Cloudsweeper Prototype

Problem Area

Solution

Guiding Observations

Common Email Accounts Contain Valuable Secrets

- "No need to delete anything!"
- Used as general data store
- Gather huge amounts of sensitive information over time

Most Security Approaches Secure Too Little

- Security is applied at perimeter
- Focus is on limiting access
- Provides no security against email providers, leaked passwords, etc.

Remaining Approaches Secure Too Much

- Maximum protection to all data
- Very difficult to use
- Negates many cloud benefits (search, monetization, etc.)

Measure and Mitigate

Heterogeneous Documents for Risk Aware Security

- Extract and protect sensitive portions of documents
- Less protection, but convenience, for remaining non-sensitive data

Public, Secure and Free

- Prototype compatible with Gmail based accounts
- No software to install
- Requires minimum permissions

Used By The Public Today

<table>
<thead>
<tr>
<th>Measure</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloudsweeper Users</td>
<td>2,386</td>
</tr>
<tr>
<td>Messages Searched</td>
<td>462,240</td>
</tr>
<tr>
<td>Passwords Found</td>
<td>200,724</td>
</tr>
<tr>
<td>Passwords Redacted</td>
<td>14,896</td>
</tr>
<tr>
<td>Passwords Encrypted</td>
<td>22,753</td>
</tr>
</tbody>
</table>

Password Sharing Stats

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Messages Per Account</td>
<td>19,177</td>
<td>7,401</td>
</tr>
<tr>
<td>Avg. Passwords Per Account</td>
<td>84</td>
<td>26</td>
</tr>
<tr>
<td>% Messages w/ Password</td>
<td>0.4%</td>
<td>0.4%</td>
</tr>
</tbody>
</table>

Cloudsweeper Application Flow

1. Server side search returns prospective messages
2. Body text search finds sensitive information
3. User chooses information to protect
4. Encrypted information
4a. Returns Key
5. System saves edited messages back on server
6. System permanently expunges original messages

This work has been funded with support from NSF under the IGERT Grant Number 1069311