ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

Covert Channel Attacks Against Anonymizer Architecture

[I]V.Swathy, [II]N.Manjubashini
[I,II]PG Scholar, Dept. of CSE, Kathir College of Engineering, Neelambur,Coimbatore, India

Abstract
Anonymizer is a proprietary anonymous communication system. It is moreover a tool that makes activity on the internet untraceable. In order to compromise the service provided by the anonymizer, packet size based covert channel attacks are made. An attacker between a malicious website and anonymizer server embeds a secret message (or) signal into the target traffic, whereas the attacker's partner at the user side snuffle the traffic and recognize the secret message. To preserve the distribution of web packet size Monte Carlo Sampling technique is adopted. To measure the TCP packet size dynamics Hurst parameter is used.

Keywords
Anonymizer, covert channel, monte carlo sampling technique.

I. Introduction

In this paper, anonymizer is explored and packet size based covert channel attacks are presented that degrades the anonymous service provided by the anonymizer. It is obviously an active traffic analysis. The main idea of this attack is to embed a secret signal (or) message at one end by the attacker and the same signal is recognized by the attacker's partner at the other end of the communication channel[1] [2]. The architecture of anonymizer consists of a reverse proxy, Secure Shell servers and web proxy servers. It is found that the size of HTTP packets in anonymizer network is very dynamic.

In order to compromise the service provided by the anonymizer covert channel attacks are designed. In basic covert channel attack, the Empirical Cumulative Distribution Function (ECDF) of each and every packet is evaluated which is then sampled by means of Monte Carlo sampling technique. In Monte Carlo sampling technique random numbers carrying the secret signals are mapped to sequence of packet sizes. But this random sampling may disturb the regularity and self similarity of TCP packet size dynamics.

To overcome the drawbacks of this basic covert channel attack, enhanced covert channel attacks are introduced. In this, the web traffic is repacketized into virtual web objects. Then the secret messages are modulated into the size of last packets of these virtual web objects. The last packet of a web object is denoted as least significant packet and it is very dynamic. The modulation of successive packets to carry message bits will disrupt the TCP packet size dynamics and that can be measured by Hurst parameter [3] [4]. Hence this least significant packet size based covert channel attack can preserve the TCP regularity and self similarity. To preserve the distribution of web packet size Monte Carlo sampling technique is used. First of all, the message which has to be sent as a secret signal is encrypted by means of a strong cipher. The generated cipher text is encoded into k-bit symbols. A k-bit symbol can then be mapped to a packet size by Monte Carlo sampling technique. But the packet size distortion caused by anonymizer and Internet traffic may cause difficulties in recovering the secret message, thus intelligent and robust detection algorithm is used to recover them.

The remainder of this paper is organized as follows: In Section II, The related works are reviewed. In Section III, basic covert channel attack against the anonymizer architecture is presented. In Section IV Enhanced covert channel attack based on least significant packets is proposed. In Section V, the analysis on the outcome is presented. The paper is concluded in Section VI.

II. Related Work

Sun et al. [2] investigated the sizes of the HTML objects transmitted over SSL connections and were able to identify the web pages based on the number and size of objects in each encrypted HTTP response. Without aggressive length padding, an algorithm was used to identify many web pages with low false positive rate based on number and size of objects alone.

Liberatore and Levine [5] examined that the packet sizes of HTTP traffic transmitted over persistent connection or tunneled via SSH port forwarding can statistically identify the web pages. The active watermarking techniques intend to embed specific secret signal into the target traffic. These techniques reduce the false positive rate significantly if the signal is long enough.

Ramsbrock et al. [6] developed a novel packet flow watermarking technique to address the botmaster traceback problem. A practical solution was given to trace low-volume botnet C&C traffic in real-time even if it is encrypted and laundered through multiple intermediate hosts. The tracer has control of a single rogue bot in the target botnet, and this bot can send a response to a query from the botmaster.

Wang et al. [1] analyzed the fundamental limitations of achieving anonymity in flow transformations. By injecting unique watermark into the inter-packet timing domain of a packet flow, they were able to make any long flow uniquely identifiable. It has been proved that the flow watermarking attack only needs about 10 minutes active web browsing traffic to penetrate the service provided by anonymizer.

III. Basic Covert Channel Attack

The basic idea of this attack is as follows: An attacker embeds a secret message into the web traffic; whereas the attacker's partner at the client side snuffles the message thereby the service provided by the anonymizer is compromised. The workflow of this attack is depicted in Fig.1

A. Generation of packet size sequence

An attacker at the malicious website controls the reverse proxy to embed a secret message into the web traffic. When the malicious website receives the HTTP requests for web objects through anonymizer, it transmits the requested web objects to reverse proxy. The attacker at the reverse proxy generates a sequence of random numbers between 0 and 1.

By using Monte Carlo sampling and Empirical Cumulative Distribution Function (ECDF) the random numbers are mapped to

sequence of packet sizes. Let the probability mass function (PMF) of the raw packet size is denoted as $\{p_1, p_2, p3....p_n\}$ corresponding to the sequence of packet sizes $\{ps_1, ps_2, ps3....ps_n\}$ The ECDF of raw HTTP packet size is given as:

$$F_e(ps_i) = P(x \le ps_i) = p_1 + p_2 + p_3 + ... + p_i \quad (1)$$

The sequence of random numbers $\{ra_1, ra_2, ra3....ra_l\}$, where ra ∈ [0, 1] and l is the length of the sequence is mapped to the packet sizes if and only if

$$F_e(ps_i) \ge ra_i \quad (2)$$

The mapping between the random number and HTTP packet size can be written as:

$$Table_1 : \{ra_1, ra_2, ra_3, ..., ra_l\} \Leftrightarrow \{p_{ra\,0}, p_{ra1}, ..., p_{ral}\}$$

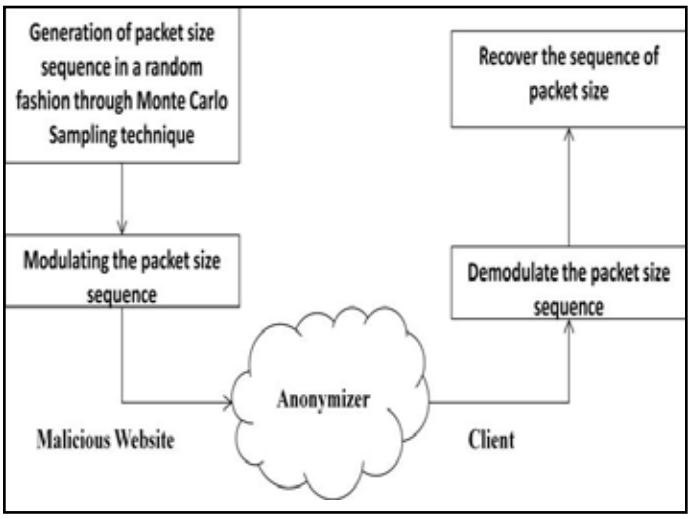

Fig. 1. Workflow of Basic Covert Channel Attack

B. Modulation and Demodulation of packet size sequence

The attacker at the reverse proxy modulates packet size variation according to $\{p_{ra0}, p_{ra1}, pra2,....p_{ral}\}$. The web traffic with these modulated HTTP packet sizes will be forwarded to anonymizer. The sequence of SSH packet sizes are deduced from the sequence of raw HTTP packet sizes.

$$Table_1' : \{p_{ra\,0}, p_{ra1}, ..., p_{ral}\} \Leftrightarrow \{ps_{ra\,0}, ps_{ra1}, ..., ps_{ral}\}$$

Algorithm to calculate SSH packet size based on raw HTTP packet.

Input: $S, S_h, S_c, S_p, S_b, S_a$

Output: S_s

Process:

$S_p = S_b - (S_h + S_c + S) \bmod S_b$

if $S_p < 4$ then

$S_p = S_b + S_p$

end if

$S_s = S_h + S_c + S + S_p + S_a$

C. Recovering the packet size sequence

The HTTP proxy of anonymizer filters HTTP header fields and the raw HTTP packet sizes are changed. To accurately recognize the embedded symbols, the attackers partner at the client side records the SSH packets and obtain the SSH packet size based on raw HTTP packets.

IV. Enhanced Covert Channel Attack Based On Least Significant Packets

To overcome the limitations of basic covert channel attack, enhanced covert channel attack is designed. The normal HTTP packets can be categorized into two classes: Class1 consists of HTTP packets whose size is greater than 1500, Class2 consists of packets whose size is less than 1500. These Class2 packets are referred to as the least significant packets. The workflow of this enhanced covert channel attack is depicted in Fig.2.

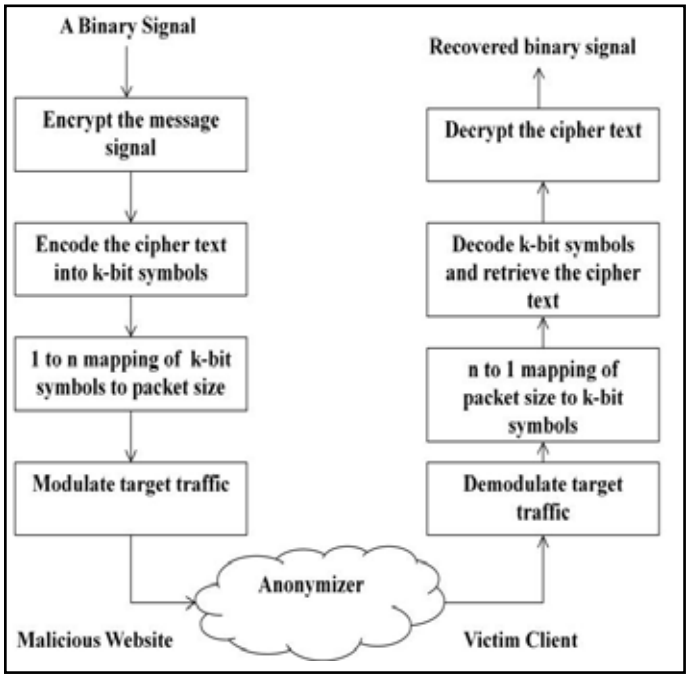

Fig. 2: Workflow of Enhanced Covert Channel Attack

A. Embedding a signal at the malicious website

1) Encrypting the signal

A message signal is being selected and encrypted (RC4 in counter mode with an initialization vector (IV) is used) by a strong cipher.

$$E_{rc\,4}(Message, IV) \rightarrow \{B_1, B_2, ..., B_{8u}\}$$

2) Encoding cipher text into k-bit symbols

The ECDF of the least significant packet size is used to find maximum k.

$$2_k \le 1/p_{max},$$

After mapping the sequence of binary bits can be translated into,

$$\{B_1, B_2, ..., B_{8u}\} \Leftrightarrow \{S_1, S_2, ..., S_{8u\,k}\}$$

3) Mapping k-bit symbols to packet sizes

Monte Carlo sampling technique is used to map the cipher text symbol to an appropriate least significant packet size.

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

$$\{s_0, s_1, \ldots, s_{2^k-1}\} \Leftrightarrow \{p'_{s0}, p'_{s1}, \ldots, p'_{s(2^k-1)}\}$$

B. Recovering the message at the client side

1. Demodulating the packet size sequence

Accomplice of the attacker at the client side sniffs the SSH traffic transmitted to the user and records the packet size. Then the size of SSH packets corresponding to raw HTTP packets are calculated.

2. Mapping packet sizes to k-bit symbols

Accomplice records the size of SSH packet carrying the symbols and obtain corresponding raw HTTP packet size. The obtained HTTP packet sizes will be mapped to the symbols.

Then the sequence of the mapped symbols is derived as $\{S_1, S_2, \ldots, S_{2u}\}$

3. Decoding k-bit symbols into signals

The recovered sequence of symbols are translated into binary bits

$$\{B_1, B_2, \ldots, B_{8u}\}$$

4. Decrypting the ciphet text

The message is decrypted using RC4 algorithm with IV as:

$$D_{rc4}(\{B_1, B_2, \ldots, B_{8u}\}, IV) \rightarrow Message$$

If the decrypted message is same as the original one, the attacker confirms the relationship between the user and malicious website.

V. Analysis On The Covert Channel Attack

1. Generation of packet size sequence

Packet Size Sequence	Probability Mass Function(PMF)	Empirical Cumulative Distribution Function(ECDF)
50	0.2	0.2
100	0.2	0.4
200	0.1	0.5
500	0.3	0.8
700	0.1	0.9
1000	0.1	1.0

2. Embedding Message : nithyasrinivasan_90@gmail.com

3. Monte Carlo sampling technique/process

Embedding Message	Random Numbers generated from embedding message	Mapped Values	Mapped sequence of packet sizes
110	0.03806228373702422	0.2	50
105	0.03633217993079585	0.2	50
116	0.04013840830449827	0.2	50
104	0.035986159169550176	0.2	50
121	0.04186851211072665	0.2	50
97	0.03356401384083045	0.2	50
115	0.039792387543252594	0.2	50
114	0.03944636678200692	0.2	50
105	0.03633217993079585	0.2	50
110	0.03806228373702422	0.2	50
105	0.03633217993079585	0.2	50

Modulated Packet Size Sequence
{160, 155, 166, 154, 171, 147, 165, 164, 155, 160, 155, 168, 147, 165, 147, 160, 145, 107, 98, 114, 153, 159, 147, 155, 158, 96, 149, 161, 159}

4. Recovery

Once the modulated sequence reaches the client, attacker at the client side recovers the embedded message (nithyasrinivasan_90@gmail.com). Thus the anonymity is compromised.

VI. Conclusion

In this paper, we investigated the packet size based covert channel attacks, which can degrade the service provided by the anonymizer. In order to make the attack more efficient, accurate and hard to detect Monte Carlo sampling technique is used. It carefully maps the ECDF of the packet size to preserve the packet size distribution. In digital forensics privacy is a dual problem, so the proposed technique can be used by law enforcement to track malicious and anonymous web users via anonymizer. It can be used by people who wishes to target marketing on the internet (ex: YouTube).

References

[1] *X. Wang, S. Chen, and S. Jajodia, " Network Flow Watermarking Attack on Low-Latency Anonymous Communication Systems," Proc. IEEE Symp. Security and Privacy (S&P), May 2007.*

[2] *Q.X. Sun, D.R. Simon, Y. Wang, W. Russell, V.N. Padmanabhan, and L.L. Qiu, "Statistical Identification of Encrypted Web Browsing Traffic," Proc. IEEE Symp. Security and Privacy(S&P), May 2002.*

[3] *W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended)," IEEE/ACM Trans. Networking, vol. 2, no. 1, pp. 1-15, Feb. 1994.*

[4] *J. Beran, Statistics for Long-Memory Processes. Chapman & Hall, Oct. 1994.*

[5] M. Liberatore and B.N. Levine, " Inferring the Source of Encrypted HTTP Connections," Proc. ACM Conf. Computer and Comm. Security (CSS), Oct. 2006.
[6] D. Ramsbrock, X. Wang, and X.Jiang, " A First Step Towards Live Botmaster Traceback," Proc. 11th Int'l Symp. Recent Advances in Intrusion Detection (RAID), Sept. 2008
[7] L. Overlier and P. Syverson, "Locating Hidden Servers," Proc. IEEE Security and Privacy Symp. (S&P), May 2006.
[8] K. Baurer, D. McCoy, D. Grunwald, T. Kohno and D. Sicker, "Low- Resource Routing Attacks against Anonymous Systems," Proc. ACM Workshop Privacy Electronic Soc. (WPES), Oct. 2007.

Authors Profie

Swathy.V, PG Scholar