Trust-based Privacy Preservation for Peer-to-peer Data Sharing

Y. Lu, W. Wang, D. Xu, and B. Bhargava
yilu, wangwc, dxu, bb @ cs.purdue.edu
Department of Computer Sciences
Purdue University

The work is supported by NSF ANI-0219110 and IIS-0209059
Problem statement

- Privacy in peer-to-peer systems is different from the anonymity problem
- Preserve privacy of requester
- A mechanism is needed to remove the association between the identity of the requester and the data needed
Proposed solution

- A mechanism is proposed that allows the peers to acquire data through trusted proxies to preserve privacy of requester
  - The data request is handled through the peer’s proxies
  - The proxy can become a supplier later and mask the original requester
Related work

- Trust in privacy preservation
  - Authorization based on evidence and trust, [Bhargava and Zhong, DaWaK’02]
  - Developing pervasive trust [Lilien, CGW’03]

- Hiding the subject in a crowd
  - K-anonymity [Sweeney, UFKS’02]
  - Broadcast and multicast [Scarlata et al, INCP’01]
Related work (2)

- Fixed servers and proxies
  - Publius [Waldman et al, USENIX’00]

- Building a multi-hop path to hide the real source and destination
  - FreeNet [Clarke et al, IC’02]
  - Crowds [Reiter and Rubin, ACM TISS’98]
  - Onion routing [Goldschlag et al, ACM Commu.’99]
Related work (3)

- $p^5$ [Sherwood et al, IEEE SSP’02]
  - $p^5$ provides sender-receiver anonymity by transmitting packets to a broadcast group
- Herbivore [Goel et al, Cornell Univ Tech Report’03]
  - Provides provable anonymity in peer-to-peer communication systems by adopting dining cryptographer networks
Privacy measurement

- A tuple \(<\text{requester ID}, \text{data handle}, \text{data content}>\) is defined to describe a data acquirement.

- For each element, “0” means that the peer knows nothing, while “1” means that it knows everything.

- A state in which the requester’s privacy is compromised can be represented as a vector \(<1, 1, y>, (y \in [0,1])\) from which one can link the ID of the requester to the data that it is interested in.
Privacy measurement (2)

For example, line $k$ represents the states that the requester’s privacy is compromised.

Point A illustrates a state that both peer identity and data handle are known.

The privacy of the requester can be compromised.

Point B illustrates a state that every detail of the data acquirement is known.
Mitigating collusion

- An operation “∗” is defined as:

\[ < c_1, c_2, c_3 > = < a_1, a_2, a_3 > * < b_1, b_2, b_3 > \]

\[ c_i = \begin{cases} 
\max(a_i, b_i), & a_i \neq 0 \text{ and } b_i \neq 0; \\
0, & \text{otherwise.} 
\end{cases} \]

- This operation describes the revealed information after a collusion of two peers when each peer knows a part of the “secret”.

- The number of collusions required to compromise the secret can be used to evaluate the achieved privacy
Trust based privacy preservation scheme

- The requester asks one proxy to look up the data on its behalf. Once the supplier is located, the proxy will get the data and deliver it to the requester
  - Advantage: other peers, including the supplier, do not know the real requester
  - Disadvantage: The privacy solely depends on the trustworthiness and reliability of the proxy
Trust based scheme – Improvement 1

- To avoid specifying the data handle in plain text, the requester calculates the hash code and only reveals a part of it to the proxy.
- The proxy sends it to possible suppliers.
- Receiving the partial hash code, the supplier compares it to the hash codes of the data handles that it holds. Depending on the revealed part, multiple matches may be found.
- The suppliers then construct a bloom filter based on the remaining parts of the matched hash codes and send it back. They also send back their public key certificates.
Trust based scheme – Improvement 1

- Examining the filters, the requester can eliminate some candidate suppliers and finds some who may have the data.
- It then encrypts the full data handle and a data transfer key $k_{Data}$ with the public key.
- The supplier sends the data back using $k_{Data}$ through the proxy.
- Advantages:
  - It is difficult to infer the data handle through the partial hash code
  - The proxy alone cannot compromise the privacy
  - Through adjusting the revealed hash code, the allowable error of the bloom filter can be determined
Data transfer procedure after improvement 1

\begin{figure}
\centering
\includegraphics[width=\textwidth]{data_transfer_diagram}
\caption{Data transfer procedure after improvement 1}
\end{figure}

\begin{itemize}
\item \textbf{R}: requester  \textbf{S}: supplier
\item \textbf{Step 1, 2}: \textbf{R} sends out the partial hash code of the data handle
\item \textbf{Step 3, 4}: \textbf{S} sends the bloom filter of the handles and the public key certificates
\item \textbf{Step 5, 6}: \textbf{R} sends the data handle and \( k_{Data} \) encrypted by the public key
\item \textbf{Step 7, 8}: \textbf{S} sends the required data encrypted by \( k_{Data} \)
\end{itemize}
Trust based scheme – Improvement 2

- The above scheme does not protect the privacy of the supplier
- To address this problem, the supplier can respond to a request via its own proxy
Trust based scheme – Improvement 2
Trustworthiness of peers

- The trust value of a proxy is assessed based on its behaviors and other peers’ recommendations.
- Using Kalman filtering, the trust model can be built as a multivariate, time-varying state vector.
Experimental platform - TERA

- Trust enhanced role mapping (TERM) server assigns roles to users based on
  - Uncertain & subjective evidences
  - Dynamic trust
- Reputation server
  - Dynamic trust information repository
  - Evaluate reputation from trust information by using algorithms specified by TERM server
Trust enhanced role assignment architecture (TERA)
Conclusion

- A trust based privacy preservation method for peer-to-peer data sharing is proposed
- It adopts the proxy scheme during the data acquisition

Extensions
- Solid analysis and experiments on large scale networks are required
- A security analysis of the proposed mechanism is required
Related publication


