VirtualIdentity: a Privacy-Preserving User Profiling Service

Sisi Wang
sisiwang@uw.edu
M.S. in Computer Science

Faculty Advisors:
Dr. Martine De Cock & Dr. Anderson Nascimento

The Secure ML team:
Wing-Sea Poon, Golnoosh Farnadi, Caleb Horst,
Kebra Thompson, Michael Nickels, Chris Allan Vishoot & Such Kamal

http://secureml.insttech.washington.edu/
American Express
Taking advantage of post-holiday sales? Use Membership Rewards points for Mobile Gift Cards to select retailers right on your smartphone! http://aexp.co/ruY

NEW! MOBILE GIFT CARDS

Browse Mobile Gift Cards
View Your Gift Card Wallet (4)
User Profiling

User Generated Content: posts, tweets, status, comments, images, videos, chats, likes, ...

Users of Online System or Service

Machine Learning

Gender, Age, Personality, Religion, Sexual Orientation, Interest,...
"On the Internet, nobody knows you're a dog."
IT USED TO BE THAT NOBODY ON THE INTERNET KNEW THAT I WAS A DOG. NOW, BECAUSE OF BIG DATA, EVERYBODY KNOWS THAT I AM A 15-YEAR OLD LABRADOODLE WHO SECRETLY LIKES CAT FOOD.
Problem?
Privacy-Preserving
VirtualIdentity -- Overview
Secure Two Party Computations

By the end of the computation each player only knows the output and its own input.

Cryptographic Protocols

- Decompose **Machine Learning Scoring** operation into smaller and simpler operations
- Use **Secure Multi-party Computation** to make each simple operation privacy-preserving
- Combine the Secure Multi-party computations to get the Machine Learning Scoring privacy-preserving
Support Vector Machine Scoring

Scoring for new instance $x_q$:

$$f(x_q) = \text{sign} \left( \sum_i \alpha_i y_i K(x_q, x_i) \right)$$

- Multiplication
- Addition
- Comparison
- ...
Let $\ell$ be the bit length of the integers to be compared. The trusted initializer pre-distributes the correlated randomness necessary for the execution of all instances of the distributed multiplication protocol. The parties have as inputs shares $[x_i]_2$ of each bit of $x$ and shares $[y_i]_2$ of each bit of $y$. The protocol proceeds as follows:

1. For $i = 1, \ldots, \ell$, compute $[d_i]_2 \leftarrow [y_i]_2 \left(1 - [x_i]_2\right)$ using the multiplication protocol $\pi_{DM}$ and locally compute $[e_i]_2 \leftarrow [x_i]_2 + [y_i]_2 + 1$.

2. For $i = 1, \ldots, \ell$, compute $[c_i]_2 \leftarrow [d_i]_2 \prod_{j=i+1}^{\ell} [e_j]_2$ using the multiplication protocol $\pi_{DM}$.

3. Compute $[w]_2 \leftarrow 1 + \sum_{i=1}^{\ell} [c_i]_2$ locally.
We hope to achieve highly practical results that allow the benefits of machine learning to be unlocked without the cost of individual privacy.

Check our website:  http://secureml.institutech.washington.edu/
Sisi Wang
M.S. in Computer Science
sisiwang@uw.edu