Computer Networks 74 (2014) 13–33

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

PACOM: Parasitic anonymous communication in the BitTorrent network

Jianming Lv [a,*], Tieying Zhang [b], Zhenhua Li [c], Xueqi Cheng [b]

[a] Department of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China
[b] Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
[c] School of Software and TNLIST, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 26 May 2013
Received in revised form 3 July 2014
Accepted 24 August 2014
Available online 16 September 2014

Keywords:
Privacy
Anonymous communication
Peer-to-Peer
Parasitic
Traffic analysis attack

ABSTRACT

Existing anonymous communication systems mask the identities of users by adopting intermediary nodes to transform message flows. However, some recently presented traffic analysis algorithms are still able to undermine the anonymity of these systems. The traditional flow transformation strategies fail to completely eliminate the traffic correlation between adjacent communication links to prevent such attacks. To address this problem, we propose a novel *parasitic anonymous communication system*, named PACOM. Each PACOM client is parasitic in the BitTorrent network which is the most popular Peer-to-Peer file sharing network, and conceals the communication path in the request driven traffic compatible with the BitTorrent protocol. The traffic patterns of adjacent communication links can be proved to be statistically independent, which effectively resists the traffic analysis attacks. Meanwhile, the "*effective anonymity set size*" of the system can be extended enormously by mixing the PACOM clients with other millions of BitTorrent clients in the Internet. To validate the PACOM solution, we analyse the anonymity of PACOM theoretically and conduct comprehensive simulations and emulations to test the scalability and effectiveness of PACOM against various attacks.

1. Introduction

As people rely on the Internet more and more in their daily life, users' anonymity has become a critical issue to protect personal privacy in a lot of Internet applications such as web browsing, file sharing, e-commerce, and electronic voting.

Existing anonymous communication systems [1–11] are generally designed to hide the relationship between the information sender and receiver from adversaries. These systems adopt intermediary nodes (often called "*mixes*") to encrypt and relay the communication messages hop by hop, so as to conceal the destinations of the messages. Besides, in order to prevent the adversaries from identifying the communication path via *traffic analysis*, some additional strategies are used to transform the network flows, such as traffic padding [1,2], cover traffic adding [4,12,13], packet dropping [14], flow mixing [15–17], batching [16,18] and rescheduling [19,20]. These transformation strategies make the network flows indistinguishable from each other.

Unfortunately, some recently presented *traffic analysis attacks* against the above mentioned systems are still applicable and can be easily conducted [21–24]. These attacks utilize the common characteristic of existing systems: the traffic patterns of adjacent links in the anonymous communication path are statistically corre-

* Corresponding author. Tel.: +86 20 39380281.
E-mail addresses: jmlv@scut.edu.cn (J. Lv), zhangtieying@ict.ac.cn (T. Zhang), lizhenhua1983@tsinghua.edu.cn (Z. Li), cxq@ict.ac.cn (X. Cheng).

http://dx.doi.org/10.1016/j.comnet.2014.08.015
1389-1286/

lated. For example, on a path $A \rightarrow B \rightarrow C$, the anonymous communication data is transferred from A to C through B. The sending time of the packets in $B \rightarrow C$ depends on the arrival time and volume of the packets in $A \rightarrow B$. Thus, the traffic patterns in $B \rightarrow C$ and $A \rightarrow B$ have strong timing correlation with each other. The traditional flow transformation strategies fail to completely eliminate such correlation information. In particular, the traffic analysis attacks [21–24] can identify the communication path effectively by exploiting the traffic timing pattern and building correlation between different links.

In this paper we present a novel *parasitic anonymous communication system*, named PACOM, which is immune to these traffic analysis attacks. Each PACOM client is parasitic in the BitTorrent network, the most popular Peer-to-Peer file sharing network containing more than 150 million active users [25]. The PACOM clients pretend to share and download files in the BitTorrent network and embed the anonymous communication data into the delivered file blocks. The communication path is concealed by the request driven traffic compatible with the BitTorrent protocol, which makes the traffic patterns of adjacent links statistically independent of each other. Compared with the state-of-the-art anonymous communication systems, PACOM achieves the following advantages:

(1) PACOM is immune to the traffic analysis attacks. No statistical correlation in the time domain can be built between different links in any anonymous communication path. The success rate of the traffic analysis attacks against PACOM is close to that of random guess without any prior knowledge.
(2) PACOM is designed for two different use cases: the *Private Use Case* for secret and hidden communication among a few special users, and the *Public Use Case* for a large number of online people to transfer messages anonymously like Tor [2]. In the *Private Use Case*, PACOM equipped with steganography techniques provides strong anonymity by mixing PACOM clients with other millions of BitTorrent clients in the Internet, and prevents adversaries from sensing when and where the anonymous communication happens. In this case, the *effective anonymity set size*[1] increases logarithmically over the total number of the online BitTorrent clients. In the *Public Use Case*, no steganography is taken and PACOM is a public and open system, in which the *effective anonymity set size* is related to number of online PACOM clients. Although the number of BitTorrent clients is not helpful to increase the anonymity of PACOM in this case, the file block request driven traffic pattern following the BitTorrent protocol still makes the system effective against traffic analysis attacks and efficient in anonymous communication.
(3) Aided by the efficient file blocks transferring mechanism of the BitTorrent protocol, PACOM is efficient to transfer communication messages. Different from the traditional inefficient batching methods adopted in mix networks to resist traffic analysis, each PACOM node transfers file blocks containing communication messages driven by file block requests following the BitTorrent protocol, which is designed for efficient P2P file sharing. The bandwidth of PACOM communication in the *Private Use Case* is about 21 kB/s on average, while the bandwidth in the *Public Use Case* is about 314 kB/s on average. The end-to-end latency in both cases is close to 2.5 s. Thus PACOM in the *Private Use Case* is competent to transfer messages or files of small size. Meanwhile, PACOM in the *Public Use Case* is efficient to support various kinds of file sharing, chatting, or accessing some web based services such as cloud storage and LBS.
(4) PACOM is decentralized and highly scalable. No centralized directory servers are needed. The PACOM clients collaborate with each other to perform the network bootstrapping and maintenance operations, and the per-client computation and traffic overhead is moderate.
(5) The cover traffic in PACOM is self-adaptive and localized, thus the load of the clients is reduced and the bandwidth utilization is improved. Comprehensive experiments show that the cover traffic occupies about 50% of the bandwidth of each PACOM client, when each client initiates one communication path. The ratio drops to about 10% when each one initiates four paths on average. Moreover, the cover traffic is only produced among the PACOM clients and has little side effect to the BitTorrent network.

The remainder of this paper is organized as follows. In Section 2, we introduce some preliminaries. Then, we present the design of PACOM in Section 3 and theoretically analyse the anonymity of PACOM in Section 4. Some discussion on the deployment of PACOM is presented in Section 5. The performance of PACOM is evaluated via comprehensive simulation and emulation experiments in Section 6. In Section 7 we survey the related work. Finally, we conclude this paper in Section 8.

2. Preliminaries

2.1. Mix networks and steganography

In the past decade, a lot of anonymous communication systems [1–11] are proposed to hide the relationship between information sender and receiver from malicious adversaries. As illustrated in Fig. 1(a), internal nodes (called mixes) are adopted in these systems to mix and relay encrypted messages. To further confuse adversaries, these mixes adopt some strategies to transform all incoming flows, such as batching [15,16,18], traffic padding [1,2], cover traffic adding [4,12,13], packet dropping [14], flow mixing [15–17], and rescheduling [19,20].

Specifically, the batching has become one basic and widely used technique since the mixes were firstly

[1] An information theoretic metric of anonymity defined by Serjantov et al. [26] that quantitatively measures the anonymity level of a communication system.