Covert Channel Detection in the ICMP Payload Using Support Vector Machine

Conference Paper *in* Lecture Notes in Computer Science · November 2003

DOI: 10.1007/978-3-540-39737-3_103 · Source: DBLP

CITATIONS
31

READS
1,438

5 authors, including:

Jongsub Moon
Korea University
92 PUBLICATIONS **1,403** CITATIONS
SEE PROFILE

Sangjin Lee
Korea University
365 PUBLICATIONS **4,751** CITATIONS
SEE PROFILE

Dong Hoon Lee
Korea University
769 PUBLICATIONS **17,529** CITATIONS
SEE PROFILE

Jongin Lim
Korea University
240 PUBLICATIONS **3,752** CITATIONS
SEE PROFILE

Covert Channel Detection in the ICMP Payload Using Support Vector Machine

Taeshik Sohn, Jongsub Moon, Sangjin Lee, Dong Hoon Lee, and Jongin Lim

Center for Information Security Technologies, Korea University,
1-ga, Anam-dong, Sungbuk-gu, Seoul, Korea
{743zh2k, jsmoon, sanjin, donghlee, jilim}@korea.ac.kr

Abstract. ICMP traffic is ubiquitous to almost TCP/IP based network. As such, many network devices consider ICMP traffic to be benign and will allow it to pass through, unmolested. So, attackers can generate arbitrary information tunneling in the payload of ICMP packets. To detect a ICMP covert channel, we used SVM which has excellent performance in pattern classification problems. Our experiments showed that the proposed method could detect the ICMP covert channel from normal ICMP traffic using SVM.

1 Introduction

Currently, internet environment has many problems in information security as its network is increasing rapidly. So, various solutions for security protection such as IDS, firewall, ESM, VPN have been evolved. Although these solutions were widely used, they have yet much vulnerability due to the problems of the protocol itself. One of the vulnerability is the possibility of hidden channel creation. The hidden channels are defined as the communication channel used in a process which transmits information by some methods violating the system's security policy. This channel is used for transmitting special information to processes or users prevented from accessing the information[1]. In this paper, we analyze the method which transmits the hidden data using the ICMP protocol payload. To detect the ICMP covert channel, we used SVM, which is known as a kind of the Universal Feed Foreword Network proposed by Vapnik in 1995, is efficient to complex pattern recognition and classification. Specially, it has the best solution for the classification problems[2][3].

The rest of this paper are organized as follows: Section 2 addresses related work of Covert Channel. Section 3 describes the background of SVM. Our detection approach is explained in section 4. Experiments are explained in section 5, followed by conclusions and future work in section 6.

2 Related Work

A security analysis for TCP/IP is found in [4]. Paper[5] and [6] describe works related to the various covert channel establishments. A general survey of

A. Yazici and C. Şener (Eds.): ISCIS 2003, LNCS 2869, pp. 828–835, 2003.

information-hiding techniques is described in "Information Hiding - A Survey." John Mchugh[5] provides a wealthy of information on analyzing a system for covert channels in "Covert Channel Analysis." Specially, "Covert Channels in the TCP/IP Protocol Suite"[6], Craig Rowland describes the possibility of passing covert data in the IP identification field, the initial sequence number field and the TCP acknowledge sequence number field. He programmed a simple proof-of-concept, raw socket implementation. "Covert Messaging Through TCP Timestamps"[7], describes a tunnel using timestamp field in TCP header. "Loki: ICMP Tunneling, daemon9, Pharack Magazine, Volume 6, Issue 49, article 6 of 16"[8] describes the possibility of generating covert channel in a ICMP protocol.

3 Support Vector Machine

3.1 Background

The SVM is based on the idea of structural risk minimization, which minimizes the generalization error, i.e. true error on unseen examples. The number of free parameters used in the SVM depends on the margin that divides the data points into classes but not on the number of input features, thus the SVM does not require a reduction in the number of features in order to avoid overfitting. The SVM provides a generic mechanism to fit the data within a surface of a hyperplane of a class through the use of a kernel function. The user may provide a kernel function, such as a linear, polynomial, or sigmoid curve, to the SVM during the training process, which selects support vectors along the surface of the function. This capability allows classifying a broader range of problems. The primary advantage of SVM is binary classification and regression that it provides to a classifier with a minimal VC(Vapnik Chervonenkis)-dimension, which implies low expected probability of generalization errors [2][3].

3.2 SVM for Classification

In this section we review some basic ideas of SVM. For the details about SVM for classification and nonlinear function estimation, see [2][3][9][10]

Given the training data set $\{(x_i, d_i)\}_{i=1}^{l}$, with input data $x_i \in R^N$ and corresponding binary class labels $d_i \in \{-1, 1\}$, the SVM classifier formulation starts from the following assumption. The classes represent by subset $d_i = 1$ and $d_i = -1$ are linearly separable, where $^{\exists} w \in R^N,\ b \in R$ such that

$$\begin{cases} w^T x_i + b > 0 \text{ for } d_i = +1 \\ w^T x_i + b < 0 \text{ for } d_i = -1 \end{cases} \tag{1}$$

The goal of SVM is to find an optimal hyperplane for which the margin of separation ,ρ, is maximized. The margin of separation ,ρ, is defined by the separation between the separating hyperplane and the closest data point. If the optimal hyperplane is defined by $w^T + b_0 = 0$, then the function $g(x) = w^T x + b_0$

gives a measure of the distance from x to the optimal hyperplane. Support Vectors are defined by data points $x^{(s)}$ that lie the closest to the decision surface. For a support vector $x^{(s)}$ and the canonical optimal hyperplane g, we have

$$r = \frac{g(x^s)}{||w_0||} = \begin{cases} +1/||w_0|| \text{ for } d^{(s)} = +1 \\ -1/||w_0|| \text{ for } d^{(s)} = -1. \end{cases} \tag{2}$$

Since, the margin of separation $\rho \propto \frac{1}{\|w_0\|}$. Thus, $\|w_0\|$ should be minimal to achieve the maximal separation margin. Mathematical formulation for finding the canonical optimal separation hyperplane given the training data set $\{(x_i, d_i)\}_{i=1}^{l}$, solve the following quadratic problem

$$\begin{cases} \text{minimize } \ \tau(w, \xi) = \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{l} \xi_i \\ \text{subject to } d_i(w^T x_i + b) \geq 1 - \xi_i \ \text{ for } \ \xi_i \geq 0,\ i = 1, ..., l \end{cases} \tag{3}$$

Note that the global minimum of above problem must exist, because $\phi(w) = \frac{1}{2}\|w_0\|^2$ is convex in w and the constrains are linear in w and b. This constrained optimization problem is dealt with by introducing Lagrange multipliers $\alpha_i \geq 0$ and a Lagrangian function given by

$$L(w, b, \xi; \alpha, \nu) = \tau(w, \xi) - \sum_{i=1}^{l} \alpha_i [d_i(w_i^T x_i + b) - 1 + \xi_k] - \sum_{i=1}^{l} \nu_i \xi_i \tag{4}$$

$$\frac{\partial L}{\partial w} = 0 \Longleftrightarrow w - \sum_{i=1}^{l} \alpha_i d_i x_i \ \ (\therefore w = \sum_{i=1}^{l} \alpha_i d_i x_i) \tag{5}$$

$$\frac{\partial L}{\partial b} = 0 \Longleftrightarrow w - \sum_{i=1}^{l} \alpha_i d_i = 0 \tag{6}$$

$$\frac{\partial L}{\partial \xi_k} = 0, \text{ for } 0 \leq \alpha_i \leq c \ , \ k = 1, ..., l \tag{7}$$

The solution vector thus has an expansion in terms of a subset of the training patterns, namely those patterns whose α_i is non-zero, called(we previous defined) *Support Vectors*. By the Karush-Kuhn-Tucker complementarity conditions, we have,

$$\alpha_i [d_i(w^T x_i + b) - 1] = 0 \text{ for } i = 1, ..., N \tag{8}$$

by substituting (5), (6) and (7) into L(equation (4)), find multipliers α_i which

$$\text{maximize } \Theta(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j d_i d_j < x_i \cdot x_j > \tag{9}$$

$$\text{subject to } 0 \leq \alpha_i \leq c,\ i = 1, ..., l \text{ and } \sum_{i=1}^{l} \alpha_i y_i = 0 \tag{10}$$

The hyperplane decision function can thus be written as

$$f(x) = sgn\left(\sum_{i=1}^{l} y_i \alpha_i \cdot (x \cdot x_i) + b\right) \tag{11}$$

where b is computed using (8).

To construct the SVM, the optimal hyperplane algorithm has to be argumented by a method for computing dot products in feature spaces nonlinearly related to input space. The basic idea is to map the data into some other dot product space (called the feature space) F via a nonlinear map ϕ, and to perform the above linear algorithm in F, i.e nonseparable data $\{(x_i, d_i)\}_{i=1}^{l}$, where $x_i \in R^N,\ d_i \in \{+1, -1\}$, preprocess the data with,

$$\phi : R^N \to F,\ x \mapsto \phi(x) \text{ where } l \ll \mathsf{dimension}(F) \tag{12}$$

Here w and x_i are not calculated. According to Mercer's theorem,

$$\langle \phi(x_i), \phi(x_j) \rangle = K(x_i, x_j) \tag{13}$$

and $K(x, y)$ can be computed easily on the input space. Finally the nonlinear SVM classifier becomes

$$f(x) = sgn\left(\sum_{i=1}^{l} \alpha_i d_i K(x_i, x) + b\right) \tag{14}$$

4 ICMP Covert Channel Detection

4.1 An Overview of ICMP Covert Channel

ICMP type 0x0 specifies an ICMP echo reply and type 0x8 indicates an ICMP echo request. This is what the ping program does. This ping traffic is ubiquitous to almost every TCP/IP based network and subnetwork. As such, many networks consider ping traffic to be benign and will allow it to pass through, unmolested. ICMP echo packets also have the option to include a payload. This data section is used when the record route option is specified, or the more common case, usually the default to store timing information to determine round-trip times. Although the payload is often timing information, there is no check by any device as to the content of the data. So, as it turns out, this amount of data can also be arbitrary in content as well[8]. The arbitrary contents of the payload can have various data according to the message types of ICMP protocol and kinds of the operating system(OS). In case of the normal ICMP packet, it has insignificant values or null values and so on. Namely, therein can lie the covert channels.

4.2 Proposing the Detection Method of ICMP Covert Channel

In this paper, we propose a model to detect covert channel in the ICMP payload. The payload of ICMP packets have generally null values or several characteristics dependent on the OS such as Windows, Linux and Solaris as illustrated in Table 1. At this time, the characteristic of payload of each OS is normally the same or it has the successive change of one byte in the payload. The rest 4 bytes of ICMP header are dependent on the each ICMP message type. Thus, we propose the detection method of ICMP covert channel using SVM with the characteristic of ICMP packet payload and the 4 bytes of ICMP header described above. First, we collect normal ICMP packets using a packet capturing tool like tcpdump and abnormal ICMP packets generated by covert channel tool like Loki2[8]. Then we preprocess the collected raw packets such as ICMP payload(13 dimensions) and ICMP payload plus the rest 4 bytes of ICMP header(15 dimensions) as illustrated in Figure 1. One dimension of preprocessed data is comprised of

Table 1. The Characteristic of ICMP payload

	ICMP Payload
Null Packet	0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000
Win Packet	0900 6162 6364 6566 6768 696a 6b6c 6d6e 6f70 7172 7374 7576 7761
Solaris Packet	50ec f53d 048f 0700 0809 0a0b 0c0d 0e0f 1011 1213 1415 1617 1819
Linux Packet	9077 063e 2dbd 0400 0809 0a0b 0c0d 0e0f 1011 1213 1415 1617 1819

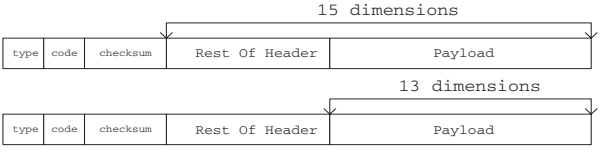

Fig. 1. The features of SVM

two bytes of raw packet. So, 26bytes ICMP payload is 13 dimension and the rest of header(4bytes) + 26bytes ICMP payload is 15 dimension. At this time, the preprocessed packets are classified as a training data set and a test data set. Each dimension, which means one input feature in SVM, is converted to decimal value, that is, the hexa values of 16bits(2bytes) are rearranged by the integer value of decimal in the raw dump values of packet. Finally, we learn the SVM using training data set and classified test data set with the SVM classifier[11].

5 Experiment

5.1 Experiment Methods

The SVM data is first comprised of two data sets : a training data set and a test data set. Also these data sets consist of the training set1, training set2 and test set1, test set2 as followed in Table 2. We preprocessed the raw packet

Table 2. SVM Training Data Set and Test Data Set

Data Set	Training Set1(Total 4000)	Training Set2(Total 4000)
Normal Packet	All typed ICMP packet(2000)	OS based ICMP packet(2000)
Abnormal Packet	Loki packet (2000)	Loki packet (2000)
Data Set	Test Set1 (Total 1000)	Test Set2 (Total 1000)
Normal Packet	All typed ICMP packet(250) + OS based ICMP packet(250)	All typed ICMP packet(250) + OS based ICMP packet(250)
Abnormal Packet	Loki packets(500)	Loki packets(500)

values in order to make training data and testing data in the SVM. Through the preprocessing, the feature is determined. The determined features are comprised of two cases : a 15 dimension including the rest 4bytes of packet header and packet payload and a 13 dimension including only the payload of packet. Each training data set is comprised of 4000 packets. Training set1 data has general ICMP packets collected by CIST server(Our Institute Web Server). Training set2 data has ICMP packets based on the characteristic of operating systems(Linux, Solaris, Windows). Also, abnormal packets are generated by Loki2 tool. Each training set contains 2000 abnormal packets. Next, the test set1 and the test set2 consist of 500 normal packets and 500 abnormal packets. Here, the normal packets of the test set have 250 general ICMP packets and 250 OS dependent packets. The abnormal packets of the test set have packets which are forged using Loki2 tool. SVM detection tool used here is the freeware package mySVM[11]. To compare the detection performance, we used the two SVM kernels : linear, polynomial type.

Table 3. The experiment results of ICMP covert channel detection

Training Set	Kernel	Features	Test Set1			Test Set2		
			FP	FN	TC	FP	FN	TC
Training Set1	Linear	13	2.5	0.4	97.1	0.7	0.7	98.6
		15	1.1	0.8	98.1	0	0.6	99.4
	Polynomial	13	0.2	0.6	99.2	0	0.8	99.2
		15	0.8	0.6	98.6	0	0.8	99.2
Training Set2	Linear	13	24.3	0.8	74.9	12.1	1.6	86.3
		15	0	0.8	99.2	0	0.6	99.4
	Polynomial	13	3.8	0.6	95.6	2.5	1.0	96.5
		15	0.8	0.6	98.6	0	0.2	99.8

*The degree of Polynomial Kernel = 3, FP = False Positive(%), FN = False Negative(%), TC = Total Correctness(%)

5.2 Experiment Results

In the detection experiments of the ICMP covert channel using an SVM, the learning data set is classified as training set1 and training set2 according to the

characteristic of the ICMP payload. We analyzed the detection results of each test set1 and test set2 as the two SVM kernel functions and the variation of the number of features. Table 3 shows overall experiment results. The resultant graph of covert channel detection using training set 1 is shown in Figure 2 and the resultant graph of covert channel detection using training set 2 is illustrated Figure 3. In case of the SVM training set, we can see that it is more efficient to classify the abnormal packets by assuming all the general types of ICMP packets are normal(The detection rate of Training set1 is 98.68%). Also, we can see in table 4 that detection performance is better in a 15 dimension and is best in polynomial kernel with degree of 3.

In this paper, we proposed the SVM method with some features of ICMP. Such an SVM could detect ICMP covert channel with the correction rate of nearly 99%(As illustrated in Table 4).

Table 4. The experiment results of each parameter(%)

	TR1	TR2	KR1	KR2	F13	F15
Detection(%)	98.68	93.79	94.13	98.34	93.43	99.04

*TR1 = Training Set1, TR2 = Training Set2, KR1 = Linear, KR2 = Polynomial, F13 = 13 Features, F15 = 15 Features

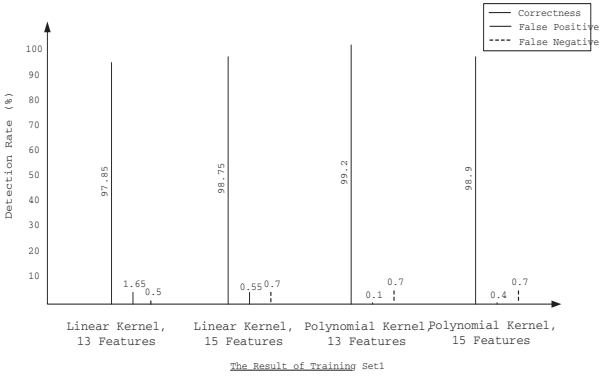

Fig. 2. The results of Training Set1

6 Conclusion and Future Work

Covert channel attacks are an increasingly potential threat to Internet environments. We did not yet have the best solution for covert channel detection. The goal of this paper was to propose the detection method for ICMP covert channels with SVM among the many covert channels. The preprocessing for SVM learning to detect a covert channel consisted of two cases: one case includes only

an ICMP payload and the other case includes an ICMP payload and the remaining 4 bytes of the ICMP header. We classified training sets into training set 1 with generally used ICMP packets and training set 2 with ICMP packets based on the characteristic of the operating system. Also, the experiment environment has been subjected to informal tests in a laboratory test bed. The experiment results show that under these conditions, the detection provided by the SVM learning described the high correction rate as illustrated in Table 4.

Future work will include the expansion of training sets, test sets and the experiments for various kernels which can be use for performance improvement and some of its constraint parameters.

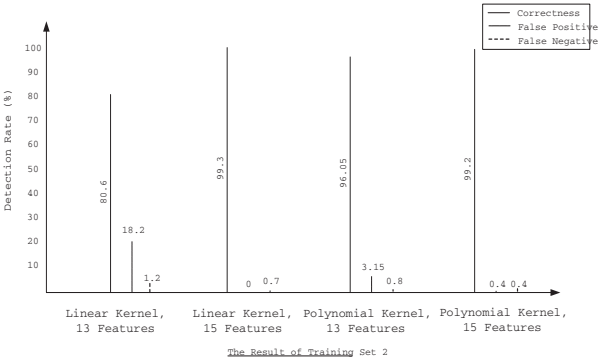

Fig. 3. The results of Training Set2

References

1. U.S. Department Of Defence, "Trusted Computer System Criteria.",1985
2. Vapnik V., "The Nature of Statistical Learning Theory", Springer-Verlag, 1995.
3. Bueges C J C., "A Tutorial on Support Vector Machines for Patter Recognition.", Data Mining and Knowledge Discovery, Boston, 1988.
4. S.M. Bellovin, "Security Problems in the TCP/IP protocol suite", Computer Communication Reviews,19(2):32–48, April 1989
5. John McHugh, "Covert Channel Analysis", Portland State University, 1995
6. Craig H. Rowland, "Covert Channels in the TCP/IP protocol suite", First Monday, 1996
7. John Giffin, "Covert Messaging Through TCP Timestamps", PET2002
8. Daemon9, "Loki: ICMP Tunneling", Pharack Magazine, Volume 6, Issue 49
9. Cristianini N., "An Introduction to Support Vector Machines.", Cambridge University press, 2000.
10. S. Mukkamala, G. Janowski, A. H. Sung, "Intrusion Detection Using Neural Networks and Support Vector Machines", IEEE IJCNN, May 2002, pp.1702–1707.
11. Joachmims T,"mySVM – a Support Vector Machine", Univerity Dortmund