Analysis on anonymity of P2P anonymous communication system

Chenfei Xu[1, a], Hua Zhang[1, b], Qiaoyan Wen[1]

[1]State Key Laboratory of networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, 100876, China.

[a]xcfei0628@163.com, [b]zhanghua_288@bupt.edu.cn

Keywords: Peer-to-Peer, Cluster Node, anonymous communication, limited strategy.

Abstract. Most of existing anonymous communication systems adopt “mixes” nodes to encrypt and relay the communication messages. Extensive works have been done on this field to find methods that can effectively resist the attacks. However, the adversaries can perform the attacks through the relationship between multiple nodes communication traffic to deanonymize the P2P anonymous systems. To address this problem, this paper presents a novel concept: “Cluster Node” in P2P system, which uses a method of “Cluster” statement and adopts a limited strategy of node selection to avoid using multiple nodes from the same cluster in the same circuit. We show the feasibility and accuracy of our approach by theoretical analysis and validate experiment.

Introduction

With the development of the Internet technology, the protection of users’ personal privacy and anonymity has become a critical issue in many Internet applications [1], and many Peer-to-Peer (P2P) based anonymous communication systems have been proposed [2]. However, some recently strategies still cannot effectively resist the attacks. The adversaries can perform the traffic related attacks through the relationship between multiple nodes communication traffic [3-5]. This paper proposes a novel Cluster Node in P2P system, and uses a limited strategy of node selection to prevent the attacks. All Cluster nodes in P2P network form many group of clusters through statement for each other. Besides, a limited strategy of node selection has been adopted to avoid using multiple nodes from the same cluster in the same circuit. We also make extensive theoretical analysis on the entropy of the anonymity of the P2P system and do evaluations experiment.

The Cluster anonymous communication system

The Cluster Node. The existing anonymous communication systems cannot effectively prevent the attacks, which can destroy the anonymity of the anonymous communication system through the relationship between multiple nodes communication traffic. Therefore, in order to solve the above problems effectively, this section presents a novel concept: “Cluster Node”. All Cluster nodes in P2P network form many group of clusters through statement for each other, and each cluster is named “Cluster”. Fig. 1 shows Cluster node division in P2P network.

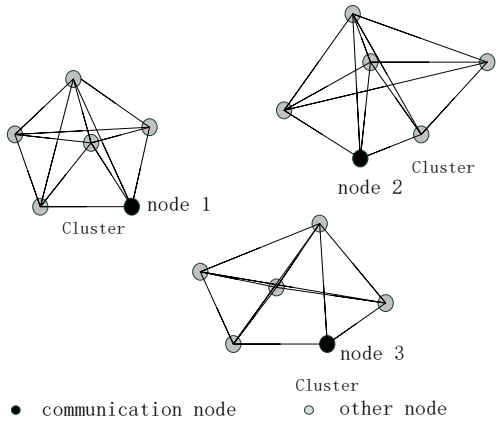


Fig. 1 Cluster node division in P2P network

The method of "Cluster" statement. The method that each cluster node can perform "Cluster" statement shows as followed: Let N denote the set of the node number in P2P network, $N = \{ n_1, n_2, n_3 \ldots n_i \}$, and for any $n_i \in N$, let $F(n_i)$ denote the Cluster statement of node n, $F(n_i) = \{ n_i \in N \mid N$ lists node n_i into its Cluster statement$\}$. If node n_i has no Cluster statement, and then $F(n_i) = \Phi$. Based on this definition, the relationship between node n_i and node n_j are as followed:

1) $n_i \leftrightarrow n_j$: If node n_i and node n_j are provided by the same volunteer, agency or operator, the relationship between node n_i and node n_j is direct Cluster statement, that is $n_i \in F(n_j)$ and $n_j \in F(n_i)$.

2) $n_i \sim n_j$: If node n_i and node n_j are provided by the same or adjacent area of the volunteers, agencies or operators, the relationship between node n_i and node n_j is indirect Cluster statement, that is $n_i \notin F(n_j)$ and $n_j \notin F(n_i)$.

Let S denote the set of the Cluster number in P2P network, $S = \{ s_1, s_2, s_3 \ldots s_k \}$, let s_k denote each Cluster of the node number, so if S meets the following conditions, it is regarded as a node Cluster:

1) $s_k \geq 2$, that is the node number in the cluster is not less than 2.

2) For any node n_i and node $n_j \in S$, both of them meet $n_i \leftrightarrow n_j$ or $n_i \sim n_j$, that is any two nodes within node Cluster meet the relationship of direct or indirect Cluster statement.

3) For any node $n_i \notin S$ and any node $n_j \in S$, $n_i \sim n_j$ is not established.

The limited strategy of node selection

In this paper, we adopt the entry node selection strategy, the middle node selection strategy and the exit node selection strategy to select the nodes as node 1, node 2 and node 3 [6][7]. Based on the node selection strategies, this paper proposes a limited strategy of node selection as followed:

1) When a user want to initiate anonymous communication, first it will connect its proxy node (proxy server), and then the proxy node will construct an anonymous tunnel with other nodes.

2) The proxy node uses directory server to collect and distribute the relay node information of the whole P2P network, and selects three nodes as node 1,node 2 and node 3 according to the released factors of each node' bandwidth, online time and set export access strategy.

3) After the proxy node constructing an anonymous tunnel with other nodes and selecting node 1according to the entry node selection strategy, node 1 needs to determine whether node 2 is in the same cluster according to the Cluster statement. If node 2 is in the same cluster, and then node 1 re-select node 2 until both of them are not in the same cluster; If not, node 2 continues to select node 3 according to the same method. Until all nodes of the anonymous communication path are not from the same cluster, the anonymous channel construction complete and the anonymous communication start.

4) In the anonymous communication process, so as to ensure the better safety of anonymous communication and resist the traffic analysis attacks and the malicious node attacks, the proxy node had better replace each node (node 1, node 2, node 3) to transmit data a period of time.

Anonymity analysis

In this section, we theoretically analyze the anonymity of P2P system. Let X denotes the anonymous communication system and $H(X)$ is its entropy value that denotes the anonymity of P2P system and is calculated as the entropy of the anonymity probability distribution:

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i = -\sum_{j=1}^{m} \sum_{k=1}^{s_k} p_k \log_2 p_k \quad (1)$$

Here p_i denotes the probability of identifying the peer node as the real sender in the system,

s_k denotes each Cluster of the node number, n denotes the node number and m denotes the Cluster number in P2P network. A larger *H(X)* means higher anonymity [8].

The probability analysis. *1*）The equal probability condition: Suppose there are *N* peer nodes in P2P system. The communication behavior of all the *N* peers looks alike under the ideal circumstance, so each node has a possibility of to be identified as the sender. The entropy can be calculated as followed:

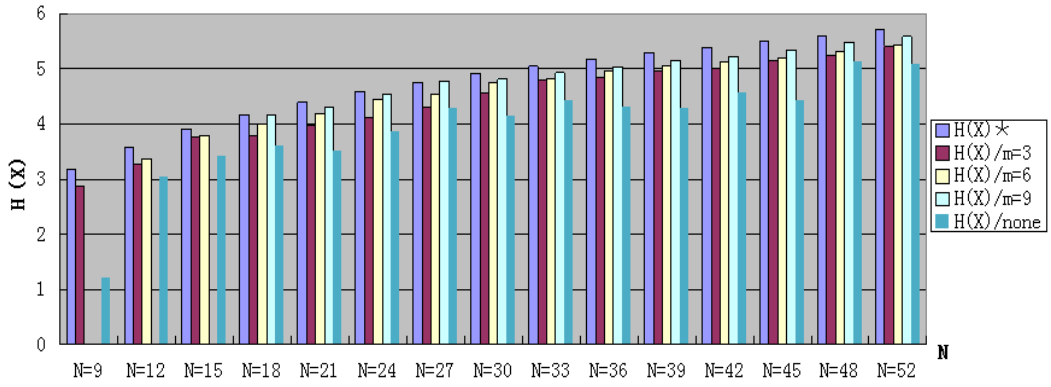
$$H(X)^* = -\sum_{i=1}^{n} p_i \log_2 p_i = \log_2 n \tag{2}$$

However, the attacks are unavoidable in a P2P system. So, a probability equation has been presented to better analyze the anonymity of P2P system.

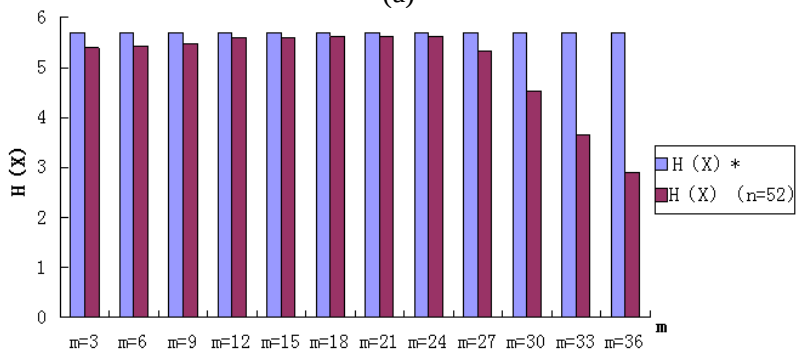
The proposed probability condition: Suppose all peer nodes are divided into the cluster number of m in a P2P anonymous communication system. Let s_k denotes each Cluster of the node number, so m = n/s_k. As mentioned in [9], the nodes with high bandwidth are more vulnerable to be attacked by the adversary, so p_k can be calculated as followed:

$$p_k = \frac{1}{m} \frac{B(n_k)}{\sum_{k=1}^{s_k} B(n_k)} \tag{3}$$

Here s_k denotes the node number of each Cluster, m denotes the Cluster number in P2P network, $B(n_k)$ denotes the peer node's bandwidth in the same cluster, and $\sum B(n_k)$ denotes the sum of all peer nodes' bandwidth in the same cluster.



(a)



(b)

Fig. 2 The entropy values of the experiments ((a) two different entropy values in the same number of cluster, (b) the entropy values in different number of cluster).

Evaluations. This subsection will analyze the anonymity of P2P system. The analysis is based on 6200 Tor nodes information and uses the above equation (1), (2) and (3) to calculate pi and *H(X)*.

This paper firstly analyzes the different entropy values in the equal probability and the unequal probability situation when different number of nodes is divided into the same number of cluster according to the above strategies and methods. Figure 2(a) shows the experimental data. From the

chart, it is known that when different number of nodes is divided into the same cluster number, the entropy value increases with the number of nodes increases and is nearly close to the ideal value in the equal probability that denotes the better anonymity. Besides, the entropy value is greater than without the cluster's that has proved the advantage of the cluster node.

Meantime, this paper also analyzes the entropy values when the same node number is divided into different cluster number. Figure 2(b) shows the experimental data. From the chart, it is known that when the same number of nodes is divided into different cluster number of, the more number of cluster, that is, the less node number of each cluster, the greater entropy value. But when the number of nodes increases to a certain extent, that is, the node number of each cluster declining, the entropy values will begin to decline.

Therefore, it is concluded that the system should be divided into a number of clusters, and it is ensured that the node number of each cluster is not too much, or the anonymity will be reduced.

Conclusion and future work

This paper proposes a novel concept: "Cluster Node" and adopts a limited strategy of node selection in P2P system. We make extensive theoretical analysis on the anonymity of P2P system and do evaluations experiment to validate the feasibility and advantage of our method and strategy. We will further our research in the future from following aspects: research deep into the division rule of *S* (the cluster division) and achieve a better division rule in P2P network.

Acknowledgments

This work is supported by NSFC (Grant Nos. 61300181, 61202434), the Fundamental Research Funds for the Central Universities (Grant No. 2015RC23).

References

[1] J. Lv, T. Zhang, Z. Li, et al. PACOM: Parasitic anonymous communication in the BitTorrent network [J]. Computer Networks, 2014, 74: 13-33.

[2] R. Dingledine, N. Mathewson, P. Syverson. Tor: The second-generation onion router [R]. Naval Research Lab Washington DC, 2004.

[3] P. N. Hoang, D. Pishva. Anonymous communication and its importance in social networking [C] //Advanced Communication Technology (ICACT), 2014 16th International Conference on. IEEE, 2014: 34-39.

[4] J. Lv, C. Zhu, S. Tang, et al. Deepflow: Hiding anonymous communication traffic in P2P streaming networks [J]. Wuhan University Journal of Natural Sciences, 2014, 19(5): 417-425.

[5] G. He, M. Yang, X. Gu, et al. A novel active website fingerprinting attacks against Tor anonymous system [C] //Computer Supported Cooperative Work in Design (CSCWD), Proceedings of the 2014 IEEE 18th International Conference on. IEEE, 2014: 112-117.

[6] X. Wang, J. Shi, B. Fang, et al. An empirical analysis of family in the Tor network [C] //Communications (ICC), 2013 IEEE International Conference on. IEEE, 2013: 1995-2000.

[7] Z. Ling, J. Luo, W. Yu, et al. Extensive analysis and large-scale empirical evaluation of tor bridge discovery [C] //INFOCOM, 2012 Proceedings IEEE. IEEE, 2012: 2381-2389.

[8] J. Zhang, H. Duan, W. Liu, et al. Anonymity analysis of P2P anonymous communication systems [J]. Computer Communications, 2011, 34(3): 358-366.

[9] C. Li, Y. Xue, Y. Dong, et al. Super nodes in Tor: existence and security implication [C] //Proceedings of the 27th Annual Computer Security Applications Conference. ACM, 2011: 217-226.