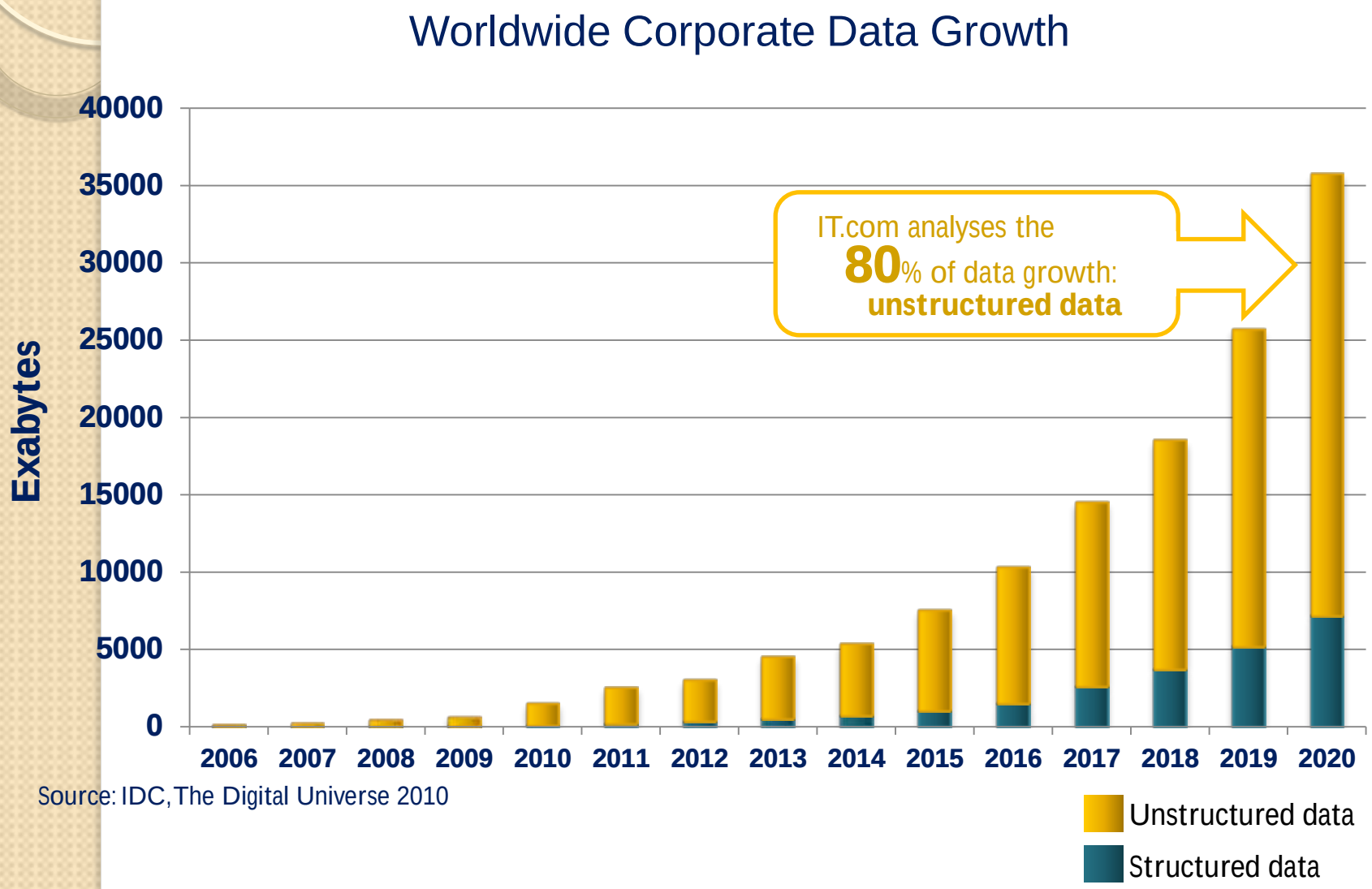
Balaji Palanisamy

Assistant Professor
School of Information Sciences
University of Pittsburgh
bpalan@pitt.edu

University of Pittsburgh

SCHOOL OF
Information
Sciences

Data Growth

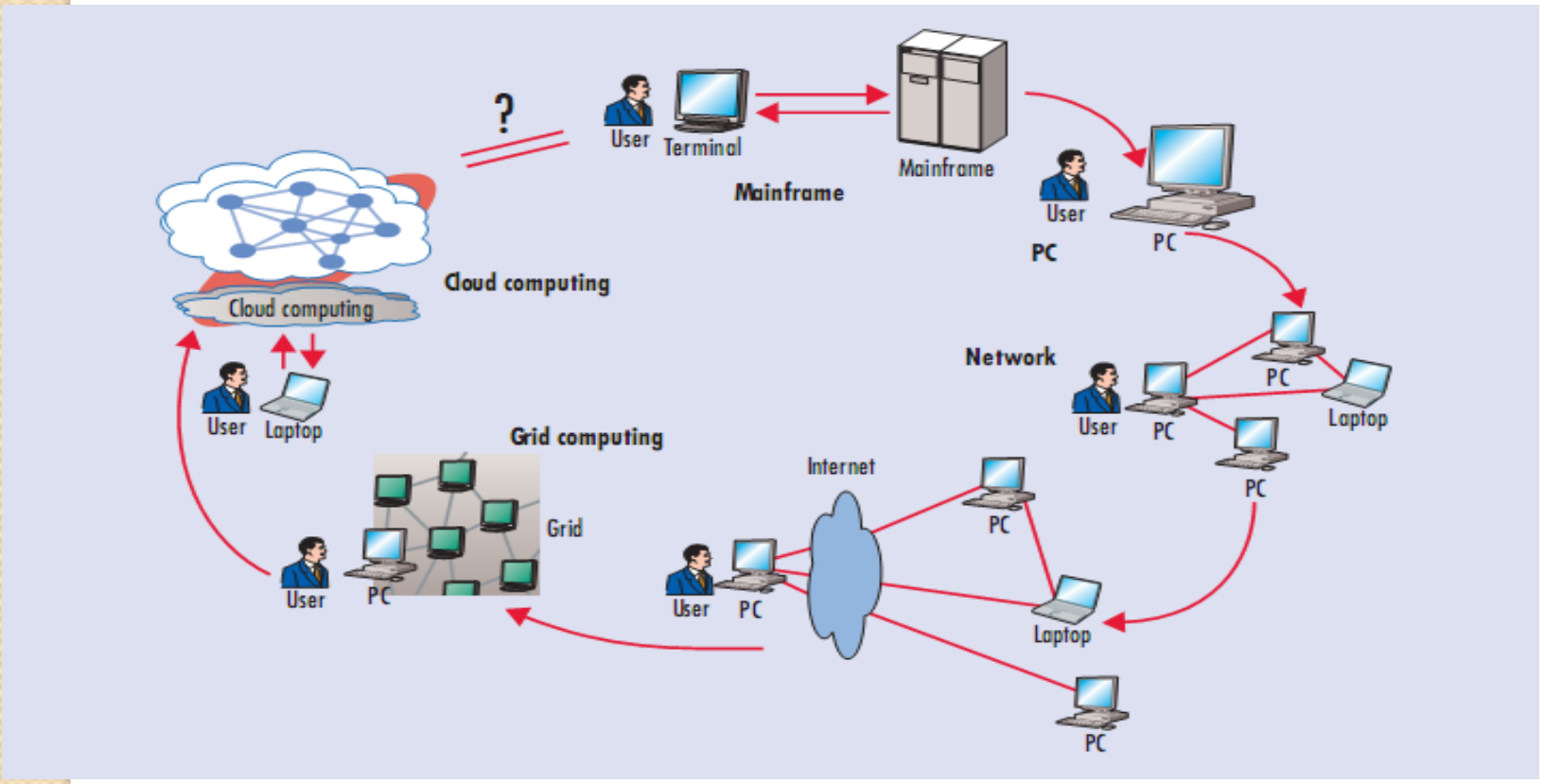


Source: IDC, The Digital Universe 2010

Computing Paradigm Shift

Cloud Challenges

- Cost-effectiveness, Performance and Security



Research Interests

- Privacy-preserving Cloud Computing
- Data and Location Privacy in Cloud/Mobile systems
- Big Data in a Cloud
  - Hadoop, NoSQL databases
  - Cost-optimized Resource management techniques
  - Performance tuning and resource optimization

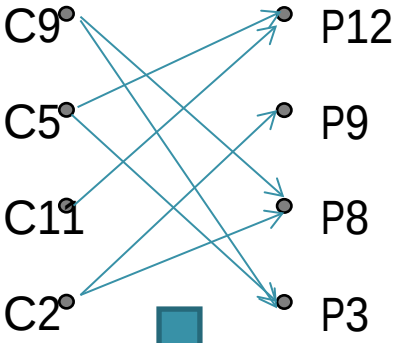
Current Projects

- **Private processing of anonymized datasets in a Cloud**
  - Existing Anonymization methods
    - Perturb the dataset to produce anonymized version (k-anonymity, l-diversity constraints)
    - Completely lose the accurate information during anonymization.
    - In other words, for even users who have access to the actual raw data, they are unable to retrieve information
  - Our techniques
    - Anonymize data in such a way that utility is not lost but retained anonymously.
    - The user provides the required anonymization secret (secret key) to obtain the accurate results of the query. (ICDCS ' 14 – sub)

Example: Privacy-preserving Access on Graph Datasets

Drug purchase associations

Cid	DOB	Sex	ZIP
C1	7/18/79	F	30323
C2	2/17/83	M	30323
C3	5/07/77	M	30327
C4	1/5/76	F	30328
C5	8/4/82	M	30330
C6	3/9/79	M	30331
C7	4/10/64	M	30331
C8	2/6/81	F	30334
C9	7/14/72	F	30337
C10	9/25/74	M	30338
C11	4/28/80	M	30338
C12	3/12/78	M	30339



Cid	Did
C2	P8
C2	P9
C5	P3
C5	P12
C9	P3
C9	P8
C11	P12

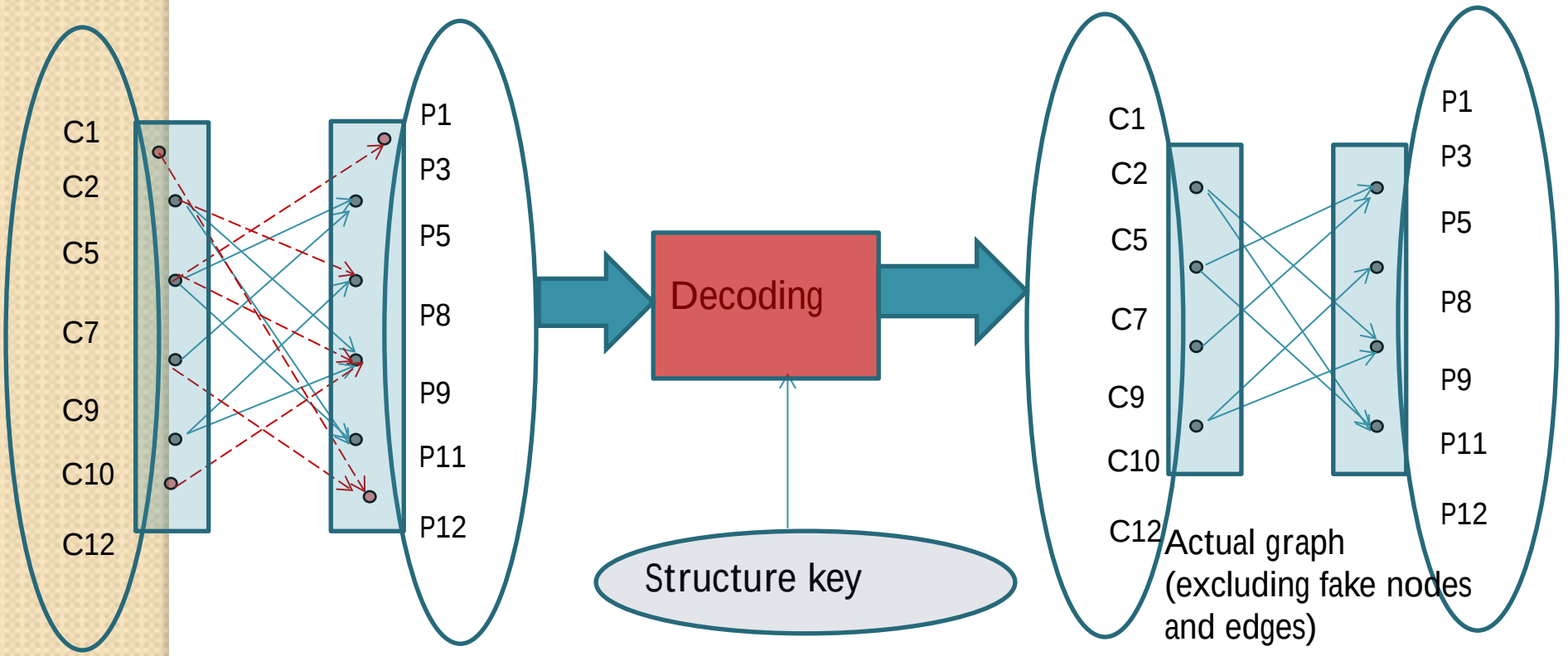
Did	Drug name	Category
P1	epinephrine	bronchodilator
P2	ibuprofen	analgesic
P3	Zovirax	antiviral
P4	Tylenol	analgesic
P5	erythromycin	antibiotic
P6	cortisone	anti-inflammatory
P7	gentamicin	antibiotic
P8	insulin	hypoglycemic
P9	sertraline	antidepressant
P10	tramadol	analgesic
P11	cetirizine	antihistamine
P12	zolpidem	hypnotic

Revealing the associations between Cid and Did violates privacy

Access Privilege levels

- Level 0: No Access:
  - Can neither obtain graph structure, aggregate utility nor personal associations
- Level 1: Graph structure only Access:
  - Can Query on graph structure (node distribution, edge distribution) – type 0 query
  - Cannot obtain information using entity attribute –
- Level 2: Aggregate query access
  - Can obtain results for queries on graph structure + entity attributes
  - Cannot obtain results on personal associations
- Level 3: Complete access
  - Can obtain results on graph structure, entity attributes as well as personal associations

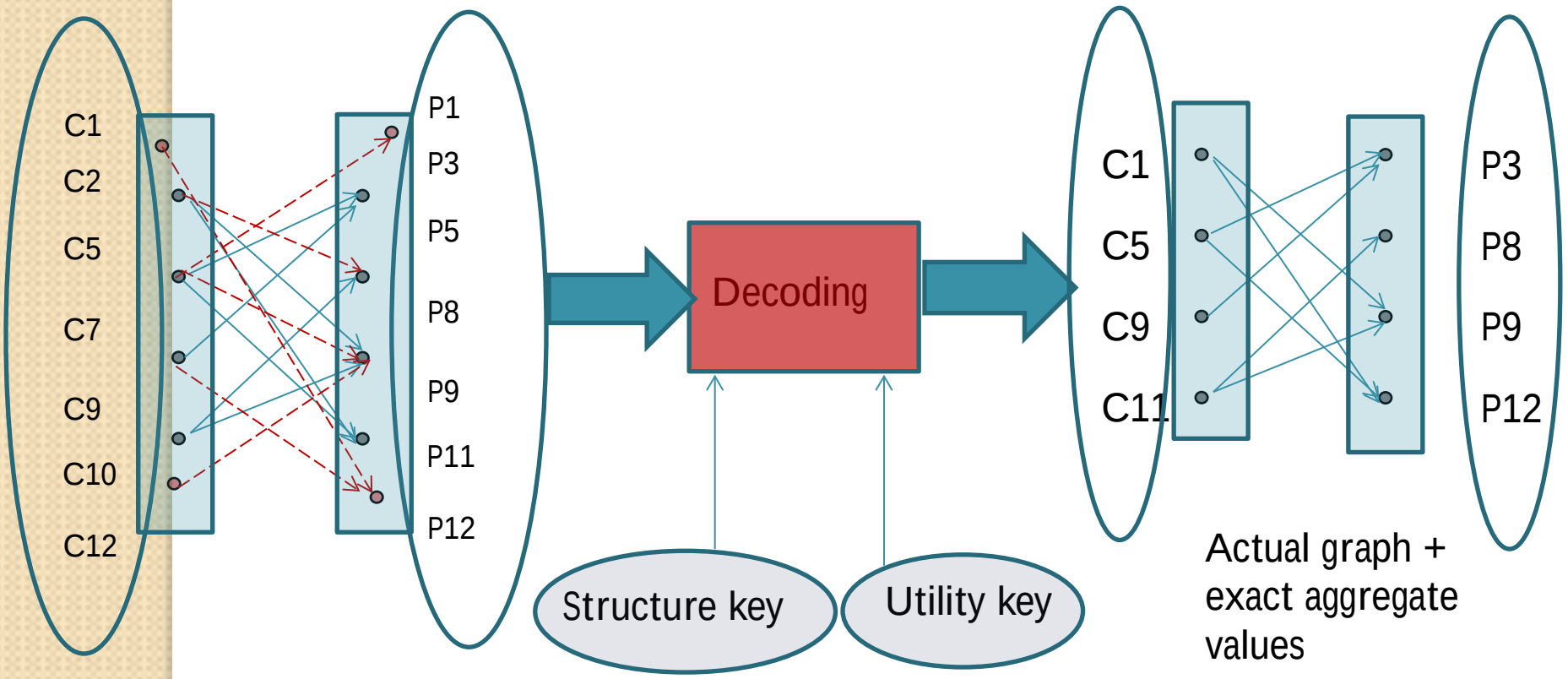
Decoding level 1 users



Level 1 users possess the graph structure key and using which they can decode the anonymized graph

However, aggregate values and individual associations can not be inferred using the structure key.

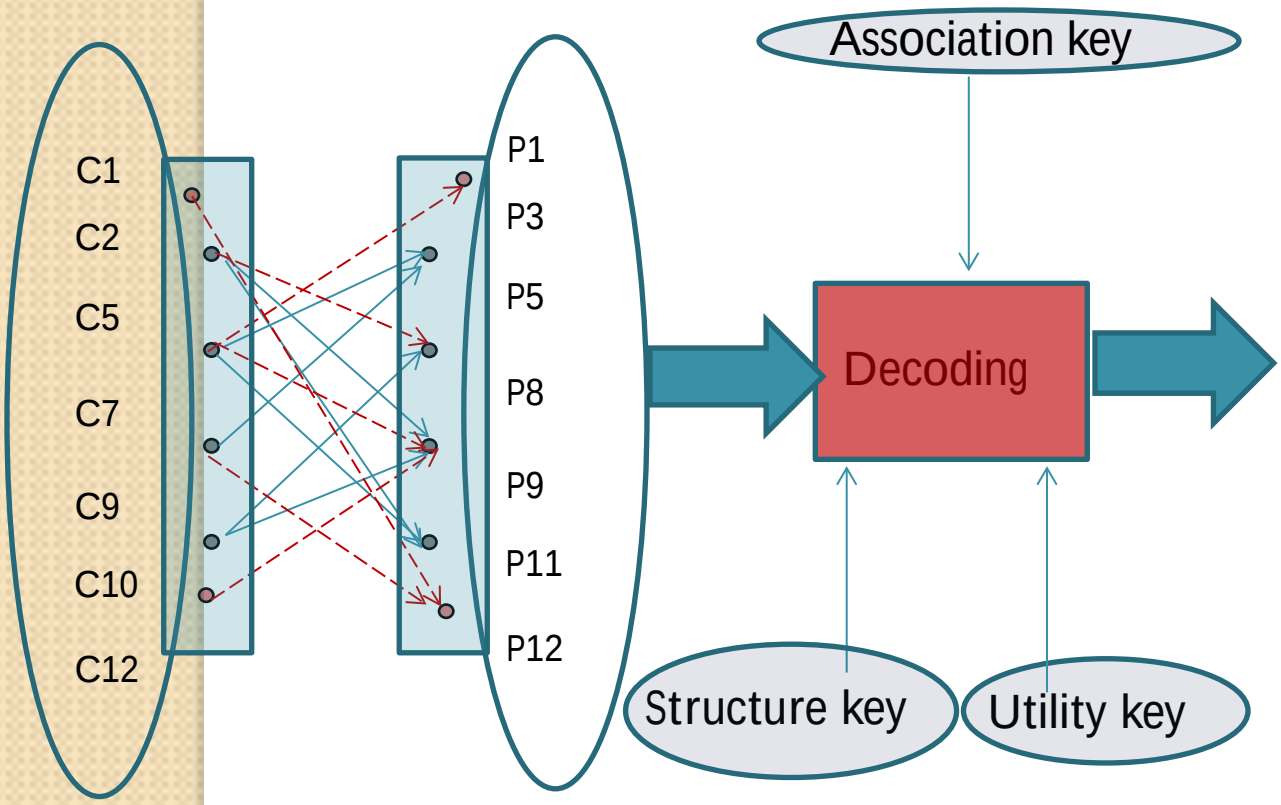
Decoding level 2 users

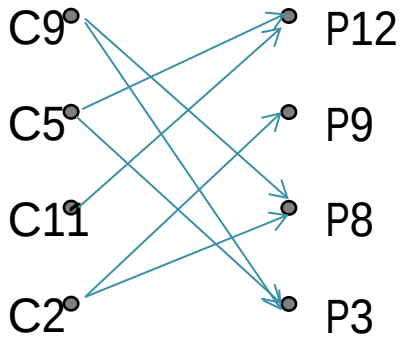


Using the structure key and the utility key, level 2 users can obtain access to both the graph structure as well as the aggregate values of the associations.

Here, the labels of the nodes are permuted on the graph after decoding and hence level 2 user can not infer the individual associations.

Decoding level 3 users





Actual graph + exact aggregate values + individual associations

Level 3 users have the highest privilege and they have access to the actual raw association graph.

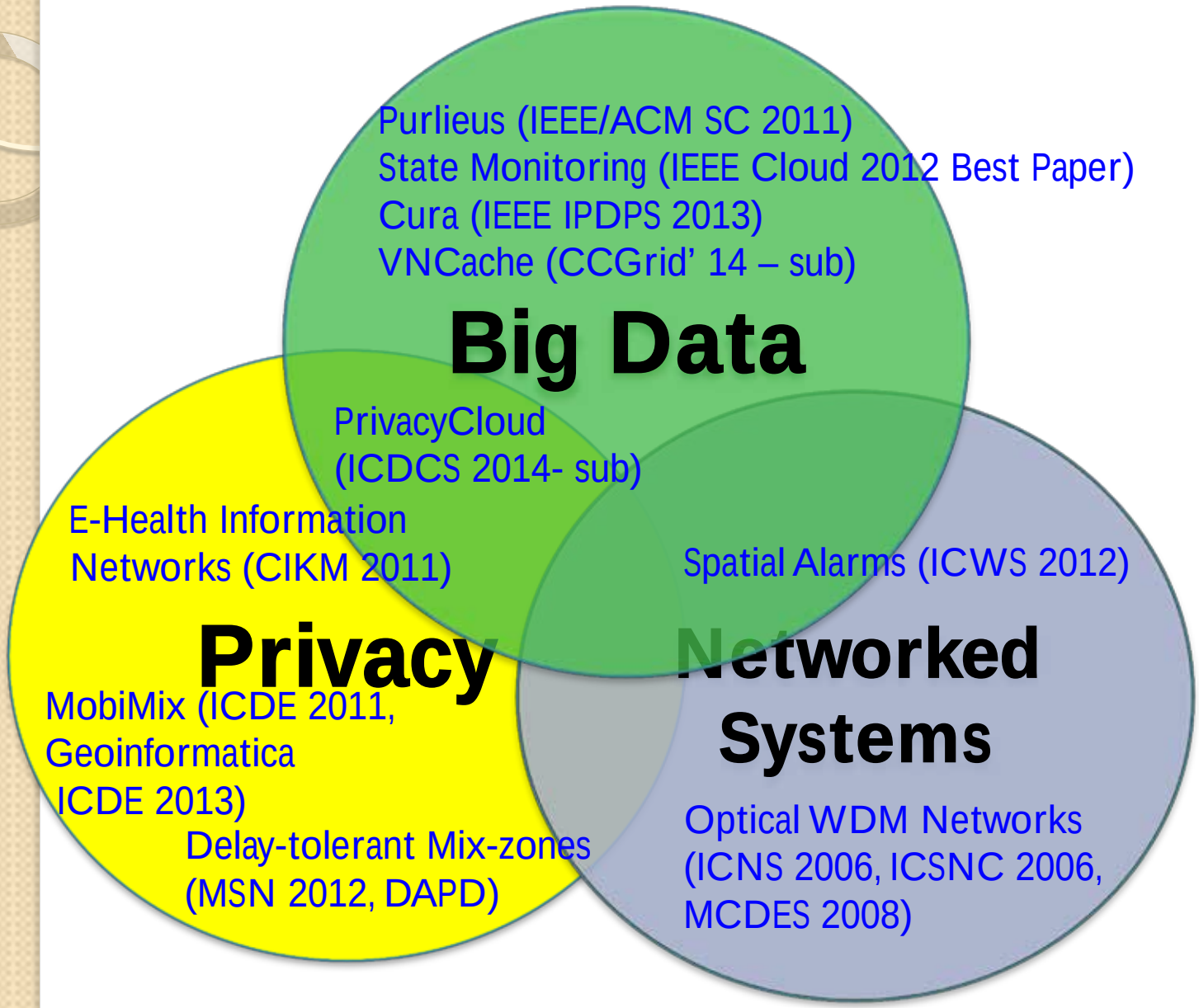
Current Projects

- Cost-optimized Resource Management for Map Reduce/Hadoop Clouds
  - Existing Hadoop Cloud services
    - Customer side Job Tuning
    - Per-job customer-side greedy optimization may not be globally optimal
    - User needs to figure out the tuning parameters and the cluster resources to use
  - Our approach
    - Achieve global resource optimization (IPDPS' 13)
    - Achieve higher data locality (SC' 11, CCGrid' 14 – sub)
    - Cloud provider manages the resources to ensure each job's service requirements
    - Requires 80% lower resources than conventional models

Current Projects

- **Location Privacy in Mobile Cloud Computing**
- Proliferation of Location-based services:
  - E.g. Finding nearest gas station?
  - "What is the present traffic congestion level on I-85 North? "
  - "How many people are inside this building?"
- Location-based Services pose new location privacy threats
  - Constant location exposure helps the adversary track user movements
  - Can learn users' medical conditions, social and political interests etc..
- Our work
  - Robust and attack-resilient location anonymization schemes (ICDE '11, ICDE' 13, MSN' 12)
  - Sophisticated attacks such as continuous query attacks and fake user attacks (DAPD 2013, Geoinformatica 2013)

Research Focus



Teaching Philosophy

- Inculcating critical thinking skills for problem solving
- Enforcing strong fundamentals
- Transforming students to become independent learners
- Top-down course organization
- Assignments/Projects/ Homeworks designed with the above goals

Teaching

- Current Courses
  - Computer Security (undergraduate) – Fall 2013
  - Information Security and Privacy (Graduate) – Spring 2014
- Plans/ideas for new courses
  - Specialized Course on Information Privacy
  - Advanced course on Cloud/Big Data Security and Privacy

Recent Publications

- Balaji Palanisamy and Ling Liu, "Effective Mix-zone Anonymization for Mobile Travelers", *Geoinformatica 2013 (to appear).*

- Balaji Palanisamy, Ling Liu, Kisung Lee, Shicong Meng, Yuzhe Tang and Yang Zhou, "Anonymizing Continuous Queries with Delay-tolerant Mix-zones on Road Networks", *Distributed and Parallel Databases (DAPD 2013).*

- Balaji Palanisamy, Aameek Singh, Ling Liu and Bryan Langston, "Cura: A Cost-Optimized Model for MapReduce in a Cloud", Proc. of *27th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2013)*.