Third-party Identity Management Usage on the Web

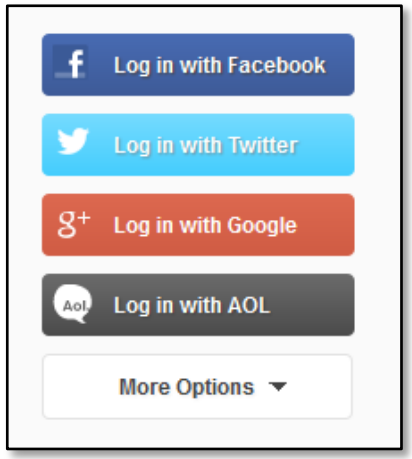
Anna Vapen[1], Niklas Carlsson[1], Anirban Mahanti[2], Nahid Shahmehri[1]

[1]Linköping University, Sweden

[2]NICTA, Australia

LiU EXPANDING REALITY

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
- Increase personalization opportunities
- Share information between websites

Motivation

- An emerging third-party authentication landscape
  - Increasing usage of third-party identity providers
  - Complex, nested relationships between RPs and IDPs
- Authorization protocol (OAuth) used for authentication
  - Applications acting on user’s behalf
  - Data transfer between parties; Less control over data
- IDP selection
  - Privacy implications

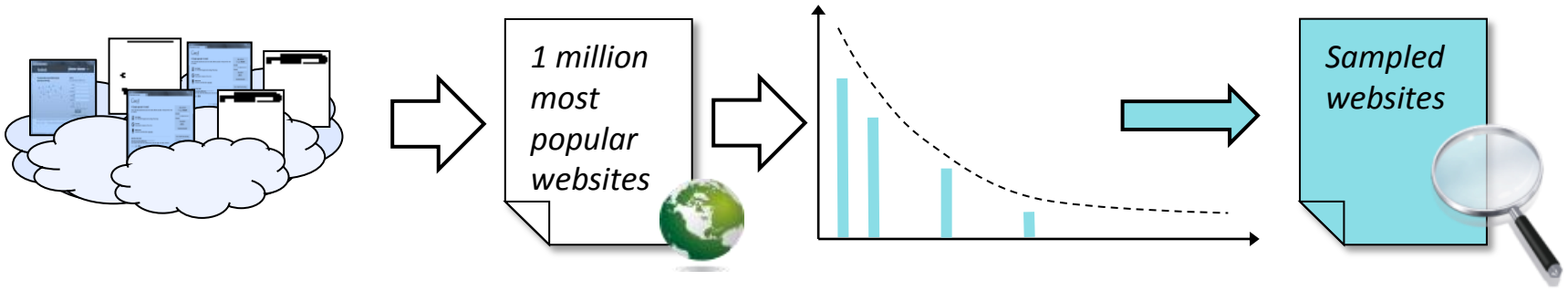
Contributions

- Novel Selenium-based data collection methodology
  - Identification and validation of RP-IDP relationships
  - Popularity-based logarithmic sampling technique

- Characterization of identified RP-IDP relationships
  - Impact on IDP selection of RP characteristics
  - Comparison to third-party content-delivery relationships

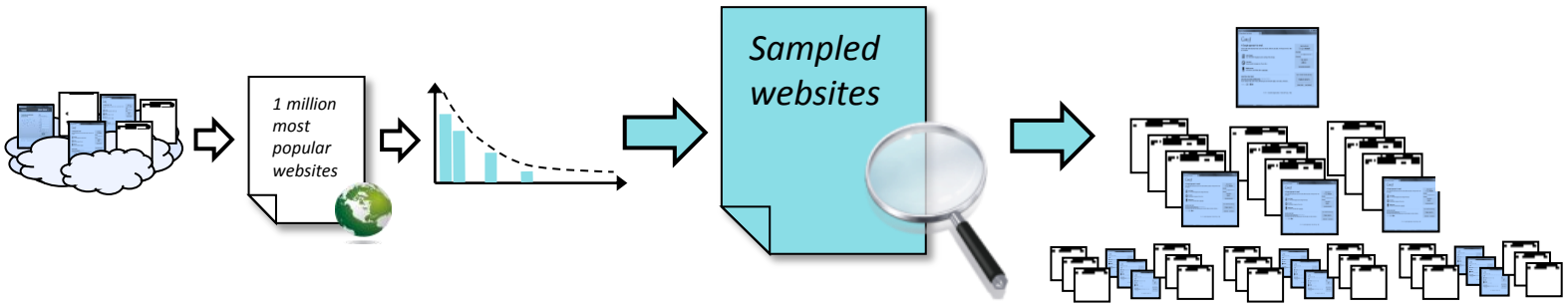
Methodology (1)

- Popularity-based logarithmic sampling
  - 80,000 points uniformly on a logarithmic range
  - Power-law distribution
  - Capturing data from different popularity segments



Methodology (2)

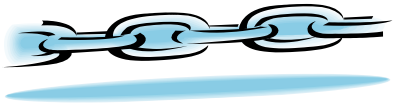
- Selenium-based crawling and relationship identification
  - Able to process Web 2.0 sites with interactive elements
  - Low number of false positives
  - Validation with semi-manual classification and text-matching

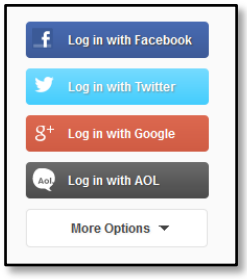


Collected Data

1,6 terabyte analyzed data

25 million analyzed links





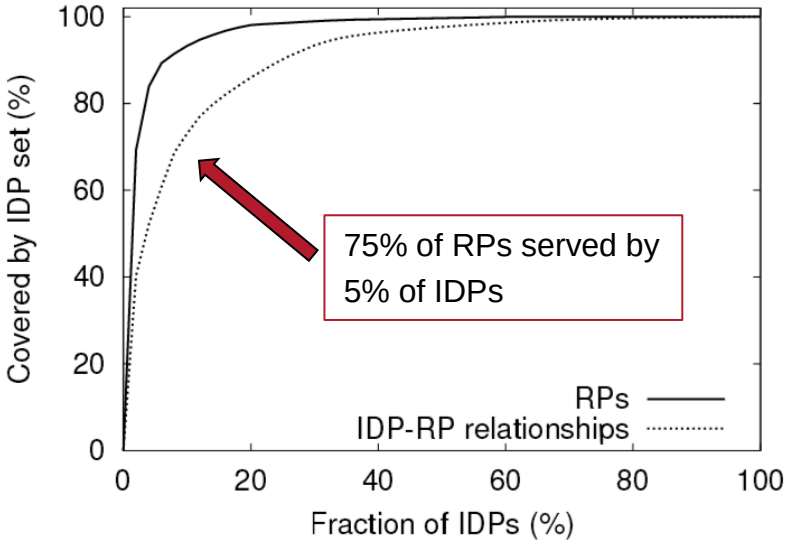
35,620 sampled sites
3,329 unique relationships
50 IDPs and 1,865 RPs

WHOIS, server location and audience location

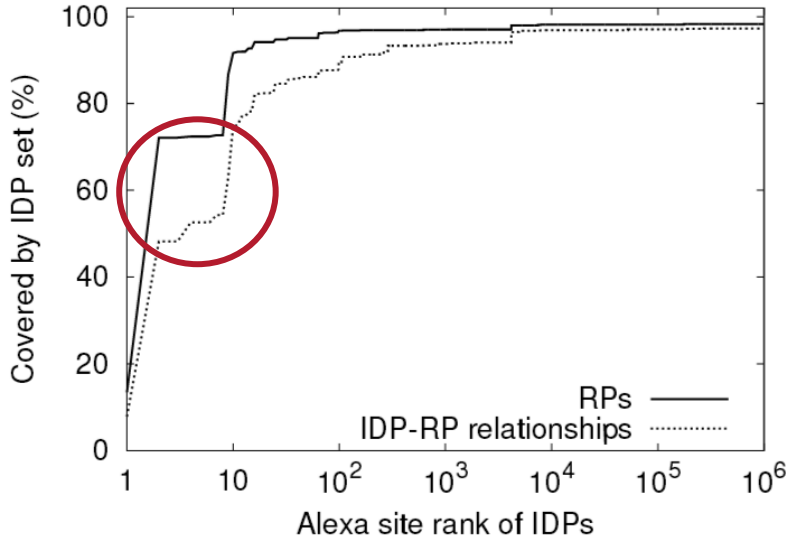
Total site size and number of links and objects

IDP Usage

- More than 75% of the RPs are served by 5% of the IDPs
- RPs tend to select popular sites as IDPs
  - Only 15 of the 44 IDPs outside top 10 on Alexa serve more than 10 sampled RPs



(a) IDP popularity



(b) Alexa rank

Top IDPs

IDP rank	Alexa rank	IDP	Protocol	Number of IDP relationships
1	2	Facebook.com	Oauth	1293
2	10	Twitter.com	OAuth	378
3	9	QQ.com	OAuth	278
4	1	Google.com	Oauth / OpenID	250
5	4	Yahoo.com	Oauth / OpenID	141
6	16 **	Sina.com.cn	Oauth	127
7	-	OpenID field	OpenID	87
8	4173 *	Vkontakte.ru	Oauth	73
9	25 **	Weibo.com	Oauth	64
10	12	Linkedin.com	Oauth	63

Login with any OpenID provider

** Domain change to vk.com*

*** Authentication with Sina.com.cn redirects to Weibo.com*

Top IDPs

IDP rank	Alexa rank	IDP	Protocol	Number of IDP relationships
1	2	Facebook.com	Oauth	1293
2	10	Twitter.com	OAuth	378
3	9	QQ.com	OAuth	278
4	1	Google.com	Oauth / OpenID	250
5	4	Yahoo.com	Oauth / OpenID	141
6	16 **	Sina.com.cn	Oauth	127
7	-	OpenID field	OpenID	87
8	4173 *	Vkontakte.ru	Oauth	73
9	25 **	Weibo.com	Oauth	64
10	12	Linkedin.com	Oauth	63

Social networks (except no. 7)

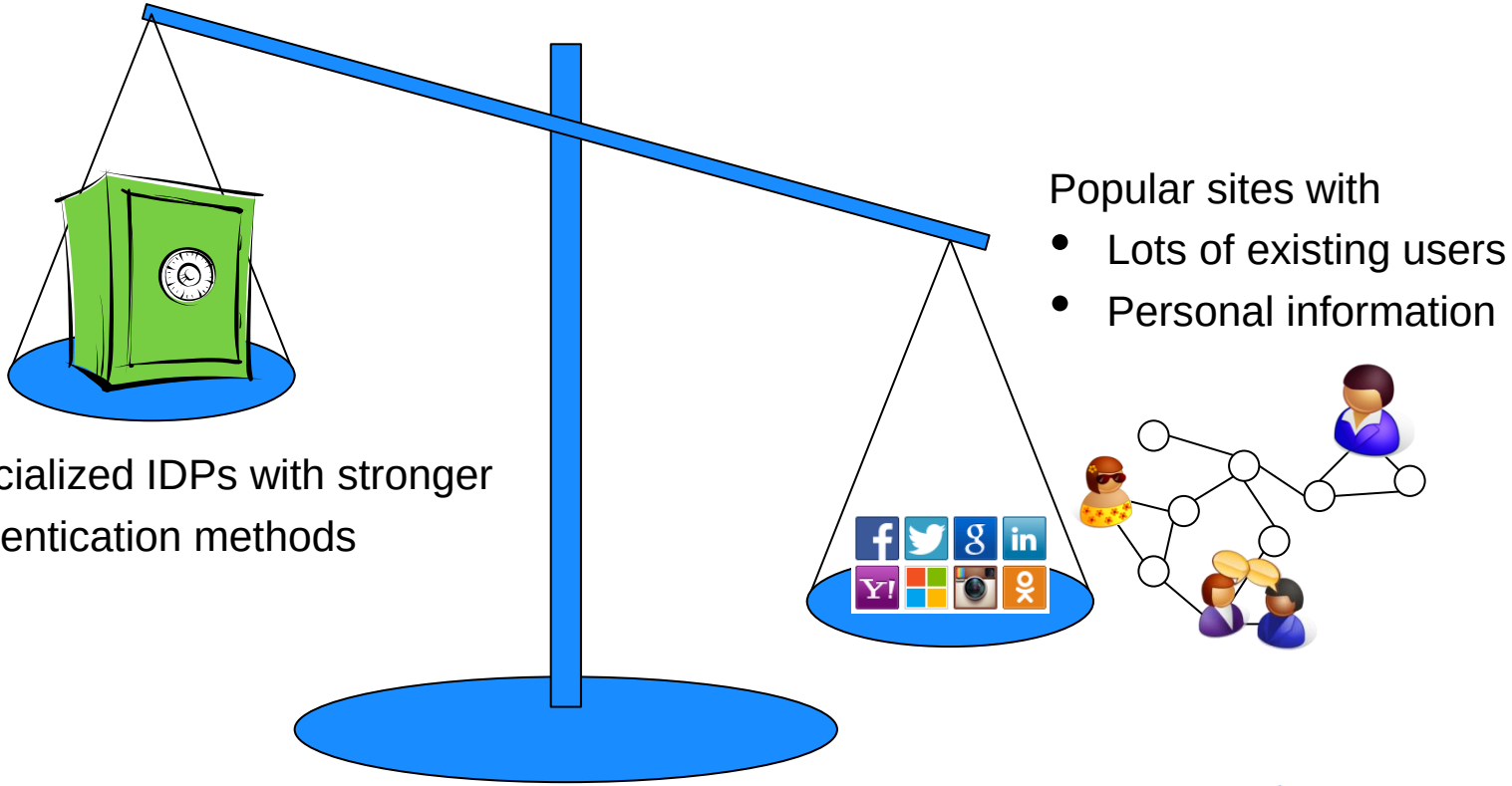
Login with any OpenID provider

** Domain change to vk.com*

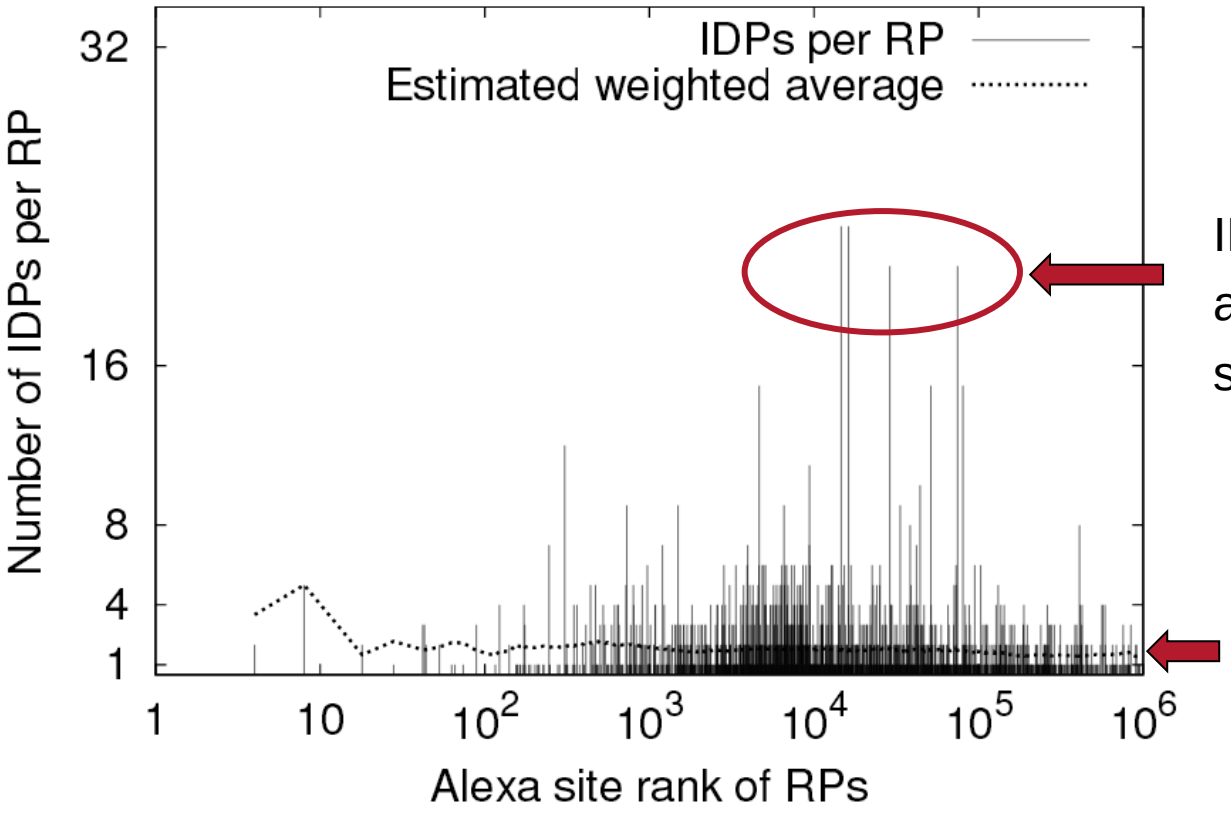
*** Authentication with Sina.com.cn redirects to Weibo.com*

IDP Selection

- Popular sites as IDPs, instead of specialized IDPs



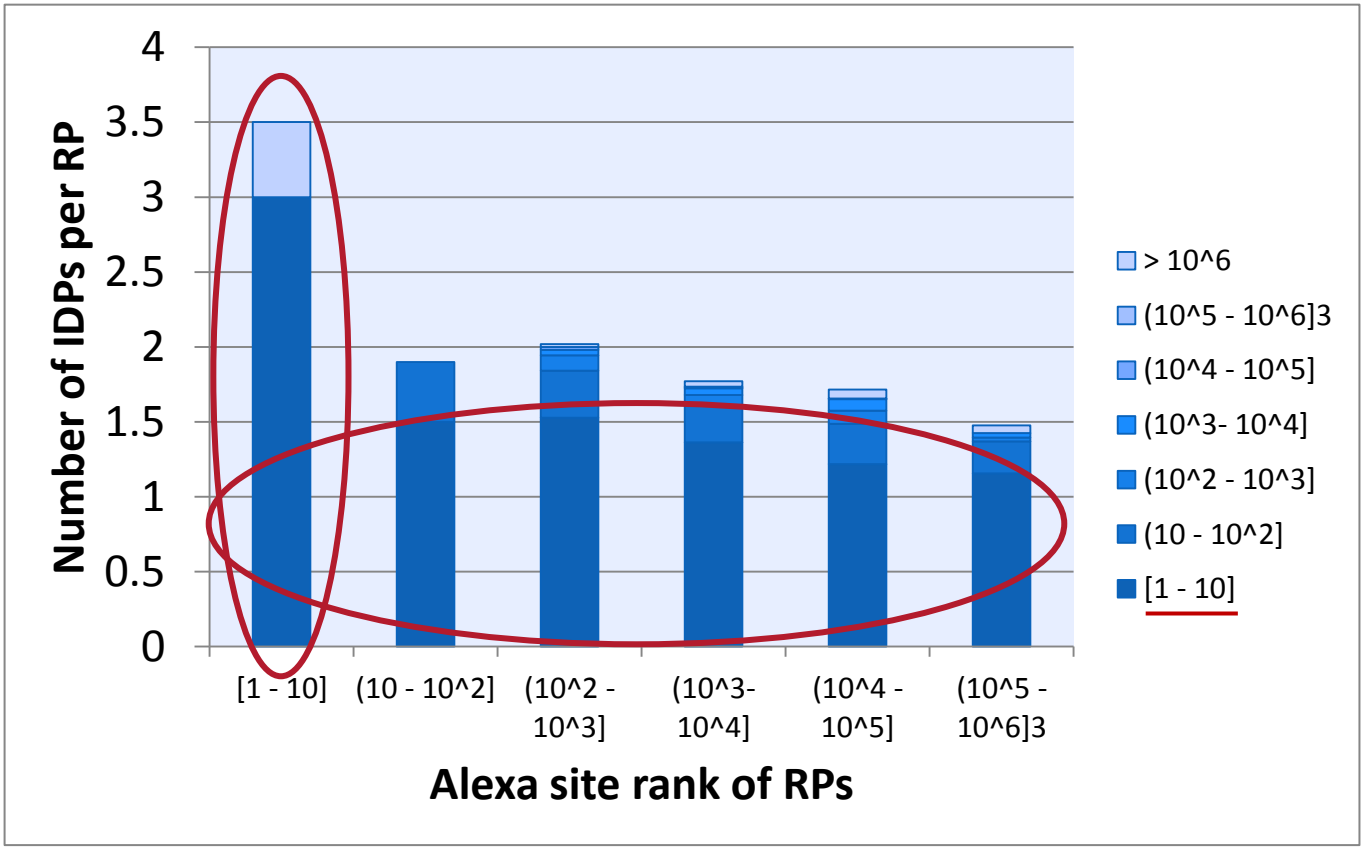
Number of IDPs per sampled RP



IDP widgets providing a pre-selected large set of IDPs

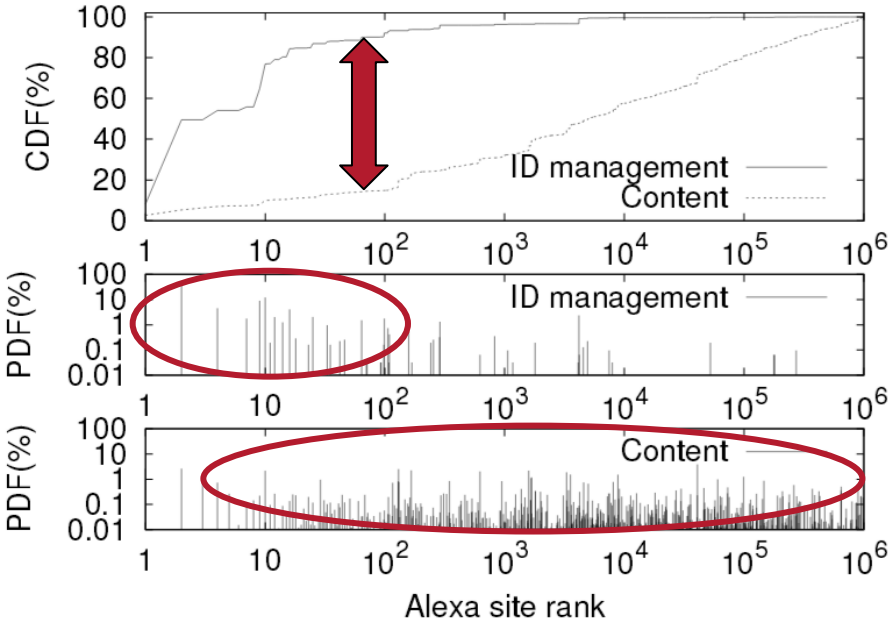
Estimated weighted average: < 3 IDPs / RP

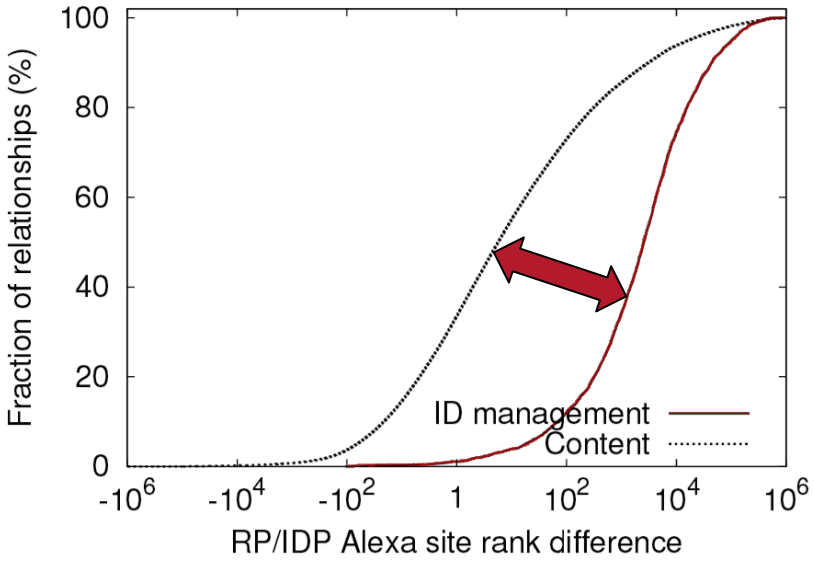
IDPs per RP Based on Popularity



Breakdown of the average number of IDPs selected per RP and popularity segment

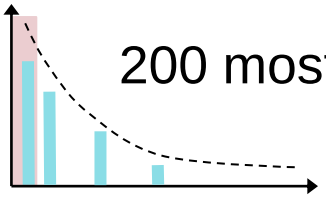
Comparison with Content Services





- Content: scripts, images and other third-party objects
- IDPs much more popular sites than content providers

Service-based Analysis



200 most popular websites

Manual classification

Likely to be RPs

News

File sharing

Info

Using social IDPs: file sharing, info

Social/portal

Likely to be IDPs; Many RPs in this category

Video

Tech

Early adopters, using several IDPs

Commerce

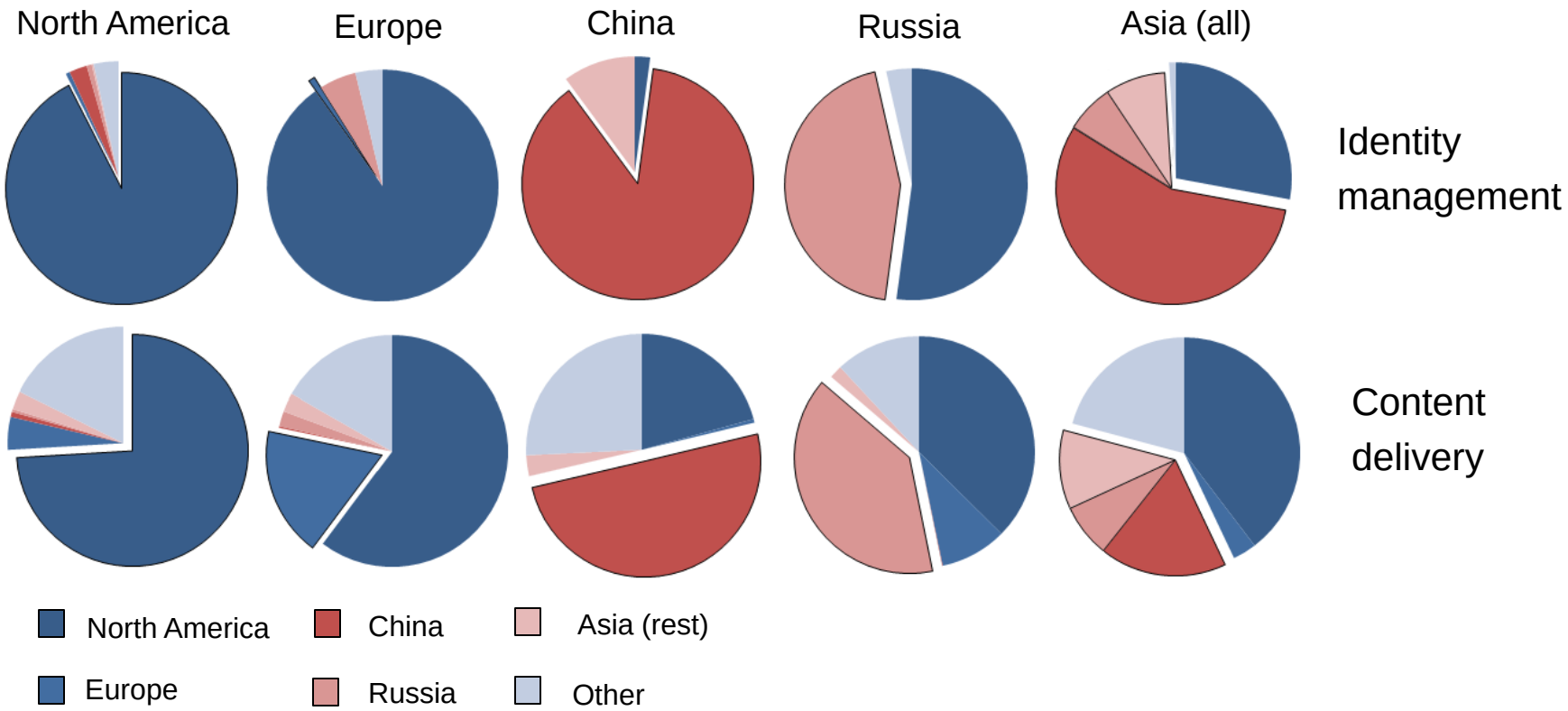
Ads

CDN

Using IDPs from same category: tech, commerce

Cultural and Geographical Analysis

- North American and Chinese RPs use local IDPs to a large extent
- Content delivery usage less biased to local providers



Summary and Conclusions

- Large-scale characterization of third-party Web authentication
- Novel data collection methodology with popularity-based sampling
- Few large third-parties serve many websites
  - Comparison with content sharing
  - IDP selection much more biased
- Risk for privacy leaks
  - Few large third-parties handling a lot of information
  - The most popular IDPs are using protocols not adapted for strong authentication