Understanding Online Social Network Usage from a Network Perspective

Fabian Schneider* ‡
fabian@net.t-labs.tu-berlin.de

Anja Feldmann‡ Balachander Krishnamurthy§ Walter Willinger§

* Work done while at AT&T Labs–Research
‡ Technische Universität Berlin / Deutsche Telekom Laboratories
§ AT&T Labs–Research

Internet Measurement Conference 2009
Motivation

- >600,000,000 users on Online Social Networks (OSNs)
  …and the number is still growing

- Open questions/challenges
  - Which features are popular among OSN users?
  - How much time do users’ spend interacting with OSNs?
  - Is there a correlation between subsequent interactions?

- Relevance of OSN usage
  - ISPs: data transport, connectivity
  - OSN providers: develop and operate scalable systems
  - R&D: Identify trends, suggest improvements, and new designs
Outline

1. Approach
2. Session Characteristics
3. Feature Popularity
4. Dynamics within Sessions
5. Conclusions

Sessions
Session = Set of interactions of one user

Features
Feature = Action a user can perform
General Approach

1. Reconstruct OSN clickstreams from anonymized packet-level traces
   - Anonymized HTTP header traces from two large ISPs
   - Used Bro\(^1\) to extract HTTP request-response pairs (rr-pairs)

2. Map rr-pairs into sessions
   - Sessions identified via SessionIDs (from HTTP Cookie header)
   - Track logins and logouts $\Rightarrow$ Authenticated or offline state
   - Cookies help if login or logout not observed

3. Classify rr-pairs
   - Active (rr-pair resulting from user action) or
   - Indirect (e.g. followup/embedded via HTTP Referer chain)
   - Determine user actions, group into 13 categories

\(^1\)www.bro-ids.org
<table>
<thead>
<tr>
<th>Time [sec]</th>
<th>Action/Click</th>
<th>No.</th>
<th>Proto</th>
<th>Method</th>
<th>URI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>a) open <a href="http://www.facebook.com">www.facebook.com</a></td>
<td>1</td>
<td>HTTP</td>
<td>GET</td>
<td>/</td>
</tr>
<tr>
<td>9.944</td>
<td></td>
<td>2</td>
<td>HTTPS</td>
<td>POST</td>
<td>/login.php?</td>
</tr>
<tr>
<td>27.696</td>
<td>b) login, enter password</td>
<td>3</td>
<td>HTTP</td>
<td>GET</td>
<td>/home.php?</td>
</tr>
<tr>
<td>29.121</td>
<td></td>
<td>4</td>
<td>HTTP</td>
<td>GET</td>
<td>/friends/?ref=tn&amp;quickling[version]=141637;0&amp;_ecdc=check</td>
</tr>
<tr>
<td>31.012</td>
<td></td>
<td>5</td>
<td>HTTP</td>
<td>GET</td>
<td>/friends/ajax/friends.php?membership=1&amp;_ecdc=check</td>
</tr>
<tr>
<td>45.513</td>
<td>c) open friend list</td>
<td>6</td>
<td>HTTP</td>
<td>GET</td>
<td>/friends/ajax/filters.php?id=XXX&amp;_ecdc=check</td>
</tr>
<tr>
<td>47.631</td>
<td></td>
<td>7</td>
<td>HTTP</td>
<td>GET</td>
<td>/profile.php?id=XXX&amp;quickling[version]=141637;0&amp;_ecdc=check</td>
</tr>
<tr>
<td>48.672</td>
<td>d) select profile of a friend</td>
<td>8</td>
<td>HTTP</td>
<td>GET</td>
<td>/ajax/profile/composer.php?_ecdc=false</td>
</tr>
<tr>
<td>95.921</td>
<td>e) write “posted something on the wall” on friends wall</td>
<td>9</td>
<td>HTTP</td>
<td>POST</td>
<td>/logout.php?h=c909dd2db7b0a83b238ea70321d2041b&amp;ref=mb</td>
</tr>
<tr>
<td>97.947</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>/index.php?1h=c909dd2db7b0a83b238ea70321d2041b&amp;</td>
</tr>
<tr>
<td>Time [sec]</td>
<td>Action/Click</td>
<td>No.</td>
<td>Proto</td>
<td>Method</td>
<td>URI</td>
</tr>
<tr>
<td>------------</td>
<td>--------------</td>
<td>-----</td>
<td>-------</td>
<td>--------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>0.000</td>
<td>a) open <a href="http://www.facebook.com">www.facebook.com</a></td>
<td>1</td>
<td>HTTP</td>
<td>GET</td>
<td>/</td>
</tr>
<tr>
<td>9.944</td>
<td>b) login, enter password</td>
<td>2</td>
<td>HTTPS</td>
<td>POST</td>
<td>/login.php?</td>
</tr>
<tr>
<td>27.696</td>
<td></td>
<td>3</td>
<td>HTTP</td>
<td>GET</td>
<td>/home.php?</td>
</tr>
<tr>
<td>29.121</td>
<td></td>
<td>4</td>
<td>HTTP</td>
<td>GET</td>
<td>/friends/?ref=tn&amp;quickling[version]=141637;0&amp;ecdc=check</td>
</tr>
<tr>
<td>31.012</td>
<td></td>
<td>5</td>
<td>HTTP</td>
<td>GET</td>
<td>/friends/ajax/friends.php?membership=1&amp;ecdc=check</td>
</tr>
<tr>
<td>45.513</td>
<td></td>
<td>6</td>
<td>HTTP</td>
<td>GET</td>
<td>/friends/ajax/filters.php?id=XXX&amp;ecdc=check</td>
</tr>
<tr>
<td>47.631</td>
<td></td>
<td>7</td>
<td>HTTP</td>
<td>GET</td>
<td>/profile.php?id=XXX&amp;quickling[version]=141637;0&amp;ecdc=check</td>
</tr>
<tr>
<td>48.672</td>
<td></td>
<td>8</td>
<td>HTTP</td>
<td>GET</td>
<td>/ajax/profile/composer.php?_ecdc=false</td>
</tr>
<tr>
<td>56.441</td>
<td></td>
<td>9</td>
<td>HTTP</td>
<td>GET</td>
<td>/logout.php?h=c909dd2db7b0a83b238ea70321d2041b&amp;ref=mb</td>
</tr>
<tr>
<td>59.199</td>
<td></td>
<td>10</td>
<td>HTTP</td>
<td>GET</td>
<td>/index.php?lh=c909dd2db7b0a83b238ea70321d2041b&amp;</td>
</tr>
</tbody>
</table>
Figure 1: State handling diagram for OSN sessions.

Table 3: OSN specific information: cookies and login/logout procedure.

<table>
<thead>
<tr>
<th>OSN</th>
<th>session cookie</th>
<th>user cookie</th>
<th>login cookie</th>
<th>login HTTPS</th>
<th>logout HTTPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>datr=</td>
<td>cuser=</td>
<td>login_x=</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>Hi5</td>
<td>JSESSIONID=</td>
<td>Userid=</td>
<td>Email=</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>LinkedIn</td>
<td>bcookie=</td>
<td>leo_authi_token=LIM:</td>
<td>n/a</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>StudiVZ</td>
<td>PHPSESSID=</td>
<td>UserID1=</td>
<td>n/a</td>
<td>✓</td>
<td>×</td>
</tr>
</tbody>
</table>
Table 4: Examples of OSN specific patterns for the classification of home and sending a message.

<table>
<thead>
<tr>
<th>OSN</th>
<th>category</th>
<th>action</th>
<th>method</th>
<th>pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>home</td>
<td>index</td>
<td>GET</td>
<td>^/index.php(???)*$</td>
</tr>
<tr>
<td>Hi5</td>
<td>home</td>
<td>home</td>
<td>GET</td>
<td>^/friend/displays/HomePage.do$</td>
</tr>
<tr>
<td>LinkedIn</td>
<td>home</td>
<td>home</td>
<td>GET</td>
<td>^/home(?*(*))$</td>
</tr>
<tr>
<td>StudiVZ</td>
<td>home</td>
<td>start</td>
<td>GET</td>
<td>^/Start(*)$</td>
</tr>
<tr>
<td>Facebook</td>
<td>messaging</td>
<td>send message</td>
<td>POST</td>
<td>^/inbox(?*(*))$</td>
</tr>
<tr>
<td>Hi5</td>
<td>messaging</td>
<td>send message</td>
<td>POST</td>
<td>^/friend/mail/sendMail.do$</td>
</tr>
<tr>
<td>LinkedIn</td>
<td>messaging</td>
<td>send message</td>
<td>POST</td>
<td>^/msgToConns(?*(*))$</td>
</tr>
<tr>
<td>StudiVZ</td>
<td>messaging</td>
<td>send message</td>
<td>POST</td>
<td>^/Messages/WriteMessage(*)$</td>
</tr>
</tbody>
</table>
OSN Selection

OSN Selection criteria:

- OSNs focussing on profiles (e.g., no YouTube, ...)
- 2 globally popular
- 2 locally popular (well represented at one ISP)

- Facebook
- studiVZ
- LinkedIn
- hi5
OSN Selection criteria:

- OSNs focussing on profiles (e.g., no YouTube, ...)
- 2 globally popular
- 2 locally popular (well represented at one ISP)
HTTP Header Traces (anonymized)

- Collected at residential broadband networks of two commercial ISPs
- Each site connects $\geq 20,000$ DSL users
- Endace monitoring cards for packet capture

**Table:** Overview of anonymized HTTP header traces.

<table>
<thead>
<tr>
<th>ID</th>
<th>start date</th>
<th>dur</th>
<th>sites</th>
<th>size</th>
<th>rr-pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISP-A1</td>
<td>22 Aug’08 noon</td>
<td>24h</td>
<td>all</td>
<td>$&gt;5$ TB</td>
<td>$&gt;80$ M</td>
</tr>
<tr>
<td>ISP-A2</td>
<td>18 Sep’08 4am</td>
<td>48h</td>
<td>all</td>
<td>$&gt;10$ TB</td>
<td>$&gt;200$ M</td>
</tr>
<tr>
<td>ISP-A3</td>
<td>01 Apr’09 2am</td>
<td>24h</td>
<td>all</td>
<td>$&gt;6$ TB</td>
<td>$&gt;170$ M</td>
</tr>
<tr>
<td>ISP-B1</td>
<td>21 Feb’08 7pm</td>
<td>25h</td>
<td>OSNs</td>
<td>$&gt;15$ GB</td>
<td>$&gt;2$ M</td>
</tr>
<tr>
<td>ISP-B2</td>
<td>14 Jun’08 8pm</td>
<td>38h</td>
<td>OSNs</td>
<td>$&gt;50$ GB</td>
<td>$&gt;3$ M</td>
</tr>
<tr>
<td>ISP-B3</td>
<td>23 Jun’08 10am</td>
<td>$&gt;7$d</td>
<td>OSNs</td>
<td>$&gt;110$ GB</td>
<td>$&gt;7$ M</td>
</tr>
</tbody>
</table>
HTTP Header Traces (anonymized)

- Collected at residential broadband networks of two commercial ISPs
- Each site connects $\geq 20,000$ DSL users
- Endace monitoring cards for packet capture

<table>
<thead>
<tr>
<th>ID</th>
<th>start date</th>
<th>dur</th>
<th>sites</th>
<th>size</th>
<th>rr-pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISP-A1</td>
<td>22 Aug’08 noon</td>
<td>24h</td>
<td>all</td>
<td>$&gt;5$ TB</td>
<td>$&gt;80$ M</td>
</tr>
<tr>
<td>ISP-A2</td>
<td>18 Sep’08 4am</td>
<td>48h</td>
<td>all</td>
<td>$&gt;10$ TB</td>
<td>$&gt;200$ M</td>
</tr>
<tr>
<td>ISP-A3</td>
<td>01 Apr’09 2am</td>
<td>24h</td>
<td>all</td>
<td>$&gt;6$ TB</td>
<td>$&gt;170$ M</td>
</tr>
<tr>
<td>ISP-B1</td>
<td>21 Feb’08 7pm</td>
<td>25h</td>
<td>OSNs</td>
<td>$&gt;15$ GB</td>
<td>$&gt;2$ M</td>
</tr>
<tr>
<td>ISP-B2</td>
<td>14 Jun’08 8pm</td>
<td>38h</td>
<td>OSNs</td>
<td>$&gt;50$ GB</td>
<td>$&gt;3$ M</td>
</tr>
<tr>
<td>ISP-B3</td>
<td>23 Jun’08 10am</td>
<td>$&gt;7$d</td>
<td>OSNs</td>
<td>$&gt;110$ GB</td>
<td>$&gt;7$ M</td>
</tr>
</tbody>
</table>

Table: Overview of anonymized HTTP header traces.
Manual Traces

Data set: Active browsing while monitoring passively

For customization
- Good faith effort to explore the feature set of the OSN
- Identify site names, relevant cookies, login/logout actions
- Identify URL patterns for action/category classification

For validation
- Provides ground truth
- 95% of observed actions covered by manual traces
- Remaining actions classified as
  - Guessed (if the URL contains a hint: /ajax/editphoto.php)
  - Unknown
Table 5: Overview of manual traces.

<table>
<thead>
<tr>
<th>OSN</th>
<th>traces</th>
<th>size</th>
<th>actions</th>
<th>rr-pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>11</td>
<td>32 MB</td>
<td>344</td>
<td>5036</td>
</tr>
<tr>
<td>Hi5</td>
<td>6</td>
<td>50 MB</td>
<td>368</td>
<td>4413</td>
</tr>
<tr>
<td>LinkedIn</td>
<td>8</td>
<td>106 MB</td>
<td>411</td>
<td>6363</td>
</tr>
<tr>
<td>StudiVZ</td>
<td>11</td>
<td>27 MB</td>
<td>354</td>
<td>3990</td>
</tr>
</tbody>
</table>
Category Examples

**Home**
All actions on the homepage once authenticated

**Profile**
Accessing and changing profiles, posting on walls, privacy settings

**Apps**
Applications (external and internal), only rr-pairs directed towards OSN servers

**Photos**
Uploading, tagging, and managing photos

**Friends**
Browsing, inviting, and accepting friends

**Offline**
All actions while unauthenticated, e.g., public profile browsing, registering
Caveats of our Approach

- No automated way for
  - producing the URL patterns or
  - extracting the relevant cookies

- External apps: Not tackled as hosted on different sites
  - Requires customization to all/top external apps
  - Navigation redirects could be leveraged

- Friendship graph: Cannot tell if two users are friends
  - Requires parsing of payload (privacy!)
  - Requires users to actually access their friend lists during observation
Outline

1. Approach

2. Session Characteristics

3. Feature Popularity

4. Dynamics within Sessions

5. Conclusions
OSN Session Characteristics

Volume of OSN sessions

- Consistent with a heavy-tailed distribution
- Facebook sessions: 200kB–10MB (StudiVZ: 50kB–5MB)
- Typical Web sessions: 100B–10kB, but heavier tail

Figure 8: CCDF of Bytes per OSN subsession for Facebook and StudiVZ for A2.
OSN Session Characteristics

OSN session durations

- Most sessions are short: 1-5 minutes
- Few lasting for more than an hour (10%–20%)
- Very long (days) sessions observed for 7d trace
OSN Session Characteristics

OSN session durations

- Most sessions are short: 1-5 minutes
- Few lasting for more than an hour (10%–20%)
- Very long (days) sessions observed for 7d trace
OSN Session Characteristics

OSN session durations

- Most sessions are short: 1-5 minutes
- Few lasting for more than an hour (10%–20%)
- Very long (days) sessions observed for 7d trace
Outline

1. Approach
2. Session Characteristics
3. Feature Popularity
4. Dynamics within Sessions
5. Conclusions
Active Facebook rr-pairs by category for ISP-A2

Percentage of RR-Pairs [%]

- Messaging: 22.9%
- Apps: 22.7%
- Home: 19.4%
- Profile: 8.9%
- Photos: 8.5%
- Offline: 5.8%
- Friends: 4.7%
- Search: 2.7%
- Groups: 1.5%
- Osnspecific: 1.2%
- Unknown: 0.9%
- Other: 0.4%
- Videos: 0.4%
- Ads: 0.1%

Categories: active - guessed vs. active - verified
Action Popularity

Active Facebook rr-pairs by category for ISP-A2

Findings

⇒ small fraction of guessed (<3%) & UNKNOWN
Action Popularity

Active Facebook rr-pairs by category for ISP-A2

Findings

⇒ small fraction of guessed (<3%) & UNKNOWN
⇒ Top categories: Messaging, Apps, Home

Percentage of RR-Pairs [%]

messaging | apps | home | profile | photos | offline | friends | search | groups | osnspecific | UNKNOWN | other | videos | ads
22.9% | 22.7% | 19.4% | 8.9% | 8.5% | 5.8% | 4.7% | 2.7% | 1.5% | 1.2% | 0.9% | 0.4% | 0.4% | 0.1%
Volume per Category

Active and indirect Facebook rr-pairs by category for ISP-A2

- Download: guessed
- Upload: guessed
- Download: verified
- Upload: verified

Percentage of HTTP Payload Bytes [%]

Categories:
- home
- profile
- photos
- apps
- offline
- friends
- messaging
- search
- videos
- groups
- UNKNOWN
- osn specific
- other
- ads

Values:
- home: 25.6%
- profile: 20.5%
- photos: 17.4%
- apps: 15.2%
- offline: 7.5%
- friends: 6.2%
- messaging: 3.5%
- search: 1.3%
- videos: 1.2%
- groups: 0.6%
- UNKNOWN: 0.5%
- osn specific: 0.4%
- other: 0.1%
- ads: 0.0%
Volume per Category

Active and indirect Facebook rr-pairs by category for ISP-A2

Findings

⇒ Home, Profile, and Photos rise in importance
⇒ Upload only for Photos and Apps
Feature Popularity: Observations

Active Facebook rr-pairs per session by category for ISP-A2

Heterogeneous user base:
Many users use only one feature category during a session.
Feature Popularity: Observations (cont’d)

OSN and all HTTP rr-pairs per hour for ISP-A2

Per hour usage:
- Time-of-day effects: similar for OSNs and all HTTP
Requested profiles

Approach:

- Profiles represent a user in an OSN.
  Requests to profiles indicate interest in a user
- We distinguish three types of profiles: own, other, and public
- Method: Count which and how often profiles are requested
Requested profiles

(a) All OSNs: Distribution of profiles
Requested profiles

(b) Facebook: Profiles per subsessions
Requested profiles

(c) Facebook: Unique profiles
Requested profiles

Approach:
- Profiles represent a user in an OSN.
  Requests to profiles indicate interest in a user
- We distinguish three types of profiles: own, other, and public
- Method: Count which and how often profiles are requested

Findings
- Types of profile requests:
  - Majority to profiles of other users, 25-35% to own profile,
  - 12% (22%) to public profiles: Facebook Pages (LinkedIn)
- Profile requests per Facebook session:
  - mean number of requested profiles: 6
  - unique profiles: **only 3**
Outline

1. Approach
2. Session Characteristics
3. Feature Popularity
4. Dynamics within Sessions
5. Conclusions
Activity vs. Inactivity Periods

Apply within session inactivity timeout of 5min:

⇒ Sessions >1min: 50% of users are active all time
⇒ Sessions >40min: >95% have inactivity periods

Action after inactivity

- Top categories: Messaging, Home, Offline
- Distribution changes with the length of the pause

Facebook action after inactivity period for ISP-A2
Feature Sequences

Click sequences of Facebook for ISP-A2: Global transition probabilities

Findings

⇒ Messaging traps users; Home, Photos and Profile attract users to stay

Similar findings as Benevenuto et al. for Orkut (IMC’09)
Summary

Findings:

- Most of the sessions are short (few minutes) and small in terms of volume (several MBytes)
- Long sessions are dominated by inactivity periods
- Top action categories are: Messaging, Apps, Home, Profile, and Photos.
- Facebook users are trapped by Messaging and Photos

Future Work

- Expand analysis to other OSNs/external apps, and overcome caveats
- Evaluate new OSN designs with OSN user model (e.g., PeerSoN\(^a\))

\(^a\)www.peerson.net
Summary

Findings:

- Most of the sessions are short (few minutes) and small in terms of volume (several MBytes)
- Long sessions are dominated by inactivity periods
- Top action categories are: Messaging, Apps, Home, Profile, and Photos.
- Facebook users are trapped by Messaging and Photos

Future Work

- Expand analysis to other OSNs/external apps, and overcome caveats
- Evaluate new OSN designs with OSN user model (e.g., PeerSoN\textsuperscript{a})

\textsuperscript{a}www.peerson.net