Journal of Theoretical and Applied Information Technology
20th June 2014. Vol. 64 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ANONYMOUS WEB BROWSING AGAINST TRAFFIC ANALYSIS ATTACKS BY REUSING THE CACHE MEMORY

JEEVAA KATIRAVAN [1], N.DURAIPANDIAN [2]
[1,2] Velammal Engineering College
E-mail: jeevaakatir@gmail.com

ABSTRACT

Anonymous web browsing is a hot topic with many potential applications for privacy reasons**.** However, there are few such systems which can provide high level anonymity for web browsing. The reason is the current dominant dummy packet padding method for anonymization against traffic analysis attacks. This method inherits huge delay and bandwidth waste, which inhibits its use for web browsing. In this paper, we propose a predicted packet padding strategy for the case of infinite cache size to replace the dummy packet padding method for anonymous web browsing systems. The proposed strategy reduces the delay and bandwidth waste. We therefore present the same predicted packet padding algorithm for the case of finite cache size. We formulated the traffic analysis attack and defense problem, and defined a metric, cost coefficient of anonymization (CCA), to measure the performance of anonymization. We thoroughly analyzed the problem with the characteristics of web browsing and concluded that the proposed strategy is better than the current dummy packet padding strategy in theory.
Keyword:- *Anonymous Web Browsing, Predicted Packet Padding and Finite Cache Memo*

1. INTRODUCTION

Anonymous communication has been explored in the last two decades with the focus on protecting user privacy in the internet age [2] Anonymous communication systems were first introduced in the seminal paper of chaum.[1]Anonymous communication systems can often be classified into two categories: High-latency systems and Low latency systems. High-latency anonymity systems are able to provide strong anonymity, but are typically only applicable for no interactive applications that can tolerate delays of several hours or more, such as the mix networks [2] for e-mail messages. Low-latency systems often provide better performance and are intended for real-time applications, particularly for web browsing. Low latency anonymity system such as the Onion routing, the Tor system, the Crowds system.

Low-Latency anonymity systems often provide better performance and are intended for real-time applications, particularly web browsing. There are few implementation mechanisms for current low-latency anonymous systems, such as onion routing, the Tor system and the Crowds system which can provide high level anonymity for web browsing because of the current dominant dummy packet padding method for anonymization against traffic analysis attacks. This method inherits huge delay and bandwidth waste.

2. EXISTING SYSTEM

The current anonymous web browsing systems are far from perfect and are vulnerable against traffic analysis attacks. The purpose of attacks on anonymous communication is to identify pair wise entities in a system rather than the content of communication.

The time constraint is a critical challenge for anonymous web browsing. The dummy packet padding strategy is widely used by defenders against traffic analysis attacks; however this method introduces huge delays and bandwidth waste.

In order to hide the fingerprint of a protected web page or website, dummy packets are usually injected into the intended traffic to disguise the unique fingerprint of the web page or web site.

2.1 Disadvantages

In this method the strategy can significantly reduce the average delay caused by dummy packet padding.

Journal of Theoretical and Applied Information Technology
20th June 2014. Vol. 64 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Web page data encryption by itself cannot defeat traffic analysis attacks, and packet padding can achieve anonymity for web browsing even though it is not practical because there are huge delays and extreme bandwidth cost caused by the dummy packet padding mechanism.

3. PROPOSED SYSTEM

In proposed system, the predicted packet padding method to replace the current dummy packet padding one for implementation of anonymous web browsing systems. This new strategy can significantly reduce the average delay caused by dummy packet padding, and therefore, makes anonymous web browsing applicable in practice.

Applying the proposed method for a given web site, we can prevent an eavesdropper from identifying the pages that a web browser has downloaded through traffic analysis. The proposed method can also be extended to protect other objects, such as web sites.

The Proposed algorithm also extended in dynamic web pages and it's applied to finite cache size for anonymous web browsing against traffic analysis attack. We explore the problem of anonymization efficiency for the first time to the best of our knowledge.

3.1 Architecture For The Proposed System

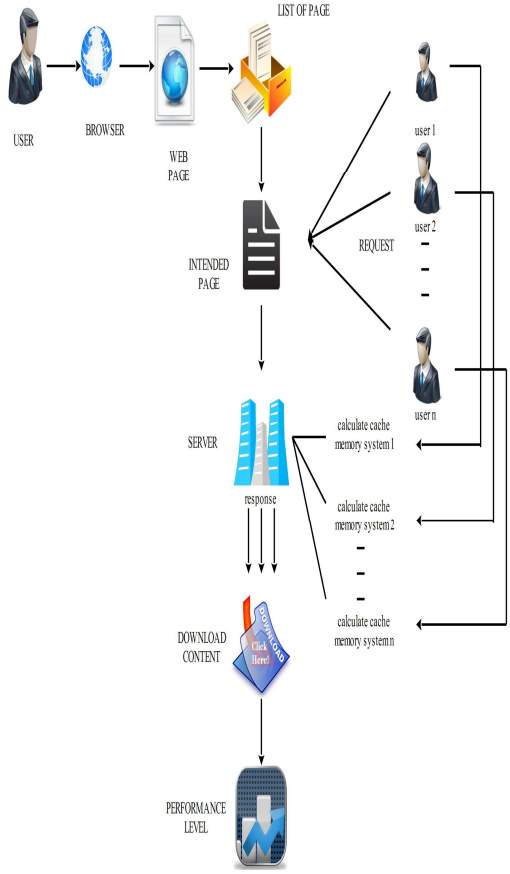

Figure 1

3.2 Dynamic web pages

The objects will be downloaded to the client one after the other. This information composes fingerprints of web pages. Moreover, an observer can clearly see packet heads, which include critical information of the traffic, such as source IP addresses and destination IP addresses. Dynamic web pages are web sites that are generated at the time of access by a user or change as a result of interaction with the user.

If a web page takes a long time for the visitor to read, the analyzer may think they are a new visitor when they finally finish the page and request a new one. Some people will download many pages at once (especially those who like tabbed browsing) and read them over a period of time. This can mean that they don't generate any new hits for so long that the analyzer assumes they have left. Then when they do access a new page they are counted as a new visit. If they don't

Journal of Theoretical and Applied Information Technology
20th June 2014. Vol. 64 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

request a new page, the length of their visit will be underestimated. Some ISPs and anonymizer services give the appearance that lots of people have the same IP address. Log analysers will count all these people as the same visitor. If a website's audience is mostly from the same area or uses the same ISP, there is a greater chance of confusing visitors.

3.3 Finite cache memory

More and more people use mobile devices, such as smart phones, to access the web and these devices have much smaller cache memories. To implement our future work in mobile device by reusing the cache memory.

4. SYSTEM MODELLING AND ANALYSIS

A typical anonymous web browsing system with data encryption (at the Internet channels) and packet padding (at the server side) is shown in Figure 2.

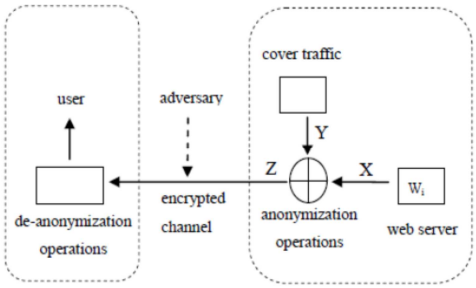

Figure 2

As a client, Alice sends a HTTP request to a web server wi via an encrypted channel. The web server also employes an encrypted channel to return the intended traffic X = {x1,x2…….xk}, where x_i(1≤i ≤k) represents the number of packets of web object i. Let ||X|| = $\sum_{i=1}^{k} x_i$ denote the total number of packets of the intended traffic X. We then extracted the fingerprint of this session as p = {p1,p2…..pk} where $p_i = x_i \cdot (\|X\|)^{-1}$; 1≤ i≤ k , and $\sum_{i=1}^{k} p_i = 1$.In order to make it anonymous to adversaries, we created cover traffic Y = {y1; y2; : : : ; yk} at the server side. Let y_i(1 ≤ i ≤k) denote the number of packets assigned to cover xi, and let ||Y|| = $\sum_{i=1}^{k} y_i$. Similar to the intended traffic X, the fingerprint of Y is q = {q1; q2; : : : ; qk}. If Z = {z1, z2…. zk} represents the mixture of the intended traffic X and the cover traffic Y , then the fingerprint of Z is r = {r1,r2….. rk}, and the total number of the mixed traffic is ||Z||. The adversary's observation T' is the mixture of Z and other background traffic on the network.

In previous works, dummy packets are employed to work as the cover traffic Y ; Once Z arrives at the client side,the dummy packet Y will be discarded; another solution is using transparent images as the cover traffic. However, in the proposed strategy the predicted web data is used as the cover traffic Y, the client decompose the received traffic, and the
intended traffic X goes to the web browser, and the prefetched data Y is stored in the cache of the local computer, and Y may be used by the following requests. In this case, the client will fetch the expected web data from the cache, rather than download it again from the server. From a long term viewpoint, the bandwidth is not wasted and the average extra delay is limited in the proposed scheme.

Cost Coefficient of Anonymity. Let function C(S) represent the cost function for a given network traffic S. For a given intended network traffic X, we inject a cover traffic Y to achieve the goal of anonymity, then the cost coefficient of anonymity is defined as

$$\beta=(C(Y \mid X) + C(X)) \div C(X)$$

This metric will be used to indicate the cost efficiency for perfect anonymity operations

5. PERFORMANCE ANALYSIS

In order to confirm the performance advantage of proposed strategy we took 24 continuous hours from the popular website as a dataset for the experiments by extracting the finger print(number of TCP packets for each web page object) of every hour. We first investigated the cost coefficient of perfect anonymity for the proposed strategy with different missing rate (namely, different prefetching accuracy). The results are shown in Figure 3. The relationship between the cost coefficients of perfect anonymity against the length of a session was also analyzed. The cost coefficient of perfect

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

anonymity of the dummy packet padding strategy was an increase function against the length of each session as shown in Figure 4.

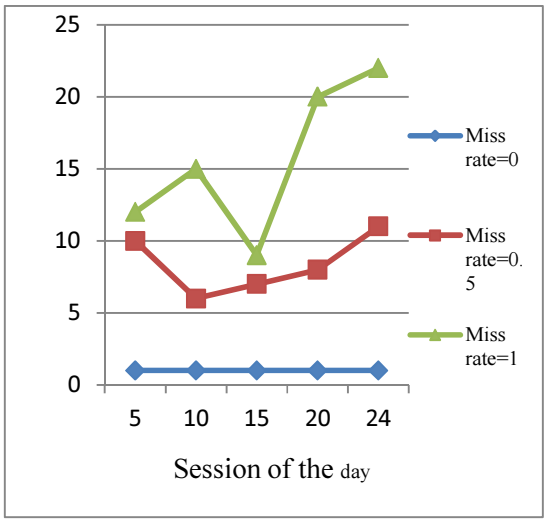

Figure 3

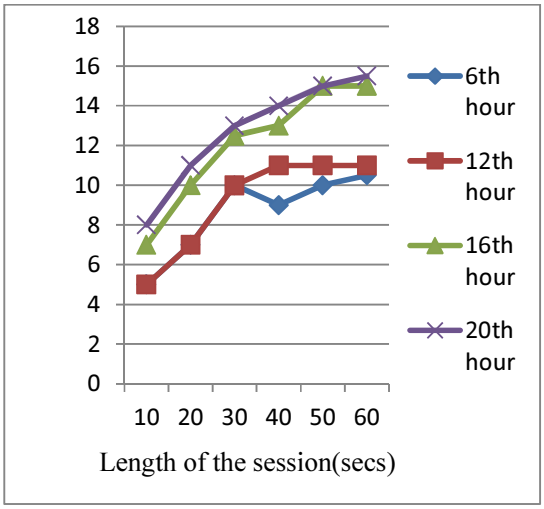

Figure 4

Every change point indicated there were a bigger objects in terms of packet number in the past, namely, the change point depended on the distribution of larger objects. In other words, the longer the session length is, the higher cost for the dummy packet padding method in order to achieve perfect anonymity. However, this is not a problem for the proposed strategy.

6. CONCLUSION

The project focused on reducing the delay and bandwidth waste of anonymous web browsing systems. In this project predicted packet padding algorithm used in the case of client computers with infinite cache size for web browsing. A simple mathematical model for the packet padding mechanism was established, followed by a thorough analysis and comparison between the proposed strategy and the traditional dummy packet padding method. In future, we extend our work to dynamic web site and also in finite cache memory. In finite space the predicted packet padding achieved by reusing the cache memory.

REFERENCES

[1] M. Edman and B. Yener, "On anonymity in an electronic society: A survey of anonymous communication systems," ACM Computing Survey, Vol. 42, no. 1, 2009.

[2] D. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," Commun. ACM, Vol. 24, no. 2, pp. 84–88, 1981.

[3] D. M. Goldschlag, M. G. Reed, and P. F. Syverson, "Hiding routing information," in Information Hiding, 1996, pp. 137–150.

[4] M. G. Reed, P. F. Syverson, and D. M. Goldschlag, "Anonymous connections and onion routing," IEEE J. Select. Areas Commun, Vol. 16, no. 2, pp. 482–494, Feb. 1998.

[5] [Online]. Available: http://www.torproject.org

[6] R. Dingledine, N. Mathewson, and P. F. Syverson, "Tor: The second generation onion router," in Proc. USENIX Security Symp., pp.303–320.,2004.

[7] M. K. Reiter and A. D. Rubin, "Crowds: Anonymity for web transactions," ACM Trans. Inform. Syst. Security, Vol. 1, no. 1, pp. 66–92, 1998.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[8] C. Diaz, S. Seys, J. Claessens, and B. Preneel, R. Dingledine and P. Syverson, Eds., "Towards measuring anonymity," in Proc. Privacy Enhancing Technologies Workshop (PET 2002), Apr. 2002,Springer-Verlag, LNCS 2482.

[9] A. Serjantov and G. Danezis, R. Dingledine and P. Syverson, Eds., "Towards an information theoretic metric for anonymity," in Proc. Privacy Enhancing Technologies Workshop (PET 2002), LNCS 2482, Apr.2002, Springer-Verlag.

[10] Q. Sun, D. R. Simon, Y.-M. Wang, W. Russell, V. N. Padmanabhan, and L. Qiu, "Statistical identification of encrypted web browsing traffic," in Proc. IEEE Symp. Security and Privacy, 2002.