Gossamer: Securely Measuring Password-based Logins

Marina Sanusi Bohuk, Mazharul Islam, Suleman Ahmad, Michael Swift, Thomas Ristenpart, Rahul Chatterjee
Modern Authentication Systems

How do we safely log information about actual passwords?

Client

username, password
+ IP address
+ user agent
+ password information

Freeman et al. 2016
Tian et al. 2019

Is hash(password) in the database?

Passwords are no longer sufficient!
Credential stuffing is a huge source of account compromise.
How do we separate benign and malicious traffic?
Logging Password-Derived Measurements

Design a measurement framework (Gossamer) for use with web login systems (1.5-year-long process).

Describe a process for assessing risk of password-based measurements.

Conduct a measurement study at two universities observing over 34M login requests.
Can we add instrumentation that looks at passwords?

Architecture

student center

email

bursar

username, password

Single-sign-on (SSO) service

sanitized login request

Measurement service (VM)

Ephemeral DB

researcher access

Analysis service (VM)

Persistent DB

Design principles

1. Safe-on-reboot (Miklas ’09)
2. Periodic deletion
3. Least privilege access
Architecture

Design principles
1. Safe-on-reboot (Miklas '09)
2. Periodic deletion
3. Least privilege access
4. Bounded leakage logging

If compromised, how could attackers use password-derived measurements to speed up attacks?

Encrypted username and pw plaintext IP…

Encryption: username & pw

Encrypted username, plaintext IP…

Pw-derived information
How can we choose safe measurements to log?

 Guess list
Gossamer logs
(Encrypted) username

Sends guess

Attacker guess list

<table>
<thead>
<tr>
<th>Guess rank</th>
<th>Password</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>qwerty</td>
</tr>
<tr>
<td>2</td>
<td>abc123</td>
</tr>
<tr>
<td>3</td>
<td>hunter</td>
</tr>
<tr>
<td>4</td>
<td>jessica</td>
</tr>
<tr>
<td>5</td>
<td>spider</td>
</tr>
</tbody>
</table>

Gossamer logs

<table>
<thead>
<tr>
<th>Encrypted username</th>
<th>zxcvbn score</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{g}1\text{V}B5\text{T}H$</td>
<td>2</td>
</tr>
<tr>
<td>$\text{g}k3\text{p}\text{P}h\text{L}$</td>
<td>1</td>
</tr>
<tr>
<td>$\text{trZQA1L}$</td>
<td>3</td>
</tr>
<tr>
<td>$\text{jNK}R3\text{Y}p$</td>
<td>2</td>
</tr>
<tr>
<td>$\text{OXJFw}2\text{r}$</td>
<td>4</td>
</tr>
</tbody>
</table>
How can we choose safe measurements to log?

Guess list
Gossamer logs
(Encrypted) username

Attacker guess list

<table>
<thead>
<tr>
<th>Guess rank</th>
<th>Password</th>
<th>zxcvbn score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>qwerty</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>abc123</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>hunter</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>jessica</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>spider</td>
<td>3</td>
</tr>
</tbody>
</table>

1 guess

Gossamer logs

<table>
<thead>
<tr>
<th>Encrypted username</th>
<th>zxcvbn score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ø1VB5TH</td>
<td>2</td>
</tr>
<tr>
<td>gk3pPhL</td>
<td>1</td>
</tr>
<tr>
<td>trZQA1L</td>
<td>3</td>
</tr>
<tr>
<td>jNKR3Yp</td>
<td>2</td>
</tr>
<tr>
<td>OXJFw2r</td>
<td>4</td>
</tr>
</tbody>
</table>
How can we choose safe measurements to log?

**Dataset:** 307 million breached passwords

Attacker’s guess list: 80% split
Target passwords: 10k passwords sampled from remaining 20%

**Problem:** Original zxcvbn score leaks too much information!

**Solution:** Bucketize score to [0, 1]

✓ Bounded leakage logging
Deploying Gossamer

Obtained approval from respective IRB and the IT offices.

Collected **34M** total login requests

Observed some high-volume attacks

- **Attack 1**
  - Accounts for 54K requests at U1
  - 7 months

- **Attack 2**
  - Accounts for 81K requests at U2
  - 3 months

- **Attack 3**
Password managers could help…
About 25% of users use password managers.

Login friction is still high

**Typos are frequent**
Over 1 in 3 failed requests at U1 were typos. Even more for mobile logins.

- Actual password: marina123
- Typo: marina1223

**Retries are common**
- 1/5 at U1
- 1/3 at U2
Eventually successful sessions required more than one attempt.

**2FA impedes usability**
Duo adds an average of 14 seconds to a user’s login.
Breached credential use is a problem.

23 U1 users and 254 U2 users were using a breached password.

Over 2K U1 users and 1K U2 users were using a tweaked breached password.

The high-volume attacks had high fractions of breached passwords.

Solution: Proactive breach alerting
Thomas et al. 2019, Li et al. 2019, Pal et al. 2022

Next: Investigate how to detect attacks better using these measurements
Gossamer

Safely record information about submitted passwords
  o Bounded leakage logging
  o Assess risk; reduce granularity

Extend with additional measurements
  o Simulate improvement in attack

Gain insight into user and attacker behavior
  o Can inform new policies
  o Develop countermeasures

https://cs.cornell.edu/~marina/gossamer
Marina Bohuk | marina@cs.cornell.edu
Mazharul Islam | mislam9@wisc.edu