layers (e.g.: HTTP, SMTP). The general idea of covert channels relies on the idea that information can be transferred in redundant or unused fields of network protocols. The reliability, speed and robustness of

communication protocols allow for the implementation of such channels over networks. Since network security analysts first started thinking about covert channel communication, two terms have been introduced, storage and timing covert channels. In storage covert channel, one of the processes directly or indirectly writes to a particular storage location whereas other process reads form that location. Number of tools employs TCP, IP, ICMP, and HTTP protocols to establish storage covert channels. In these protocols unused fields are used to transmit the information. In a way steganography can be seen as a form of storage covert channel. The timing covert channel involves modifying the time characteristics to hide information. Specifically it can be done by modulating inter-packet delays. In this paper, we pay our attention to detect covert channels related to TCP ISN and IP ID fields.

The remainder of this paper is organized as follows. Sec. 2, provides various ways for embedding covert information in TCP/IP protocol. Sec. 3, presents some covert channel detection mechanisms. Sec. 4 and 5 presents the problem statements and the proposed solution respectively. Sec. 6 , concludes the whole paper

2. Storage covert channel in TCP/IP protocols

In the next subsections, a *non-exhaustive* summary is given of known techniques to establish covert channels over TCP/IP protocols. We give an estimation of the theoretical efficiency of each mechanism and provide empirical observations for some of them.

2.1. IP

2.1.1. Mechanism

Version 4 of internet protocol i.e. IP_{v4} is a network layer protocol. This is an interesting carrier for covert channels. The header of IP datagram is depicted in fig. 1. As specified in RFC 791 [4] (some fields of interest are marked).

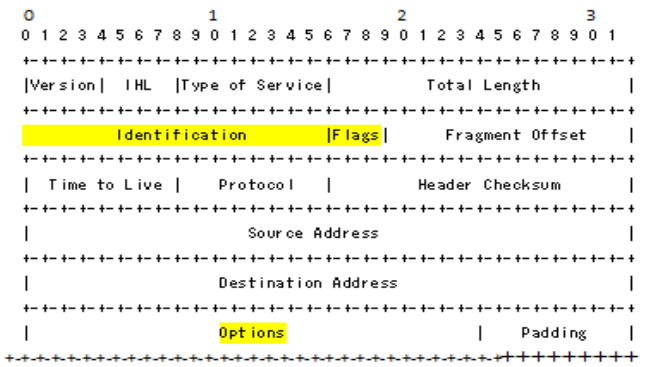

Figure 1: IP_{v4} header from RFC 971

➢ *Field IP Identification*

The 16- bit identification field is used to uniquely identify the IP datagram for reassembling of datagram in case of fragmentation. The value of this field should be chosen randomly by the source. An adversary may conceal 16 bits of data in this field and send it to any other networked system. *Rowland* proposed using the *IP identification (IPID)* field to construct covert channels [5]. Figure 2 shows a unidirectional channeling process over IP Identification field.

➢ *Field IP flags*

The IP 3-bit flag is used to handle fragmentation issues. As discussed in [6] the *'Don't Fragment' (DF)* flag may actually be considered to be a redundant bit. Hence it is an interesting carrier target for covert channels, even though it can only hold one bit per IP packet.

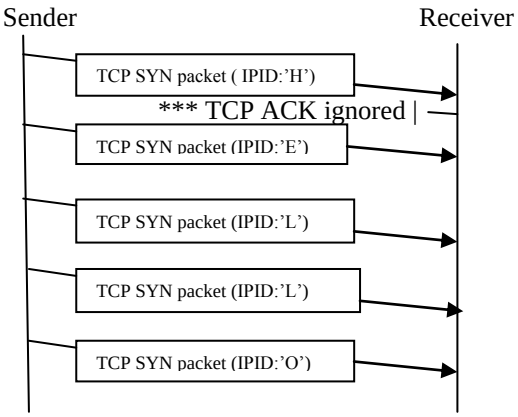

Figure 2: covert channel using IP Identification Field

➢ *Field IP option*

The 24-bit option field is optional for each IP datagram. IP option includes provisions for timestamps, security, and special routing. Adversaries may use this field to transfer information in covert form.

2.2 TCP

2.2.1 Mechanism

The Transmission control protocol is used for reliable data transmission in transport layer. It considered to be an equally evident carrier target for covert channels as IPv4. The header of a TCP packet, as specified in RFC 793 [7], is depicted in figure 3.

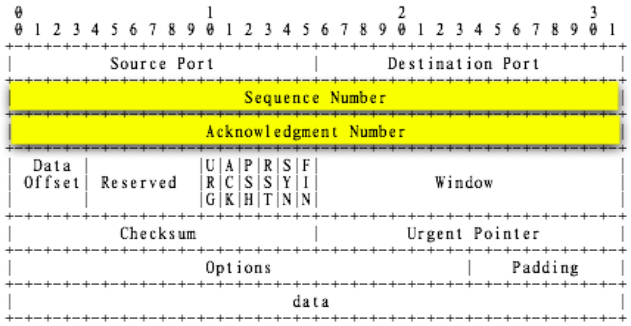

Figure 3: TCP packet header from RFC 793

➢ *Field TCP Sequence Number*

The 32-bit TCP sequence number field is used as an identification number to provide for packet reordering on arrival at receiver end and aid reliability through request for retransmission of individual packets. The first packet in TCP session contains a random initial sequence number, or ISN. The receiving host typically acknowledges it's receival by responding with a SYN/ACK packet, using ISN+1 as an acknowledgment number. Instead of using a random ISN, however, this field can also contain a non-random value without disrupting the TCP mechanism. An adversary may

conceal up to 32 bits of data in this field and send it to any other networked system. In [5] Rowland used ISN field to construct a covert channel. Figure 4 shows an example scenario for covert communication using TCP sequence number field.

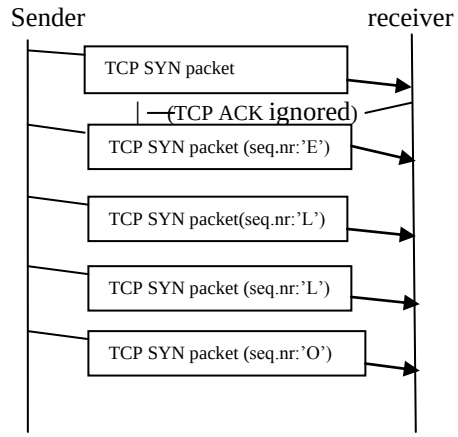

Figure 4: Covert channel using TCP sequence number

➢ *Field TCP Acknowledgement Number*

The 32-bit acknowledgment number field (byte 9-12) is used to acknowledge the receival of a TCP packet to its source. This field must always contain the sequence number of the sender, increment by 1. It has been demonstrated that adversaries may spoof the sender IP of a TCP packet, making the receiving host acknowledge to an arbitrary host with the (incremented) input bytes encoded into this field. Rowland outlined an indirect channel called the bounce channel (Fig. 5). Instead of sending the ISN directly in a TCP SYN packet to the receiver, the sender sends a TCP SYN to a bounce host with a spoofed IP source address set to the intended destination. On receiving the SYN packet the bounce channel sends the SYN/ACK or SYN/RST to the receiver with the acknowledged sequence number equal to the ISN+1. The receiver decrements the ACK number and decode the hidden information.

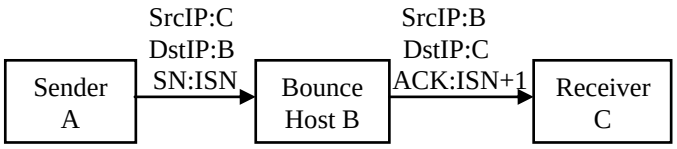

Figure 5: Covert channel using TCP Acknowledgement numbers through a bounce host.

➢ *Field TCP Urgent Pointer*

The 16-bit TCP Urgent pointer field is used when a segment contains urgent data. It is significant only if *URG* control flag is set. Whenever URG is not set the urgent pointer field becomes unused in this case any adversary may use this field as a covert channel. Hints [8] proposed to transmit the covert data in urgent pointer field of TCP packet that is unused if URG bit equals to 0.

3. TCP/IP Covert Channel Detection Mechanisms

In this section we provide related work done for TCP/IP storage covert channel detection.

In [9] J.Zhai *et.al* proposed a MAP based detection method to detect covert channel in TCPISN and RST fields of TCP packets under various applications such as HTTP, FTP, TELNET, SSH and SMTP. The behaviors of TPC flows are modeled by the Markov chain composed of the states of TCP packets. And the abnormality caused by covert channel is described by the difference between the overt and covert TCP transition probability matrix. Figure 6 presents the diagram of detection scheme proposed in this paper.

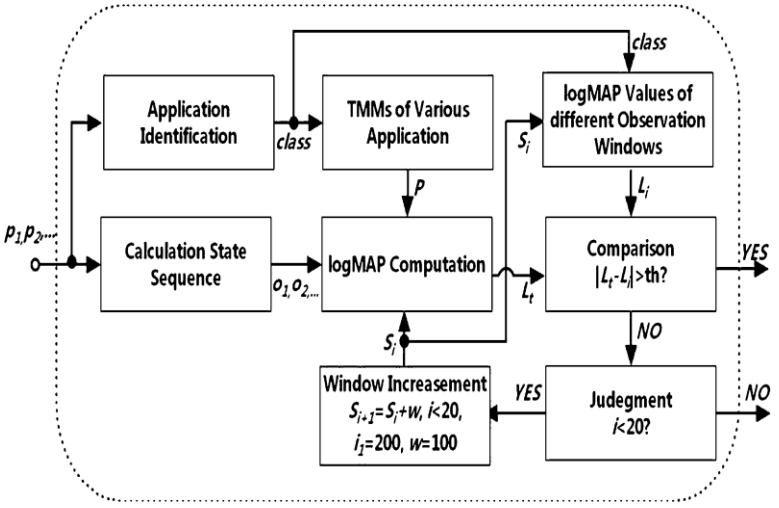

Figure 6: MAP based detection scheme form [9]

➢ This method relies on application identification based on TCP port. It might meet failure when applications without specific port are used.
➢ The detector must store double feature in an observation window because the TCP Markov model is based on two-way TCP link.

In [10] Hong Zhao *et.al* proposed a detection method to detect covert channel in TCP ISN field. In this scheme first, the phase space reconstruction is used to reveal the dynamic feature of ISNs. The statistical feature model is proposed based on reconstructed phase space dataset. The high order statistic analysis is conducted to construct the classifier which separates steganographic ISNs from normal ISNs. Figure 7 presents the classifier they used for covert channel detection.

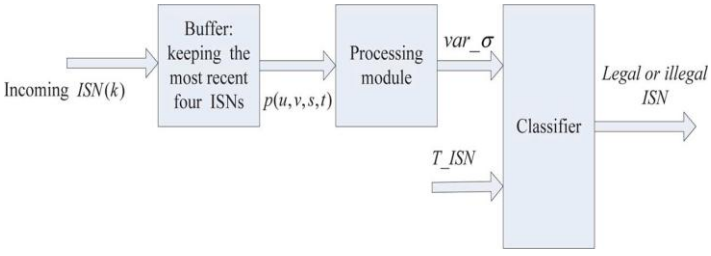

Figure 7: The proposed classifier in [10]

➢ This online malicious detection method is time consuming because of the first-come-first-serve concept and a very complex detection system.

In [11] T. Sohn *et.al* proposed a detection mechanism based on Support Vector Machine to detect covert channel in forged IP Identification and TCP ISN field. They used the Covert_tcp program as a covert channel generation tool for TCP/IP

header. The Trojan packets are themselves are masqueraded as common TCP/IP traffic. At the SVM training time they collected normal TCP/IP packets using a tcpdump tool and abnormal TCP/IP packets(including covert fields) generated from covert_tcp and then tested it for IP Identification field of IP header and sequence number field of TCP header.

- In this method author evaluated a three feature data set which increases its complexity to $O(n\_sample^2 \times m\_features)$ and time consumption as discussed in[10].

In [12] E. Tumoian *et.al* proposed a covert channel detection scheme based on ISN generation model of original OS. Whenever any statistical deviations of ISN network packet from the ISN model are discovered; it is considered that this ISN packet is generated by malicious software, which tries to create covert channel. The method was tested using experimental data generated by NUSHU covert channel creation tool.
Table 1 provides the summary of all the techniques that has been surveyed.

Table 1: comparison of storage covert channel detection techniques in TCP/IP

Detection Techniques	Fields used for creating covert channel	Method used for covert detection	Mode of detection	Detection accuracy	Responsible for network congestion	Feature Extraction
[9]	TCP ISN, RST	Markov model, MAP based detection	Offline	94.32%	No	No
[10]	TCP ISN	PRM model	Online	100%	Yes	Yes
[11]	TCP ISN, IP ID	SVM classifier	Offline	76.30% [9]	No	Yes
[12]	TCP ISN	Neural network based detection	Offline	60%	No	Yes

4. Problem Statement

Today internet is facing a many difficulties in protecting the data from theft, these difficulties comes from rapid improvement in technology and continuous enhancement in hacking techniques. This competition between attackers and information security experts makes protection of data very difficult to be achieved.

Network covert channel is way of transmitting hidden information that violates system security policies. Third party does not know the existence of such a channel. This activity by an attacker is an increasing potential security threat to internet. So it is necessary to detect such malicious activity which uses some mandatory fields of TCP/IP protocol header to leak/transfer information form one party to another.

While detecting covert channels in computer network, existing systems have following problems:

- Especially the online detection mechanism is based on First-come-first-serve concept. Due to this, these mechanisms are more time consuming.
- Existing systems for detecting covert channels are complicated.
- Traffic congestion may occur due to these covert detection methods.

All of the challenges given above are concerns that must be addressed in order to effectively detect the use of covert channels. In light of this eminent threat, we developed a framework to detect the leak of confidential information via covert channels. We limit our focus to instances where the users of covert channels modulate the information that is being sent by some form of encoding in two mandatory fields TCP ISN and IP ID of TCP/IP protocol headers.

5. Proposed method

The proposed system consists of following three processes. Such as:

- Covert channel analysis
- Attacker prediction
- Classifier method

5.1. Covert channel analysis and attacker prediction

TCP/IP covert channel requires alteration of packet header fields to transfer information without impacting the normal communication. Information can be embedded using modification to some header fields or using header fields which require random numbers. One can used unused header fields to encode information to transmit covertly.

For our purpose, we select two mandatory fields, IP ID field of IPv4 and TCP ISN field of TCP to embed covert data. The randomness of these fields makes attackers difficult to predict these numbers.

At first data packets are captured through our laboratory gateway *Wireshark 1.8.6* and stored in a database. Then a feature data set is created and used for training. The feature data set is obtained using above method and a tool named

NetScan Pro tool which is a packet generator tool used to generate TCP packets with embedded covert information in to ISN and IP header fields.

5.2. Classifier

The packet captured through wireshark form any network interface and covert information embedded packets created through NetScan packet generator, are first merged in to a single database using MergePcap facility provided by wireshark. To classify efficiently both types of packets (covert or normal), we proposed to use Naïve-Bayesian classifier in this work.

The Naïve-Bayesian classifier is a statistical classifier which predicts class membership probabilities, such as the probability that a given tuple belongs to a particular class. It is based on Bayes` theorem. This classifier exhibits a very high accuracy and speed when applied to a large dataset (database).

Let X is a data tuple. In Bayesian terms, X is considered "evidence". It is described by measurements made on a set of n attributes.

Let H be a hypothesis, such the data tuple X belongs to a specified class C.

For classification problem, we want to determine P(H/X), the probability that the hypothesis holds the given evidence or observed data tuple X. In other words we estimate the probability that tuple X belongs to a class C, given that we know the attribute description of X. Here P(H/X) is the posterior probability or a posteriori probability of H conditioned on X.

5.3. Block Diagram of Proposed System

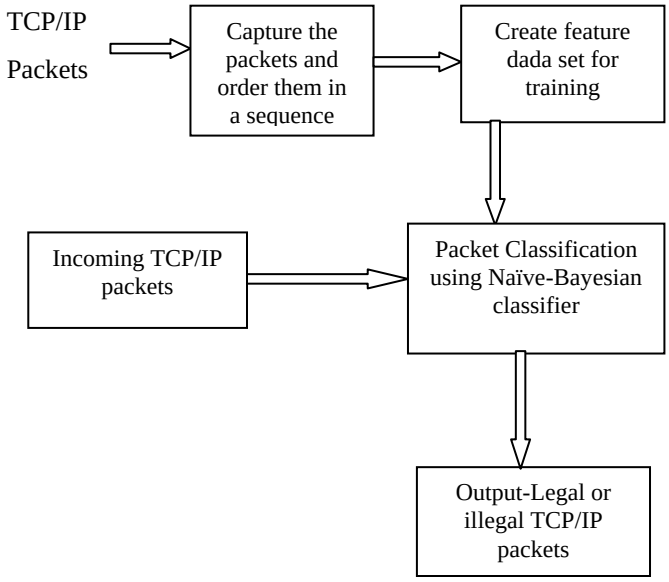

Figure 8: Block Diagram of Covert Channel Detection System

6. Conclusion

The huge amount of data transmitted over internet using TCP/IP protocols makes it ideal as carrier for covert information. Covert channel attacks become a potential threat to the internet. Covert channel using unused combination of flag fields of TCP/IP header, reserved fields or modification of some header fields can be easily detected. Detecting covert channels embedded in ISNs and IP IDs are the most difficult covert channels to be detected due to their random behavior. Proposed system detects covert channel in TCP ISN and IP ID, using the Naïve-Bayesian classifier in more efficient manner

.

References

[1] Lampson, B. (1973). A note to the confinement problem. Communications of the ACM, *16*(10), 613–615

[2] U.S. Department of Defense. Trusted Computer System Evaluation "The Orange Book". Publication DoD 5200.28-STD. Washington: GPO-1885, http://www.radium.ncsc.mil/tpep/library/rainbow/5200.28-STD.html

[3] G. J. Simmons, "The Prisoners' Problem and the Subliminal Channel," in Proceedings of Advances in Cryptology (CRYPTO), pp. 51–67, 1983R.

[4] Postel, Jon: RFC 791 - Internet Protocol Specification, 1981, http://www.ietf.org/rfc/rfc791.txt

[5] C. H. Rowland, ―Covert Channels in the TCP/IP Protocol Suite,‖ First Monday, Peer Reviewed Journal on the Internet,July 1997.

[6] D. Kundur and K. Ahsan, ―Practical Internet Steganography: Data Hiding in IP,‖ Proc. Texas Wksp. Security of Information Systems, Apr. 2003

[7] Postel, Jon: RFC 793 - Transmission Control Protocol, 1981, http://www.ietf.org/rfc/rfc793.txt

[8] Hintz, A. (2003). Covert channels in TCP and IP headers. URL: http://www.defcon.org/images//dc10-hintz-covert.ppt.

[9] J. Zhai, G. Liu,Y. Dai. " Detection of TCP covert channel Based on Markov Model", Telecommun Syst(2013) 54:333-343, DOI 10.1007/s11235-013-9737-7.

[10] Hong Zhao, Yun-Quing Shi(2013)- IEEE Transactions on On Information Forensics and Security. Vol. 8. No.2.

[11] Sohn, T., Seo, J., & Moon, J. (2003). A study on the covert channel detection of TCP/IP header using support vector machine. In Proceedings of the 5th international conference of information and community security (pp. 313–324).

[12] Eugene Tumoian, Maxim Anikeev, ―Network Based Detection of Passive Covert Channels in TCP/IP‖, Proceedings of the IEEE Conference on Local Computer Networks 30th Anniversary, 2005.