Carnegie Mellon University

Formal Analysis of Privacy Requirements Specifications for Multi-tier Applications

21st IEEE International Requirements Engineering Conference

Travis Breaux, Ashwini Rao

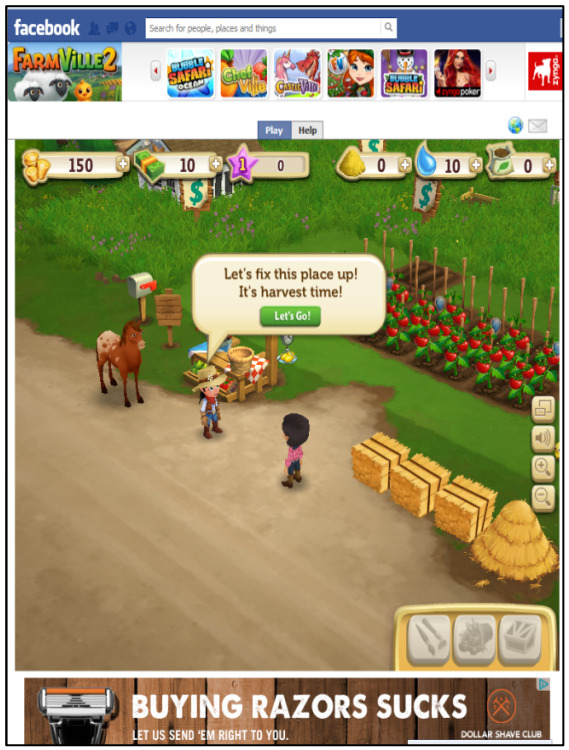
Carnegie Mellon University

July 17, 2013

institute for SOFTWARE RESEARCH

Privacy in multi-tier applications

What the user sees?



What the policy says?

Facebook: *You will not directly or indirectly transfer any data you receive from us to any ad network, even if a user consents to such transfer*

Zynga: *We do not actively share personal information with third party advertisers for their direct marketing purposes unless you give us your consent*

AOL Advertising *uses the information collected on Network Participating Sites to better target advertisements to people across different websites*

Who should read the policy?

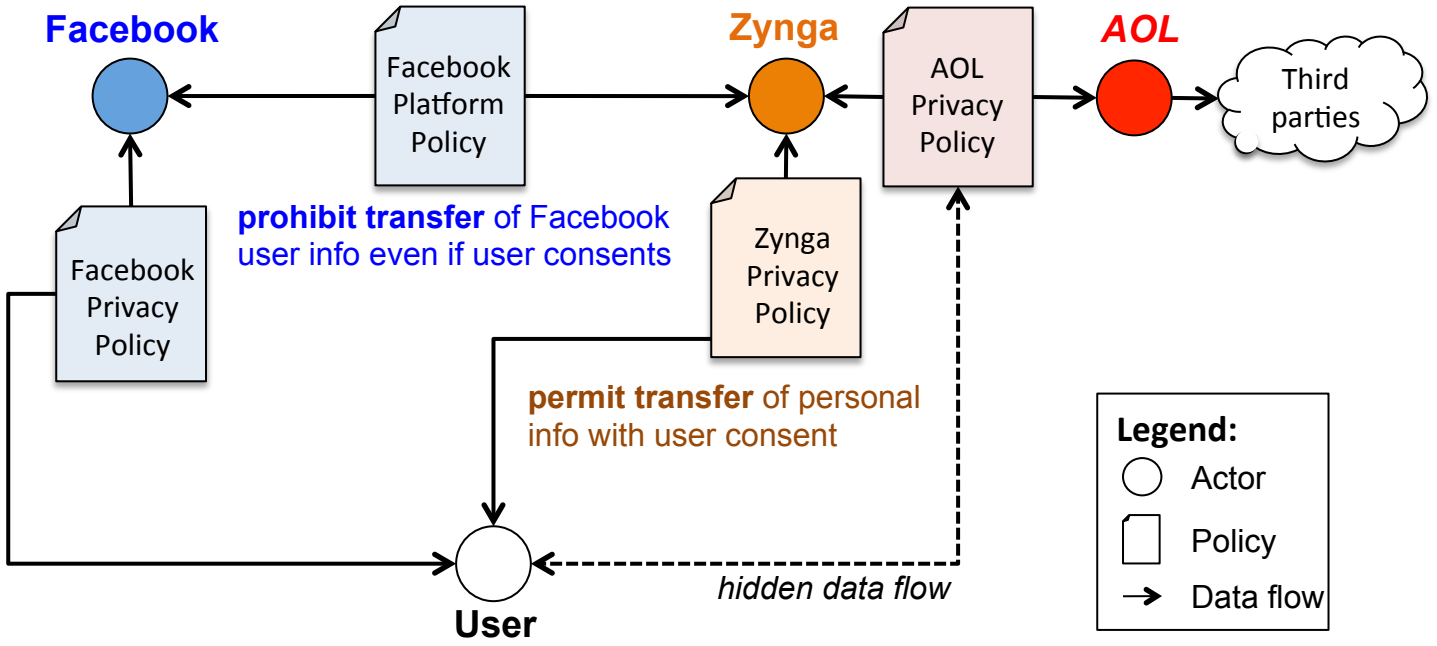
App Developer

App User

Everyone

Key: <--- Data flow } Content owner

Privacy and data supply chain



Privacy policies contain privacy requirements for data that flow within a data supply chain; conflicts can exist among these requirements; repurposing can be an issue

Requirements specification language

Discover a *privacy RSL* to…

- Express a critical subset of privacy policy statements (requirements) in formal logic
- Reason about interactions between policy statements, such as conflicts and repurposing
- Enable verification across different policies in a data supply chain

Approach and research method

- Exploratory case study design [Yin08]
  - Data: Facebook Platform Policy (for developers)
  - Developed specification language from results
- Extended evaluation
  - Data: Zynga privacy policy, AOL privacy policy
- Applied content analysis [Sal13] to extract phrases to formalize data requirements in logic

R. Yin, *Case Study Research: Design and Methods*, 4th ed. SAGE, 2008.
J. Saldaña, *The Coding Manual for Qualitative Researchers, 2nd ed.* SAGE, 2013

Mapping policy statements to types

- ***Policy Statements*** *describe events or states outside the app*

  "You must not violate any law or the rights of any individual or entity."

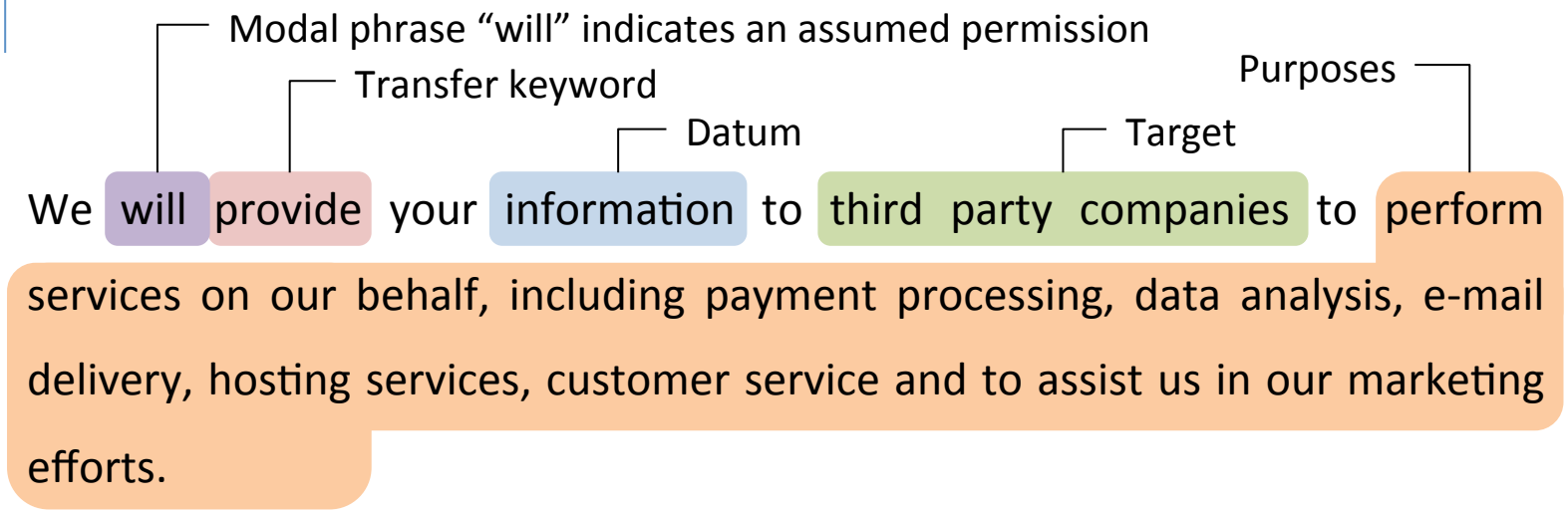
- **Non-data Requirements** describe non-data functionalities

  "You will include your privacy policy URL in the App Dashboard."

- **Data Requirements** describe actions on data

  "You must not include functionality that proxies, requests or collects Facebook usernames or passwords."

Step 1: Manually annotate policy text



Step 2: Write expression in specification language

P TRANSFER information TO third-party-companies FOR performing-services

Using context in annotation

- [Zynga] *"may access and store some or all of the following information, as allowed by you, the SNS and your preferences"*

  Action is COLLECT

- [AOL] *"Personal information such as name, address and phone number is never accessed for this purpose"*

  Action is USE

- [AOL] *"In that the case, the acquiring (or merging) company will have access to your information"*

  Action is TRANSFER

Specifying privacy requirements

- Expressing Modality in Description Logic (DL)
  - Obligation $\sqsubseteq$ Permission
  - *Conflict* $\equiv$ *Permission* $\sqcap$ *Prohibition*
- Actions
  - Collect, Use and Transfer
- Actions have following DL Roles
  - hasObject.Datum – the object of the action (data element)
  - hasSource.Actor – the source of the object (an actor)
  - hasPurpose.Purpose – the purpose of the action
  - hasTarget.Actor – the recipient of the object (an actor)

T. Breaux, A. Antón, J. Doyle. "Semantic Parameterization: A Process for Modeling Domain Descriptions." ACM TOSEM, 18(2): 5, November 2008

Expressing role values in hierarchies

Datum

- information
  - public-information
    - zynga-user-id
    - user-name
    - …
  - *personal-information*
    - *billing-information*
    - *user-age*
    - *…*
  - technical-information
    - ip-address

…

Purpose

- payment-processing
- communicating-with-user
  - notifying-game-activity
  - *customer-support*
    - *technical-support*
    - *…*
- *delivering-advertisement*
  - *marketing-zynga*
  - *marketing-third-party*
  - *target-advertising*

…

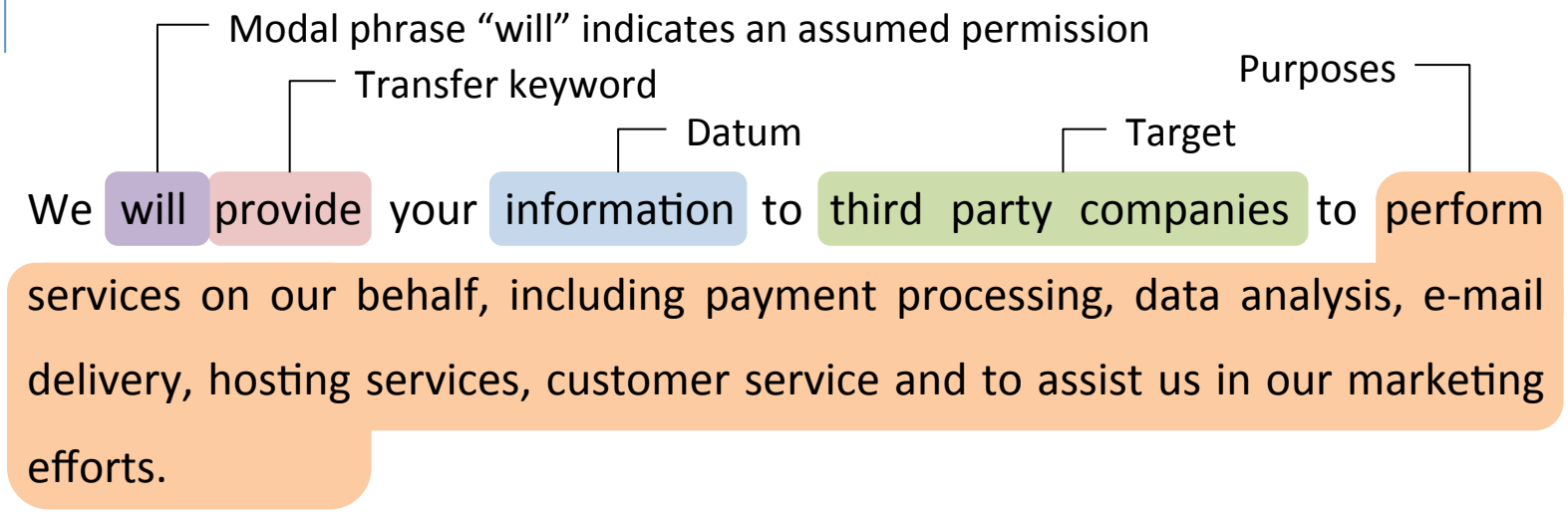
Actor

- zynga
  - zynga-inc
  - affiliate
    - subsidiary
    - joint-venture
    - …
- service-provider
  - google-analytics
- third-party-advertiser
- user

…

Example of a DL concept hierarchy from Zynga privacy policy. Inner bullet concepts are subsumed by (contained within) outer bullet concepts.

Step 1: Manually annotate policy text



Step 2: Write expression in specification language

P TRANSFER information TO third-party-companies FOR performing-services

Step 3: Compile language into Description Logic

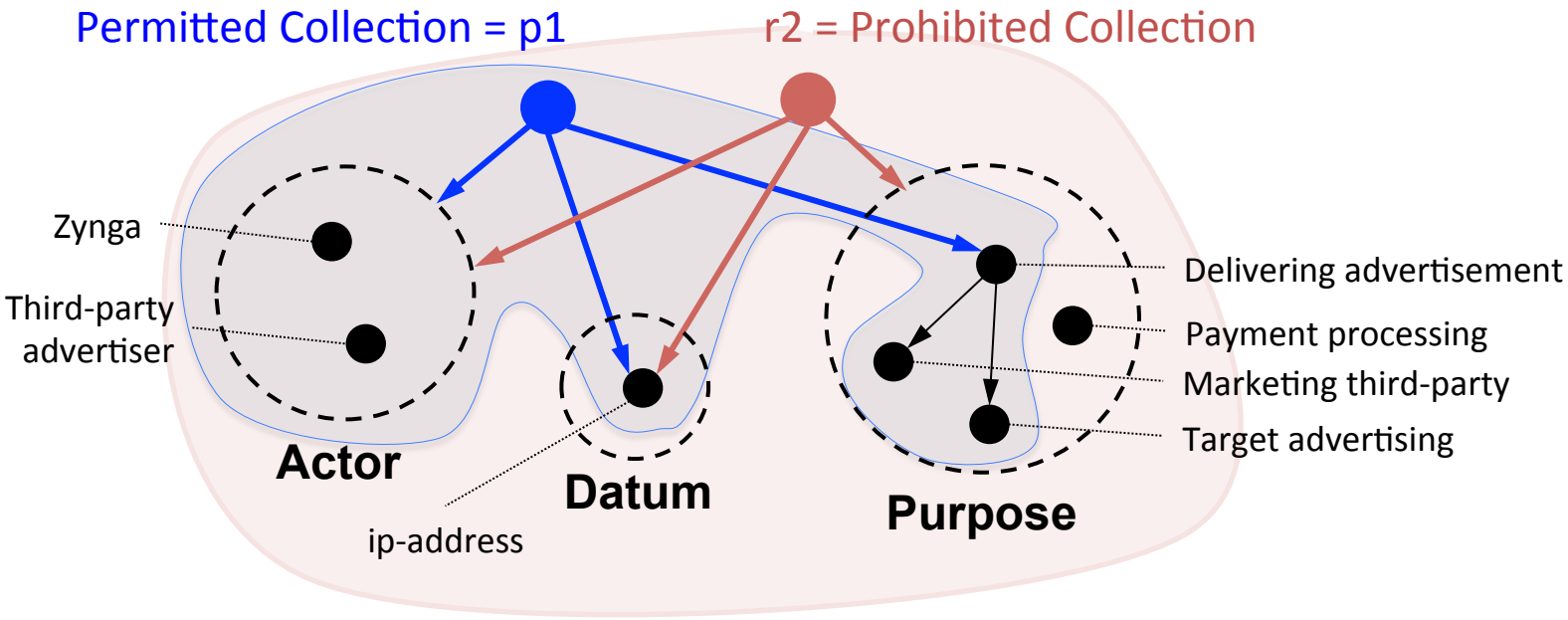
$p_2 \equiv \text{TRANSFER} \sqcap \exists \text{hasObject.information} \sqcap$
$\exists \text{hasTarget.third-party-companies} \sqcap \exists \text{hasPurpose.performing-services}$

$p_2 \sqsubseteq \text{Permission}$

How we identify conflicts – 1

p1: Permitted to collect IP address from anyone for advertising

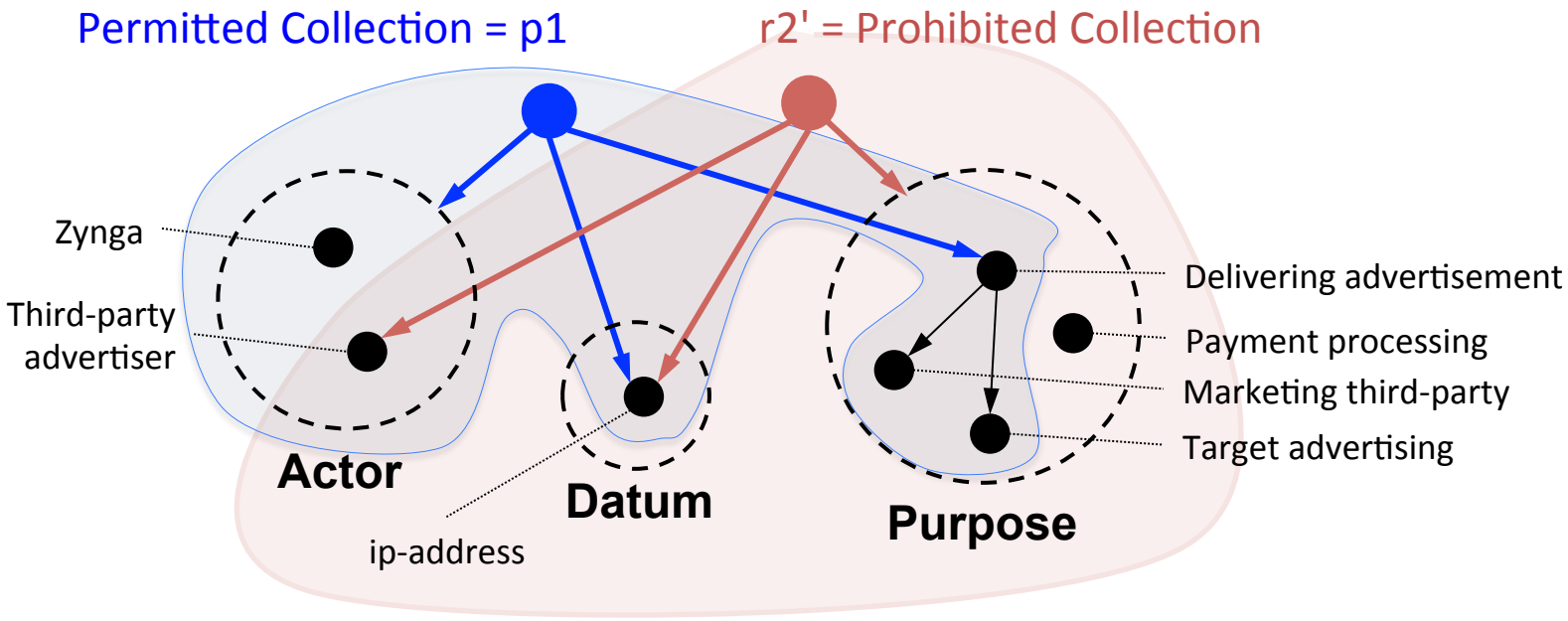
r2: Prohibited from collecting IP address from anyone for anything



How we identify conflicts – 2

p1: Permitted to collect IP address from anyone for advertising

r2': Prohibited from collecting IP address from third-party advertisers for anything



How we trace data

- Characterizing data flows using subsumption
  - *Underflow*, occurs when the data target subsumes the source
  - *Overflow*, occurs when the data source subsumes the target
  - *Exact flow*, occurs when the data source and target are equivalent
  - Identify repurposing, visualize dependencies etc.

AOL-16: Collect name, contact information, payment method from site visitor for business purposes

AOL-48: Transfer personally identifiable information to key partners

$$contact\_info \sqsubseteq personally\_identifiable\_info$$
$$business\_purposes \sqsubseteq anything$$

RESULTS OF CASE STUDY

Results of extended evaluation

Policy	S	D	Modality			Action		
			P	O	R	C	U	T
Facebook	105	39	15	4	25	6	15	14
Zynga	195	64	58	1	8	22	8	15
AOL	74	41	43	0	4	12	15	10

Extracted: (S)tatements, (D)ata requirements

Modalities: (P)ermission, (O)bligation, (R) prohibition

Actions: (C)ollection, (U)se), (T)ransfer

Results of extended evaluation

Policy	S	D	Modality			Action		
			P	O	R	C	U	T
Facebook	105	39	15	4	25	6	15	14
Zynga	195	64	58	1	8	22	8	15
AOL	74	41	43	0	4	12	15	10

Extracted: (S)tatements, (D)ata requirements

Modalities: (P)ermission, (O)bligation, (R) prohibition

Actions: (C)ollection, (U)se), (T)ransfer

Results of extended evaluation

Policy	S	D	Modality			Action		
			P	O	R	C	U	T
Facebook	105	39	15	4	25	6	15	14
Zynga	195	64	58	1	8	22	8	15
AOL	74	41	43	0	4	12	15	10

Extracted: (S)tatements, (D)ata requirements

Modalities: (P)ermission, (O)bligation, (R) prohibition

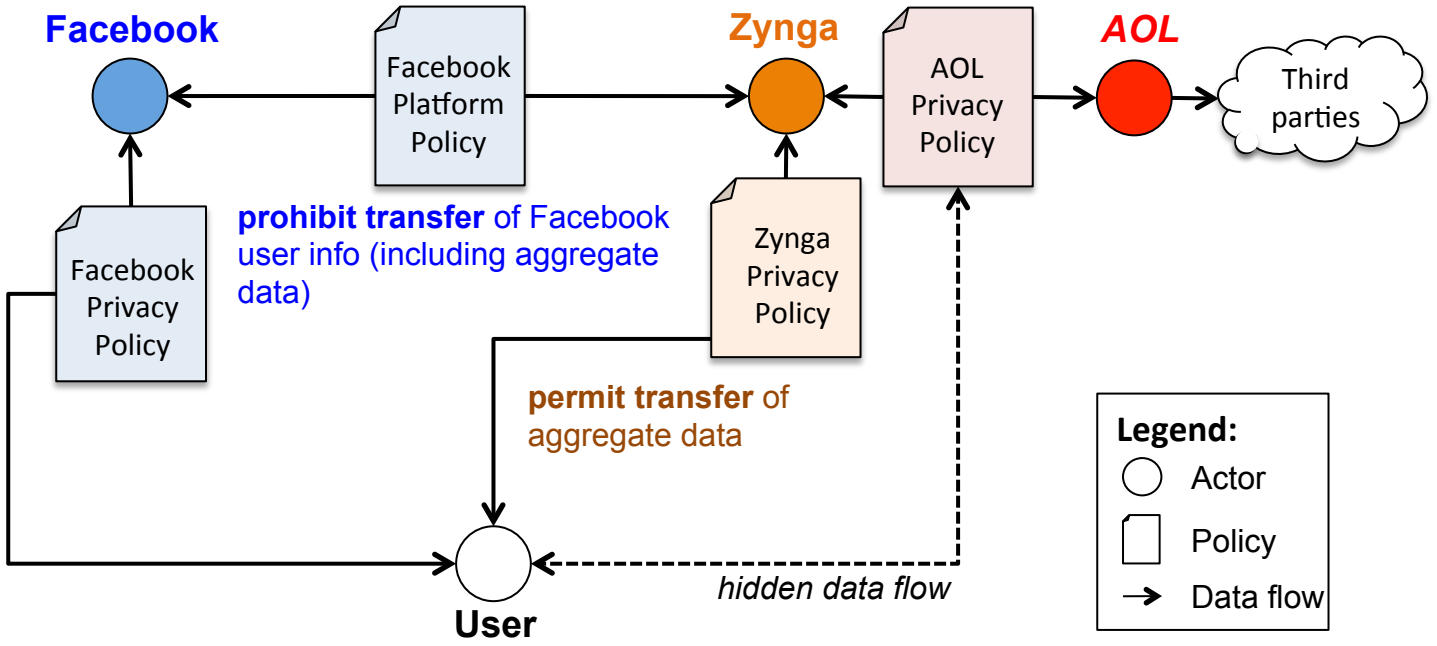
Actions: (C)ollection, (U)se), (T)ransfer

Phrase heuristics used in mapping

Action keywords indicate when a statement was coded as a collection, use or transfer requirement

DL Action	Action keywords
COLLECT	Access, assign, collect, collected, collection, collects, give you, import, keep, observes, provide, receive, record, request, share, use
USE	Access, accessed, communicate, delivering, include, matches, send, use, used, uses, using, utilized
TRANSFER	Access, disclose, disclosed, disclosure, give, in partnership with, include, make public, on behalf of, provide, see, share, shared, transfer, use, used with, utilized by

Identifying conflicting requirements



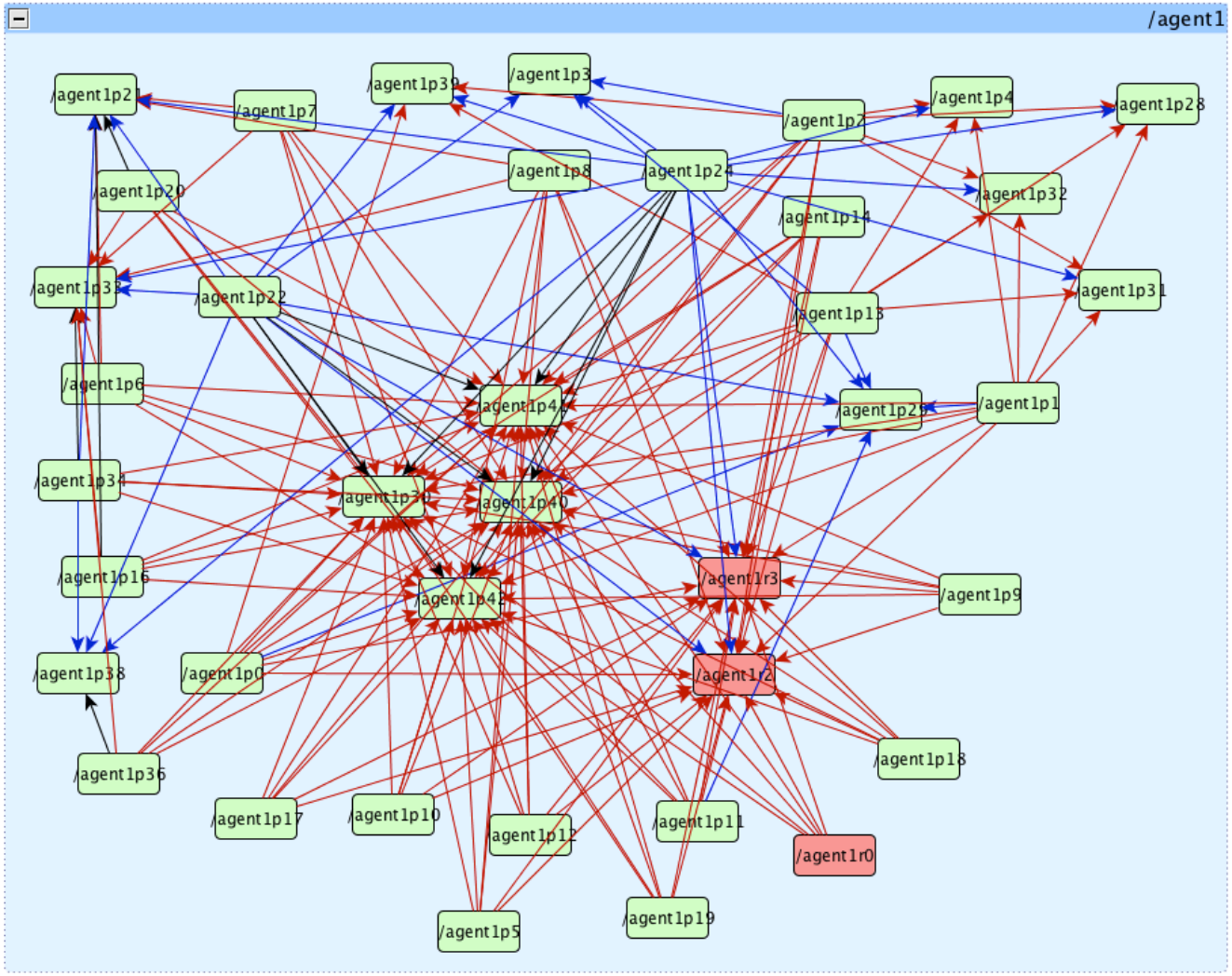
In a multi-tier application, conflicts can exist between privacy requirements in policies governing data flow in a data supply chain

Conflicts identified in our study

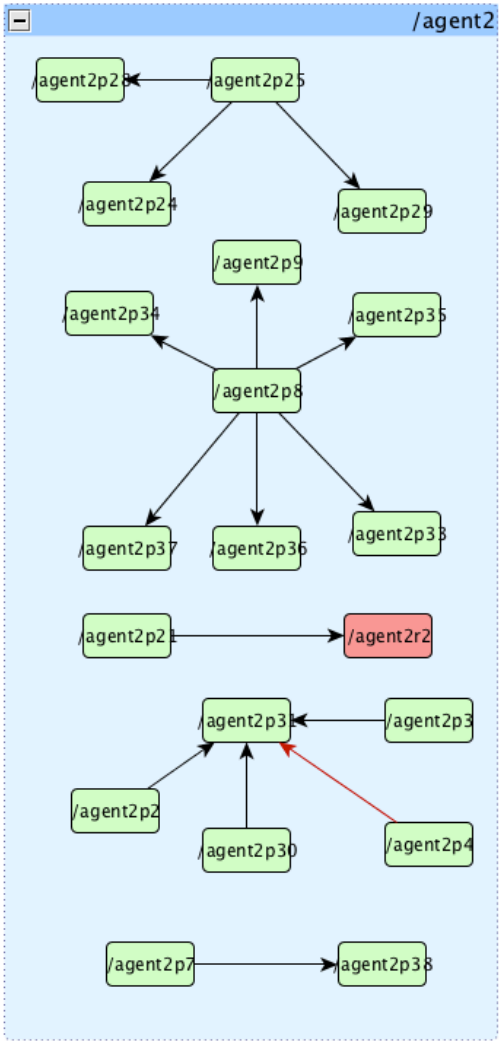
- Conflicts between Facebook and Zynga (3 conflicts)
  - sharing of aggregate or anonymous data
  - transfer of unique user IDs to third party advertisers
  - sharing data for the purposes of merger and acquisition by a third-party
- Conflict within AOL Advertising (1 conflict)
  - collection and use of personally identifiable information

Carnegie Mellon University

Zynga



AOL



isr institute for SOFTWARE RESEARCH

Threats to validity

- Lexicon alignment
  - Is customer service = customer support?
- Reliability of mapping methodology
  - Variability in interpretation
- Human work load and resource requirements

Related work

- L. Cranor et al., “Platform for Privacy Preferences (P3P) Specification,” W3C Working Group Note, 2006
- C. Powers, M. Schunter, “Enterprise Policy Authorization Language,” Version 1.2, W3C Member Submission, Nov. 2003
- C. Hanson, T. Berners-Lee, L. Kagal, G.J. Sussman, D. Weitzner, “Data-purpose algebra: modeling data usage policies.” *8th IEEE Work. Pol. Dist. Sys. & Nets.*, 2007, pp. 173-177
- M.J. May, *Privacy APIs: Formal Models for Analyzing Legal and Privacy Requirements*, Ph.D. Thesis, U. of Pennsylvania, 2008

and others…

Differences: underlying formalism, computational guarantees, semantics for permissions, and focus

Questions?

- Research funded by Naval Postgraduate School
  ONR Award #N00244-12-1-0014