A Novel Packet Size Based Covert Channel Attack against Anonymizer

Zhen Ling*‡, Xinwen Fu† Weijia Jia‡, Wei Yu§ and Dong Xuan¶
*Southeast University, Nanjing 211189, P. R. China. zhenling@seu.edu.cn
†University of Massachusetts Lowell, Lowell, MA 01854, USA. xinwenfu@cs.uml.edu
‡City University of Hong Kong, Kowloon, Hong Kong SAR. wei.jia@cityu.edu.hk
§Towson University, Towson, MD 21252, USA. wyu@towson.edu
¶The Ohio State University, Columbus, OH 43210, USA. xuan@cse.ohio-state.edu

Abstract - Anonymizer is a proprietary anonymous communication system. We discovered its architecture and found that the size of web packets through Anonymizer are very dynamic at the client. Motivated by this fnding, we investigated a novel packet size based covert channel attack, against the anonymity service. In the attack, one attacker manipulates the web packet size between the web server and Anonymizer and embed signal symbols into the target traf c. An accomplice at the user side can snif the traf c and recognize the secret signal. We developed intelligent and robust algorithms to cope with the packet size distortion incurred by Anonymizer and Internet. We developed several techniques to make the attack harder to detect: (i) We pick up right packets of web objects to manipulate in order to preserve the regularity of the TCP packet size dynamics; (ii) We adopt the Monte Carlo sampling technique to preserve the distribution of the web packet size despite manipulation. We have implemented the attack over Anonymizer and conducted extensive analysis and experimental evaluations. It is observed that the attack is highly e f cient and requires only tens of packets to compromise the anonymous web surfng. The experimental results are consistent with our theoretical analysis.

Index Te r s - Anonymizer, Covert Channel, TCP dynamics.

I. INTRODUCTION

Anonymizer is a commercial anonymous communication system. In this paper, we present a novel covert channel attack that may drastically degrade the Anonymizer service. This covert channel exploits the varying size of packets through Anonymizer and is one type of active traffic analysis [1], [2], [3], [4]. Such active attacks can reduce the false positive rate significantly and don't require massive traffic training required in passive traffic analysis attacks [5], [ ].

We will present the first exposure of the Anonymizer architecture, which consists of anonymizing server and client. The server consists of reverse proxy/NAT, SSH server and HTTP proxy, while the client software is a SSH port forwarding configuration tool. We found that the size of HTTP packets through Anonymizer is very dynamic and random at the client. Motivated by this finding, we designed the novel covert channel attack against the Anonymizer service. In this attack, the attacker between the malicious web site and the victim client can embed a secret message into the packet size variation of target traffic. This attacker can be the owner of the malicious web server or one manipulating (repacketizing) the traffic between the web server and Anonymizer server. Without loss of generality, we use the former case as the example in this paper. An accomplice at the client side can sniff the traffic and recognize the secret message. Given the small size of the Anonymizer network, such sniffing is feasible to organizations or people with modest power. To cope with packet size distortion caused by Anonymzier and Internet (e.g., packet padding, packet merging, limited TCP buffer and various MTU), we design intelligent and robust detection algorithms to recover the message. In this way, the anonymity service provided by Anonymizer is compromised.

The attack can be made hard to detect. (i) To attack a HTTP session, we repacketize the web traffic into virtual web objects, and modulate secret messages bits into the size of last packets of these virtual web objects. The last packet of a web object is denoted as the least significant packet for brevity and clarity. The size of a least significant packet is very dynamic in comparison with other packet sizes. Modulation of successive packets to carry message bits will disrupt TCP packet size dynamics (as illustrated in Figure 7), which can be measured by Hurst parameter from R/S plot [7], [8]. This least significant packet based covert channel approach can preserve TCP regularity and self-similarity (as illustrated in Figure 8) while the attacker can control the number of virtual objects to control the number of message bits. (ii) To preserve the size distribution of web packets of virtual web objects, we apply the Monte Carlo sampling technique to carefully sample the empirical cumulative distribution function (ECDF) of the least significant packet size of real web objects. This requires the input of the Monte Carlo method should be random and uniformly distributed. To this end, we first encrypt the message. The generated ciphertext bits are uniformly distributed and encoded into k-ary symbols. A k-ary symbol can then be mapped to a packet size by a Monte Carlo sampling technique.

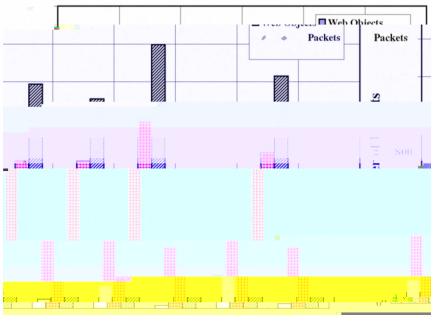
We implemented this novel covert channel attack against Anonymizer and performed extensive theoretical analysis and real-world experiments over Anonymizer. The attack achieves high detection rate with ve r low false positive rate. The experimental results are consistent with our theoretical analysis. To the best of our knowledge, the attack presented in this paper is the first exploiting the Anonymizer architecture and degrading its anonymity via packet size based covert channel. It is simple, efficient, and hard to detect. Compared with related attacks [1], our attack requires just tens of packets to achieve high

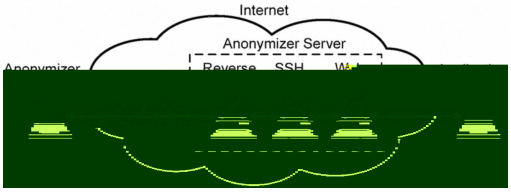
detection rate and low false positive rate.

The remainder of this paper is organized as follows: We explore the components of both Anonymizer server and client, and report our finding that size of web packets in Anonymizer network is very dynamic in Section II. In Section III, we introduce the covert channel based on least significant packets. Extensive experimental results are presented in Section IV. We review related work in Section V and conclude this paper in Section VI.

II. EXPLORATION OF ANONYMIZER

In this section, we first present the Anonymizer architecture discovered by our passive inspection of traffic into and out of





III. PACKET SIZE BASED COVERT CHANNEL

Because of web traffic packet size dynamics at the Anonymizer client shown in Section II-B, the packet size variation can be explored for a covert channel over Anonymizer to compromise the anonymity service. In this section, we first introduce the least significant packets and then present the basic idea and workflow of the attack. We discuss some practical issues and present our solutions at last.

A. Least Significant Packets

Normal HTTP packets (not through Anonymizer) can be roughly categorized into two classes. A Class I packet is defined as the largest packet in normal HTTP traffic, i.e., 1500 bytes in case of Ethernet, including an IP header of 20 bytes and a TCP header of 32 bytes. The size of HTTP content in the TCP payload is 1448 bytes. A Class I packet has a size less than 1500 bytes. Such packets are usually generated by the "tail" of a web object, i.e., the last packet when the web object is downloaded. If the web object size with the HTTP header is w bytes, the size of the web object "tail" is $(w \ mod \ 1448) + 20 + 32$ bytes. In this paper, we denote Class I packets as least significant packets for brevity and clarity.

By analyzing the traffic from 30 well known web sites including CNN, Yahoo and YouTube, we obtain the the empirical cumulative distribution function (ECDF) of the size of raw HTTP packets in Class I as shown in Figure 3. The raw HTTP packet size does not include the IP header and the TCP header. Also, the ACK packet is ignored because of its zero length. The mass probability function (MPF) of the least significant packets is $\{p_1^{'}, p_2^{'}, \ldots, p_m^{'}\}$ and the corresponding packet sizes are $\{pc_1^{'}, pc_2^{'}, \ldots, pc_m^{'}\}$. $p_i^{'}$ is the probability of the packet size $pc_i^{'}$. Therefore, the ECDF of the least significant packet size can be formalized as follows,

$$F_{lsp}(pc_i^{'}) = P(x \leq pc_i^{'}) = p_1^{'} + p_2^{'} + \ldots + p_i^{'}. \quad (1)$$

B. Basic Idea of Covert Channel over Anonymizer

Without loss of generality, we assume that the attacker is between a malicious web site and Anonymizer server and will embed a secret message into the target traffic packet size variation. This attacker can be the other

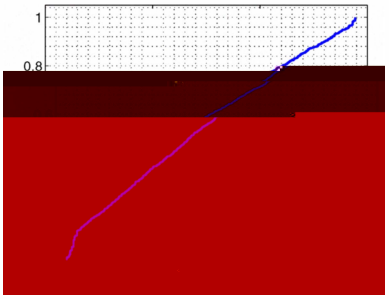


Fig. 3. ECDF of Least Significant Packet Size

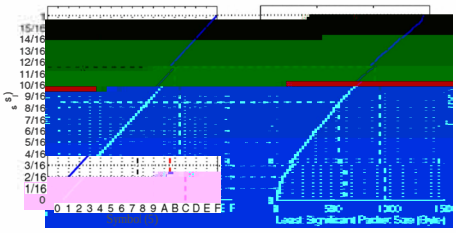


Fig. 4. Mapping between the Symbols and the Least Significant Packet Sizes

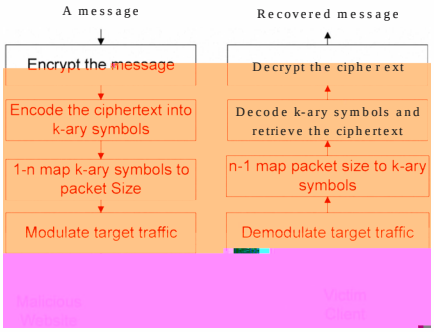
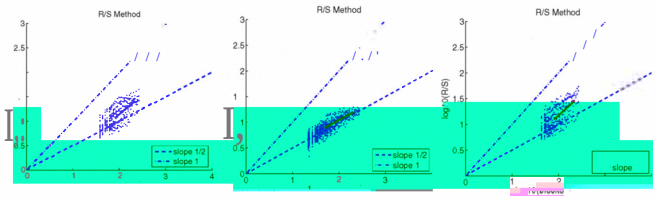


Fig. 5. Workflow of Encrypted Covert Channel Attack based on Least Significant Packets

B. TCP Packet Size Dynamics in Attacks



detection algorithm proposed in [11] to detect 20 random symbols from the clean traffic and calculate the false positive rate.

Figure 10 illustrates the relationship between the detection rate and the d la i al, ell h th h ld S i [11]. Notice th: t $_t$ i : thr: h d t: :t ain iffer nc

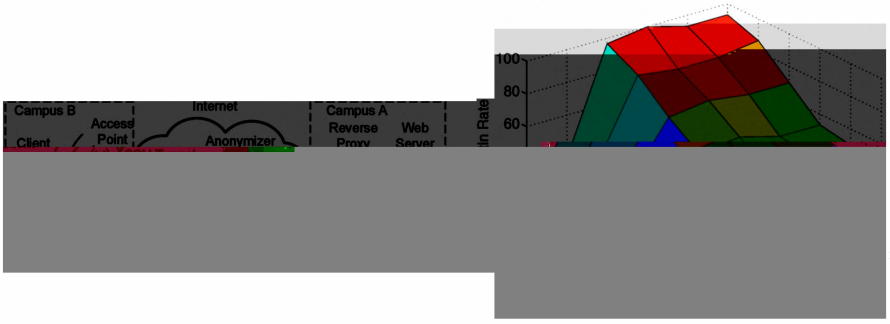


Fig. 9. Experiment Setup

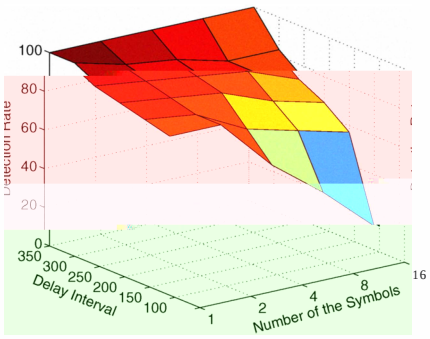


Fig. 10. Detection Rate vs. Delay Interval and St

Fig. 11. Detection Rate vs. Delay Interval and Number of Symbols

embed a signal into the variation of cell counter of the target traffic by varying the counter of cells in the target traffic at the malicious exit onion router and did not explore weakness of Anonymizer.

VI. CONCLUSION

In this paper, we discovered the architecture of Anonymizer and investigated a novel covert channel attack based on least significant packet size variation to drastically degrade the anonymity service provided by Anonymizer. We developed several techniques that make the attack efficient, accurate, and hard to detect. In particular, we applied the Monte Carlo sampling technique to carefully sample the least significant packet size ECDF in order to preserve its distribution. We designed techniques to choose right packets of web objects in order to preserve the regularity of the TCP packet size dynamics measured by the Hurst parameter and R/S plot. All these efforts make the attack practical and more undetectable. We also designed intelligent and robust detection algorithms to recover the distorted symbols caused by Anonymizer and Internet traffic dynamics. Extensive analysis and experiments were conducted to validate the effectiveness and feasibility of the proposed attack. Our data show that the covert channel attack could dramatically and quickly degrade the anonymity service by Anonymizer. Defending against the proposed attack remains a challenging task. We plan to work with Anonymizer developers and investigate the solution in our future work.

ACKNOWLEDGMENT

This work was supported in part by USA NSF grants 0942113, 0958477, 0943479, 09079 4, 054 8 and 091 584, by the Army Research Office (ARO) under grant AMSRD-ACC-R50521-CI, by Research Grants Council of Hong Kong SAR, No. (CityU 114 09), CityU Applied R & D Funding (ARD-(Ctr)) No. 9 81001, ShenZhen-HK Innovation Cycle Grant No. ZYB200907080078A, and NSFC under grants 1070222/F020802, 09031 2, 09031 1, 90912002 and 1003311, National Key Basic Research Program of China under grant 2010CB328104, China National Key Technology R&D Program under grants 2010BAI88B03, China Specialized Research Fund for the Doctoral Program of Higher Education under grant 2008028 0031, China National S&T Major Project under grant 2009ZX03004-004-04, Jiangsu Provincial Natural Science Foundation of China under grant BK2008030, Jiangsu Provincial Key Laboratory of Network and Information Security under grant BM2003201, and by Key Laboratory of Computer Network and Information Integration of Ministry of Education of China under grant 93K-9. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsor agencies.

REFERENCES

[1] X. Wang, S. Chen, and S. Jajodia, "Network fow watermarking attack on low-latency anonymous communication systems;' in P r ceedings of the IEEE Symposium on Securit & Privacy (S&P), May 2007.

[2] W. Yu, X. Fu, S. Graham, D. Xuan, and W Zhao, "Dsss-based fow marking technique for invisible traceback," in P r ceedings of the 2007 IEEE Symposium on Security and Privacy (S&P), May 2007.

[3] A. Houmansadr, N. Kiyavash, and N. Borisov, "Rainbow: A robust and invisible non-blind watermark for network fows," in P r ceedings of the 16th Network and Distributed System Security Symposium (NDSS), February 2009.

[4] Z. Ling, J. Luo, W. Yu, X. Fu, D. Xuan, and W Jia, "A new cell counter based attack against tor," in P r ceedings of 16th ACM Conference on Computer and Communications Securit (CCS), November 2009.

[5] B. N. Levine, M. K. Reiter, C. Wang, and M. Wright, "Timing attacks in low-latency mix-based systems," in Proceedings of Financial C r p-tog r phy (FC), February 2004.

[6] Y Zhu, X. Fu, B. Graham, R. Bettati, and W Zhao, "On fow correlation attacks and countermeasures in mix networks," in P r ceedings of Workshop on Privacy Enhancing Technologies (PE T, May 2004.

[7] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V Wilson, "On the self-similar nature of ethemet tra f c (extended)," lEEEIACM T r nsactions on Networking, vol. 2, February 1994.

[8] J. Beran, Statistics for Long-Memo r Processes. Chapman & Hall, October 1994.

[9] Anonymizer, Inc., http:/ www.anonymizer.co m , 2010.

[10] J. Soto and L. Bassham, "Randomness testing of the advanced encryption standard fnalist candidates," in N ST lR 6 83, National Institute of Standards and Technology, 1999.

[I I] Z. Ling, X. Fu, W. Jia, W. Yu, and D. Xuan, "A novel packet size based covert channel attack against anonymizer," Computer Science Department, Texas A&M Universi[illegible]