DEAKIN UNIVERSITY AUSTRALIA

Networking for Big Data
- Challenges and Opportunities

Dr Shui Yu (余水)
School of Information Technology
Deakin University, Melbourne, Australia
http://www.deakin.edu.au/~syu
Email: syu@deakin.edu.au

About Deakin, SIT, and Melbourne

- Deakin University is ranked 214 worldwide by ARWU.
- School of IT is ranked 123 worldwide by ARWU.
- At SIT, we are very good at networking and cybersecurity.
- Melbourne is the most liveable city in the world.
- We welcome outstanding students for various scholarship applications and visits.

DEAKIN UNIVERSITY AUSTRALIA

Outline

- Introduction
- Big Data Modelling
- Networking for Big Data
- Big Data Security
- Big Data Privacy
- Q&A

1. Big Data: an Introduction

Our understanding of a subject.

Englishman : What is the name of the animal?

Aboriginal: Kangaroo

1. Big Data: an Introduction

- What is Big Data?
- 3V, 4V, 5V…
- Big data, big problem (Science, 2014)
- The end of privacy in big data era (Science, 2015 January)

1. Big Data: an Introduction

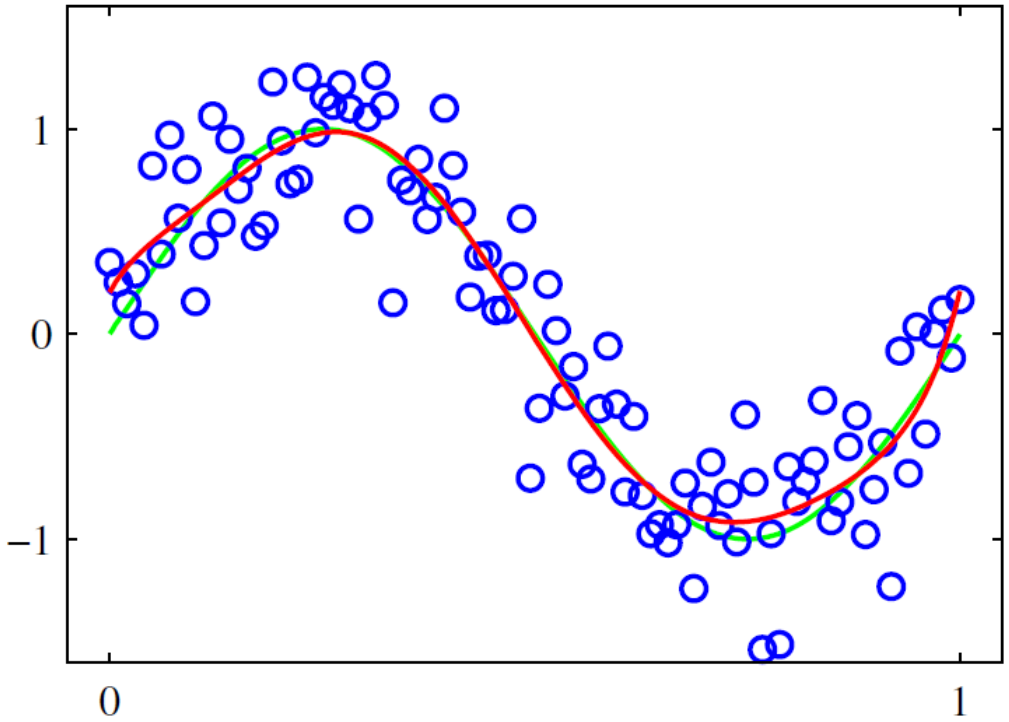
Big data applications

– Artificial Intelligence, Machine Learning, …

- Neural networks
- Deep learning
- Reinforce learning
- …

1. Big Data: an Introduction

My understanding of big data



DEAKIN UNIVERSITY AUSTRALIA

1. Big Data: an Introduction

What are we doing in terms of Big Data?

- at the application layer.

Is it enough?

No

1. Big Data: an Introduction

Preconditions for current mining and learning tools

- Data are “local”
- Data are homogeneous

Shui Yu, Meng Liu, Wanchun Dou, Xiting Liu, and Sanming Zhou, “Networking for Big Data: A Survey," IEEE Communications Surveys and Tutorials, 2017.

DEAKIN UNIVERSITY AUSTRALIA

1. Big Data: an Introduction

Networking is the “hardware” for Big Data

- Big data must be distributed.
- Big data is heterogeneous
- Big data is expected to be real-time

DEAKIN UNIVERSITY AUSTRALIA

1. Big Data: an Introduction

Old stories in a new environment

- Resource management
- Job scheduling
- Networking
- Security
- …

DEAKIN UNIVERSITY AUSTRALIA

2. Big Data Modelling

Features of Big Data

- Big data representation
- Big data property
- Big data analysis

DEAKIN UNIVERSITY AUSTRALIA

2. Big Data Modelling

From graph theory

- Big data to Big graph
- Make big graph smaller, simpler until computable.
- Tools: graph summary, others?

DEAKIN UNIVERSITY AUSTRALIA

2. Big Data Modelling

From matrix

- Matrix processing
- Tools: tensor

DEAKIN UNIVERSITY AUSTRALIA

2. Big Data Modelling

From statistics

- Large n, small p (traditional)
- Large n, large p (big data era)
- Research on small probability, rare events.
- This research somewhat contradicts the foundation of statistics

- Large deviation

DEAKIN UNIVERSITY AUSTRALIA

3. Networking for Big Data: Infrastructure

Big Data infrastructure research

- Cloud (fog)
- data center networks
- Space-territorial networks
- Software defined networks

DEAKIN UNIVERSITY AUSTRALIA

3. Networking for Big Data: Platforms

Big data software platform

- MapReduce
- Hadoop
- New concepts, new strategies

 - big task division

 - co-flow scheduling

DEAKIN UNIVERSITY AUSTRALIA

3. Networking for Big Data: Platforms

Big data division

- Division metrics? Feature-based?
- Graph division + big data features.
- Power law distribution ?

 - Size of data may somewhat power law.

 - Size of tasks not uniform (power law?)

DEAKIN UNIVERSITY AUSTRALIA

3. Networking for Big Data: Platforms

Big data scheduling

- feature: multiple subtask, multiple phase
- coflow concept

DEAKIN UNIVERSITY AUSTRALIA

3. Networking for Big Data: Platforms

Big data scheduling mathematical tools

- Fork-join system
- Network calculus

- Revised queueing theory ? feedback

4. Big Data Security

- Cryptograph
  - Performance
  - Cost
- non-cryptograph
  - Anomaly detection (mining methods)
  - Privacy protection (noise, cost)

DEAKIN UNIVERSITY AUSTRALIA

4. Big Data Security

Many questions to be studied, answered, and updated

1. Who are the bad guys?
2. How many bad guys?
3. What are the structures of their networks?
4. How do they recruit members and commit crimes?
5. Where are they?
6. Is there a lower bound for security?

…

4. Big Data Security

We use our research on attacking and defense (mainly in Distributed Denial of Service) as an example to study the landscape of cybersecurity in the age of big data

There are three categories in cybersecurity

- Detection
- Mitigation
- Traceback

4.1. Big Data Security - Detection

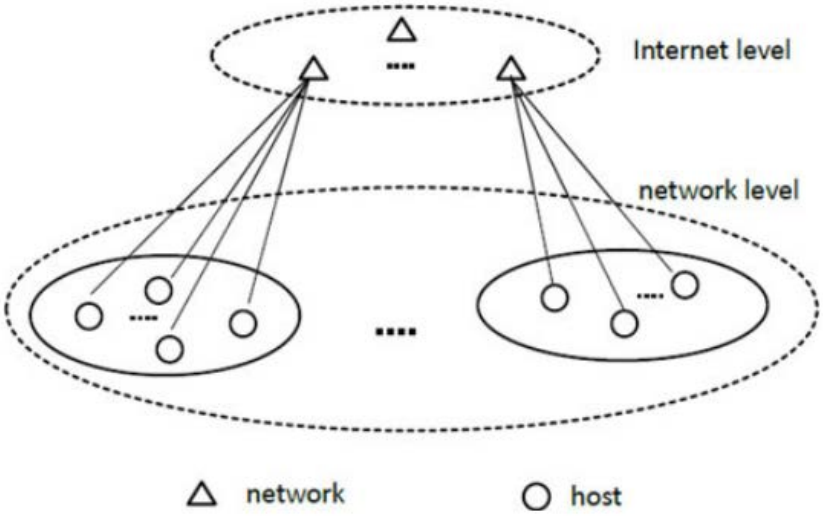
We found the rule of malware distribution

- Exponential distribution at the early stage.
- Power law distribution with a short exponential tail at the late stage
- Power law distribution at the final stage.

Shui Yu, Guofei Gu, Ahmed Barnawi, Song Guo, and Ivan Stojmenovic, "Malware Propagation in Large-Scale Networks," IEEE Transactions on Knowledge and Data Engineering, Vol. 27, Issue 1, 2015, pp. 170-179.

4.1. Big Data Security - Detection

We used the Epidemic theory and a proposed two layer model.



Shui Yu, Guofei Gu, Ahmed Barnawi, Song Guo, and Ivan Stojmenovic, "Malware Propagation in Large-Scale Networks," IEEE Transactions on Knowledge and Data Engineering, Vol. 27, Issue 1, 2015, pp. 170-179.

4.1. Big Data Security - Detection

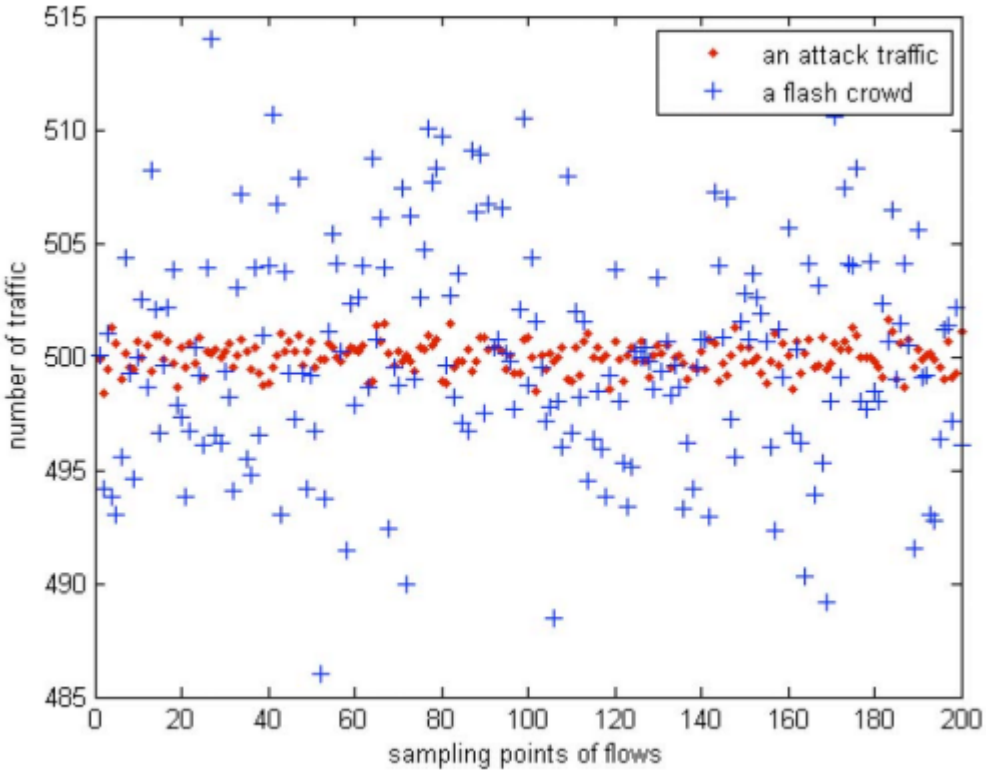
Discriminate mimicking attack from flash crowd

- It is hard to deal with mimicking attacks
- The resources that hackers have usually limited
- It is critical to find the features to differentiate them

Shui Yu, Wanlei Zhou, Weijia Jia, Song Guo, Yong Xiang, and Feilong Tang, “Discriminating DDoS Attacks from Flash Crowds Using Flow Correlation Coefficient,” IEEE Transactions on Parallel and Distributed Systems, Vol. 23, Issue 6, June 2012, pp. 1073-1080.

4.1. Big Data Security - Detection

We invented a second order statistics method to discriminate mimicking attack from flash crowd



4.2. Big Data Security - Mitigation

We found the essential of cyber battle.

- It is a competition of resources
- the winner is the party who has relatively more resources than the other party.

Shui Yu, Song Guo, and Ivan Stojmenovic, "Fool Me If You Can: Mimicking Attacks and Anti-attacks in Cyberspace," *IEEE Transactions on Computers,* Vol. 64 Issue 1, 2015, pp.139-151.

4.2. Big Data Security - Mitigation

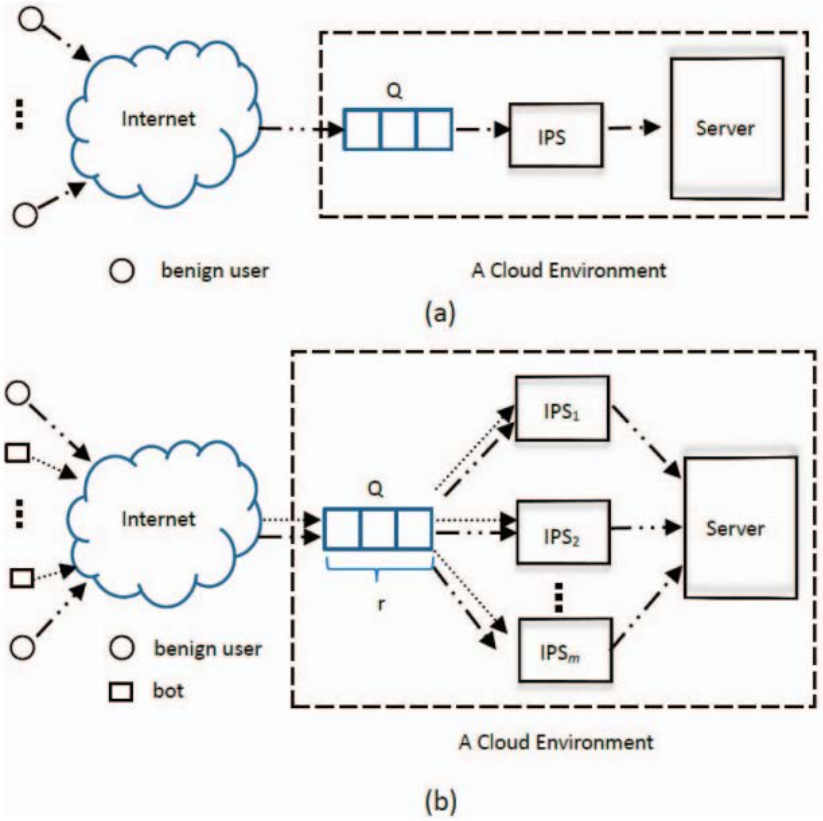
Following the previous finding, can we beat DDoS attacks or not?

- It is very hard in the traditional Internet
- But, we can in clouds.

Shui Yu, Yonghong Tian, Song Guo, and Dapeng Oliver Wu, "Can We Beat DDoS Attacks in Clouds?" *IEEE Transactions on Parallel and Distributed Systems,* vol 25, no 9, 2014, pp.2245-2254.

4.2. Big Data Security - Mitigation

We can beat DDoS attacks in clouds in terms of resource and cost



4.2. Big Data Security - Mitigation

- The strategy has been adapted by Amazon, please see Amazon white paper 2016.
- They named it “Auto Scale”

Shui Yu, Yonghong Tian, Song Guo, and Dapeng Oliver Wu, "Can We Beat DDoS Attacks in Clouds?" *IEEE Transactions on Parallel and Distributed Systems,* vol 25, no 9, 2014, pp.2245-2254.

4.3. Big Data Security - Traceback

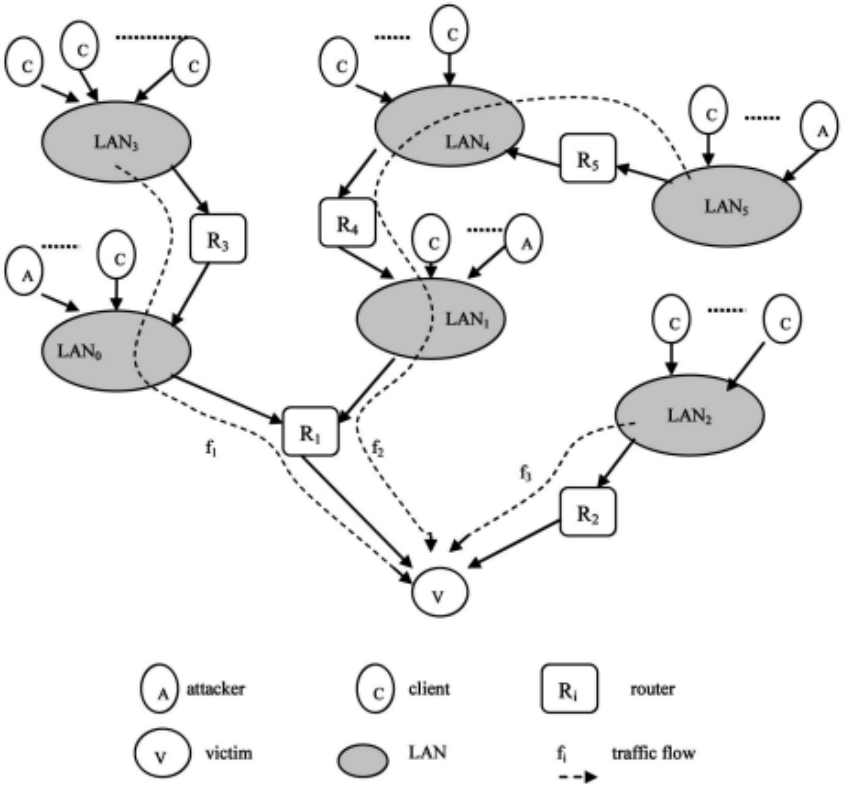
We invented a flow entropy based traceback method

- New methodology besides PPM and DPM
- But hard to implement in the Internet environment

Shui Yu, Wanlei Zhou, Robin Doss, and Weijia Jia, “Traceback DDoS Attacks using Entropy Variations,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, No. 3, March, 2011, pp. 412-425.

4.3. Big Data Security - Traceback

We invented a flow entropy based traceback method



4.3. Big Data Security - Traceback

We invented a feasible packet marking strategy for traceback

- Addressed the scalability problem of DPM
- Not every internet router is evolved in an attack.
- Using a round-robin method to utilize the marking space resource.

Shui Yu, Wanlei Zhou, Song Guo, and Minyi Guo, "A Feasible IP Traceback Framework through Dynamic Deterministic Packet Marking," IEEE Transactions on Computers, vol 65, no 5, 2016.

4.3. Big Data Security - Traceback

Some other traceback methods we developed.

Shui Yu, Keshav Sood, and Yong Xiang, "An Effective and Feasible Traceback Scheme in Mobile Internet Environment," IEEE Communications Letters, vol 18, issue 11, 2014, pp. 1911-1914.

Jiaojiao Jiang, Sheng Wen, **Shui Yu**, Yang Xiang, and Wanlei Zhou, "K-center: An Approach on the Multi-source Identification of Information Diffusion," IEEE Transactions on Information Forensics and Security (accepted on August 2, 2015).

Jiaojiao Jiang, Sheng Wen, Shui Yu, Yang Xiang, and Wanlei Zhou, "Identifying Propagation Sources in Networks: State-of-the-Art and Comparative Studies," IEEE Communications Surveys and Tutorials, 2017.

5. Big Data Privacy

- Data Mining community- privacy aware mining
- Statistics community – large n, large p; small probability events.
- Social science community – policy, psychology, etc.

M. Jordan and T. Mitchell, “Machine learning: trends, perspectives, and prospects,” Science, vol. 349, no. 6245, pp. 255–260, 2015.

5. Big Data Privacy

Essential problem of big data privacy

- Data utility for public
- Privacy protection of data generators

Goal

A trade-off between the two perspectives

Shui Yu, “Big Privacy: Challenges and Opportunities in Privacy Study in the Age of Big Data,” *IEEE Access, 2017*.

5. Big Data Privacy

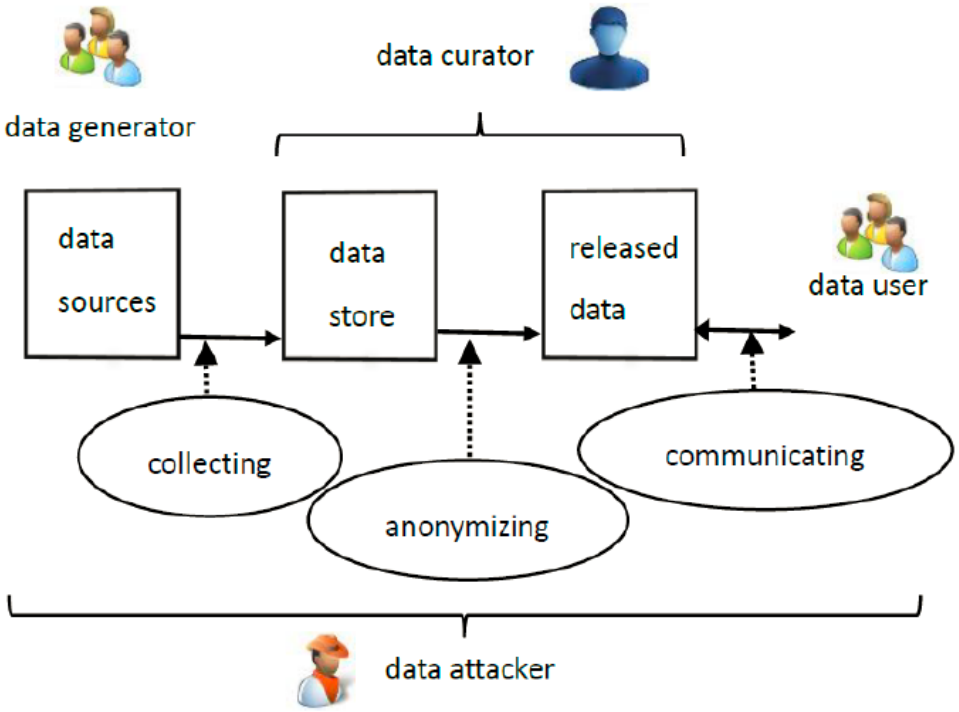
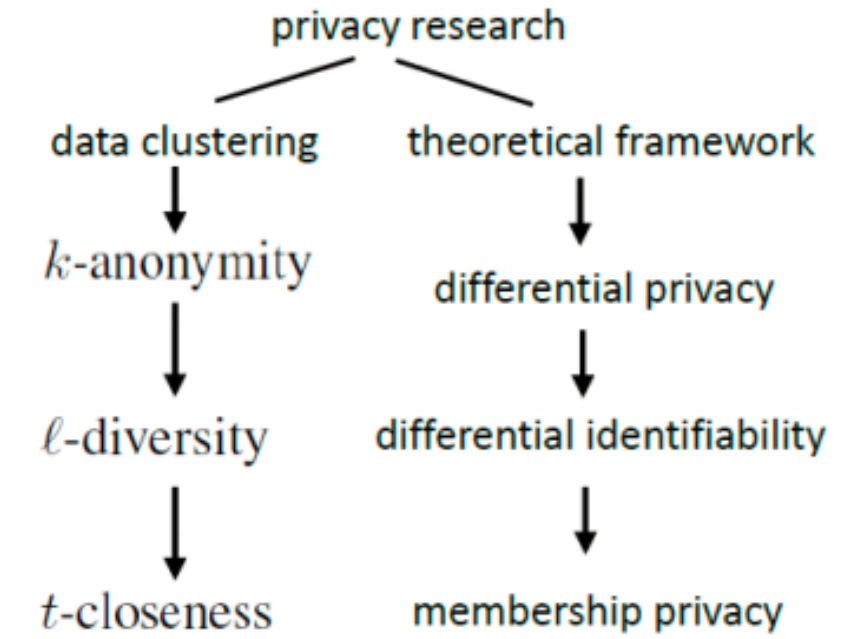


Fig. 1. The roles and operations of a privacy system.

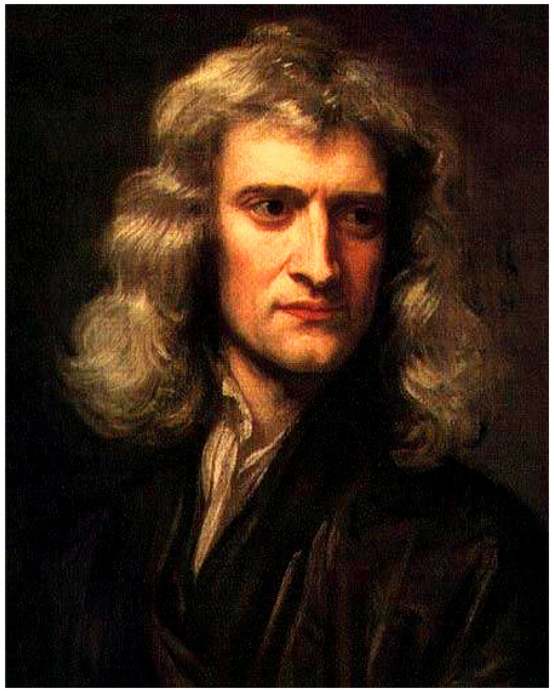
Shui Yu, "Big Privacy: Challenges and Opportunities in Privacy Study in the Age of Big Data," *IEEE Access, 2017*.

5. Big Data Privacy



Shui Yu, “Big Privacy: Challenges and Opportunities in Privacy Study in the Age of Big Data,” *IEEE Access, 2017*.

5. Big Data Privacy



First Challenge: privacy measurement.

“I can calculate the movement of stars, but cannot measure the madness of men”

DEAKIN UNIVERSITY AUSTRALIA

5. Big Data Privacy

- Other challenges in big data privacy

 - personalized privacy

 - theoretical tools for privacy (mathematical tools, models)

 - privacy for trading

 - …

DEAKIN UNIVERSITY AUSTRALIA

5. Big Data Privacy

- Cross discipline directions
  - psychology + CS
  - social science + CS
  - …

Thank you
&
Questions

http://www.deakin.edu.au/~syu
email: syu@deakin.edu.au