Privacy-Aware Collaborative Spam Filtering

Collaboration is a natural defense against spam attacks. However, any large-scale collaborative anti-spam approach is faced with a fundamental challenge, ensuring the privacy of the messages among collaborative but untrusted users.

Our Approach

1. Design a cryptographic transformation that satisfies two competing requirements:
   a) hide the actual content for privacy protection.
   b) retain important features of the message so that effective similarity comparison can still be performed on the encrypted messages.

2. Design a novel query-response protocol to avoid privacy breaches by minimizing the information revealed during the collaboration.

Approach and Impact

<table>
<thead>
<tr>
<th>New approach</th>
<th>Research Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Classification with shingle-based message digests</td>
<td>- Robust defense against camouflage attacks</td>
</tr>
<tr>
<td>- Privacy-aware distributed query protocols</td>
<td>- Metrics for privacy breach measurement in collaboration</td>
</tr>
</tbody>
</table>

Feature-preserving Transformation

- Shingle based message digests
  - Robust similarity detection even under camouflage attacks
- Message shuffle to prevent term-level privacy breaches

Privacy-aware Query Protocols

- Approximate Query
  - Partial digests and digest ranges
- Asymmetric Response
  - Spam/ham dichotomy: return entire spam digests but partial ham digests