Privacy Attacks

Ashwin Machanavajjhala
ashwin@cs.duke.edu

Privacy breaches on the rise…

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr.
Published: August 9, 2006

SIGN IN TO E
THIS

NETFLIX

Why 'Anonymous' Data Sometimes Isn't

By Bruce Schneier 12.13.07

Last year, Netflix published 10 million movie rankings by 500,000 customers, as part of a challenge for people to come up with better recommendation systems than the one the company was using.

The New York Times

Business Day
Technolo

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HE

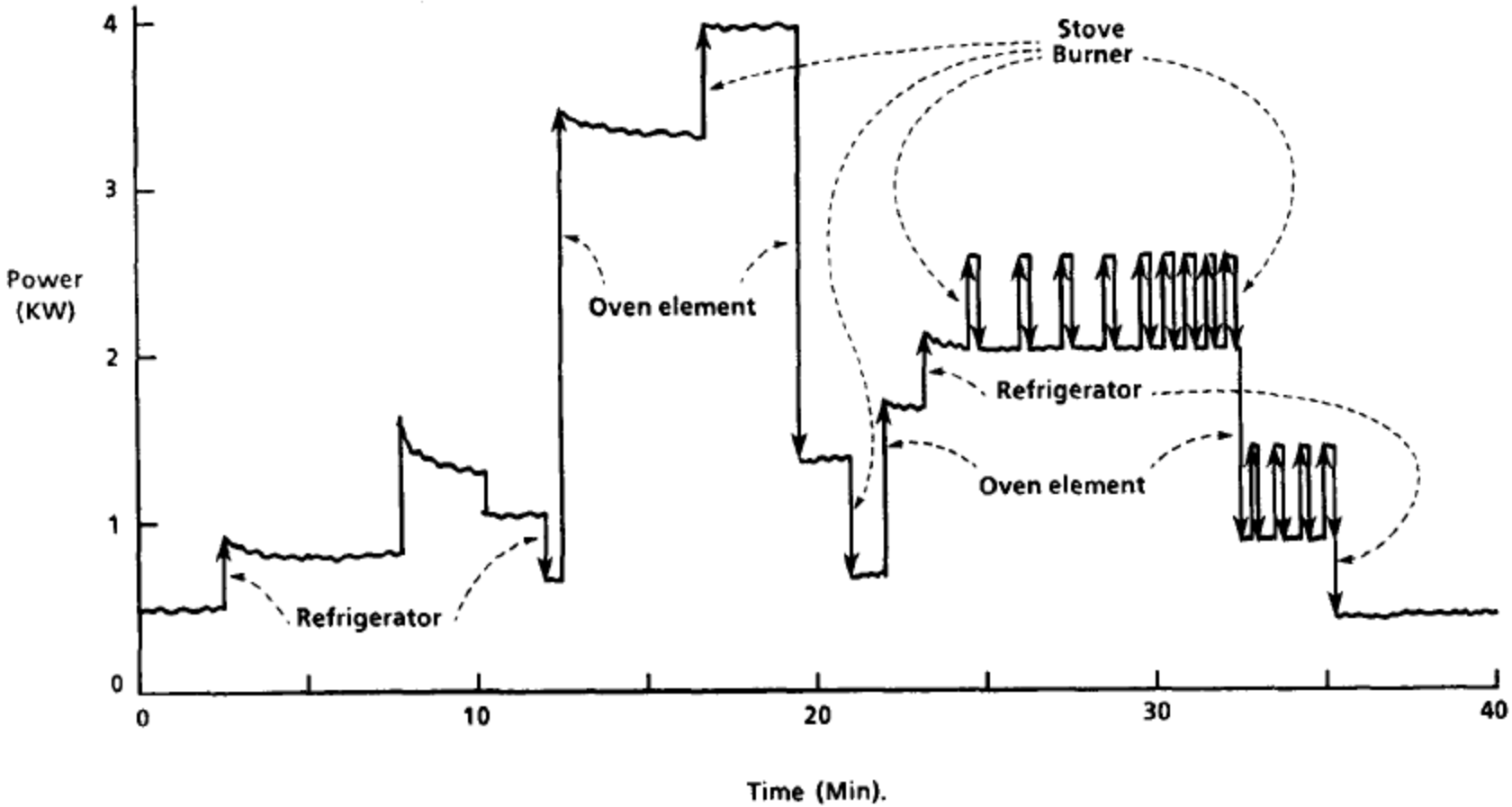
Marketers Can Glean Private Data on Facebook

TARGET

TECH | 2/16/2012 @ 11:02AM | 837,678 views

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

Energy patterns



[Hart IEEE S&P '92]

Energy Patterns disclose Private Information

Question	Pattern	Granularity
Were you home during your sick leave?	Yes: Power activities during the day No: Low power usage during the day	Hour/Minute
Did you get a good night's sleep?	Yes: No power events overnight for at least 6 hours No: Random power events overnight	Hour/Minute
Did you watch the game last night?	Yes: Appliance activity matching TV program No: No power event in accordance with game showtime	Minute/Second
Did you leave late for work?	Yes: Last power event time later than Google maps estimated travel time No: Last power event time leaves enough time for commute	Minute
Did you leave your child home alone?	Yes: Single person activity pattern No: Simultaneous power events in distinct areas of the house	Minute/Second
Do you eat hot or cold breakfast?	Hot: Burst of power events in the morning (microwave/coffee machine/toaster) Cold: No power event matching hot breakfast appliances	Second

[Molina et al BuildSys '10]

Outline

- Removing identifiers is not sufficient
  - Massachusetts Governor Privacy Breach
- Releasing “unsafe” data (to the public) can be a (PR) disaster
  - AOL Search Log Fiasco
  - Netflix Prize Data Deanonymization
- Aggregated Data also leak Private Information
  - Background knowledge attacks
  - Active Attacks

The Massachusetts Governor Privacy Breach

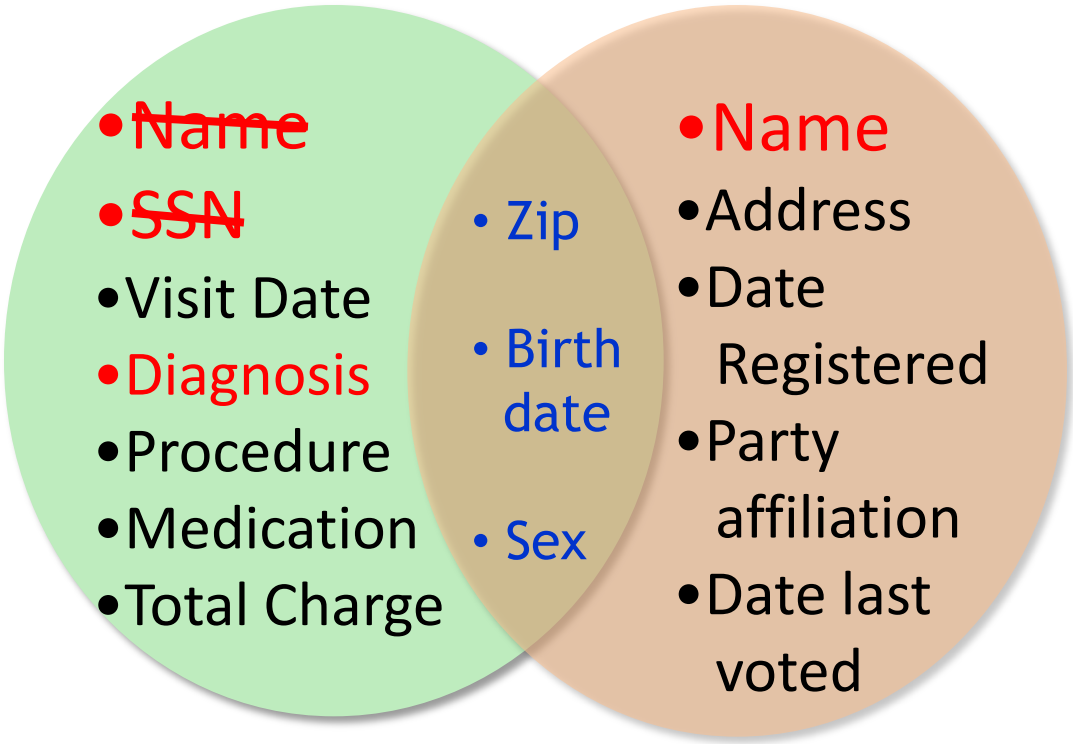
- ~~Name~~
- ~~SSN~~
- Visit Date
- Diagnosis
- Procedure
- Medication
- Total Charge

- Zip
- Birth date
- Sex

Medical Data

[Sweeney IJUFKS 2002]

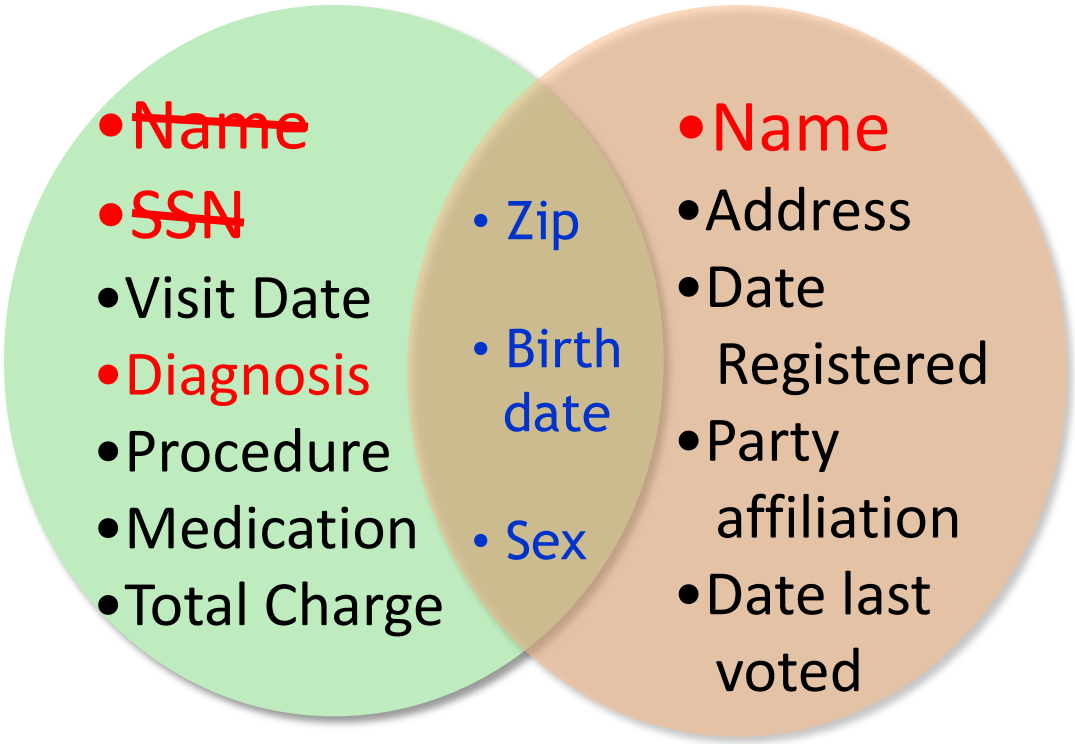
The Massachusetts Governor Privacy Breach



Medical Data **Voter List**

[Sweeney IJUFKS 2002]

The Massachusetts Governor Privacy Breach

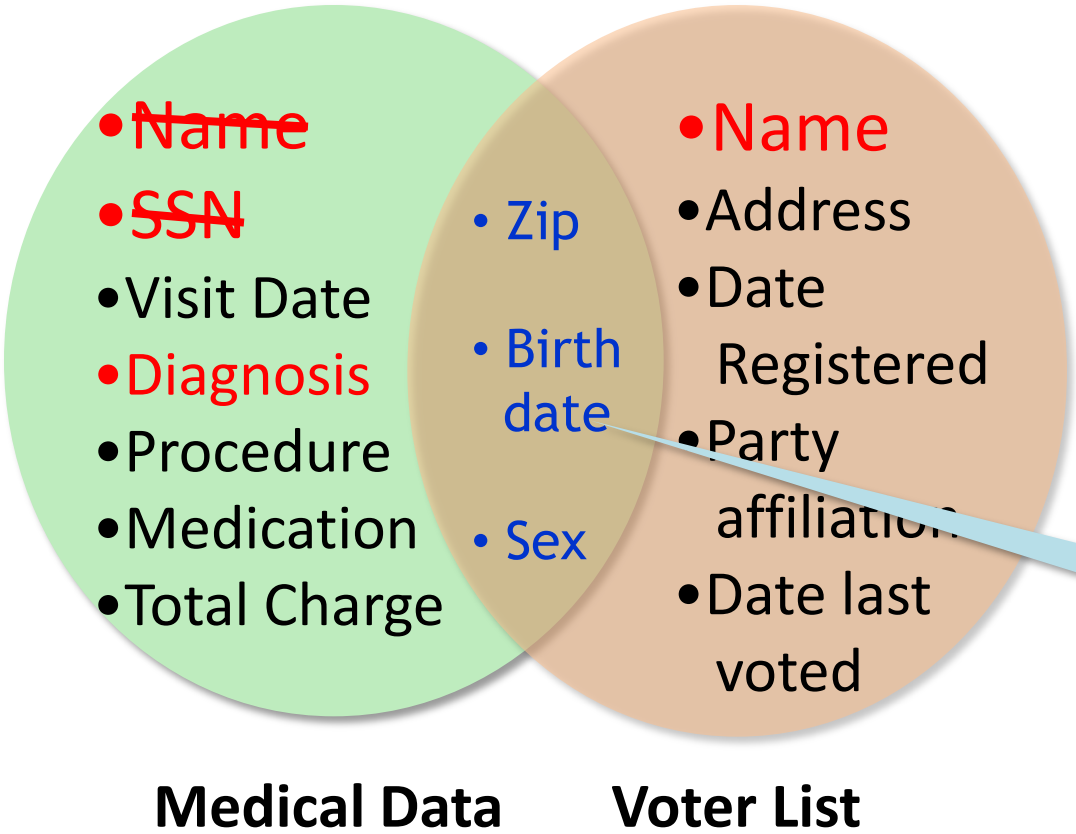


- Governor of MA **uniquely identified** using ZipCode, Birth Date, and Sex.

Name linked to Diagnosis

[Sweeney IJUFKS 2002]

The Massachusetts Governor Privacy Breach



- 87 % of US population **uniquely identified** using ZipCode, Birth Date, and Sex.

Quasi Identifier

[Sweeney IJUFKS 2002]

Quasi-identifiers in Energy

- Set of appliances
- Pattern of appliance usage
- Sleep patterns
- …

AOL data publishing fiasco …

AOL “anonymously” released a list of 21 million web search queries.

Ashwin222	Uefa cup
Ashwin222	Uefa champions league
Ashwin222	Champions league final
Ashwin222	Champions league final 2007
Pankaj156	exchangeability
Pankaj156	Proof of deFinitti’s theorem
Cox12345	Zombie games
Cox12345	Warcraft
Cox12345	Beatles anthology
Cox12345	Ubuntu breeze
Ashwin222	Grammy 2008 nominees
Ashwin222	Amy Winehouse rehab

AOL data publishing fiasco …

AOL "anonymously" released a list of 21 million web search queries.

UserIDs were replaced by random numbers …

865712345	Uefa cup
865712345	Uefa champions league
865712345	Champions league final
865712345	Champions league final 2007
236712909	exchangeability
236712909	Proof of deFinitti's theorem
112765410	Zombie games
112765410	Warcraft
112765410	Beatles anthology
112765410	Ubuntu breeze
865712345	Grammy 2008 nominees
865712345	Amy Winehouse rehab

Privacy Breach

[NYTimes 2006]

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr.
Published: August 9, 2006

SIGN IN TO E-
THIS

Netflix Prize Data

Why 'Anonymous' Data Sometimes Isn't

By Bruce Schneier 12.13.07

Last year, Netflix published 10 million movie rankings by 500,000 customers, as part of a challenge for people to come up with better recommendation systems than the one the company was using. The data was anonymized by removing personal details and replacing names with random numbers, to protect the privacy of the recommenders.

Arvind Narayanan and Vitaly Shmatikov, researchers at the University of Texas at Austin, de-anonymized some of the Netflix data by comparing rankings and timestamps with public information in the Internet Movie Database, or IMDb.

NETFLIX

[Narayanan-Shmatikov S&P 2008]

Aggregation

- One possible solution is to aggregate energy data from sets of users.

- Such simple aggregation could also leak sensitive information
  - Attacker may know a unique combination of appliances in target individual's house
  - Attacker can monitor power patterns based on these unique appliances, and hence infer private information

  - *Similar to the attacks on Genome databases*

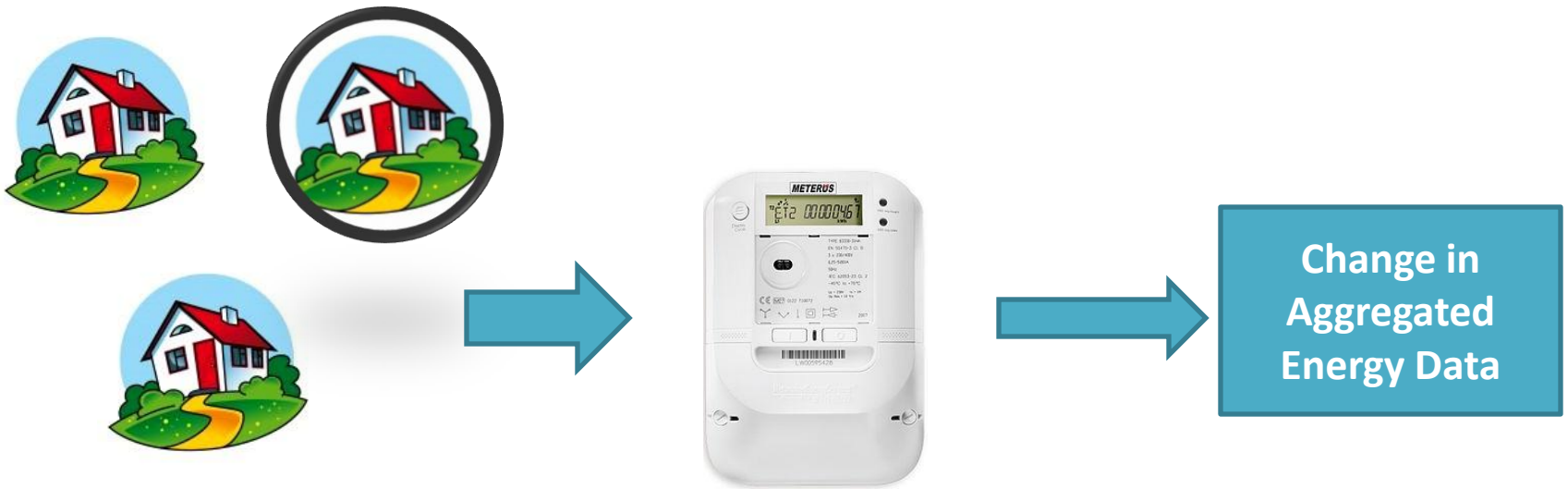
Composition Attack

- If the aggregation set changes, then individuals can be uniquely identified.

Aggregated Energy Data

Composition Attack

- If the aggregation set changes, then individuals can be uniquely identified.



- Attacker can use this change to infer new house's energy data.

Summary

- Removing identifiers is not sufficient
  - Massachusetts Governor Privacy Breach
- Releasing “unsafe” data (to the public) can be a (PR) disaster
  - AOL Search Log Fiasco
  - Netflix Prize Data Deanonymization
- Aggregated Data also leak Private Information
  - Background knowledge attacks
  - Active Attacks

References

- Sweeney, “K-Anonymity”, International Journal of Uncertainty Fuzzy Knowledge Systems 2002
- Hart, “Non-intrusive load monitoring”, Proceedings of IEEE 1982
- Molina-Markham, Shenoy, Fu, Cecchet, Irwin, “Private Memoirs of a Smart Meter”, BuildSys 2010
- Quinn, “Smart metering and privacy: Existing laws and competing policies” Colorado Public Utilities Commission 2009
- Narayanan, Shmatikov, “Robust Deanonymization of Sparse Datasets”, IEEE S&P 2008