Anonymous Peer-to-Peer Systems

Wenbing Zhao
Cleveland State University, USA

INTRODUCTION

A peer-to-peer (P2P) system refers to a distributed system in which the role played by each member is roughly equivalent, that is, a member both consumes and provides services to other members. Therefore, a member in such a system is often referred to as a *peer*. The primary design goal of P2P systems is to facilitate sharing of resources, including files, processing power, and network bandwidth.

A Brief History of P2P Systems

The first major P2P system, Napster (Taylor, 2004), was introduced in 1999. Since then, P2P systems have evolved very rapidly. The first generation of P2P systems, represented by Napster, used a centralized index server to provide peer and resource discovery. This design makes the systems vulnerable to attacks on the central server. The second-generation P2P systems, such as Gnutella, avoided such problems by adopting a fully decentralized design (Taylor, 2004). However, these systems do not provide anonymity to their users, in that a resource owner is exposed to its requester, and vice versa. A user's activities are also easily observable by the neighboring nodes. The privacy concerns triggered the development of the third generation of P2P systems (Rossi, 2006; Rohrer, 2006). These systems followed a variety of strategies to achieve anonymity, which will be discussed in detail in this article. Even though most of such systems are in the early development phase, we believe they will soon take over the second generation of P2P systems.

Why Should We Care About Anonymous P2P Systems?

First of all, P2P systems have become an essential part of the Internet, as evidenced by the fact that P2P traffic constitutes more than 40% of the total TCP traffic and it is still growing (Saroiu, Gummadi, Dunn, Gribble, & Levy, 2002). P2P systems have revolutionized the Internet by enabling direct communication and resource sharing among the end users operating inexpensive PCs. The security and ethical use of such systems are crucial to the health of the Internet and the commerce of many sectors, especially the music and movie industries.

Secondly, by allowing anonymous information flow, anonymous P2P systems protect users' privacy and freedom of speech. It allows one to voice unpopular opinions, to report misconduct of one's superiors, or simply to discuss freely on controversial issues, without being threatened. As an end user, to select the best P2P system that protects users' rights, one needs to know how anonymity is achieved in such systems.

Thirdly, from the government and copyright owners' perspective, we need to understand the anonymity techniques to devise an effective surveillance method for illegal activities in P2P systems. Indeed, existing P2P systems are commonly used for trading copyrighted materials. It is also a legitimate concern that such systems might aid terrorism and many other illegal activities. Virtually every practical anonymity technique has its limitations. Such limitations offer the possibility for surveillance.

BACKGROUND

Put simply, *anonymity* means the state of not being identified. To avoid confusion, anonymity must be said with respect to some observing entity. Whether a subject can remain anonymous depends on how much effort the observing entity spends to uncover the identity of the subject. Unfortunately, such contextual information is often omitted or not clearly stated (Chothia & Chatzikokolakis, 2005).

In P2P systems, the main subjects of concern related to anonymity are the peer that initiated a request (i.e., *requester,* also termed as sender or initiator) and the peer that responded to the request (i.e., *responder,* also referred to as receiver or recipient). For a requester (responder), the observing entity could be the responder

(requestor), or some other entity within or external to the P2P system.

If the observing entity has the intention to remove the anonymity of the subject, we call such an entity an attacker, or an *adversary*. An adversary is not restricted to a single node. It can consist of multiple nodes that may *observe* part or all communications coming in and out of a subject. A more powerful adversary may compromise or introduce malicious nodes into the system and consequently be able to *control* part or all network traffic.

Consequently, we can categorize adversaries into *local* and *global* adversaries, based on the scale of network traffic they can observe or control, and also into *passive* and *active* adversaries, based on if they actively compromise nodes in the system and/or generate network traffic to help uncover the identity of the subject.

In P2P systems, there exist several forms of anonymity:

- *Requester anonymity* means that the system can hide the requester's identity from the responder and some adversaries.
- *Responder anonymity* refers to the fact that the responder's identity is hidden from the requester and some adversaries.
- If the system ensures that neither the requester nor the responder knows with whom it is communicating, we have achieved a form of *mutual anonymity*.
- It is also possible that the relationship between a requester and a responder needs to be hidden from some observers, although the requester and the responder themselves might be seen to have made requests and responses. We call this type of anonymity *unlinkability of requester and responder*.

We should note that with respect to a determined local or global adversary, *absolute anonymity* is often not achievable (unless a subject never communicates with others). Therefore, in practice, anonymity is described in terms of probabilistic metrics and relative to other peers in the system (Reiter & Rubin, 1998). A more appropriate term perhaps is pseudonymity rather than anonymity in the context of P2P systems.

ANONYMOUS COMMUNICATION IN P2P SYSTEMS

There have been a large number of proposals for anonymous communication in P2P systems. These approaches can be roughly classified into three categories: (1) using a pre-determined indirect path from a requester to a responder with layered encryption; (2) mixing a new request (or response) with relayed traffic; and (3) using transient pseudo identities and multi-path routing. All three approaches depend, to different degrees, on the strength of cryptography. In general, the more honest the peers that join the communication, the higher the degree of anonymity that can be achieved. We do not include multicast-based approaches (Chaum, 1988; Dolev & Ostrovsky, 2000) because they are not practical in large-scale P2P systems.

Indirect Path with Layered Encryption

Mixes (Chaum, 1981), *onion routing* (Reed, Syverson, & Goldschlag, 1998), and many of their derivatives belong to this category. The strategy is to set up an indirect path between the requester and the responder, so that: (1) the relationship between the requester and the responder is hidden, and (2) the requester is made anonymous to the responder.

Mixes are proposed by Chaum (1981). The idea is to use one or more computers to serve as intermediaries to relay a requester's message to the destination. To defend against traffic analysis (Back, Moller, & Stiglic, 2001) that might link an input with its corresponding output, every computer batches and reorders its inputs. Hence, this method is referred to as mixes and so are the computers used for that purpose.

As shown in Figure 1, each message is signed by the requester and recursively encrypted using the public keys of the responder and the mixes on the path. When this message reaches the first mix P1, P1 decrypts the outermost layer of encryption using its private key, retrieves the next node on the path (P2) as chosen by the requester, and passes the remaining ciphertext to P2 after batching and reordering. This process continues until the message is relayed to the final destination B. The reply message is passed along the same path in reverse direction.

Because the path information, including the responder address, is directly encrypted in the message,