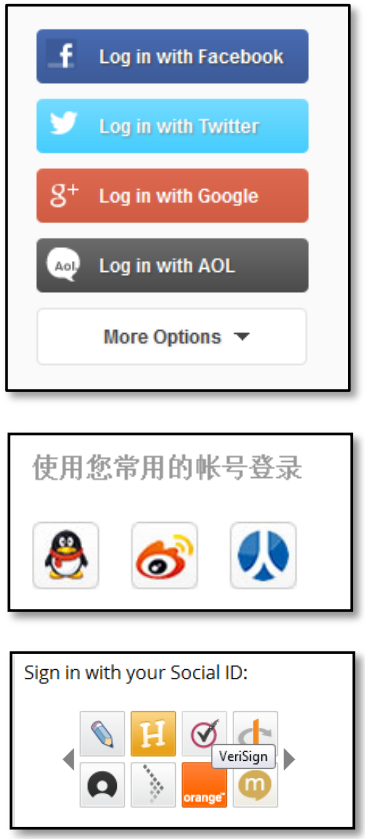
Longitudinal Analysis of the Third-party Authentication Landscape

Anna Vapen, Niklas Carlsson, Nahid Shahmehri

Linköping University, Sweden

LINKÖPING UNIVERSITY

Background: Third-party Web Authentication



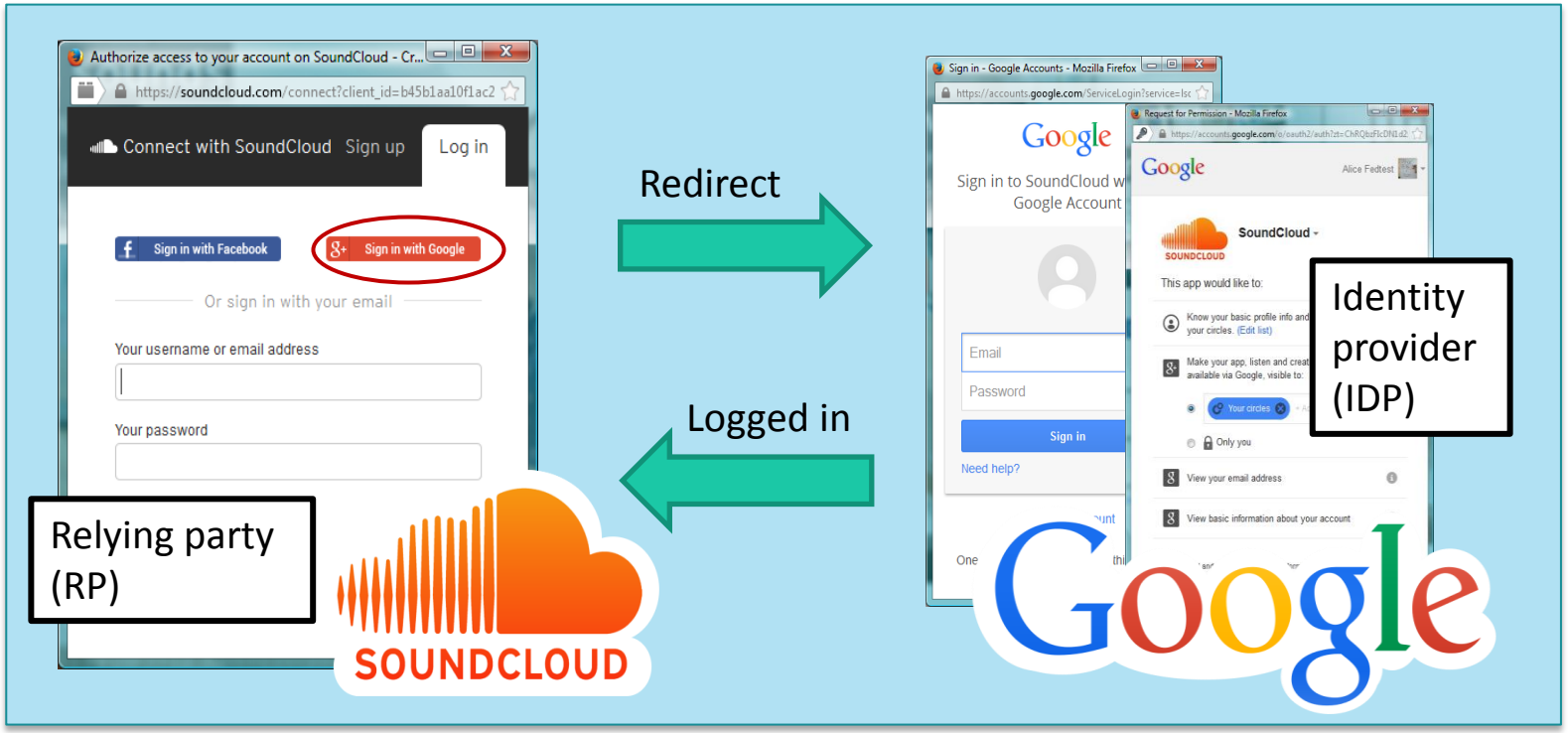
Web Authentication

- Registration with each website
- Many passwords to remember

Third-party authentication

- Use an existing **IDP** (identity provider) account to access an **RP** (relying party)
- Log in less often; Stronger authentication
- Share information between websites
- Information sharing → **privacy leaks!**

Third-party Authentication Scenario



Relationship between RP and IDP

Putting the Work in Context

- Our previous work
  - Large-scale study on the RP-IDP landscape (PAM'14)
  - Categorization of RPs (IEEE IC'16)
  - Detailed study on information flows (SEC'15)

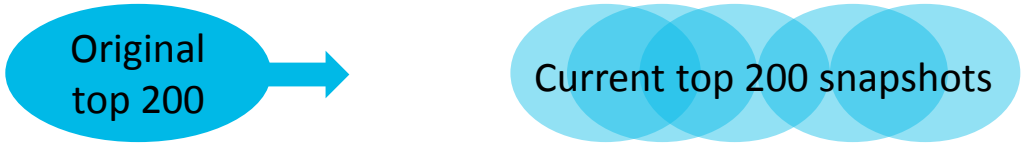
- Current longitudinal study
  - How has the RP-IDP landscape changed over time?
  - Privacy implications of landscape structure?
  - Changes in information flows over time?

Contributions

1. **Structural dynamics**
   - Structural model of the RP-IDP landscape
2. **Protocol-based analysis**
   - Protocol- and IDP changes vs. popularity changes
3. **Flow-based analysis** of privacy risks
   - Information leaks between RPs and IDPs

Methodology

- Top 200 most popular websites
  - Measured at ten points in time, April 2012 to April 2015
  - Original top 200 sites from April 2012, over time
  - Current top 200 at a specific time of measurement



- Data flow analysis of sites using top IDPs (2014-2015)
- Facebook permission agreements

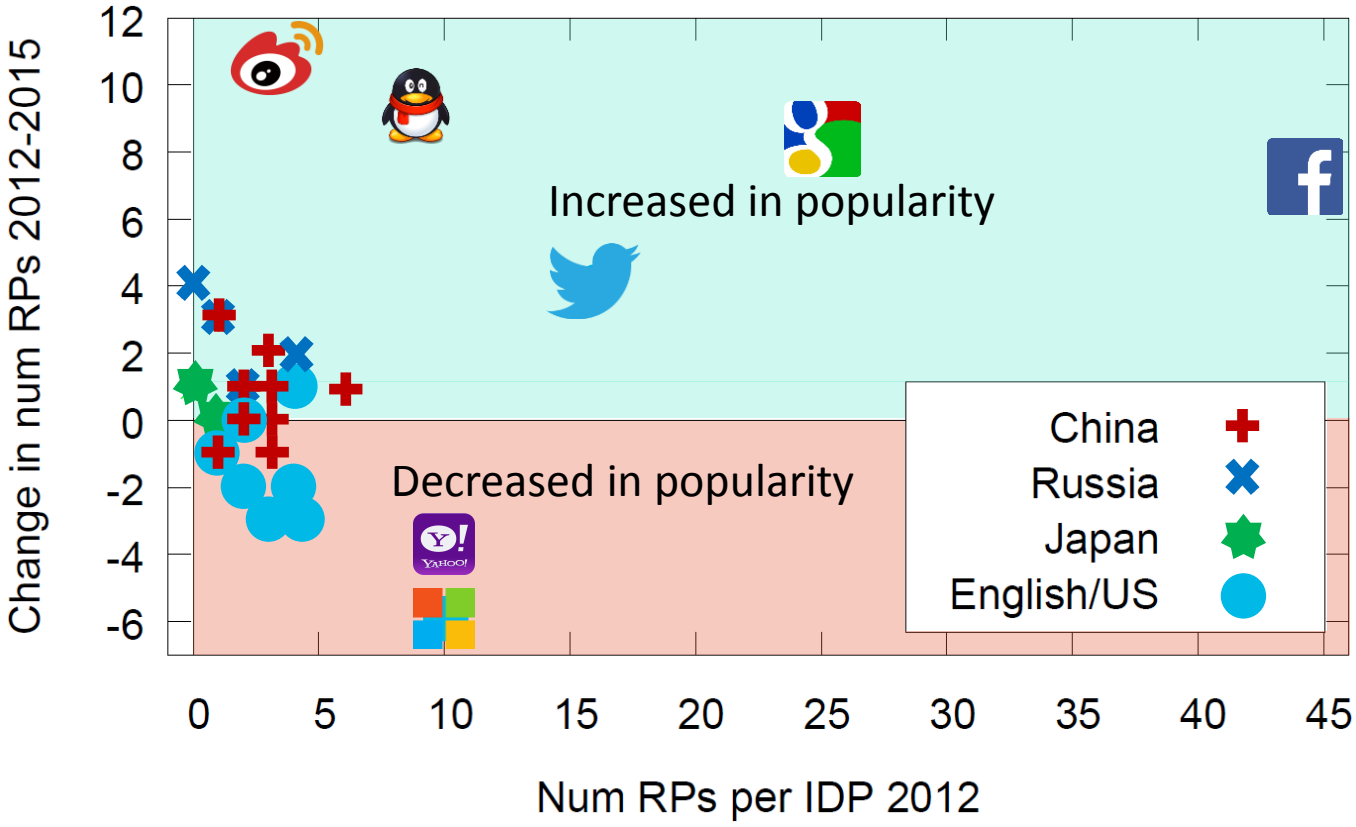
Popular IDPs

Top 200 April 2012: 69 RPs and 180 relationships

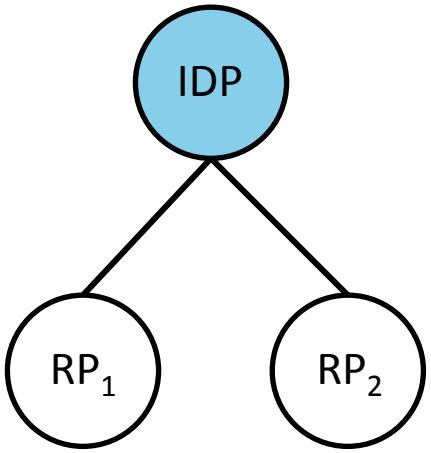
Same sites, April 2015: **+15** RPs and **+33** relationships

Num. relationships with	April 2012	April 2015
Facebook	45	52
Google	25	33
Twitter	16	20
QQ	9	18
Weibo	3	14
Non-top IDPs	82	76
% rels. with top IDPs	54.44%	64.32%
% RPs using top IDP(s)	86.96%	90.48%

Popular IDPs

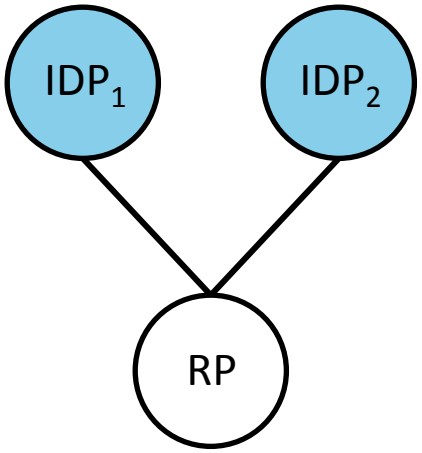


Structures in the RP-IDP Landscape



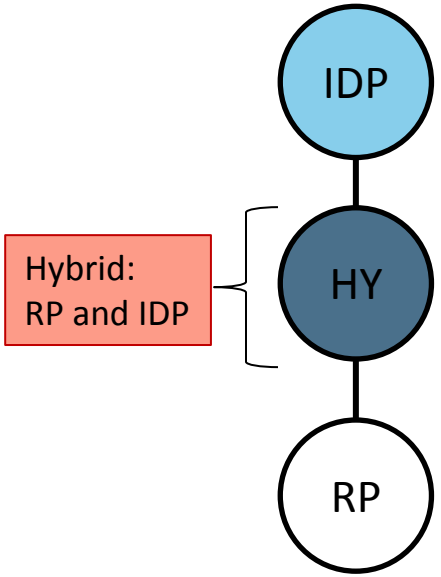
High-degree IDP case

- IDP having many RPs
- Top IDPs



High-degree RP case

- RP having many IDPs
- Specialized IDPs

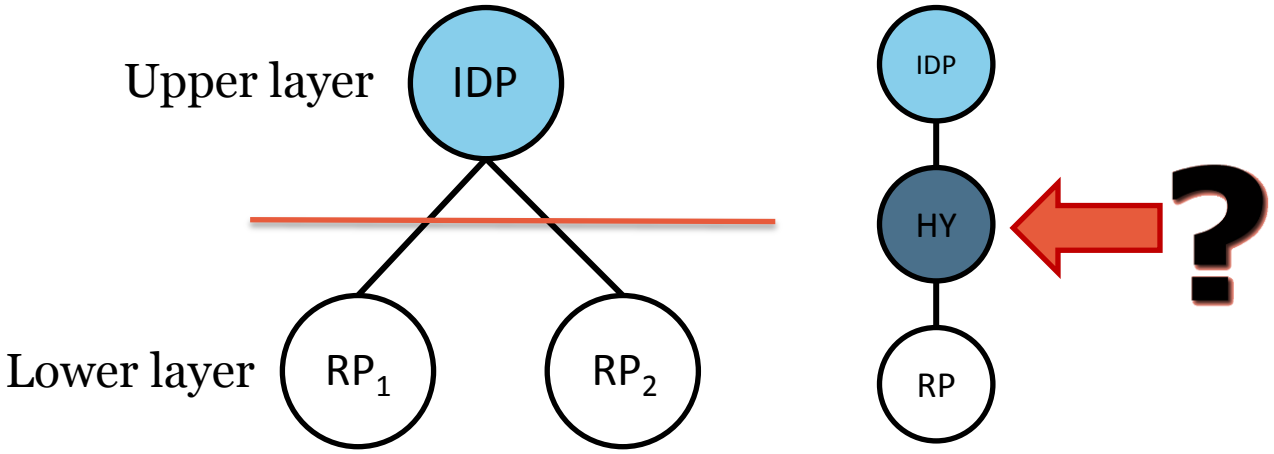


Hybrid case

- Hybrids are both RP and IDP

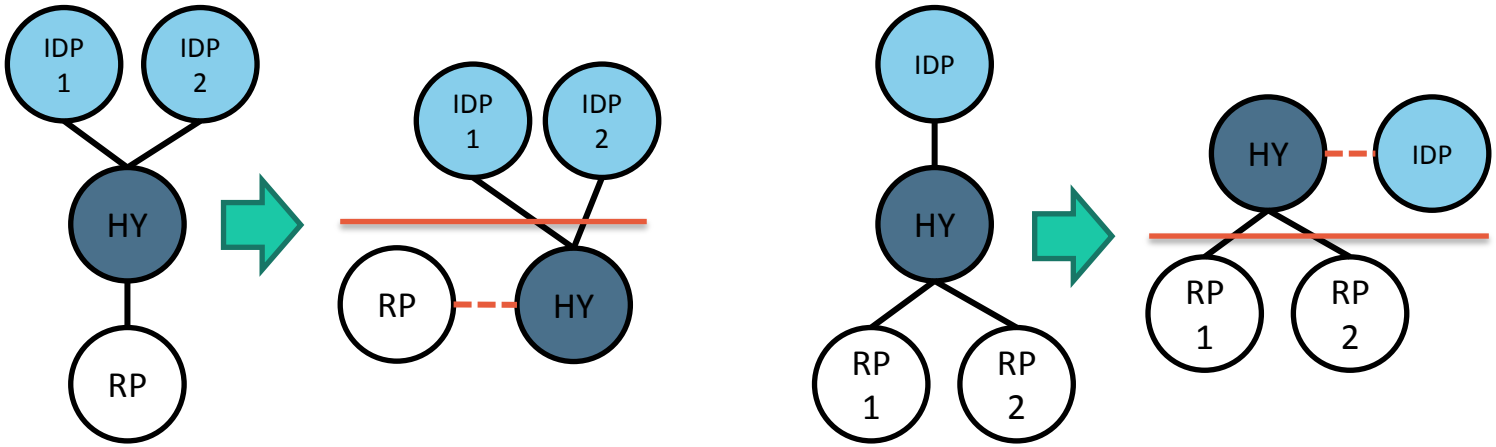
Structural Model

- We have modeled the landscape as a bipartite graph
  - Mainly high-degree IDP structures



Structural Model

Place HY nodes in layers, based on their main feature

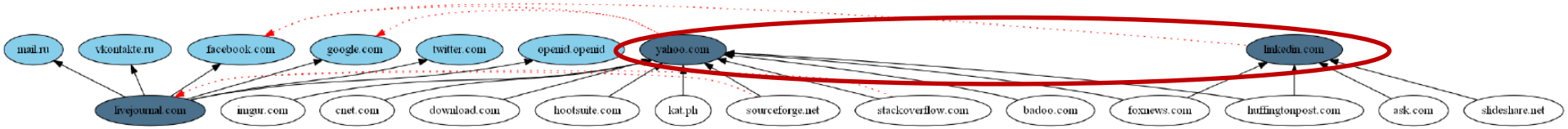


Structural Changes

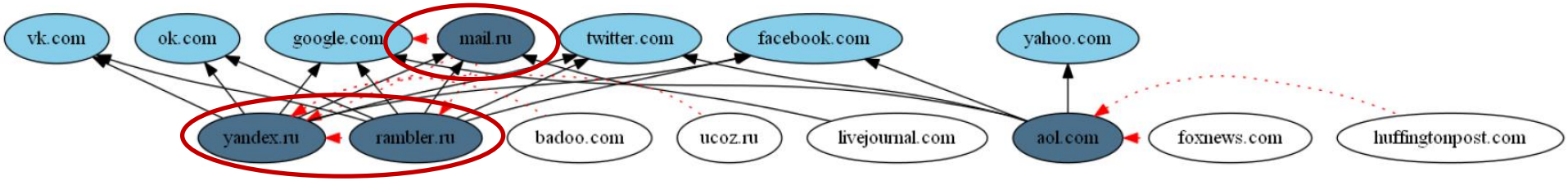
- Three stages of the landscape:
  1. Adding many IDPs (trying out new technology)
  2. Nested landscape with many hybrids
  3. Simplified landscape

- Regional and language-based differences:
  - English/US Web: Stage 3 with few IDPs
  - Chinese Web: Stage 3, still with many hybrids
  - Russian Web: Entering stage 2!

Example: Structural Changes



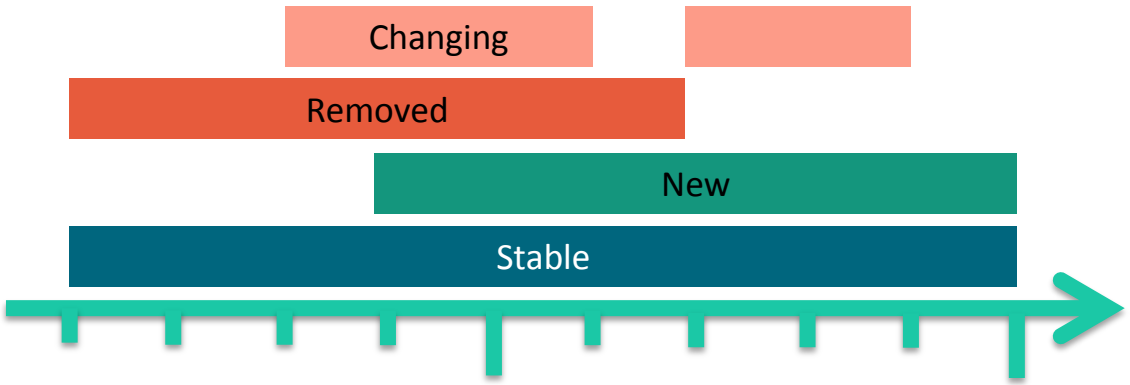
Non-Chinese Web April 2012: IDP-like hybrids (few)



Non-Chinese Web April 2015: Emerging Russian HY-structures

Relationship Types

- Relationship types:
  - **Stable:** Kept by the RP, during all 10 snapshots
  - **New:** Added after the first snapshot
  - **Removed:** Observed in the 1st snapshot and later removed
  - **Changing:** Added and removed one of more times



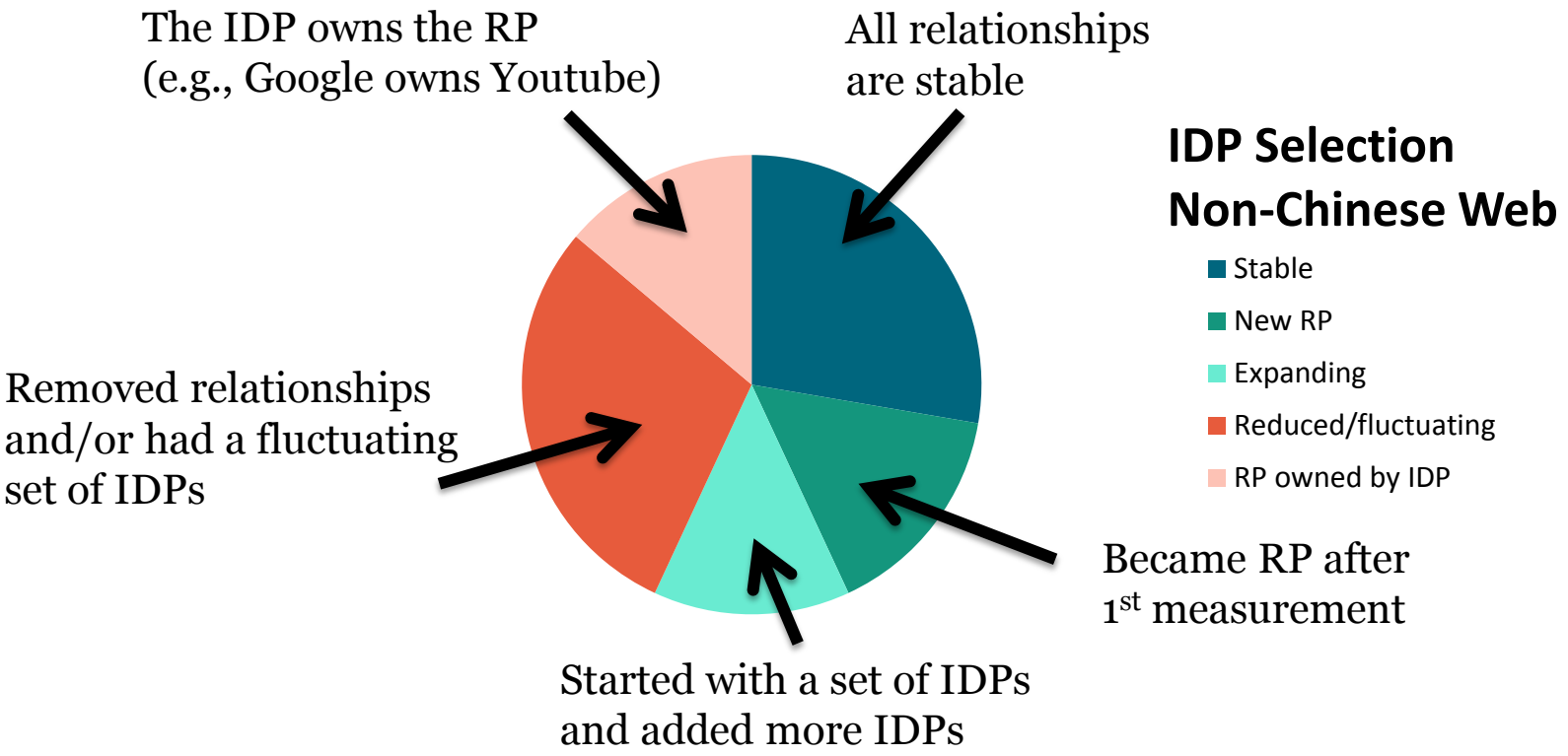
Protocol Usage per Relationship Type

Protocol	Total	Stable	New	Removed	Changed
OAuth	140	46%	33%	10%	11%
OAuth* China	102	25%	28%	15%	31%
OpenID	40	5%	15%	68%	13%
OpenID to OAuth	7	86%	0%	0%	14%
Internal/unknown	14	71%	7%	0%	21%

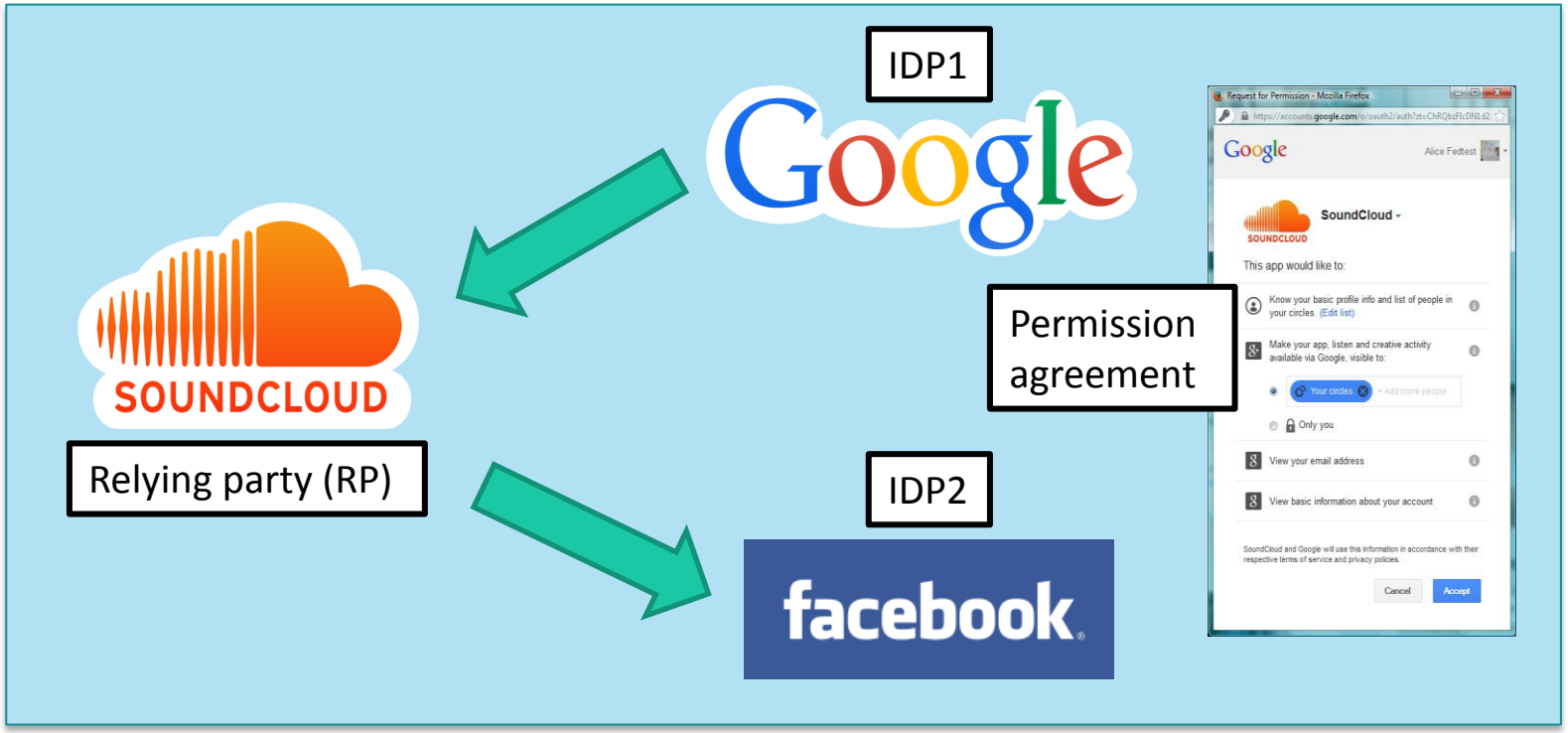
OAuth protocol: Less privacy preserving than OpenID!

* Parts of the Chinese OAuth relationships may be internal

RP Behavior

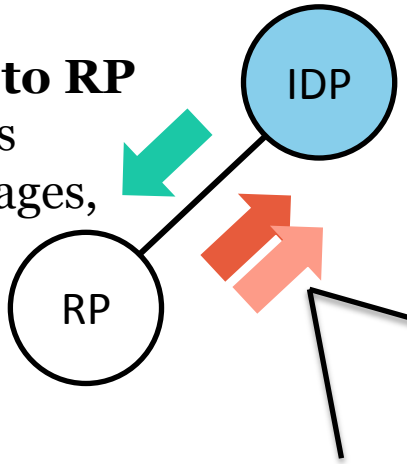

The IDP owns the RP
(e.g., Google owns Youtube)
All relationships
are stable
Removed relationships
and/or had a fluctuating
set of IDPs
Became RP after
1st measurement
Started with a set of IDPs
and added more IDPs
IDP Selection
Non-Chinese Web
Stable
New RP
Expanding
Reduced/fluctuating
RP owned by IDP

Information Sharing Between RP and IDPs



Types of Information Flows

READ:
Data read from IDP to RP
Rich user data, contents created by the user (images, videos, “likes” etc).



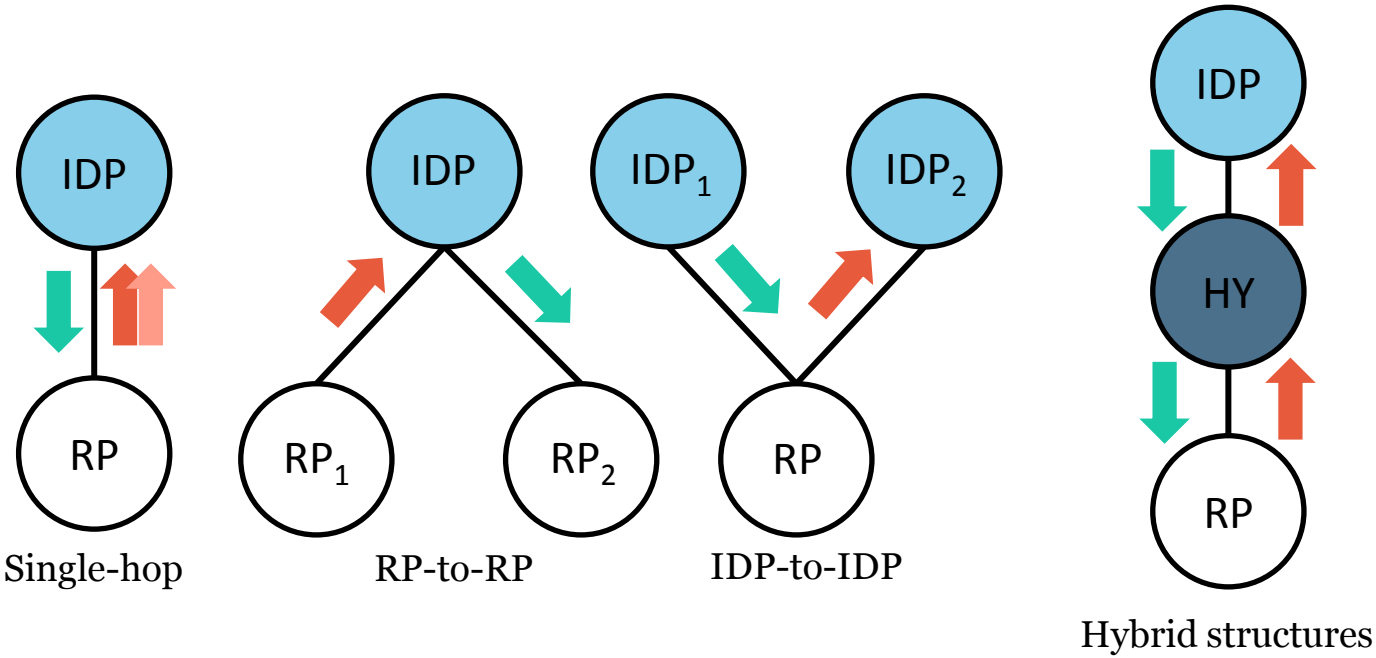
RP acts on behalf of the user on the IDP

WRITE:
Data posted by RP on IDP
Notifications, or created contents

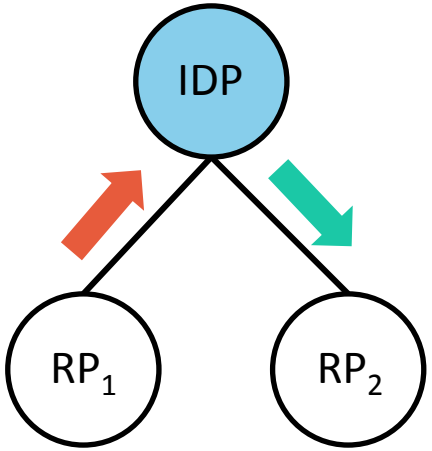
UPDATE/REMOVE:
Other actions taken on the IDP
The RP can add the user to groups and modify the user’s IDP account

Potential Information Leaks

- **Single-hop data transfer:** RP to IDP (or IDP to RP)
- **Multi-hop leak:** Indirect leak via proxy node(s)



RP-to-RP Leakage Example



RP-to-RP

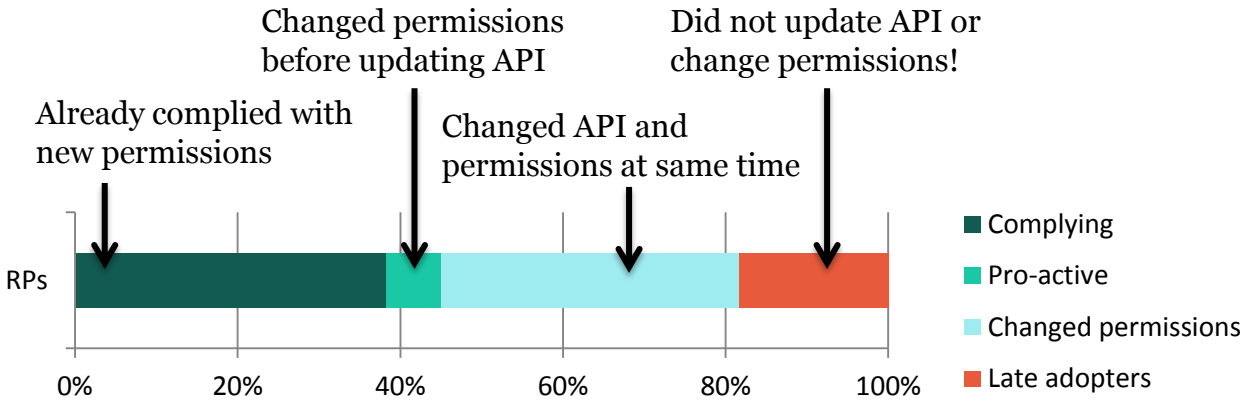
RP-to-RP leaks	February 2014		April 2015	
IDP	**All**	**Severe**	**All**	**Severe**
Facebook	645	150	473	66
Twitter	110	110	110	110
Google	91	0	91	0

Dataset with 44 RPs using Facebook, 14 using Twitter and 12 using Google

- Potential RP-to-RP leaks
  - Information written/posted from RP1 to IDP
  - Information read from IDP to RP2
  - Leak only possible with Write(RP1-IDP) + Read(IDP-RP2)

Facebook Use-case

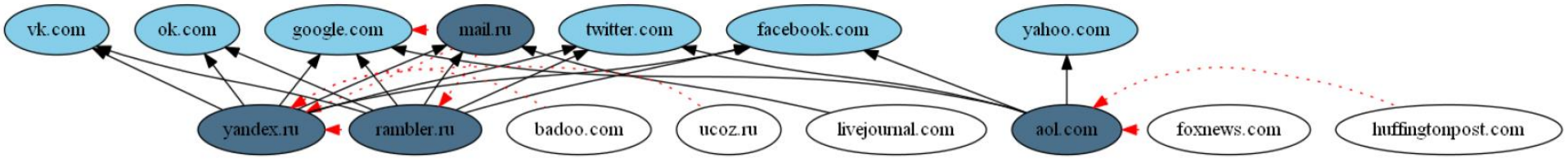
- Facebook API changes in 2015 to strengthen privacy
  - Most RPs needed to change to more privacy-preserving data sharing permissions to comply
  - Four measurements: Sept. 14 – May 2015
  - 63 top-200 RPs using Facebook as their IDP



Contributions and Findings

- Showed that the RP-IDP landscape can be modeled as a bipartite graph
  - Designed a model for RP-IDP structures
  - Identified structural changes over time
- Protocol- and IDP selections made by RPs
  - A few popular IDPs increasingly used
  - More data sharing – less user privacy
- Identified privacy leakage risks
  - Multi-hop, enabled by the structures

Longitudinal Analysis of the Third-party Authentication Landscape



Anna Vapen, Niklas Carlsson, Nahid Shahmehri

anna.vapen@liu.se

LINKÖPING UNIVERSITY