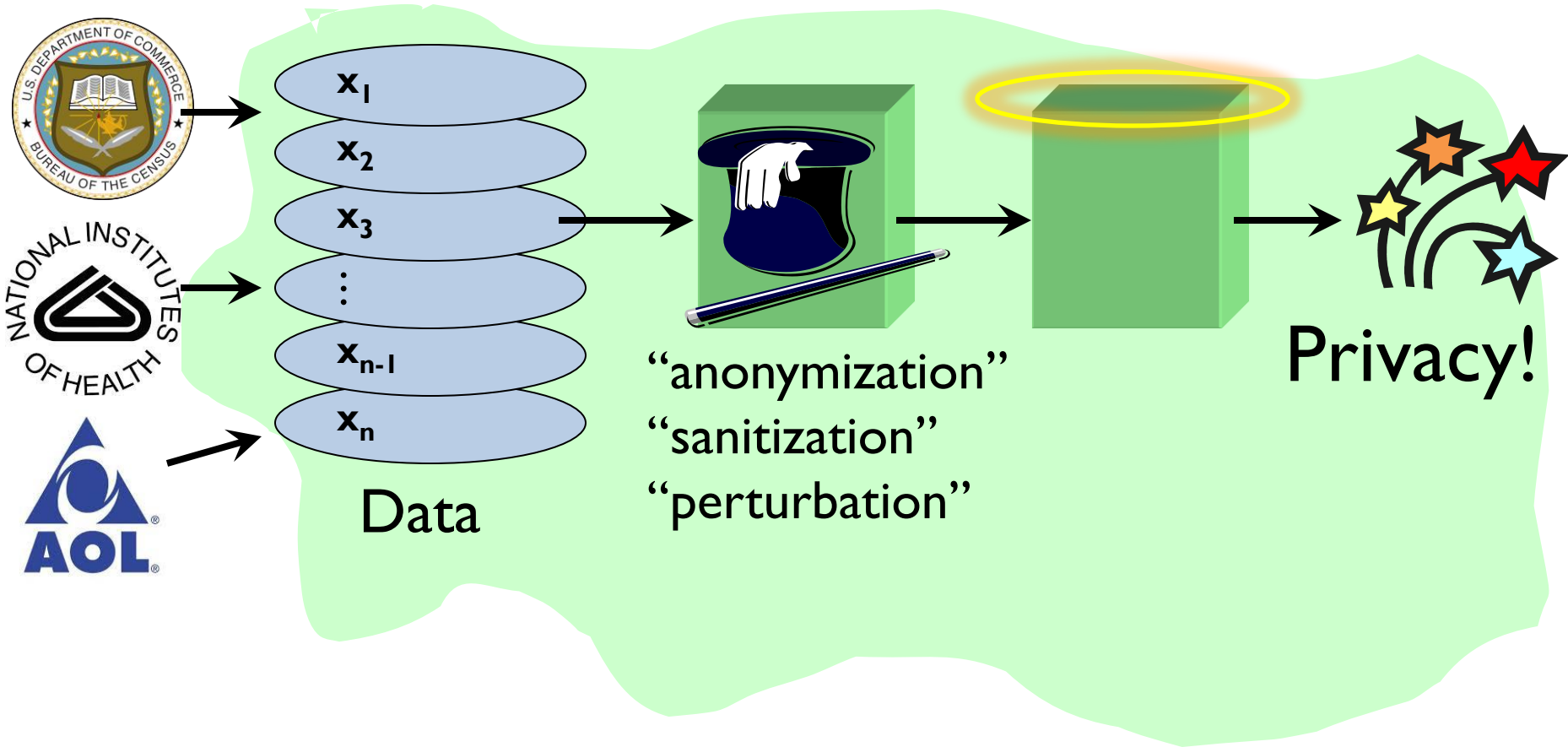


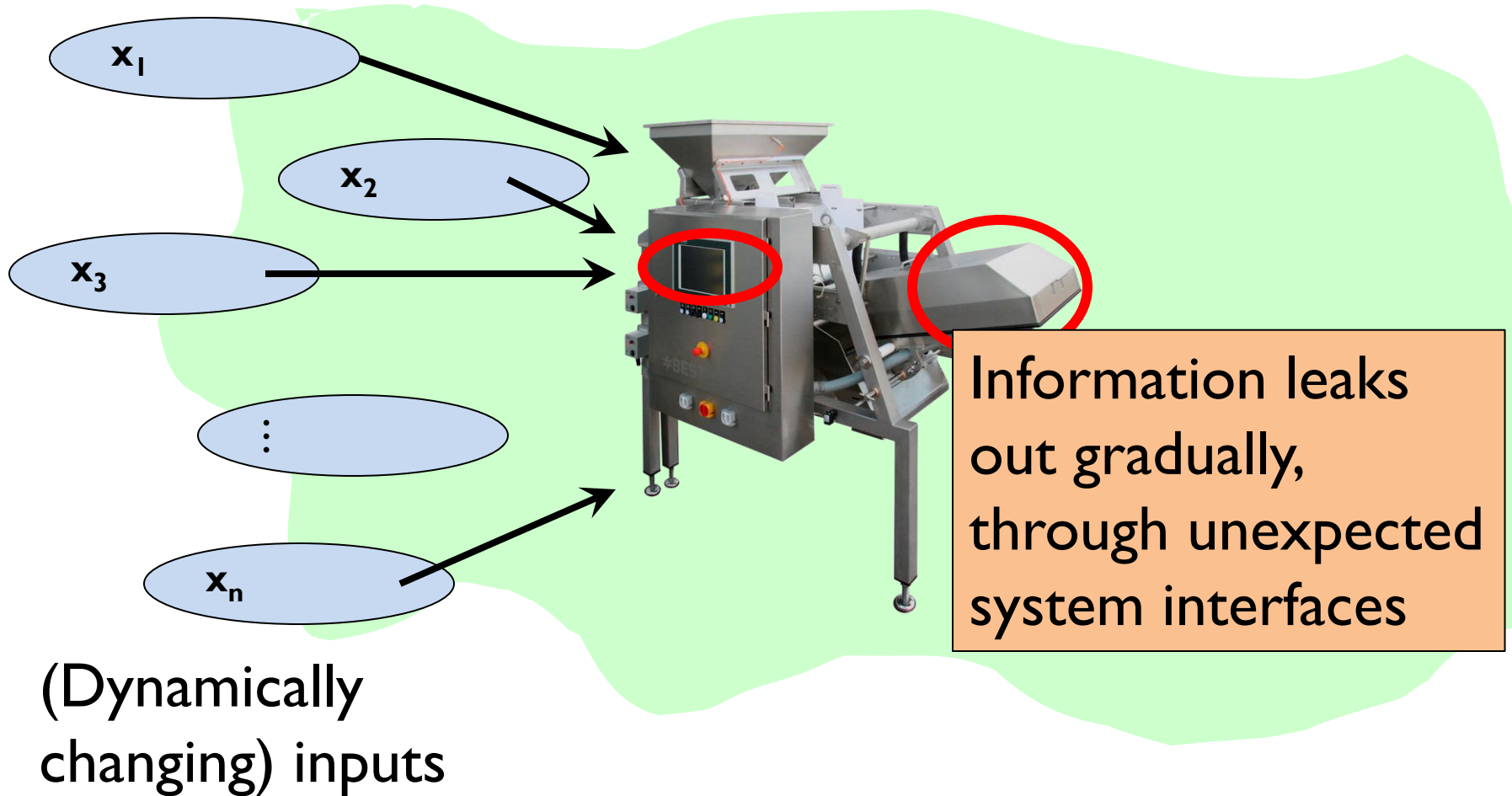
# Inference Attacks

Vitaly Shmatikov

# No Data Released = No Privacy Problems?



# Welcome To The Machine



# Reading Material

Calandrino, Kilzer, Narayanan, Felten, Shmatikov

“You Might Also Like:” Privacy Risks of Collaborative Filtering

**Oakland 2011**

# Recommender Systems

amazon.com

NETFLIX

hulu™

facebook

The New York Times

monster®

hunch

tripadvisor®

You Tube

LibraryThing

last.fm

PANDORA®

yelp.✿

iTunes

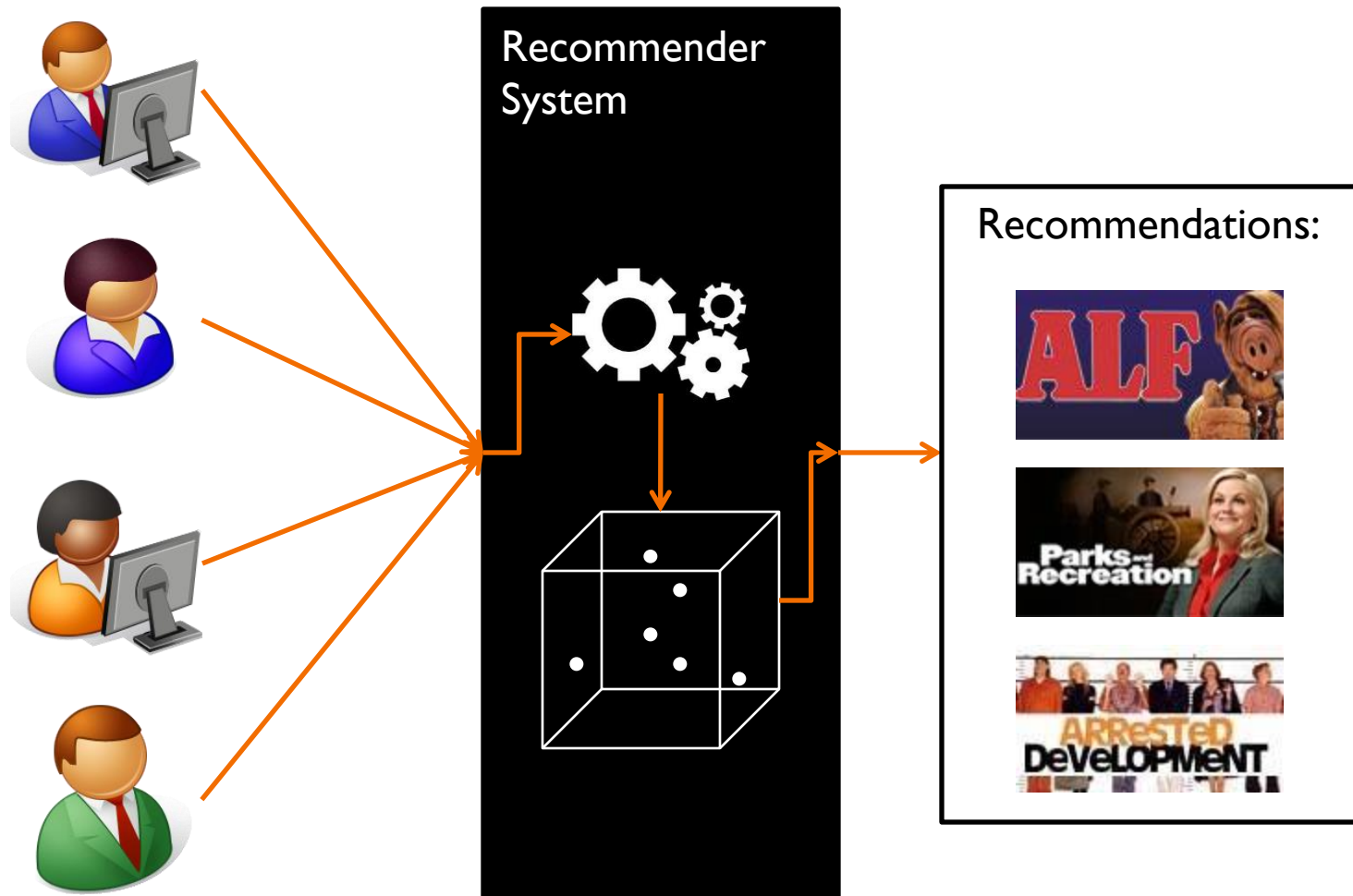
Google reader

twitter

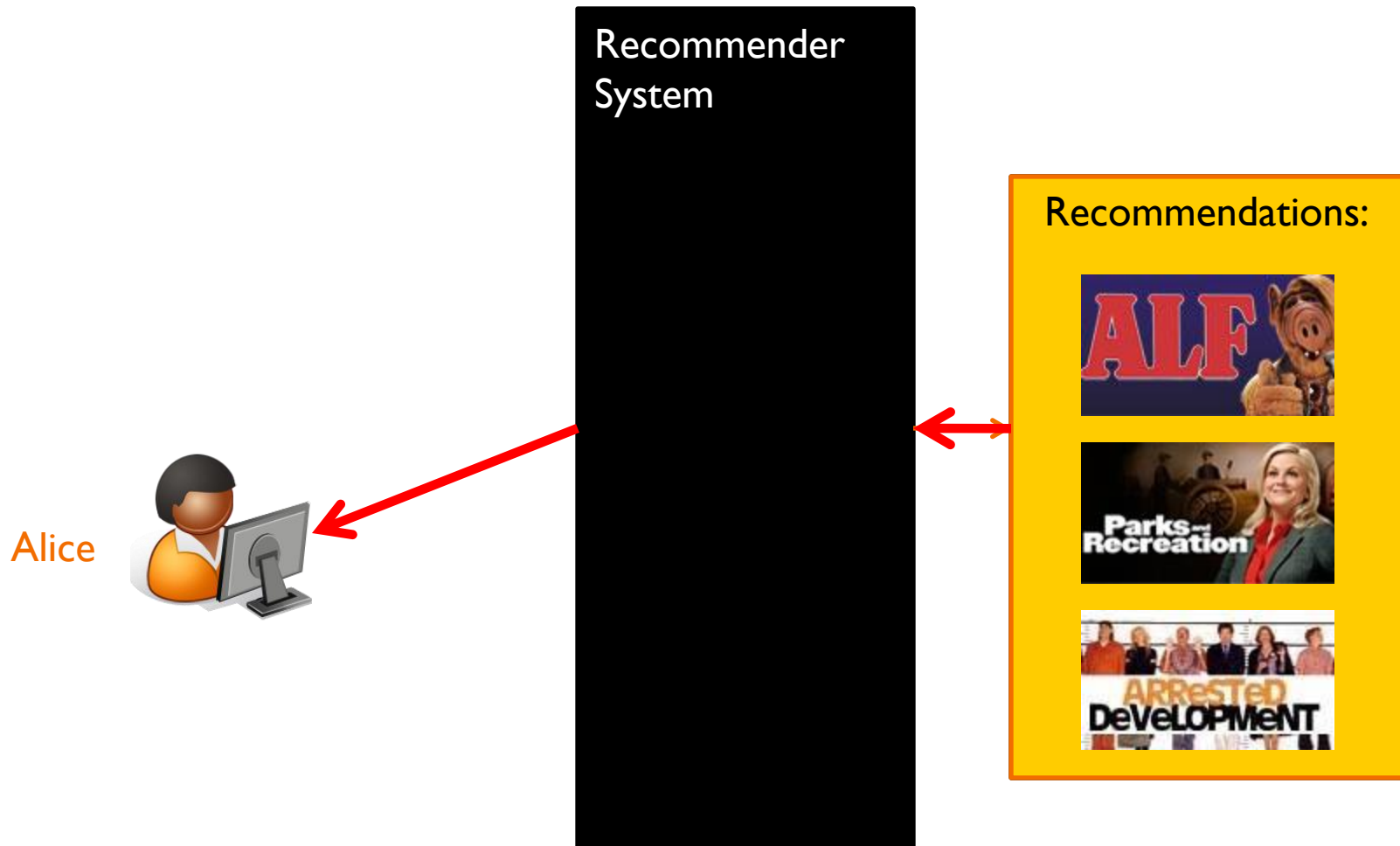
IMDb  
The Internet Movie Database

The Motley Fool.  
To Educate, Amuse & Enrich™

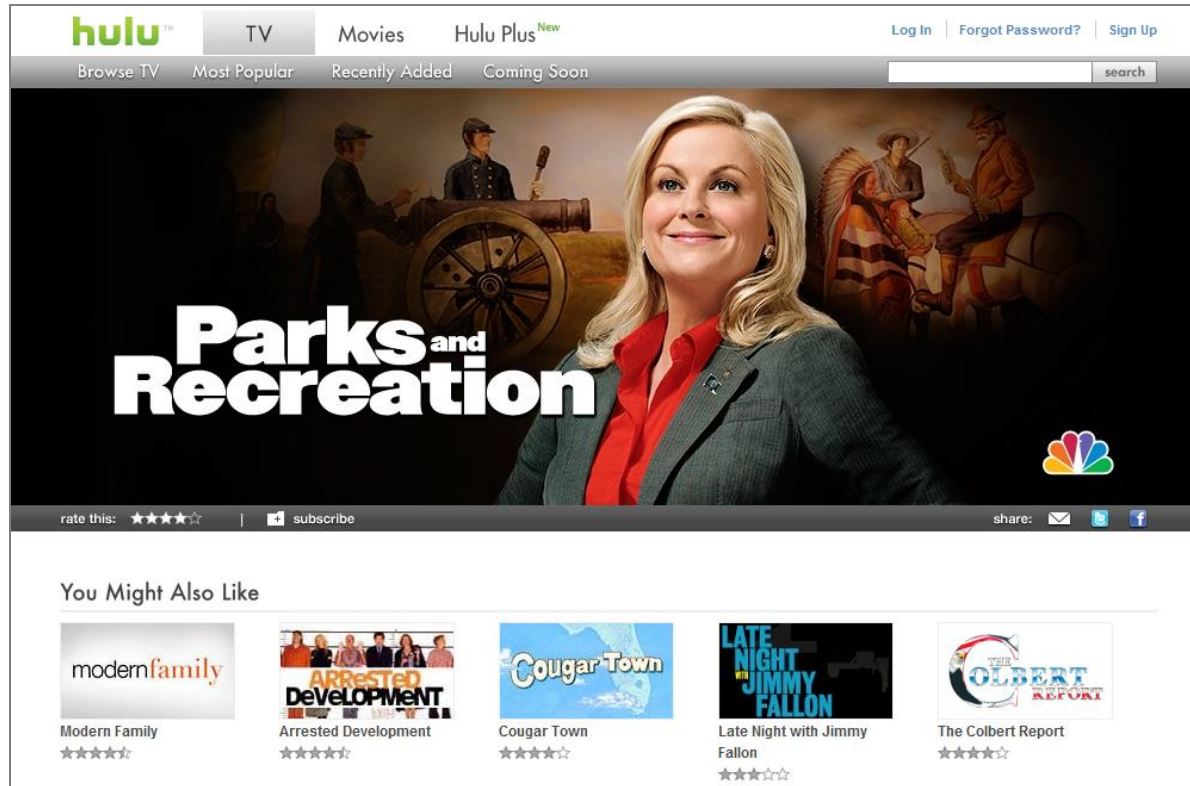
# Collaborative Filtering



# Can We Invert This?



# Item-to-Item Recommendations



**...versus user-to-item recommendations**

**This allows for passive adversaries**



# Our Tools

**Auxiliary information  
and  
dynamics of public outputs**

# Auxiliary Information

7 of 8 people found the following review helpful:

★★★★★ **Love it!**, May 24, 2011

By **John Doe** (Springfield, IL United States) - [See all my reviews](#)

VINE™ VOICE REAL NAME

This review is from: [Jump \(Audio CD\)](#)

This is one of my favorite albums of all time and includes two songs of 2007. They clearly spent a great deal of time honing their well-reviewed but largely forgettable debut album in 2005. Not only are the tunes far more refined, but the lyrics are both more mature. Besides the third track ("Quiet" - which isn't bad, but not as good as the other tracks), the album is solid from start to finish. "Elephant" and "Faint Bounce" in particular show that they're one of the most talented bands on the scene.

I recommend listening through the whole thing twice before you can appreciate the complexity of some of the songs like "Yellow Wall" take on a new meaning. Also, be aware that this is a true album, not just a collection of songs. While many can stand on their own - I expect to hear "Electric Fan" on the radio all summer - the album is best listened to start to finish like a finely told story. In an era increasingly dominated by simple, catch three-minute ones, albums like this prove that bands can think in sixty-minute pieces.

One of the biggest challenges of this album will be its genre. 2005's *Margins* was clearly targeting more of a dance crowd with its indie-rock undertone. Don't get me wrong, if you want to dance, there's plenty to love - from "Shark Hound" to "Red T-Shirt." "Swim" is also appreciated during moments of quiet reflection.

**John Doe** ▶ **Edit Profile**

**Music:**

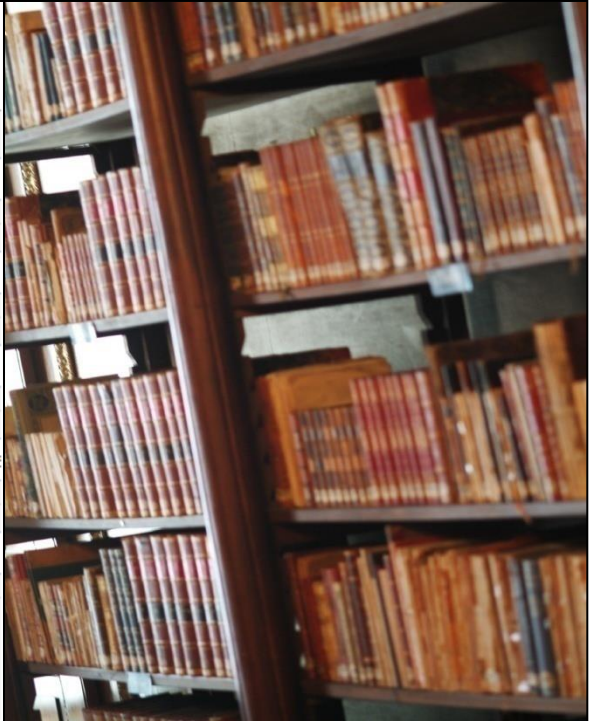
**Books:**

**Movies:**

**Television:**

**Games:**

**Save Changes**



An attacker can sometimes learn partial history

# Alice's TV Taste



Transaction history tends to be  
(relatively) unique

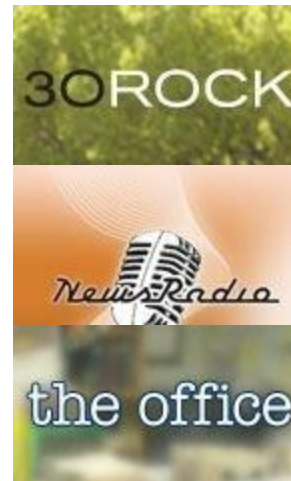
# Modern Collaborative Filtering

## Item-Based and Dynamic



Selecting an item makes it and past choices more similar  
Thus, output changes in response to transactions

# Inferring Alice's Transactions



Today, Alice watches a new show (we don't know this)  
We can see the recommendation lists for auxiliary items  
...and we can see changes in these lists  
Based on these changes, we infer transactions

# Prediction vs. Inference



1. 30 ROCK
2. COMMUNITY
3. the office

Predictions (recommendations) seek to impose order

# Prediction vs. Inference



1. 30 ROCK
2. COMMUNITY
3. the office

Inferences are based on temporal changes in order

# More Recent Work

McPherson, Shokri, Shmatikov

“Defeating Image Obfuscation with Deep Learning”

**arXiv 2016**



# deep learning revolution

Web

News

Videos

Images

Sho

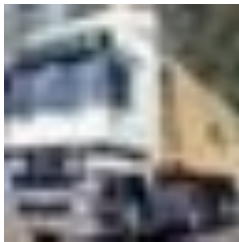
About 1,890,000 results (0.23 sec)

What does this  
revolution mean for  
a privacy researcher?

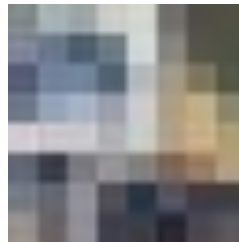


# Outperform Humans, Eh?

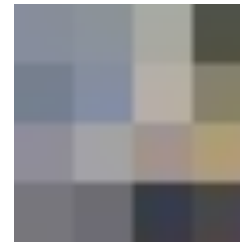
Does this extend to obfuscated images?



Truck



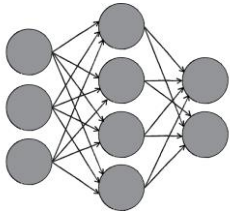
Truck?



???



??????????



Truck

Truck

Truck

Truck

# Image Obfuscation Techniques



Pixelation



Blurring



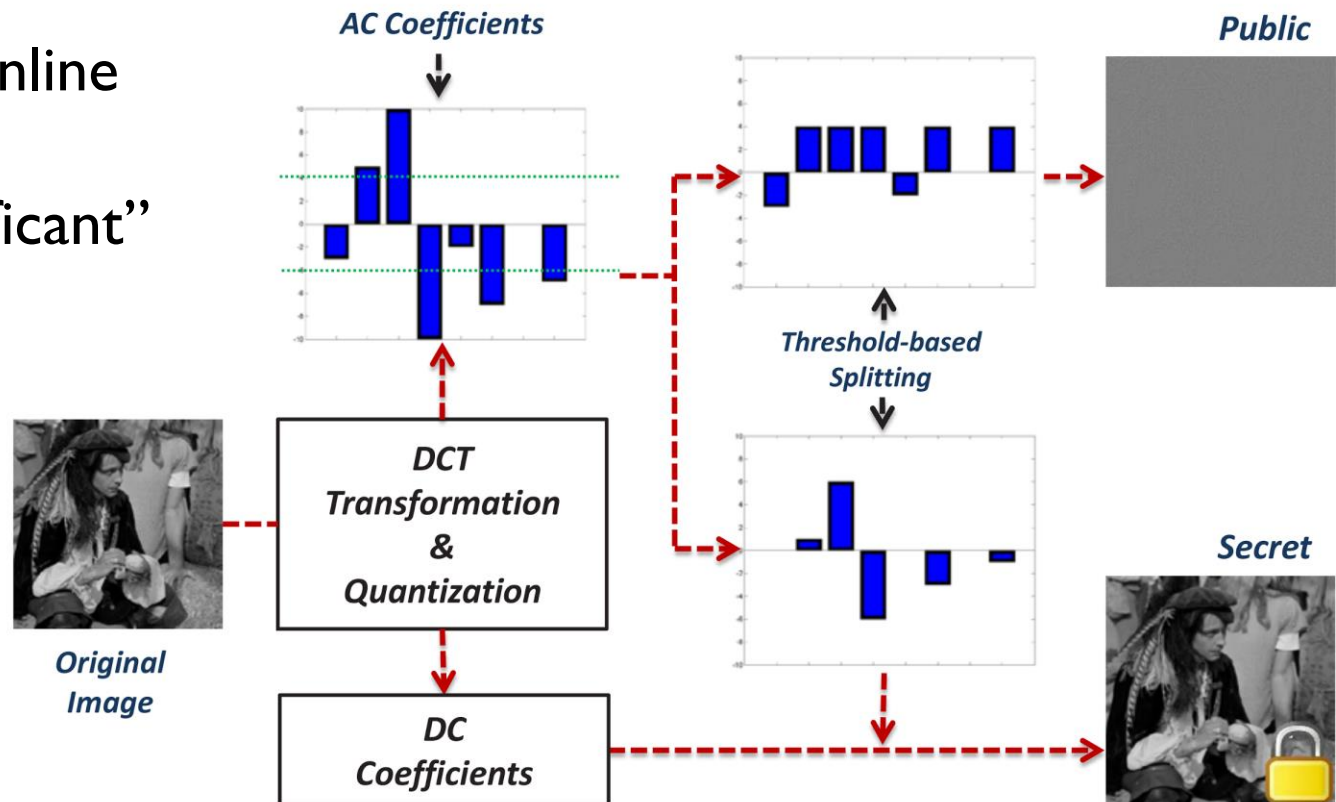
P3

As deployed  
by YouTube

# P3: Privacy-Preserving Photo Sharing

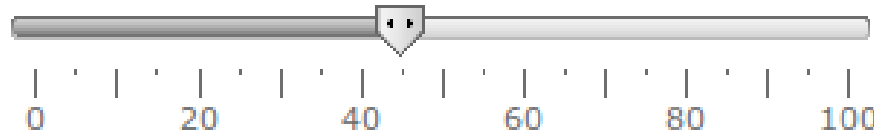
Ra et al. (NSDI 2013)

Add privacy to online image storage by encrypting “significant” JPEG coefficients



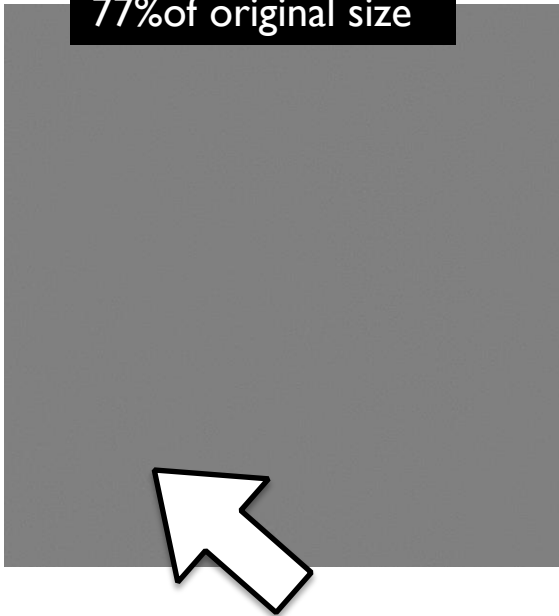
# Privacy/Size Tradeoff in P3

More privacy



Smaller images

77% of original size

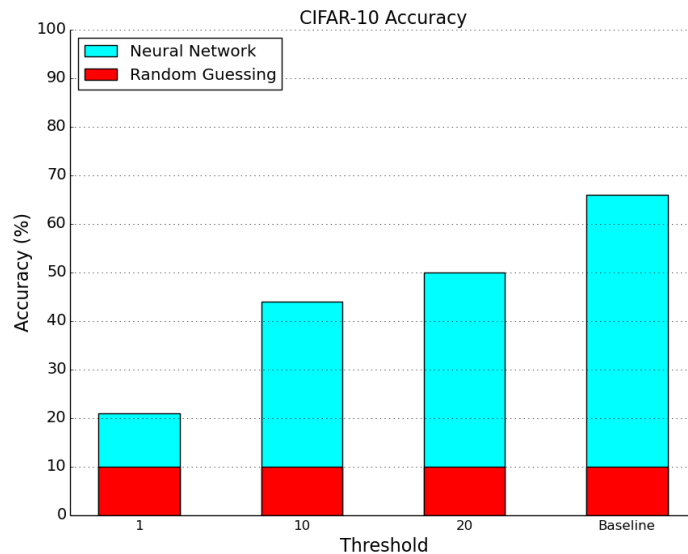


43% of original size

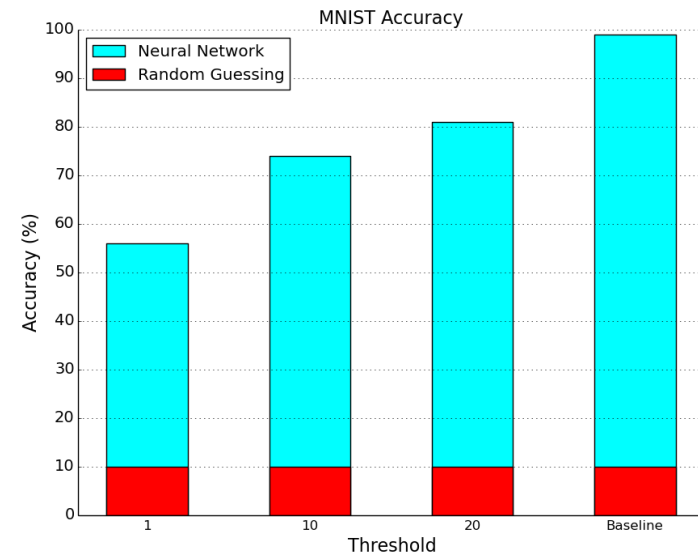


Threshold of 10 for encrypting JPEG coefficients  
)P3 recommends threshold of (10-20

# Training a Neural Network to Classify P3-Protected Images



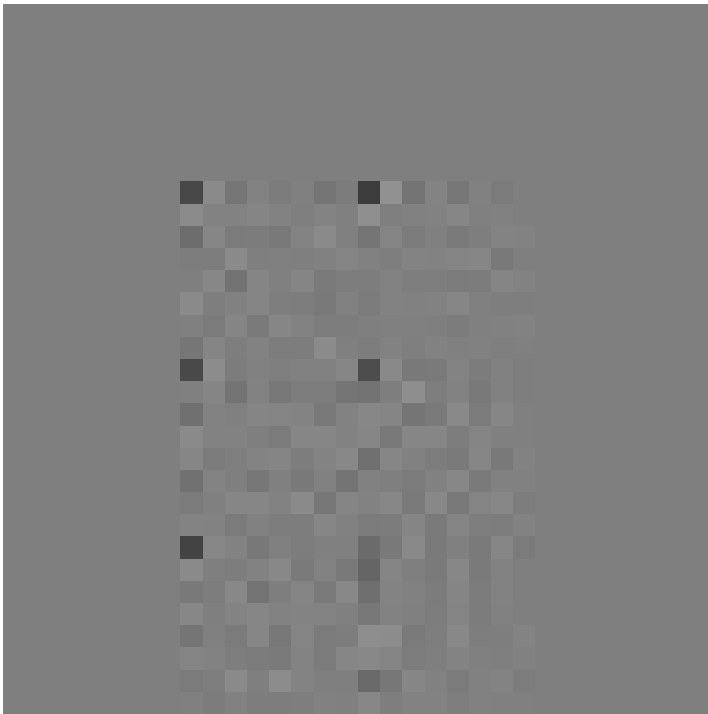
60,000 colored images  
in 10 categories  
(e.g. ship, car, frog, cat)



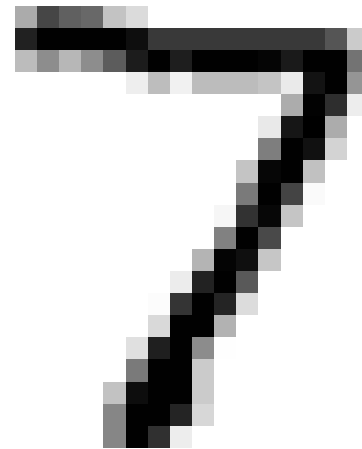
70,000 handwritten digits

# Which Digit Is This?

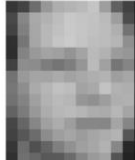
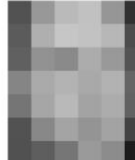
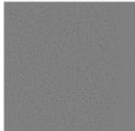
The neural network knows...



32)x32 pixels(



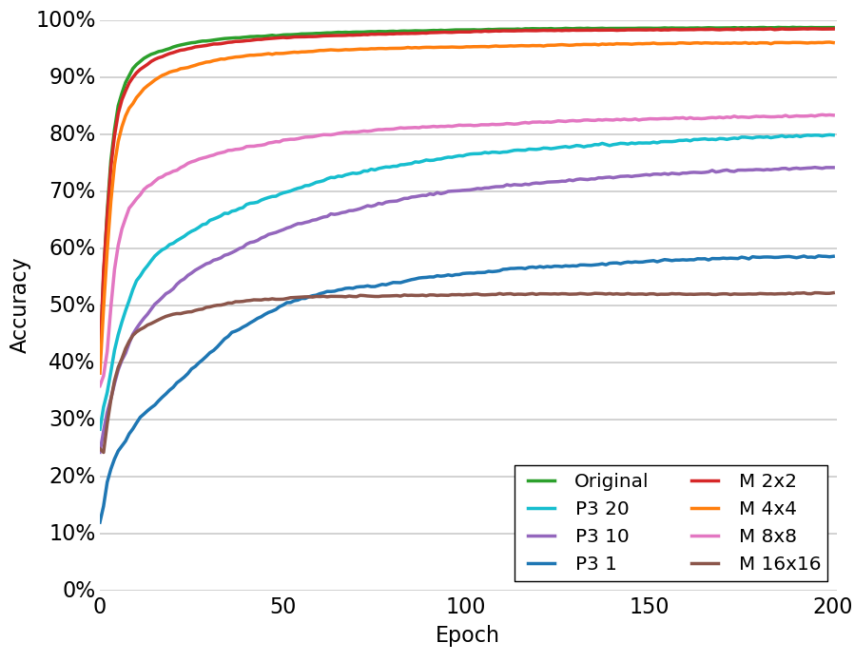
# More Targets

| Dataset   | Original                                                                           | Mosaic                                                                             |                                                                                    |                                                                                     |                                                                                      | P3                                                                                   |                                                                                      |                                                                                      |
|-----------|------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|
|           |                                                                                    | $2 \times 2$                                                                       | $4 \times 4$                                                                       | $8 \times 8$                                                                        | $16 \times 16$                                                                       | 20                                                                                   | 10                                                                                   | 1                                                                                    |
| MNIST     |   |   |   |   |   |   |   |   |
| CIFAR-10  |   |   |   |   |   |   |   |   |
| AT&T      |   |   |   |   |   |   |   |   |
| FaceScrub |  |  |  |  |  |  |  |  |

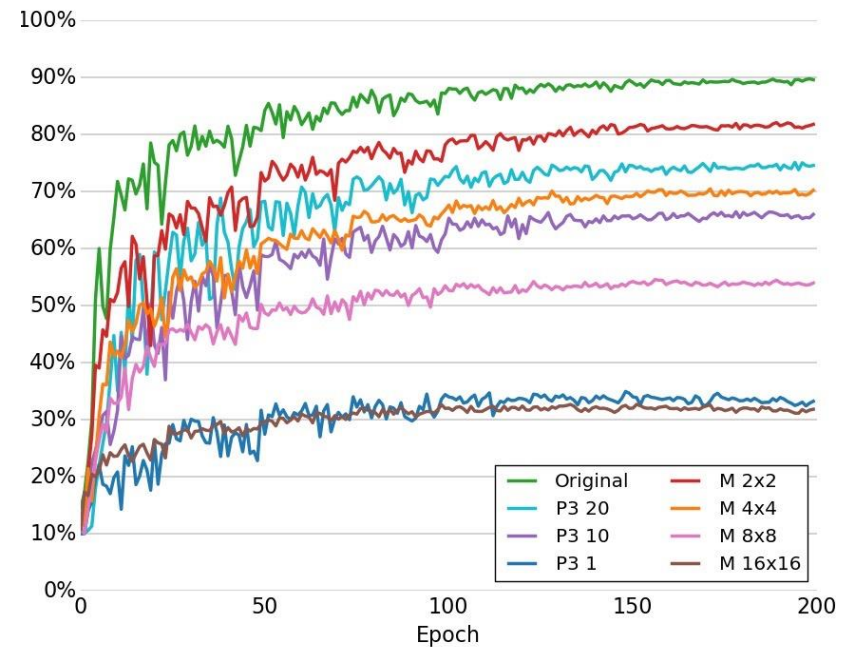
Handwritten digits, objects, faces...



# Results for Object Recognition

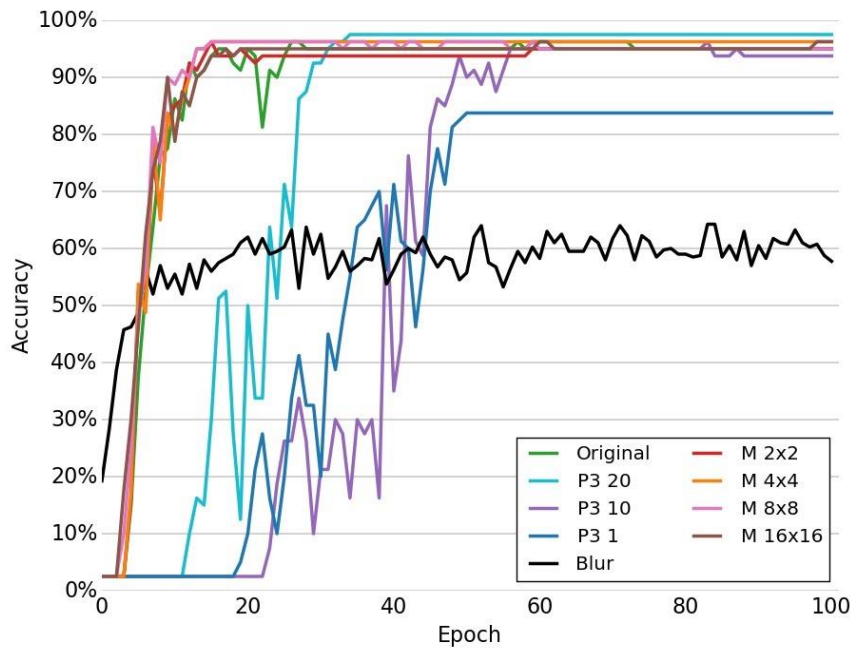


**MNIST**  
handwritten digits

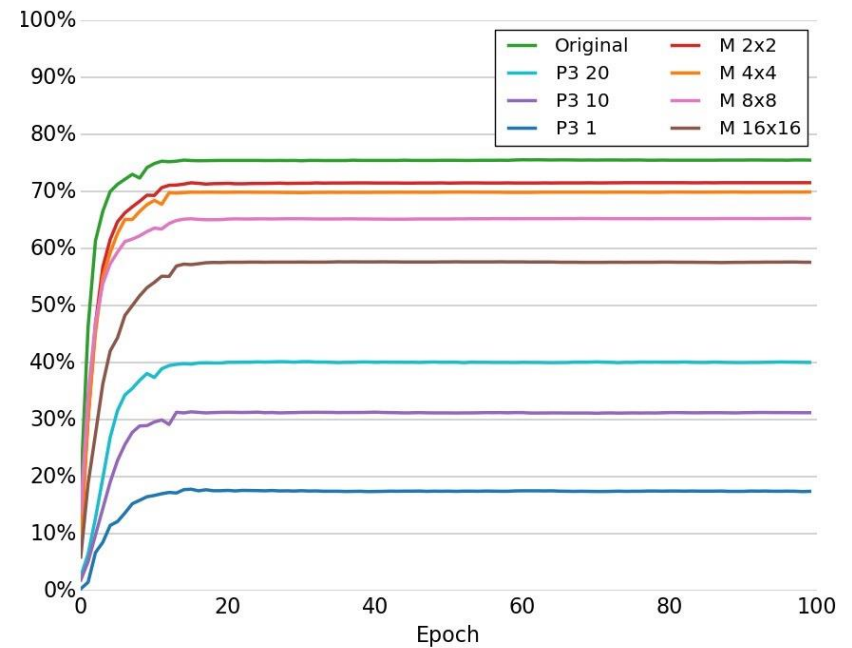


**CIFAR-10**  
animals and vehicles

# Results for Face Recognition



AT&T  
database



FaceScrub  
database