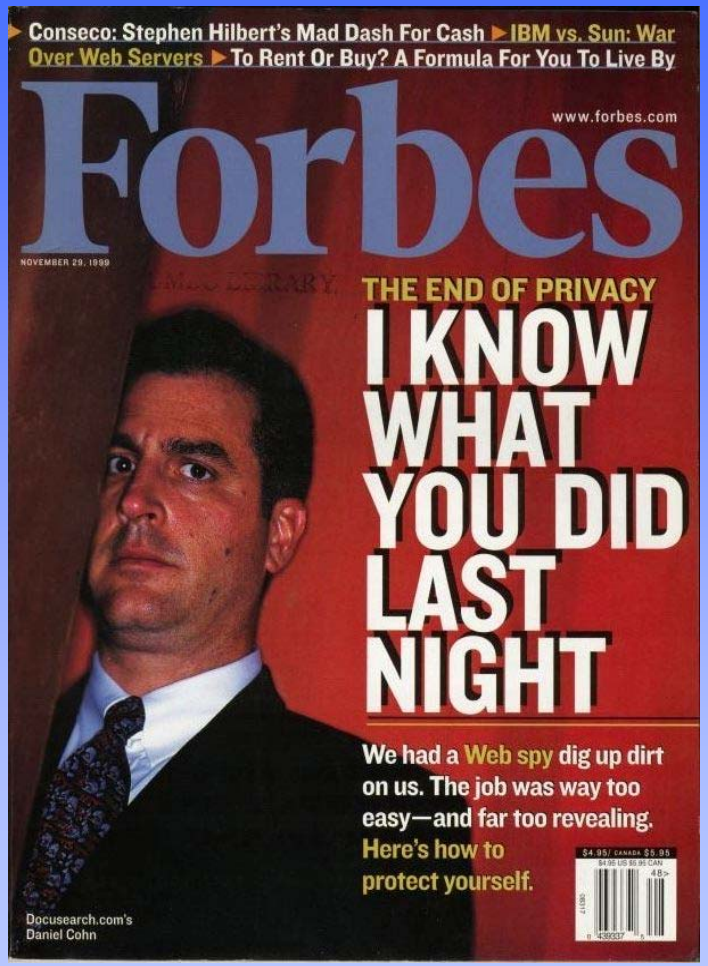
Economics of Privacy Preserving Data Mining

Hillol Kargupta
University of Maryland, Baltimore County
&
Agnik

Web: http://www.cs.umbc.edu/~hillol

Privacy: From Concerns to Value Creation



- **CDT Faults Guidelines for Terror Information Sharing** -- "A CDT analysis finds that privacy guidelines issued by the Bush Administration for the Information Sharing Environment are inadequate……."
  -- *Center for Democracy and Technology, February 5, 2007*

- **Surveillance Cameras Begin to Judge Pedestrians** – "Researchers and security companies are developing cameras that not only watch the world but also interpret what they see. Soon, some cameras may be able to find unattended bags at airports, guess your height…..
- *--- www.privacy.org, February 25, 2007.*

Types of Privacy Violation

- 1890 Harvard Law Review article written by Samuel D. Warren and Louis D. Brandeis on The Right of Privacy.
  - **Intrusion of solitude** - physical or electronic intrusion into one's private quarters.
  - **Public disclosure of private facts** -- the dissemination of truthful private information which a reasonable person would find objectionable
  - **False light** - the publication of facts which place a person in a false light, even though the facts themselves may not be defamatory.
  - **Appropriation** -- the unauthorized use of a person's name or likeness to obtain some benefit

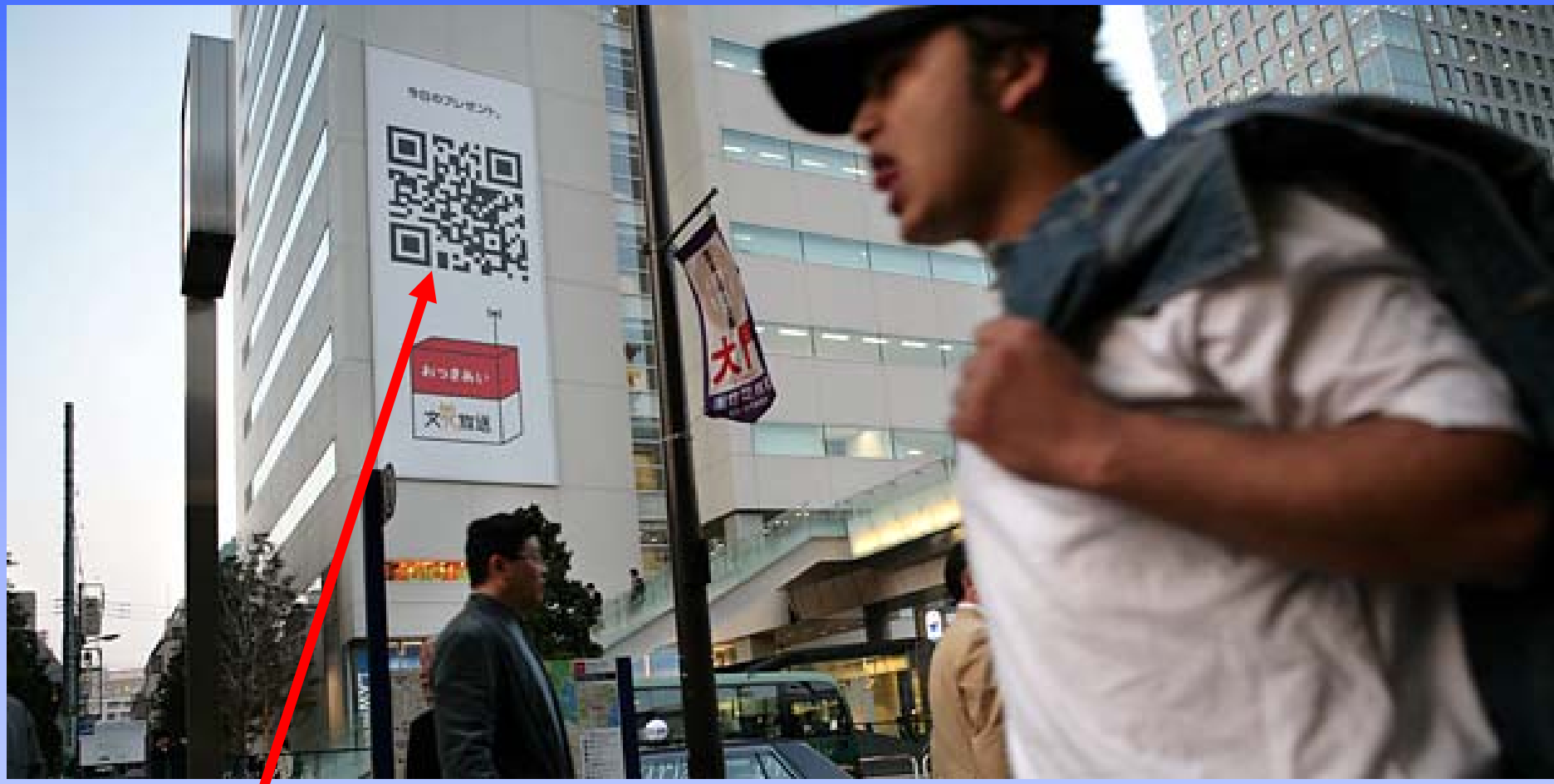
Predictions from Circa 1890

- Writing before the era of electronic eavesdropping, telephoto lenses, and other modern technology, Warren and Brandeis prophesied that

  "mechanical devices threaten to make good the prediction that `what is whispered in the closet shall be proclaimed from the housetops'."

  -- *Warren and Brandeis , 1890*

Ubiquitous Sensing from “Rooftops”: Circa 2007

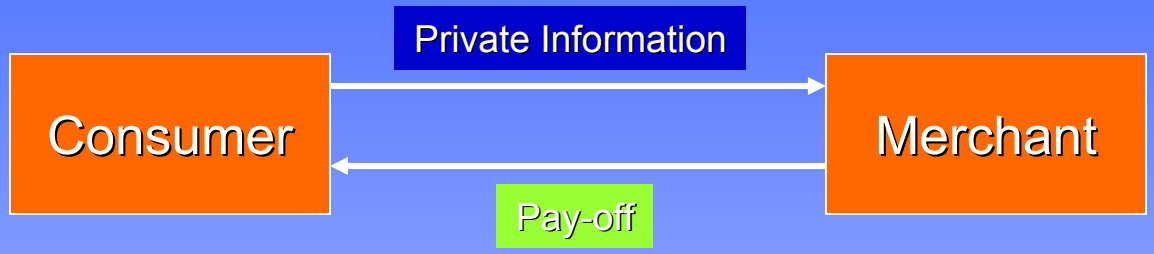


Barcode on top of a building that can be read by a cellphone camera

Anonymity and Privacy

- Can we really have anonymity in the world of ubiquitous computing?
- How about Privacy?

Privacy and Knowledge Discovery from Data



- Merchant offers some benefit (e.g. a coupon or personalized service)
- Consumer offers private data in return

What Makes a PPDM Technology Fly in Business?

$$B_u - C_u \gtrless B_n - C_n$$

- B_u = Benefit of using PPDM
- C_u = Cost of using PPDM
- B_n = Benefit of not using PPDM
- C_n = Cost of not using PPDM

Cost and Benefit of Using PPDM

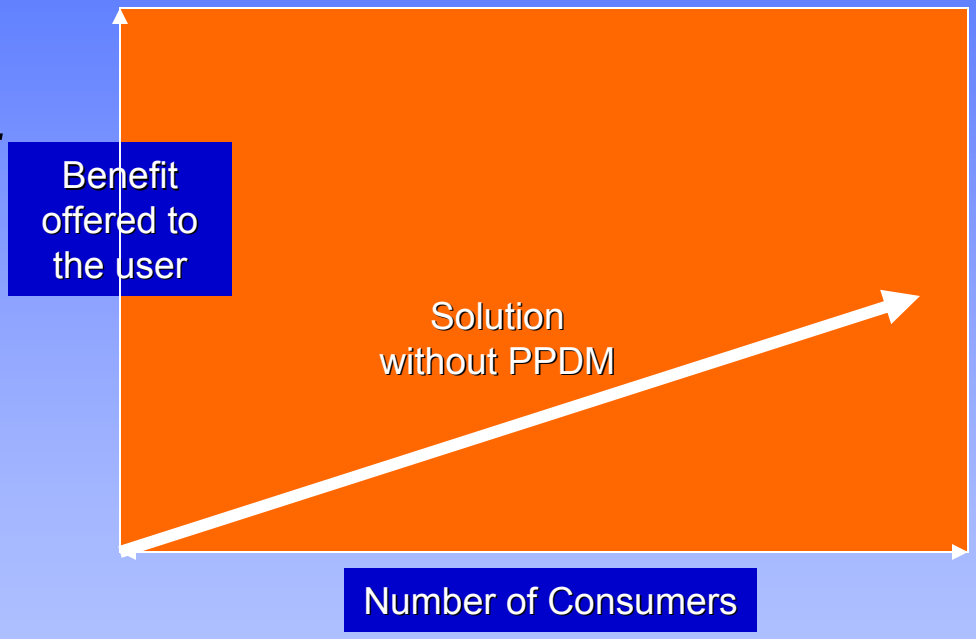
- Cost
  - Usually low running cost
  - But high initial investment for changing existing system and support mechanisms
- Benefit
  - Improved consumer confidence/satisfaction
  - Satisfying regulations (laws and self-imposed policies)

Cost and Benefit of not Using PPDM

- Cost
  - Less consumer confidence
  - High initial investment
- Benefit
  - Personalized services
  - No additional cost

How Concerned Are We About Privacy?

- Spiekermann, Sarah, Grossklags, Jens, Berendt, (2002) *Privacy preferences versus actual behavior.*
  - Individuals are less concerned about privacy then what they claim to be
  - Willing to share “very private” information in exchange of small rewards



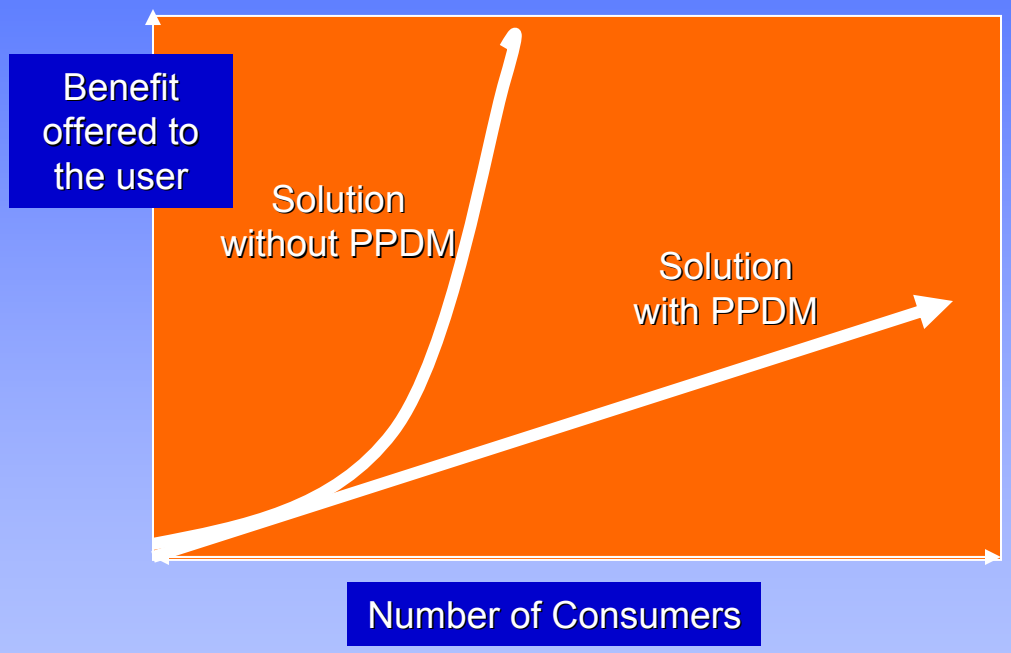
A Rational Analysis

- We think that the likelihood of a major harm (e.g. financial) from privacy-breach is very low.
- Myopic; need values in short-term

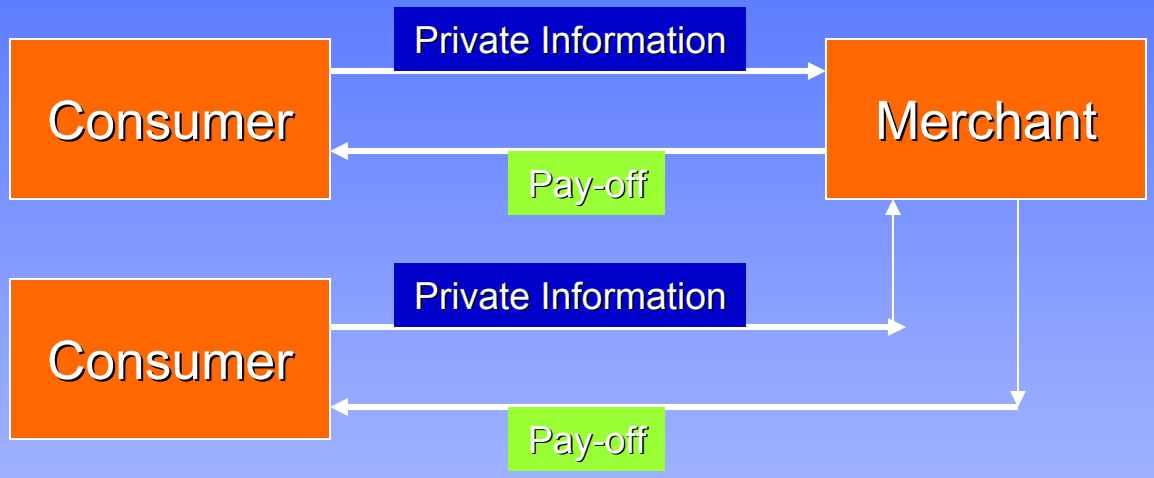
Hope for Market Equilibrium and be Happy!

- Acquisti & Varian, 2002
  - Market laws alone can produce pareto optimality in the business of selling customer information about customers' taste and purchase history to retailers.
- Taylor, 2002
  - Merchants naturally tend to protect the privacy of strategic customers
- Theory of natural convergence to equilibrium with optimal blend of privacy-protection and benefits from knowledge discovery

How About Making PPDM Products Fly High?



The Pay-Off Model: One Possibility



- Pay-Off is a function of the information received from multiple parties who also own the data.
- Also a function of multiple parties

Case Study: PURSUIT

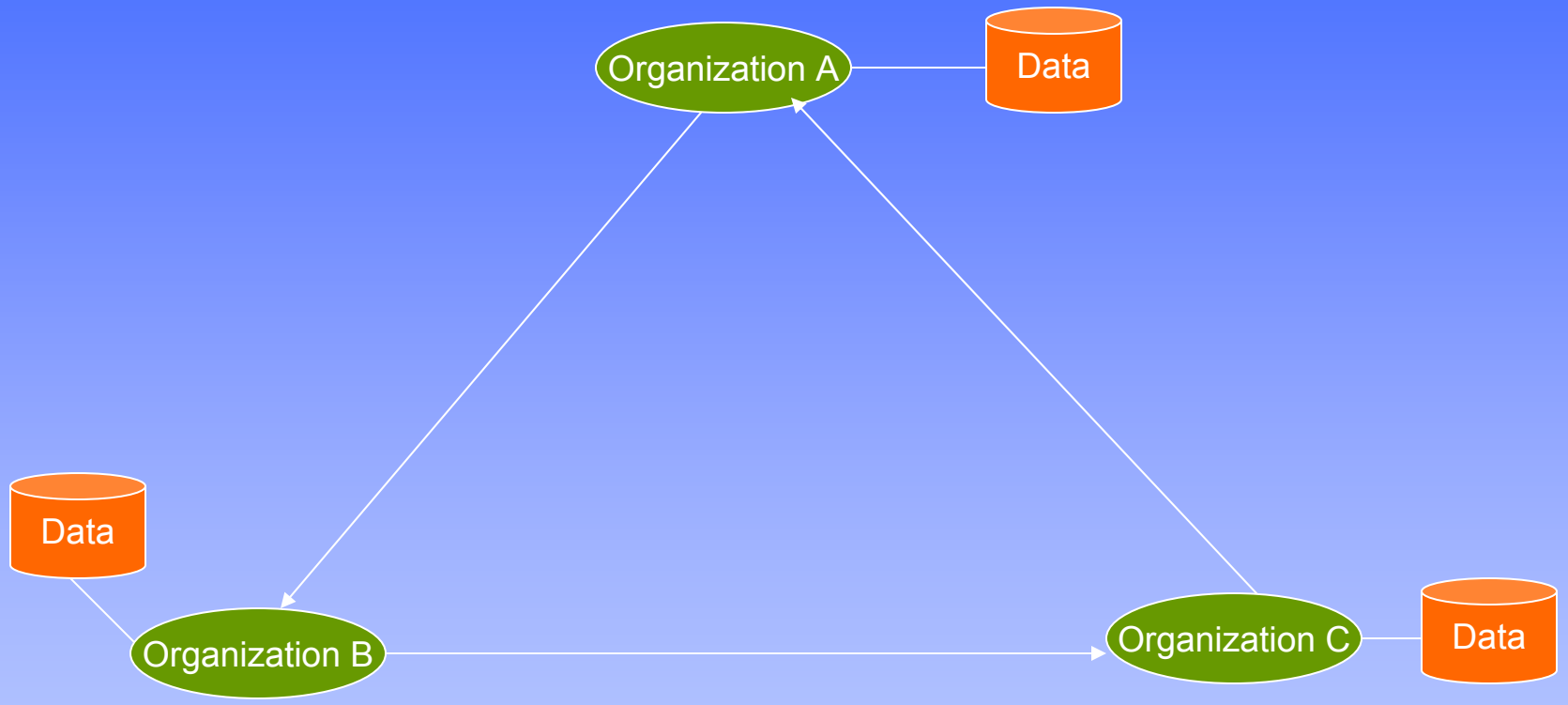
- **PURSUIT: Privacy-Sensitive Cross-Domain Network Threat Detection**

- **Cross-Domain Network Attack Detection system using Privacy-Preserving Distributed Data Mining**

- **Partners:**
  - Agnik, University of Minnesota, Tresys Technology, LLC.

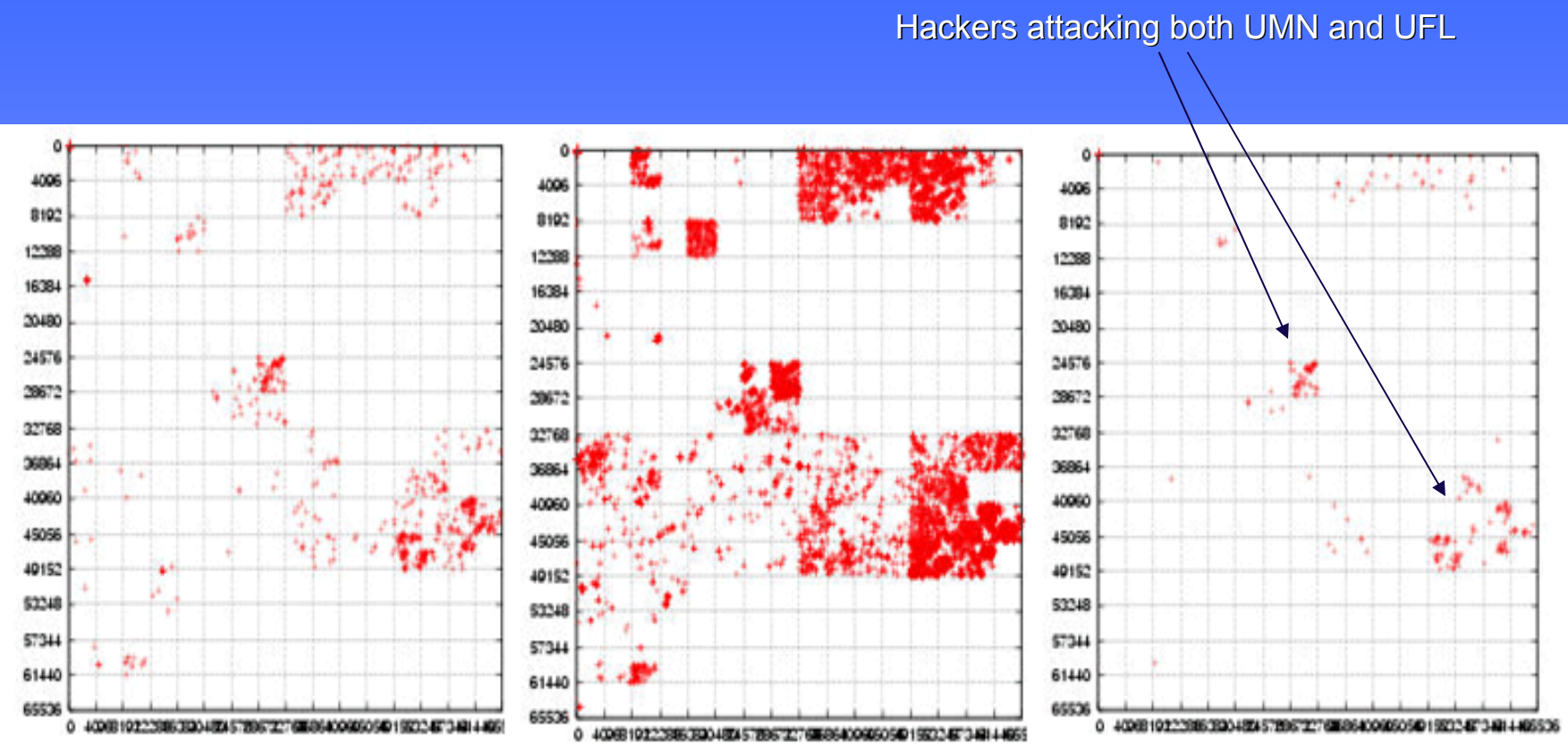
- **PURSUIT Early Consortium:**
  - Purdue University, Ohio State University, Stevens University, SRI International, University of Illinois at Urbana-Champaign

PURSUIT for Cross-Organizational Privacy-Preserving Data Analysis



Compare, match, and analyze data from different organizations without disclosing the private data to any other party

Motivation: Cross-Domain Attacks



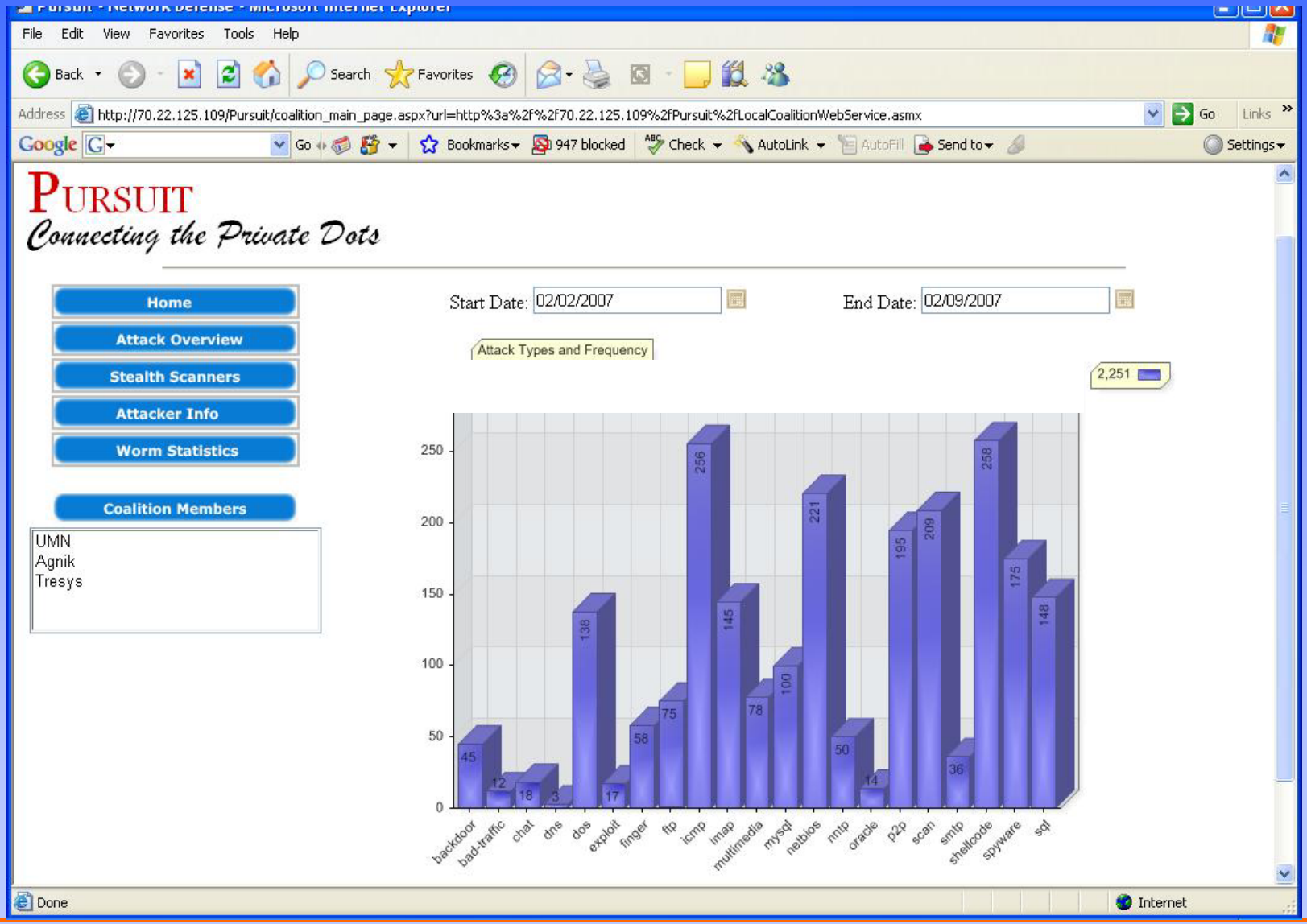
- Spatial Attack Distribution of IPs on the Same Day: (Left) IPs attacking the UFL network on 12/09/04 (712 scanners). (Middle) IPs attacking the UMN network on 12/09/04 (14,938 scanners). (Right) Intersection of the IPs attacking UFL and UMN (201 scanners). *Courtesy: Vipin Kumar, UMN*

How PURSUIT Works for the User

- **Download PURSUIT plug-in for the existing network monitoring sensors such as SNORT, MINDS and install**

- **PURSUIT plug-in offers**
  - A stand-alone interface for processing your alerts from the sensor and cross-domain analysis
  - Web account for detailed cross-domain statistics
  - Optional distributed collaboration management module for managing the threats and archiving forensics

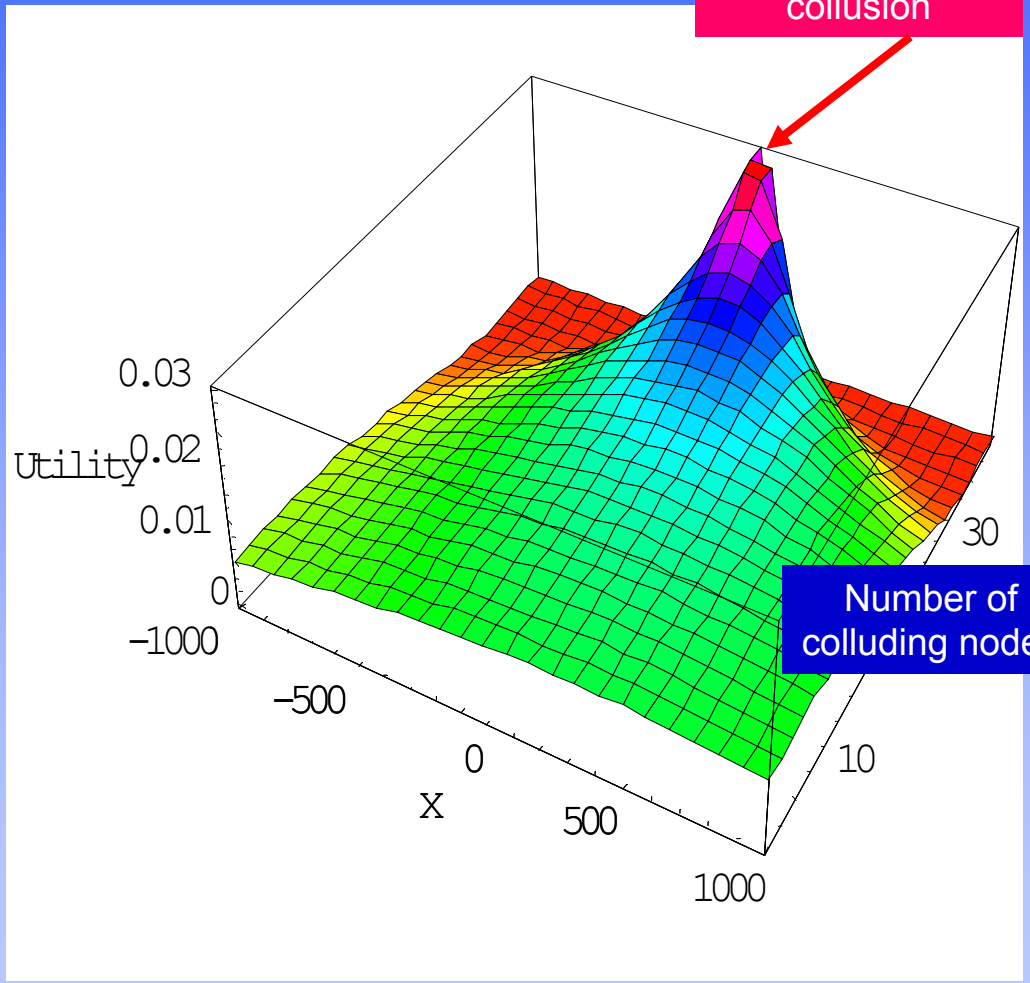
PURSUIT Web Site



Some Technical Challenges for PPDM that may Work

- Multiparty environment
- Communication efficiency
- Forget assumptions about the user behavior.
  - Performs computations and communications as expected
  - Semi-honest
  - No collusion

Utility Function for Secure Multi-Party Sum with Collusion



If you implement Secure Sum in a real distributed multi-party environment then you will most likely have collusion

Game Theoretic Perspective

- Multi-Party PPDM as games
- Related Work:
  - Halpern and Teague, 2004 explored Shamir's secret sharing problem from game theoretic perspective
  - Zhang et al, 2005
  - Abraham et al., 2006 extended their earlier work introduced generalized notion of nash equilibrium
  - Kargupta, Das, Liu, 2006

Conclusions

- Time to produce “killer apps” and products.
- Need to review the current effort on PPDM algorithmic research and identify what makes sense in the real-world.
- Get rid of assumptions that we have been using for long time.
- Economics, game theory may be useful.