Secure Anonymous Database Search

Mariana Raykova, Binh Vo, Steven Bellovin, Tal Malkin
Columbia University

Problem Statement

Secure Anonymous Database Search – controlled data sharing between untrusting parties. Applications:
- Intelligence agencies sharing information
- Police investigation on sensitive information
- Medical records search
- Detecting attack behavior from log files
- Automatic email filtering

Security Architecture

PH-DSAEP+

- Group property – reencrypt under new key, allows user authorization and revocation.
- PH-DSAEP+ Construction - Pohlig-Hellman function (group property) + SAEP padding (security guarantee); deterministic transformation PH-DSAEP+ (Bellare et al. 2007)

Private Key Deterministic Encryption

PH-DSAEP+ - private key deterministic encryption

- Group property – reencrypt under new key, allows user authorization and revocation.
- PH-DSAEP+ Construction - Pohlig-Hellman function (group property) + SAEP padding (security guarantee); deterministic transformation PH-DSAEP+ (Bellare et al. 2007)

Bloom Filter Search and Storage

PH-SAEP+ Construction

- Pohlig-Hellman function, randomness with hash of inputs.
- Security guarantee – reencrypt under new key, allows user authorization and revocation.
- PH-SAEP+ construction – replace random herm with hash of inputs.

Privacy

Server’s database:

- IS cannot link BFs to documents; QR enforces client authorization; IS and QR cannot search;
- C receives only relevant results (adjustable FP rate).
Client’s query:

- IS cannot link results to documents and queries of the same client;
- QR learns only equality of queries, QR does not learn anything about results.

Boolean Queries

- AND queries unioned in query indices.
- OR queries processed in parallel; query indices are handled in order of frequency in queries, improves cache behavior.
- Efficient search for boolean queries representable in monotone disjunctive normal form.

Performance

Performance Results for Different Corpus Sizes

<table>
<thead>
<tr>
<th>Capacity</th>
<th>1k</th>
<th>2k</th>
<th>3k</th>
<th>5k</th>
<th>10k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size(bits)</td>
<td>16384</td>
<td>32768</td>
<td>65536</td>
<td>131072</td>
<td>262144</td>
</tr>
</tbody>
</table>

Fixed false positive rate = 0.001