Herd: A Scalable, Traffic Analysis Resistant Anonymity Network for VoIP Systems

Reviewed by Xinghao Li and Yibo Pi

Citation:

Le Blond, Stevens, et al. "Herd: A Scalable, Traffic Analysis Resistant Anonymity Network for VoIP Systems." *ACM SIGCOMM Computer Communication Review*. Vol. 45. No. 4. ACM, 2015.

Summary

In this paper, the authors introduce Herd, which is an anonymity network designed for VoIP systems to defend eavesdropping and analyses for calling activities. The motivation comes from the severe threat of information collection from authorities or adversaries. Although modern cryptographic algorithms are sufficient to secure the contents of the conversation, the calling information such as the caller/callee, time, duration and location is still exposed to the vision of attackers. Thus, in additional to adopting modern advanced encryption technology from Tor with layered encryption, this paper introduces new mechanisms for Herd to ensure zone anonymity and to resist the calling information analysis.

Insights of the paper

1. **Zone anonymity is introduced**

The first thing we like about this paper is its design to ensure the zone anonymity. Herd guarantees that the only factor of users' anonymity is their selection of providers and the

anonymity of one client does not affect that of its partners. This is achieved by layered encryption and rendezvous mechanism. Contents are encrypted in layers so that any individual node cannot decrypt it completely. Besides, each user only needs to choose a trusted mix, specified by the administrators of Herd. The connections between mixes are established via the rendezvous mixes, which hide the mixes users attach to and ensure the anonymity. Thus, if caller and callee are in different zones, even though one zone is compromised, the one in another zone will not be affected. This design buffers the attacks from one zone of the circuit and ensures the anonymity of the users in another zone. Also, such a buffer enables administrators to have time to recognize and fix the security issues.

2. Using superpeers (SPs) to increase scalability and reduce operational cost

The second highlight of Herd is its scalability. With SPs, the scalability of Herd can be linearly increased by inviting more partners to provide their bandwidth and servers. Also, it significantly reduces the operational cost by reducing the data used by mixes. According to the observations by the authors, only a relatively small portion of clients are active at any time point. According to the SP architecture, as shown in Figure 2 in the paper, an SP is connected to several clients and one mix, where the packet from clients are XORed at the corresponding SP and then forward to the mix. Some SPs may be untrusted, but it does not reduce the overall security level because SPs do not know if a certain client is active or not and SPs cannot manipulate the traffic without penalty. Thus, administrators of Herd do not need to consider too much about SPs, and they do not need to spend money to evaluate their security level.

3. Traffic-analysis resistance

The third point we like about Herd is its innovative design to confuse the traffic analyses from adversaries. Herd uses padding mechanisms among the entire circuit. The key idea is to ensure a constant link rate to make such activity analyses inaccurate. What is seen by the attacker is

only the constant rated, encrypted packets. Under this mechanism, traffic analyses only tell the number of maximum possible active users instead of detailed individual activities. This mechanism fixes the potential activity leaks in other designs of VoIP systems.

Limitations of the paper

1. **Using historical information on traffic obfuscation may affect performance**

The link rates between mixes and SPs may change according to the call volume and updates once every several hours. An hour-level change may not be enough to accommodate successive volume increments or may waste some bandwidth when call volume decreases continually.

2. **How to find SPs remains a problem**

The low operational cost heavily relies on the fare policy of SPs. According to the result derived from the authors, the number of SPs is much smaller than the number of clients. Thus the communication cost between mixes and clients is greatly reduced by using SPs as a relay. However, even though the authors provide some incentives to SPs, how to guarantee enough SPs is still a problem. If not enough SPs are available, the communication from clients to mixes would have some probability to be blocked under high channel utilization. In this case, to guarantee the availability, the administrator needs to build more mixes and this increases the operational cost and reduces the profit.

3. **Herd is not portable**

According to the designs of Herd, clients must be constantly online to maximize the anonymity because clients need to send padded chaffing data at a constant rate to confuse the adversaries. Therefore, it is not feasible for mobile devices since sending data constantly will

significantly reduce the battery life. Also, cellular data is far more expensive than wired data in some countries.

Discussion and Improvement

1. **Resistance to round-trip time analysis from compromised participants**

As mentioned by the authors, if one of the two participants is compromised, the adversary can estimate the distance between caller and callee by analysing the round trip time of packets in the circuit. The authors are planning to add an artificial delay to fix this issue. However, this mechanism has two weaknesses. First, as mentioned by the authors, it will degrade the calling quality. Second, the artificial delay will reduce the data efficiency of the VoIP circuit. In our opinion, it is better to develop an updated routing algorithms for mixes to confuse the distance analysis while maintaining the utilization efficiency. For instance, the mixes may deliberately choose a non-optimal path while ensuring the utilization efficiency still at a high level.

2. **Dealing with the compromise of the trusted mix**

Even though the decoupling mechanism of Herd ensures that the compromise of one client does not affect another, the compromise of mixes still matters. We wonder if there is a way to free clients from choosing a trusted mix. In other words, is there possible to have a communication between caller and callee with no trusted mixes? For example, is it possible for a mix communicate with clients, but do not know the identity of the clients?

3. **Discussion about the portability**

As mentioned in the previous section, the original design of Herd limits its portability. The key point of the issue is the constant chafing traffic. Since the cellular modules in mobile devices tend to be silent at most of the time [1]. If someone using the Herd in a mobile device, at the time of the start of the application, the adversaries may detect the unusual constant data traffic and trace the location of the caller. In our opinion, it may be helpful to use random padding

instead of constant padding. In this case, the adversaries may think it is normal web data traffic instead of Herd VoIP data traffic.

References

[1] Fukuda, Kensuke, Hirochika Asai, and Kenichi Nagami. "Tracking the Evolution and Diversity in Network Usage of Smartphones." *Proceedings of the 2015 ACM Conference on Internet Measurement Conference*. ACM, 2015.