Privacy Engineering in DRM Systems

ACM Workshop on Security and Privacy in Digital Rights Management

November 5, 2001

Michael J. Freedman
MIT Lab for Computer Science

Joan Feigenbaum, Yale CS Dept
Tomas Sander, InterTrust STAR Lab
Adam Shostack, Zero-Knowledge Labs

The reality of web privacy…

CHEAP TICKETS inc

- To search for fares:
  - unique subscriber ID
  - full name
  - e-mail address
  - home phone
  - work, fax number (opt)
  - traveling partners (opt)
  - preferred airport (opt)

The reality of web privacy…

CHEAPTICKETS inc

- To search for fares:
  - unique subscriber ID
  - full name
  - e-mail address
  - home phone
  - work, fax number (opt)
  - traveling partners (opt)
  - preferred airport (opt)

The New York Times ON THE WEB

- To browse news content:
  - unique subscriber ID
  - e-mail address
  - country
  - zip code
  - age
  - sex
  - household income (optional)

The reality of web privacy…

CHEAPTICKETS inc

- To search for fairs:
  - unique subscriber ID
  - full name
  - e-mail address
  - home phone
  - work, fax number (opt)
  - traveling partners (opt)
  - preferred airport (opt)
- Valuable

The New York Times ON THE WEB

- To browse news content:
  - unique subscriber ID
  - e-mail address
  - country
  - zip code
  - age
  - sex
  - household income (optional)
- Why not? It's easy…

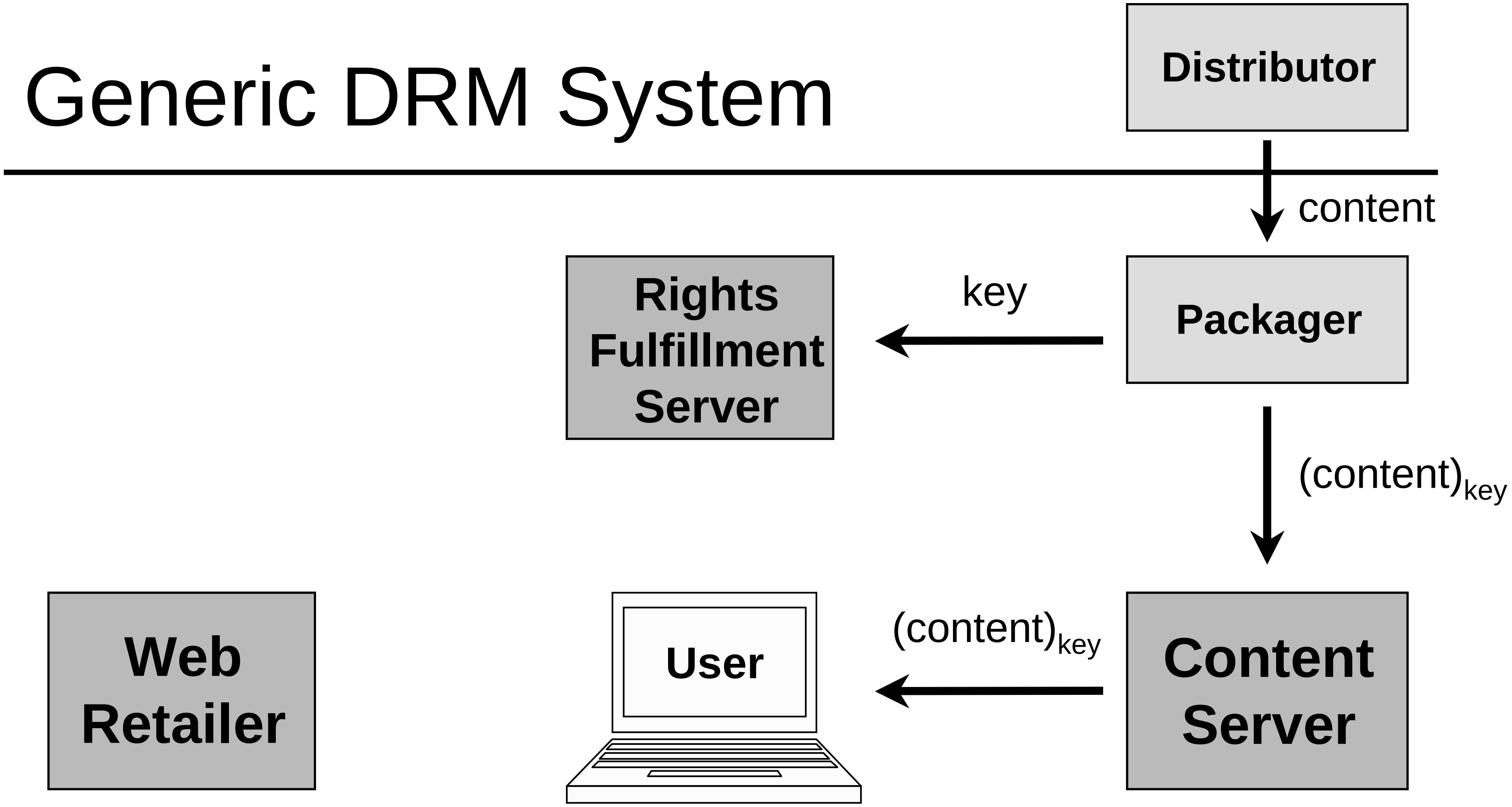
DRM: a hard privacy playground

- Focus of this talk
  - Mass-market DRM-enabled content distribution on the Internet
- Inherent tension
  - Copyright enforcement goals of **copyright owners** vs.
  - Privacy goals of **consumers**
- Privacy threats
  - Information about consumed content is privacy-sensitive
  - Information centrally aggregated by few players
  - Actual usage information collected
  - Devices linked to personally-identifiable information (PII)

Outline

- A generic DRM architecture
- Assertion: Crypto doesn’t solve privacy in DRM
- Real goal: Privacy abstractions for good practices
- Needed:
  - Practical methodology for privacy engineering
  - Enforcement procedures

Generic DRM System



Distributor

content

Packager

key

Rights Fulfillment Server

(content)$_{key}$

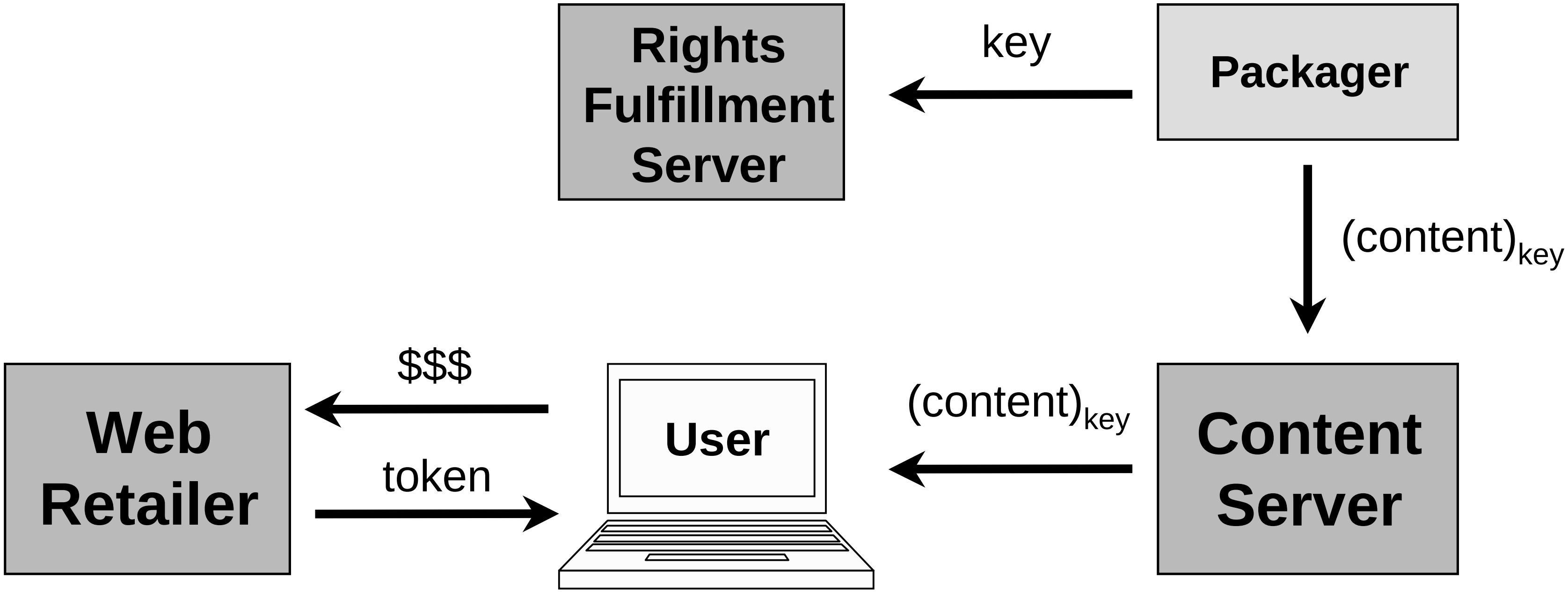
Content Server

(content)$_{key}$

User

Web Retailer

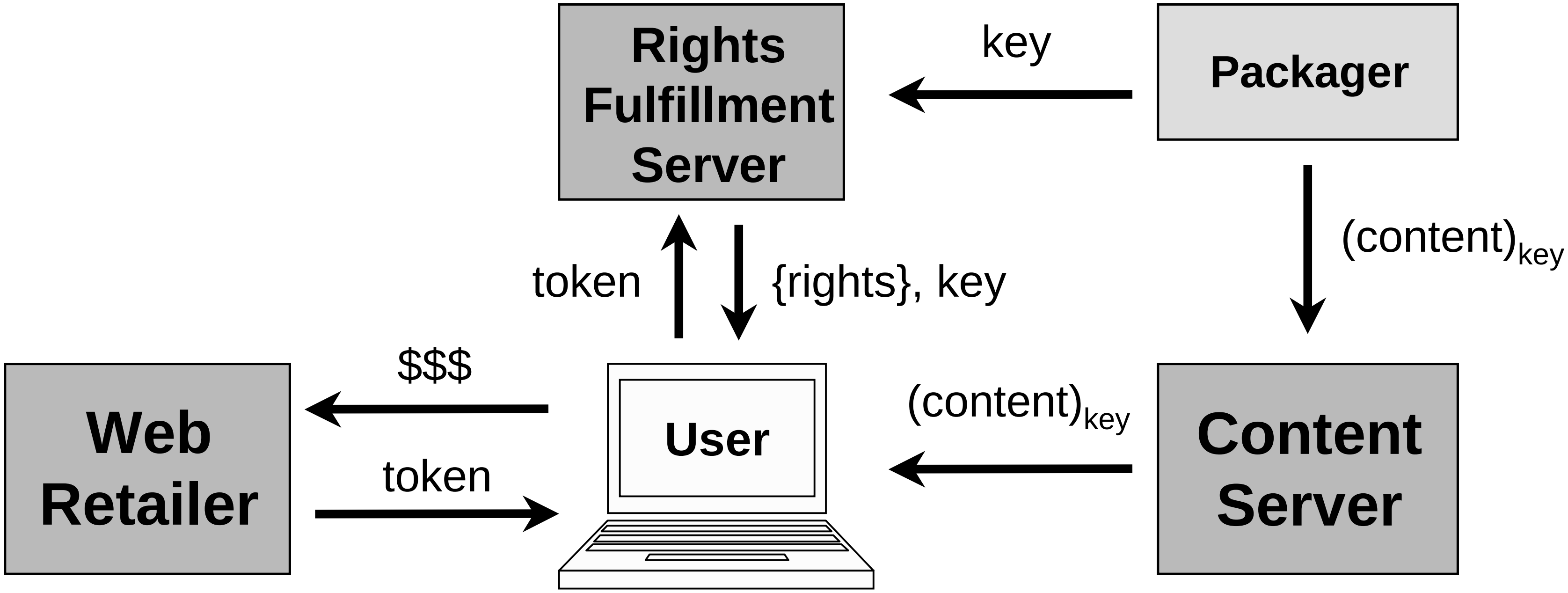
Generic DRM System



Different options for purchase:

- Pay-per-use content-specific token
- Subscription token

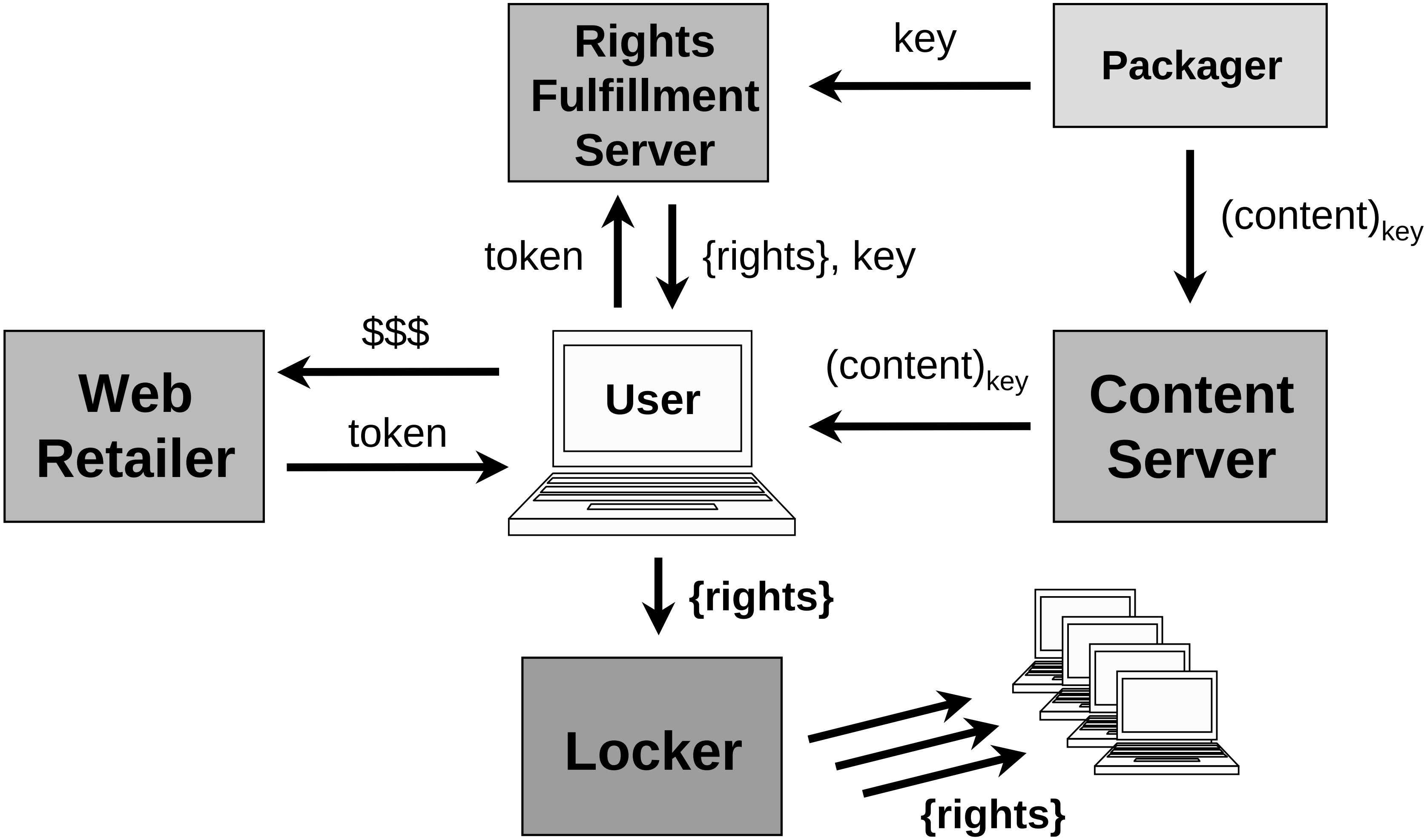
Generic DRM System



Key concept:

Separation of content and rights to access/use

Generic DRM System



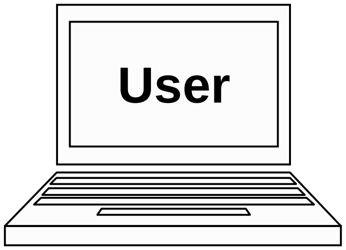
But what about privacy?

Cryptography to the rescue!

- Pay for content with anonymous ecash
- Connect to all servers via anonymous mixnets
- Authenticate with anon credentials or ZK protocols
- Download content via PIR or OT
- Use SFE when services require information

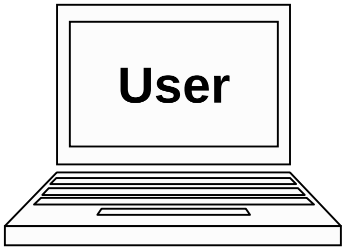
…voila!?!

Forgetful Alice…



```
alice@foo%   play ThatSongILike.mp3
Passphrase for alice@foo/1024:  ***********
Incorrect passphrase.  Try again.
Passphrase for alice@foo/1024:  ***********
Incorrect passphrase.  Try again.
```

Forgetful Alice…



```
alice@foo%   play ThatSongILike.mp3
Passphrase for alice@foo/1024:  ***********
Incorrect passphrase.  Try again.
Passphrase for alice@foo/1024:  ***********
Incorrect passphrase.  Try again.
```

What is Alice to do?

Call customer service…
give them some information…
get a new passphrase?

But she just hid all her information from them!

The cryptographic pixie-dust fallacy

Let’s assume…

cryptography is cheap…

even secure operating systems…

Can crypto just be sprinkled throughout system?

→ Cryptographic abstractions do not adequately model “reality” and what people want to do

Mismatch of cryptographic abstractions

- Privacy-enhancing techniques typically offer:
  - Hiding of information
  - Anonymity
  - Require a clearly defined “Alice and Bob” threat model
- In “real” world:
  - Cryptographic protocols can not address “purpose binding”: How *already learned* information is used
  - Business world in flux: mergers, acquisitions, etc.
  - Know thy enemy? good guy vs. bad guy, trusted vs. untrusted, private vs. public

“Legitimate” uses for information

- Risk management: misuse and anomaly detection, revocation, fraud deterrence
- Profiling and counting, e.g., for artist compensation
- Targeted marketing and recommendation services
- Depersonalized data for trend-spotting, mining
- Customer service and retention
- Backup and archiving
- Traffic modeling for infrastructure, QoS

Asymmetry of power

- Alice = consumer
- Bob = content provider

- Crypto paradox:
  - Crypto protocol protects Alice's info from Bob
  - But against his Bob's will?

    Bob needs to agree to run it in the first place.

- Consequence:
  - Providing privacy requires Bob's buy in
  - There may be technically much-easier solutions:

    Bob may favor over complicated cryptographic protocols

Asymmetry of knowledge

- Consumers cannot measure or differentiate
  - Earthlink vs. Zero-Knowledge
- Consumers are not willing to “pay” for privacy
  - No commercially-successful privacy technology on Internet
  - Even free software (e.g., cookie blockers) not adopted
- Consequence:
  - Business-incentives to offer cryptographically-strong privacy?

Costs of privacy

- Economics argument:
  - Switch only if benefits > costs
  - Network externalities
  - Asymmetries, demand…
- Technical argument:
  - Engineering costs, system complexity grows dramatically
  - High computational costs for privacy
  - Opportunity costs for not learning the “legitimate” info

⟶ Internet businesses typically want to leverage existing infrastructure

Outline

- A generic DRM architecture
- Assertion: Crypto doesn’t solve privacy in DRM
- Real goal: Privacy abstractions for good practices
- Needed:
  - Practical methodology for privacy engineering
  - Enforcement procedures

Practical privacy engineering

- Provocative claim:
  - – Crypto is a distraction for actually improving privacy
  - – Crypto will play important role only for simple tasks
    - SSL, authentication, etc…

- Needed: reasonable notion of privacy goals
- Needed: practical methodology for privacy engineering
  - – Low business costs
    - Provides “necessary” information
  - – Low consumer costs
    - Low latency
    - Easy to use
    - No initiative required: built into the DRM infrastructure

Goals for practical privacy engineering

- Based on Fair Information Principles (FIPs)
  - Notice
  - Choice
  - Access
  - Security
  - Enforcement
- Advantages:
  - Do not prescribe technical implementation
  - Underlie most privacy friendly legislation, e.g., Europe
  - Becoming *de facto* measure
    - good enough for businesses
    - strong for consumers

Simple privacy engineering principles

1. Collection limitation:
   - Only collect the information really needed
   - Step 1: Analyze precisely what is needed
   - Many tasks may not require PII or UID

2. Database design
   - Separation of duty: split databases
   - Easy pseudonymization
   - Data erasure, esp. for long-term storage

3. Client-side data aggregation
   - Transfer preprocessed data
   - Control info flow by granularity

Simple privacy engineering principles

4. Notice and purpose disclosure
   - What's being collected
   - How it's being used

5. Reasonable choice
   - "No privacy" vs. "No service" not sufficient

⟶ System capabilities

⟶ Language expressibility

(similar goals as P3P)

Simple privacy engineering principles

- Why would business want to follow?
  - – Adhere to privacy laws
  - – Market differentiation
  - – Costs of managing collected data
  - – Process issues dominate privacy failures

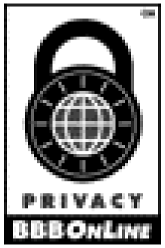
- But this is a "trust me" solution:

Cryptographer: I want math and proofs!

Audits and enforcement

Users: Math? I don’t understand math…

A seal is proof for me!



Combining notice and auditability is strong!

- Notice enforcement:

  Legally binds companies to specific practices

- Auditing process, tools:

  Verify that agreed-upon practices are enacted

Main lessons from DRM…

- DRM can be a key enabler of privacy
  - Exchange of content for money
  - Reduces the need to rely on privacy-intrusive revenue generations

- Meta Lesson

  Privacy should be part of initial design phase

Main lessons from DRM…

- Cryptography does not effectively address complex privacy concerns in DRM

- Likely similar for many "web services"

- Needed:

  1. Methodology for practical privacy engineering
  2. Catalog of best privacy practices
  3. Tools and standard components

     a. System development for notice/choice
     b. Techniques to reduce leakage
     c. Process auditing