Enhancing the Security of Yioop Discussion Board

Masters Defense

Prajna Gururaj Puranik
Importance of security

Protection of user data
Safety of sensitive user information

Preserving user trust
Instill confidence in users

Safeguarding against attacks
Web portals are vulnerable to attacks

Compliance with regulations
Comply with laws like CCPA and GDPR

Existing features
Differential Privacy
Encrypted groups
External database
## Newly added Security Features

### Extending Differential Privacy
- Hide number of users

### Flag and Moderation
- Flag button and Moderation group

### Secret Sharing
- Restrict access to encryption key
Security considerations

- Security in social media sites similar to Yioop
- Study by Social Network Analysis and Mining journal [1] → 60% of respondents support use of flags
- Moderation success in Reddit → banned several subreddits in 2015
- Banning posts saw an 80% decrease in hate speech usage [2]
Security considerations

- The California Consumer Privacy Act (CCPA) and the General Data Protection Regulation (GDPR) → keep sensitive user data safe and private [3]
- Compelled businesses to be more open about how they gather user data and to give users greater control over it, which has resulted in the adoption of better privacy measures on the internet
Differential Privacy

- Statistical attacks → Extraction of private information by analyzing patterns/statistical properties of the data.
- Differential privacy → Mathematical framework for protecting users’ privacy in datasets.
- Adds noise so that individual data points cannot be distinguished → causes statistical attacks to fail
- Yioop uses $\epsilon$-differential privacy → $\epsilon$ is the privacy parameter
- Extended to mask the number of users in the group

Benefits:
- User anonymity
- Avoiding Bias or Prejudice
- Protection against Targeted Attacks
Differential Privacy

- 3 UI instances where the group user count is displayed:
  - Group owner - Edit Group
  - Group owner - Edit Members
  - Group user - Manage Group
Yioop allows creation of encrypted groups.

Title and description of threads are encrypted and stored.

Key stored in a separate database.

GROUP_ID attribute is used to access the key.
GROUP_ID is stored in public database - Not safe to access key

So use secret sharing scheme to securely compute keys

Sharing a secret among a group of participants in a way that no individual can deduce the secret by themselves

Linear secret sharing is used here, where a line is used to generate shares that will be distributed to the users

Note: Group owner adds users to the group

Secret Sharing
(random1, random2) from private DB

OWNER_ID, GROUP_ID, random3 and random4 from public DB

Use X = random4 to get Y value
Flagging

- Flagging → marking content that violates guidelines
- Flagged posts sent to moderators for review

**Benefits:**
- Early detection of harmful content
- Transparency and user empowerment

**Considerations:**
- A user can't flag a post more than once
- Appropriate dialog boxes to confirm choice to flag
- Appropriate message if threshold reached
- Care taken to ensure encrypted groups have masked flag values
A group of moderators to review flagged posts

Root user can add other moderators to the group

Each flagged post appears as a separate thread

Each thread allows moderators to:
- Comment
- Approve
- Delete
- Check original posts

Benefits:
- Ensures a safe atmosphere for users
- Risk mitigation: Detect malicious content, phishing attempts, spams, harassment etc
<table>
<thead>
<tr>
<th>Differential Privacy</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test Type</strong></td>
<td><strong>Baseline</strong></td>
<td><strong>Post Implementation</strong></td>
</tr>
<tr>
<td>System Load Time</td>
<td>0.091s</td>
<td>0.091s</td>
</tr>
<tr>
<td>Page load time - Edit Group Page</td>
<td>0.19s</td>
<td>0.19s</td>
</tr>
<tr>
<td>Page load time - View Group Page</td>
<td>0.15s</td>
<td>0.16s</td>
</tr>
<tr>
<td>Page load time - Manage Group Page</td>
<td>0.1s</td>
<td>0.1s</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Flagging</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test Type</strong></td>
<td><strong>Baseline</strong></td>
<td><strong>Post Implementation</strong></td>
</tr>
<tr>
<td>System Load Time</td>
<td>0.091s</td>
<td>0.092s</td>
</tr>
<tr>
<td>Page load time - Group Thread page</td>
<td>0.29s</td>
<td>0.32s</td>
</tr>
<tr>
<td>Response Time - Flag post</td>
<td>1.6s</td>
<td>0.92s</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Moderation</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test Type</strong></td>
<td><strong>Baseline</strong></td>
<td><strong>Post Implementation</strong></td>
</tr>
<tr>
<td>System Load Time</td>
<td>0.091s</td>
<td>0.095s</td>
</tr>
<tr>
<td>Page load time - Root login and load</td>
<td>0.513s</td>
<td>0.515s</td>
</tr>
<tr>
<td>Time taken to approve/delete</td>
<td>-</td>
<td>0.3s</td>
</tr>
<tr>
<td>Time taken to view original thread</td>
<td>0.14s</td>
<td>0.12s</td>
</tr>
<tr>
<td>Time taken to comment</td>
<td>0.13s</td>
<td>0.13s</td>
</tr>
<tr>
<td>Time to add new users</td>
<td>0.16s</td>
<td>0.18s</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Secret Sharing</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test Type</strong></td>
<td><strong>Baseline</strong></td>
<td><strong>Post Implementation</strong></td>
</tr>
<tr>
<td>System Load Time</td>
<td>0.091s</td>
<td>0.096s</td>
</tr>
<tr>
<td>Page load time - Group Creation</td>
<td>0.24s</td>
<td>0.25s</td>
</tr>
<tr>
<td>Time taken to generate a key</td>
<td>-</td>
<td>0.72s</td>
</tr>
</tbody>
</table>
Conclusion

❖ Implemented new mechanisms to enhance security
❖ Incorporated content moderation features to elevate user experience and increase security of the platform
❖ Enhanced security of encryption keys
Future Work

- Content moderation can be extended to include features like banning users, locking threads etc
- New avenues to extend the implementation of differential privacy
- Encryption techniques like homomorphic encryption can be explored to protect the existing upvote/downvote feature
Thank you!
References
