

Bootstrapping Privacy Compliance in Big Data Systems

Giulia Fanti

Based on slides by Anupam Datta
CMU

Fall 2019

Administrative

- ▶ HW2 will be released this week
 - ▶ Stay tuned

- ▶ How did Docker recitation go?

Quiz on Canvas

- ▶ Take the quiz on your laptops/tablets/devices
- ▶ Please do not look back at your notes
 - ▶ these quizzes do not affect your grade, so you should just try to do your best without gaming the grade
- ▶ 10 minutes

Today's Lecture

Bootstrapping Privacy Compliance in Big Data Systems

S. Sen, S. Guha, A. Datta, S. Rajamani, J. Tsai, J. M. Wing

Proceedings of 35th IEEE Symposium on Security and Privacy

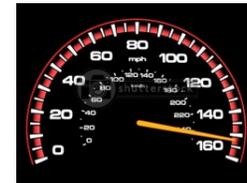
May 2014.

Privacy Compliance for Bing

The image shows a side-by-side comparison of two web browser windows. The left window displays the Bing homepage at <http://www.bing.com/>. It features a yellow search bar with the Bing logo, a navigation menu with links for IMAGES, VIDEOS, MAPS, NEWS, SEARCH HISTORY, and MORE, and a large image of a cartoon detective character. Below the image are several news thumbnails, including 'Hoffman case arrests', 'Newton-John show', and 'iPhone catches fire'. The right window displays the Bing Privacy Statement page at <http://www.microsoft.com/privacystatement/en-us/b>. The page title is 'Bing Privacy Statement'. It contains sections for 'Cookies & Similar Technologies', 'Collecting Your Information', and 'How We Use Your Personal Information'. A vertical sidebar on the right side of the page lists various topics: Cookies, Collecting Your Information, Using Your Information, Sharing Your Information, Accessing Your Information, Mobile and Location Services, Facebook Personalization, Bing Applications, Children, Advertising, Communications, Microsoft Account, and Other Information. The 'Cookies' section is currently selected and highlighted.

Setting:

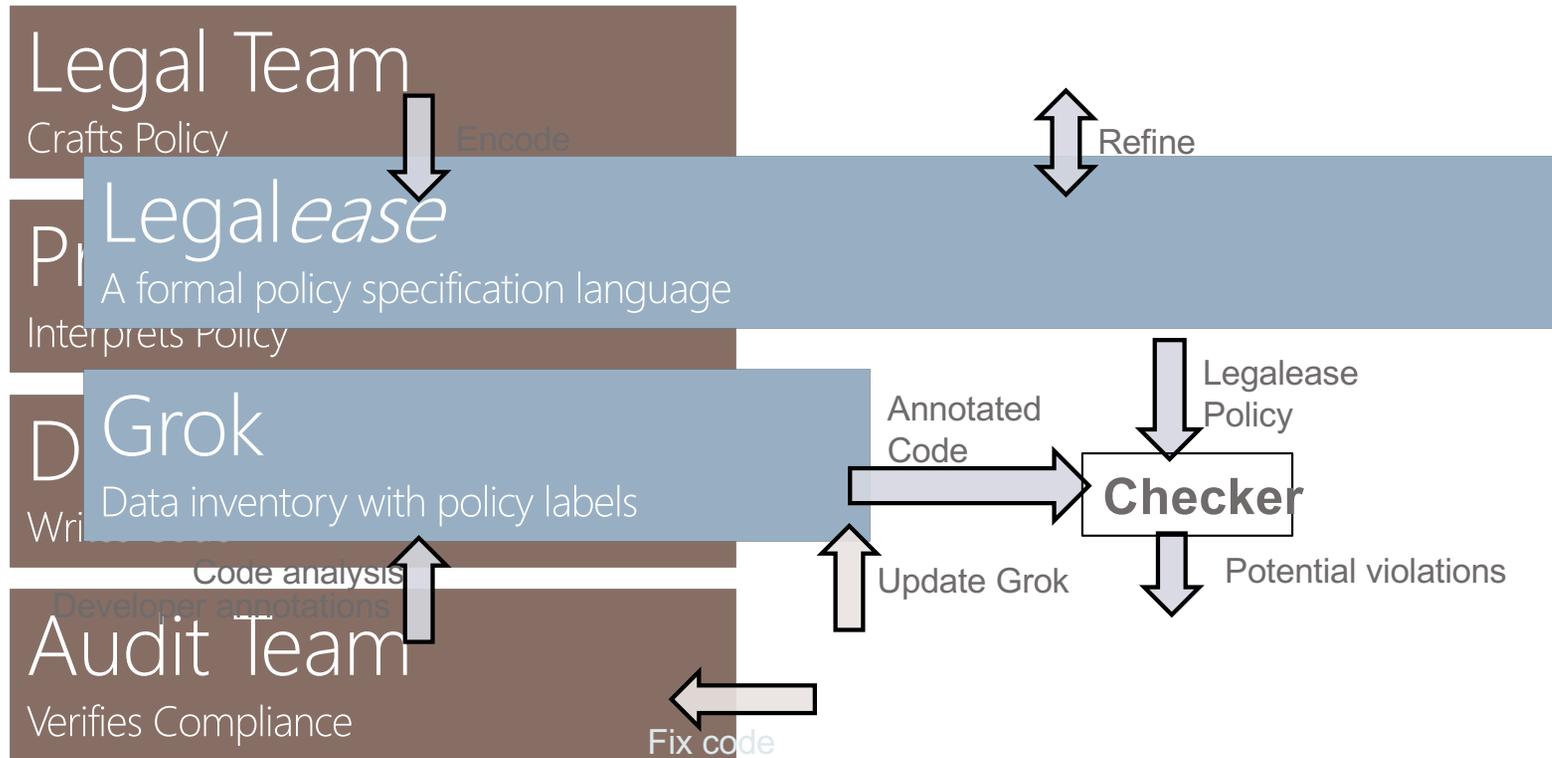
- ▶ Auditor has access to source code



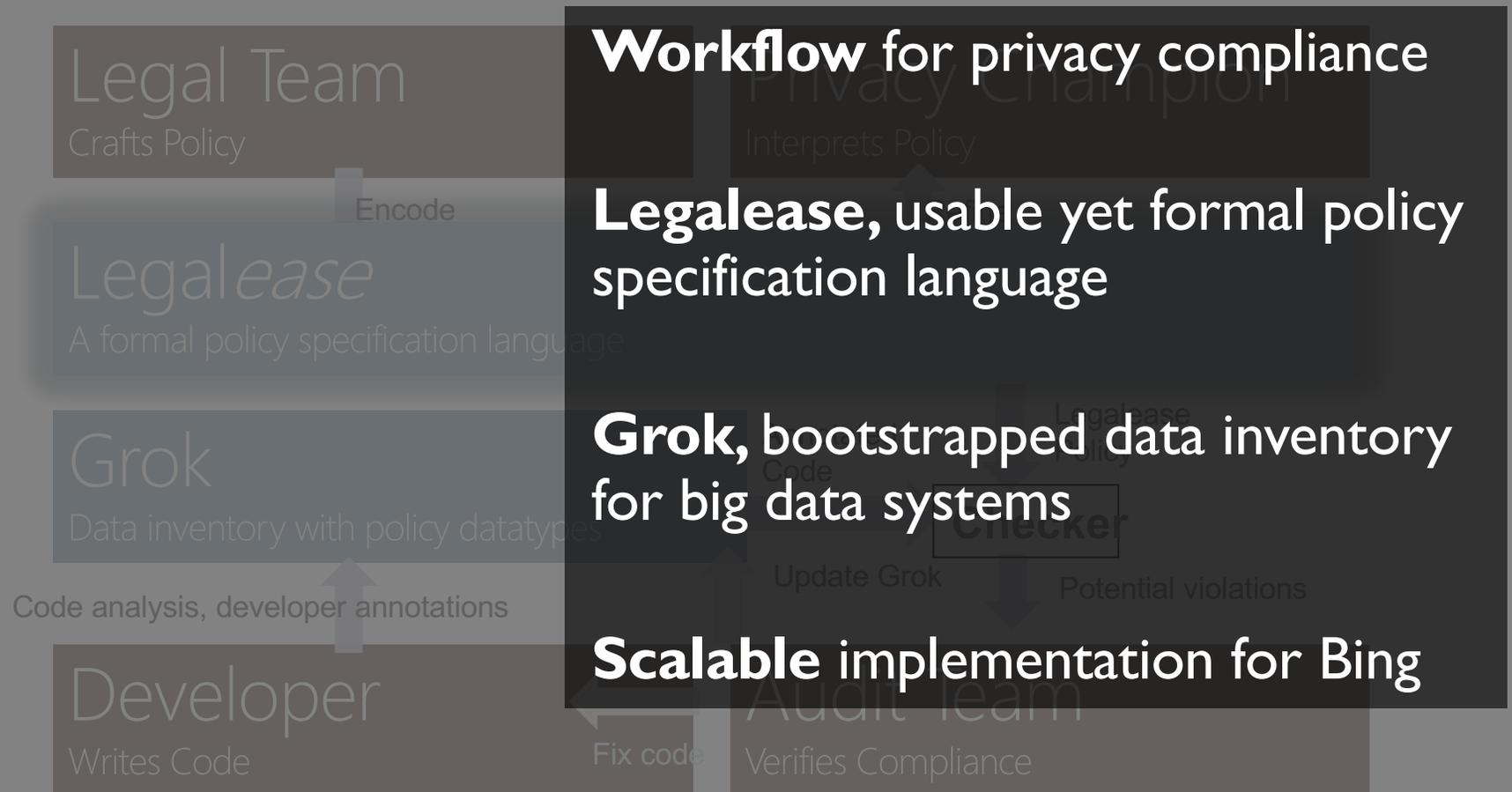
The Privacy Compliance Challenge



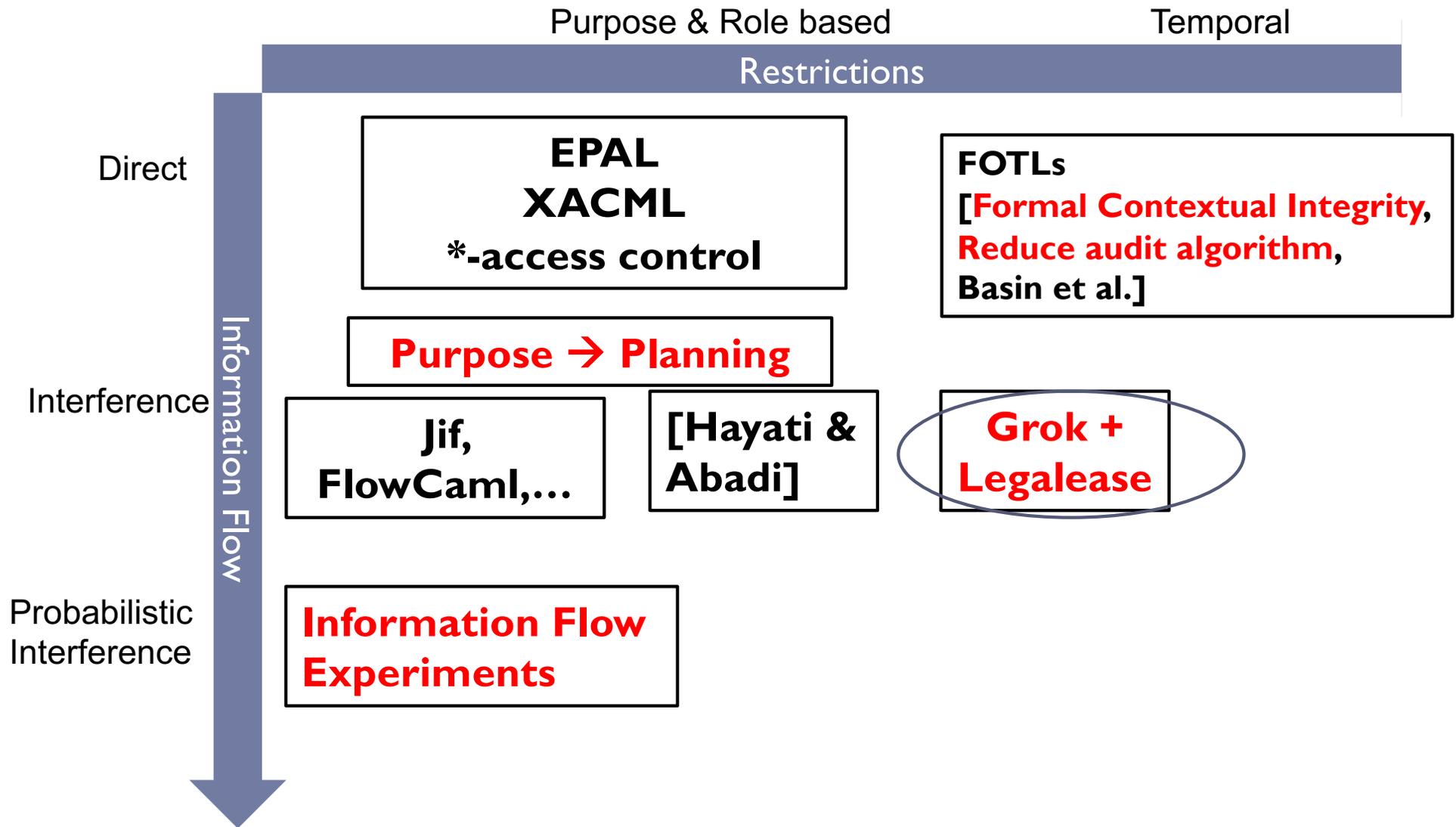
A Streamlined Audit Workflow



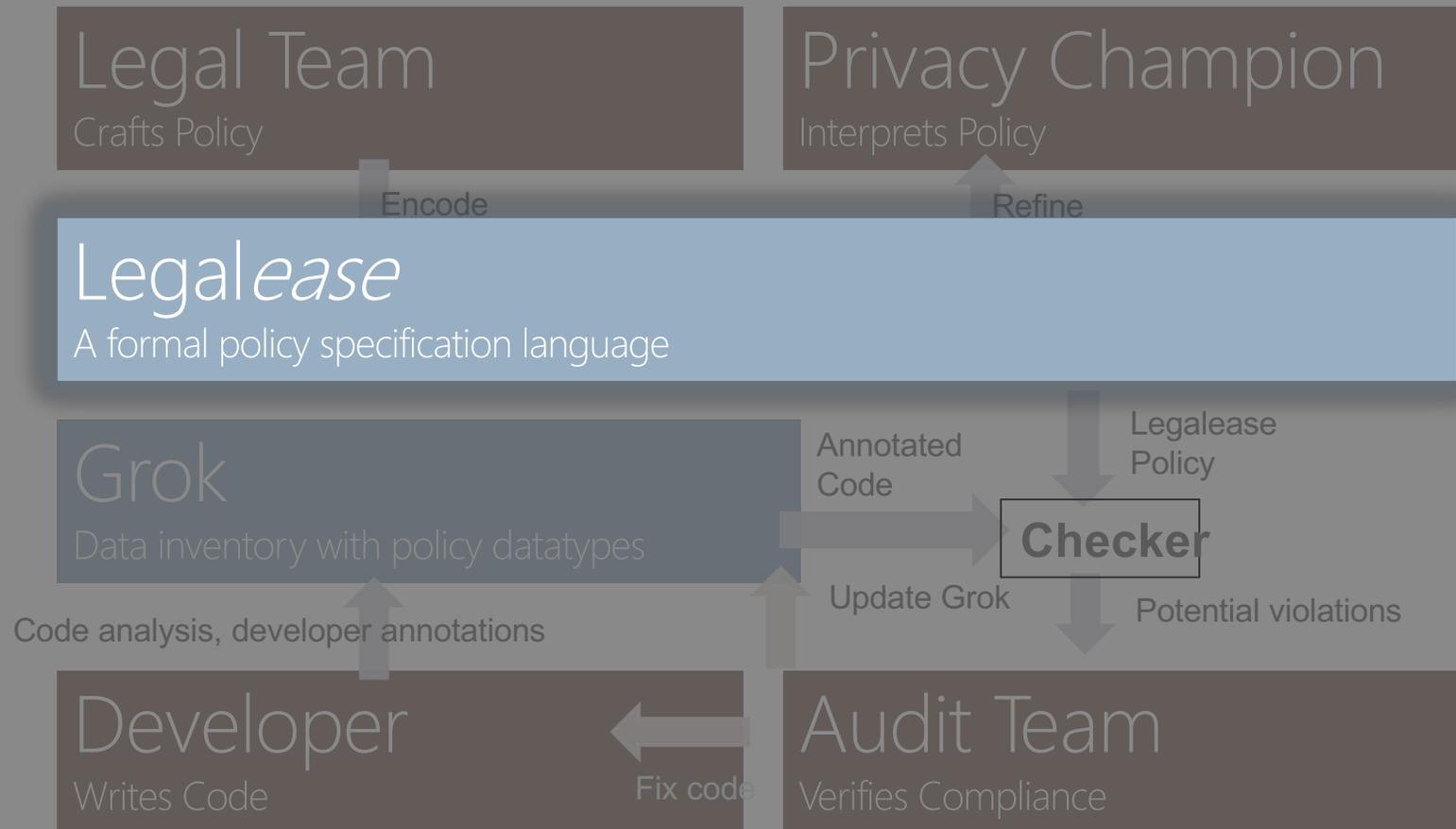
A Streamlined Audit Workflow



Privacy as Restrictions on Personal Information Flow



A Streamlined Audit Workflow



Specification: *Legalease*

Usable.
Expressive.
Precise.

Usable by
lawyers
and
privacy
champs.

Expressive
enough for
real-world
policies.

Precise
semantics
for local
reasoning.

Legalease: Components

- ▶ Each statement is a **clause** reasoning about **attributes**
- ▶ **Clauses**
 - ▶ Allow
 - ▶ Deny
 - ▶ Except
- ▶ **Attributes**
 - ▶ InStore
 - ▶ UseForPurpose
 - ▶ AccessByRole
 - ▶ DataType

Legalease : Syntax

Policy Clause C ::= $D \mid A$
Deny Clause D ::= $\text{DENY } T_1 \cdots T_n \text{ EXCEPT } A_1 \cdots A_m$
| $\text{DENY } T_1 \cdots T_n$
Allow Clause A ::= $\text{ALLOW } T_1 \cdots T_n \text{ EXCEPT } D_1 \cdots D_m$
| $\text{ALLOW } T_1 \cdots T_n$
Attribute T ::= $\langle \text{attribute-name} \rangle v_1 \cdots v_l$
Value v ::= $\langle \text{attribute-value} \rangle$

Legalease

DENY *Datatype* IPAddress
UseForPurpose Advertising

We will **not** use **full IP**
Address for **Advertising**.

Designed for Usability

DENY *Datatype* IPAddress
UseForPurpose Advertising

EXCEPT

ALLOW

Datatype IPAddress:Truncated

ALLOW

UseForPurpose AbuseDetect

EXCEPT

DENY *Datatype*
IPAddress, Account

Exceptions

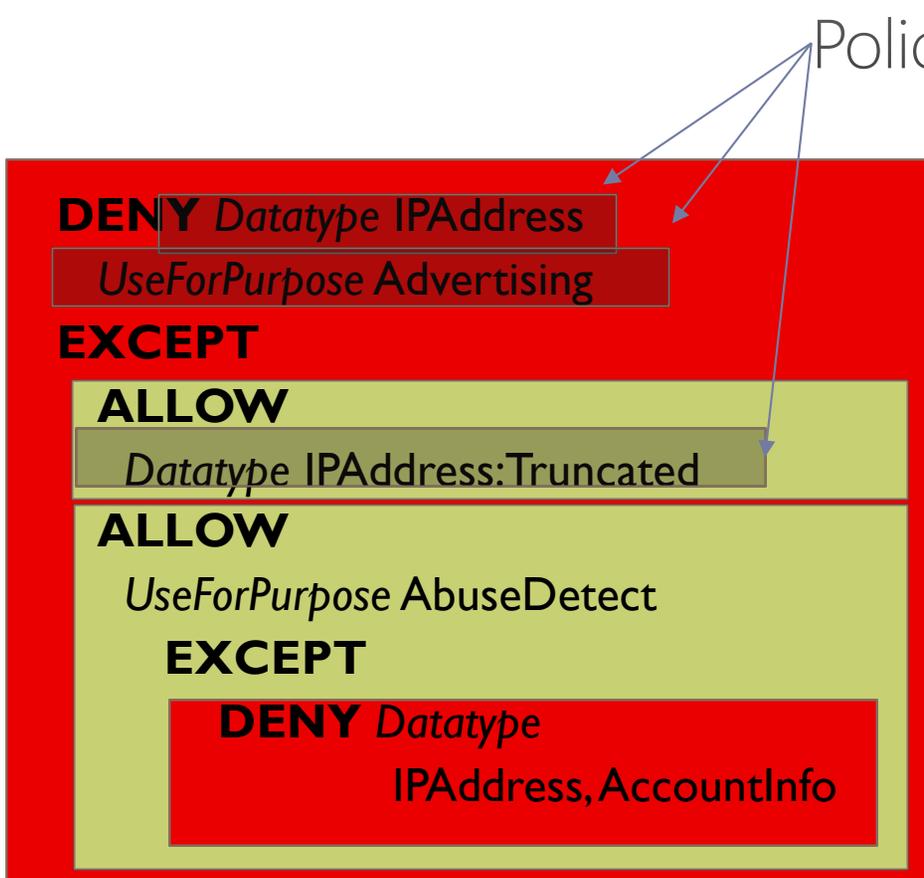
How legal texts are structured
One-to one correspondence

Local Reasoning

Each exception refines its
immediate parent
Formally proven property

H. DeYoung, D. Garg, L. Jia, D. Kaynar, and A. Datta, "Experiences in the logical specification of the HIPAA and GLBA privacy laws"

Legalease : In Action



Program

Datatype: IPAddress, AccountInfo
UseForPurpose: AdsAbuseDetection

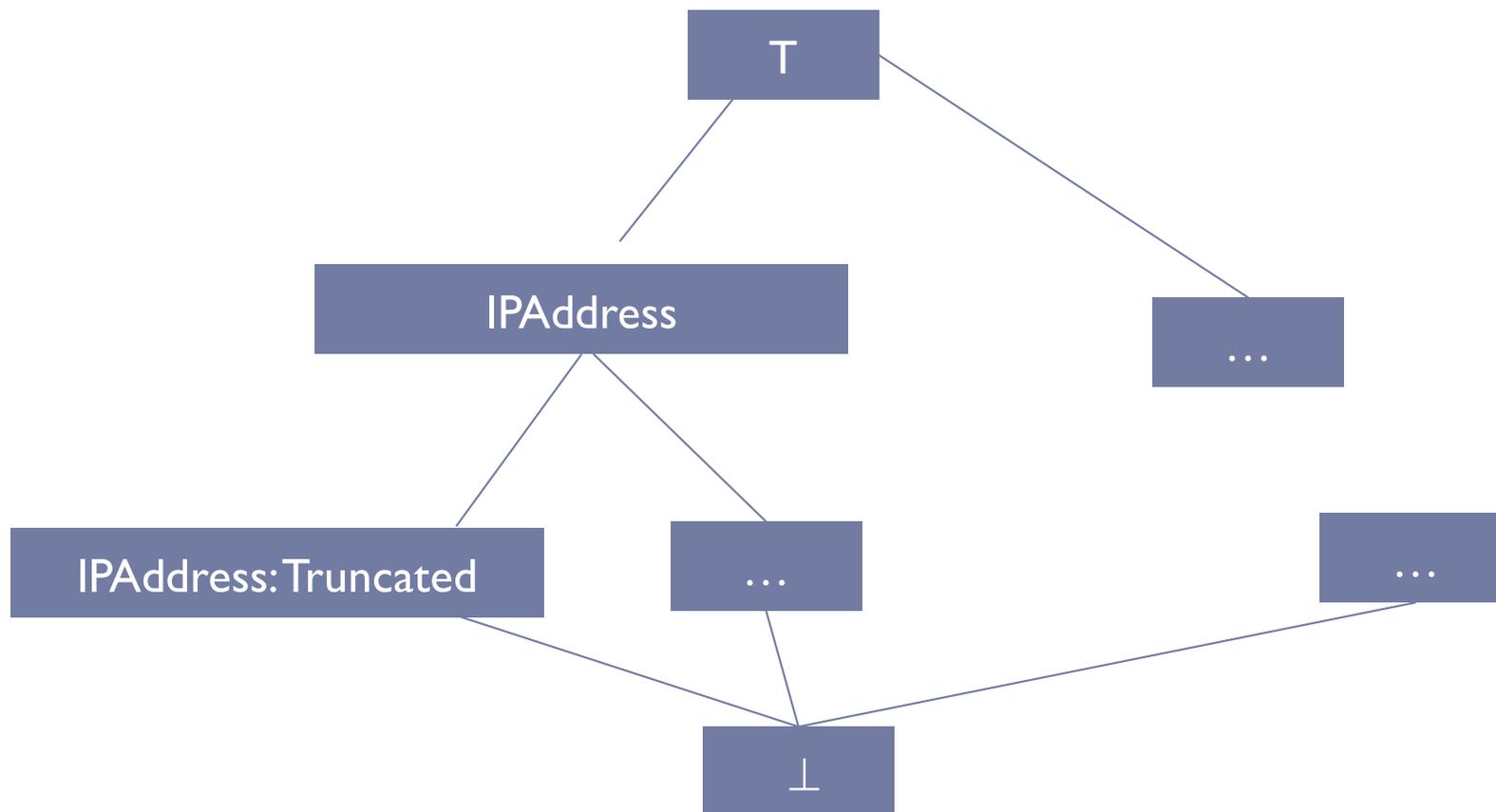
We will **not** use **full IP Address** for **Advertising**. IP Address may be used for **detecting abuse**. In such cases, it will not be combined with **account information**.



Primer on Lattices

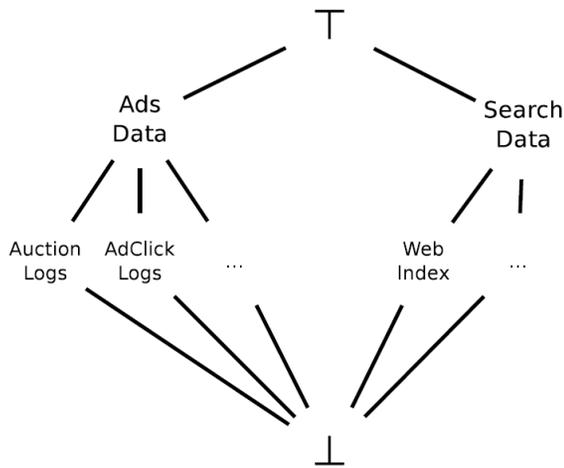
- ▶ On document camera
- ▶ Posets
- ▶ Lattices
- ▶ Rules for ALLOW and DENY
 - ▶ Examples

A Lattice of Policy Labels

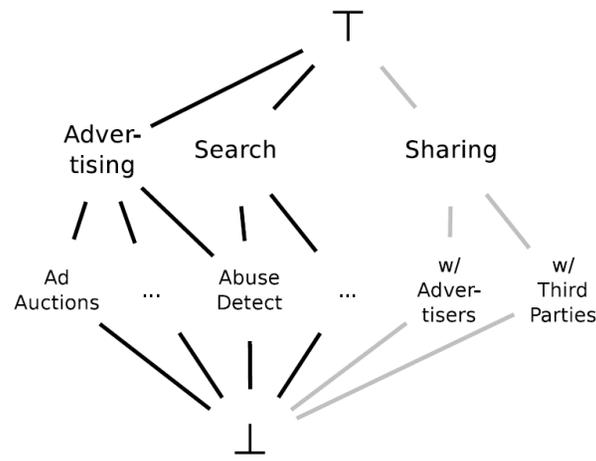


- If “IPAddress” use is allowed then so is everything below it
- If “IPAddress:Truncated” use is denied then so is everything above it

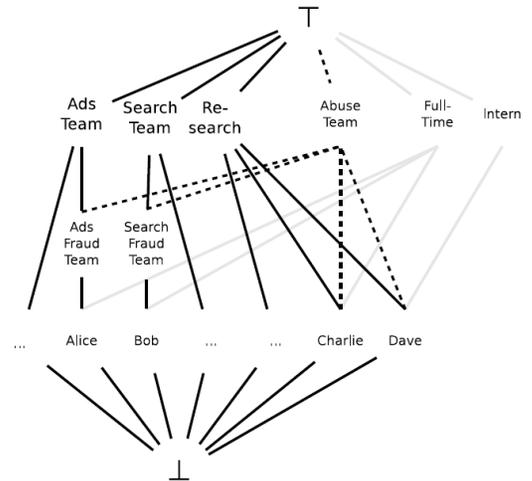
Policy Types : Concept Lattices



InStore Lattice



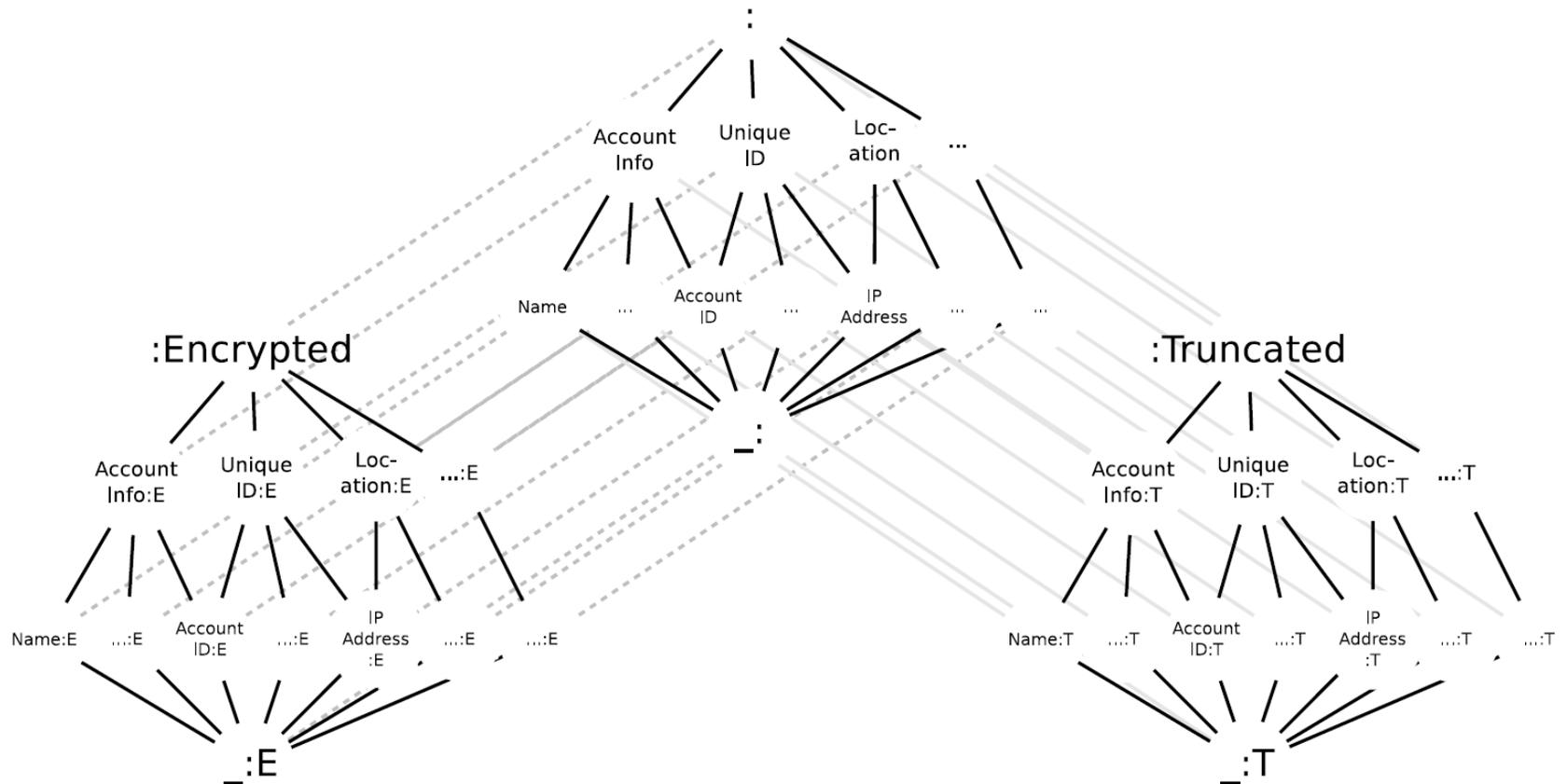
UseForPurpose Lattice



AccessByRole Lattice



Policy Labels : Datatypes



Formal Semantics

$$\frac{\boxed{T^G \sqsubseteq T^C} \exists_i D_i \text{ denies } T^G}{\text{ALLOW } T^C \text{ EXCEPT } D_1 \cdots D_m \text{ denies } T^G} \text{ (A}_2\text{)}$$

Based on Lattice Orderings on Policy Types

Formal Semantics

$$\frac{T^G \sqsubseteq T^C \quad \exists_i D_i \text{ denies } T^G}{\text{ALLOW } T^C \text{ EXCEPT } D_1 \cdots D_m \text{ denies } T^G} \quad (A_2)$$

Recursively check exceptions

ALLOW clauses have DENY clauses as exceptions

Top Level clause determines Blacklist/Whitelist