Inference Attacks

Vitaly Shmatikov
No Data Released = No Privacy Problems?

Data

- $x_1$
- $x_2$
- $x_3$
- $\ldots$
- $x_{n-1}$
- $x_n$

“anonymization”
“sanitization”
“perturbation”

Privacy!
Welcome To The Machine

(Dynamically changing) inputs

Information leaks out gradually, through unexpected system interfaces
Calandrino, Kilzer, Narayanan, Felten, Shmatikov
“You Might Also Like:” Privacy Risks of Collaborative Filtering
Oakland 2011
Recommender Systems
Collaborative Filtering

Recommender System

Recommendations:

- ALF
- Parks and Recreation
- Arrested Development
Can We Invert This?

Recommender System

Recommendations:
- ALF
- Parks & Recreation
- Arrested Development
Item-to-Item Recommendations

...versus user-to-item recommendations

This allows for passive adversaries
Our Tools

Auxiliary information and dynamics of public outputs
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
Predictions (recommendations) seek to impose order
Inferences are based on temporal changes in order.
More Recent Work

McPherson, Shokri, Shmatikov
“Defeating Image Obfuscation with Deep Learning”
arXiv 2016
What does this revolution mean for a privacy researcher?
Outperform Humans, Eh?

Does this extend to obfuscated images?

Truck  Truck  Truck  Truck

Truck  Truck?  ???  ?????????

Truck  Truck  Truck  Truck
Image Obfuscation Techniques

- Pixelation
- Blurring
- P3

As deployed by YouTube
Add privacy to online image storage by encrypting “significant” JPEG coefficients
Privacy/Size Tradeoff in P3

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…
## More Targets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Original</th>
<th>2 × 2</th>
<th>4 × 4</th>
<th>8 × 8</th>
<th>16 × 16</th>
<th>P3</th>
<th>10</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNIST</td>
<td><img src="image" alt="2" /></td>
<td><img src="image" alt="2" /></td>
<td><img src="image" alt="Pixel" /></td>
<td><img src="image" alt="Mosaic" /></td>
<td><img src="image" alt="20" /></td>
<td><img src="image" alt="10" /></td>
<td><img src="image" alt="1" /></td>
<td></td>
</tr>
<tr>
<td>CIFAR-10</td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td>AT&amp;T</td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td></td>
</tr>
<tr>
<td>FaceScrub</td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td><img src="image" alt="Image" /></td>
<td></td>
</tr>
</tbody>
</table>

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