Data Centric Security

Vijay Varadharajan

Global Innovation Chair Professor and Chief Cyber Strategist
The University of Newcastle
vijay.varadharajan@newcastle.edu.au

Honorary Professor and Former Microsoft Chair Professor
Macquarie University
vijay.varadharajan@mq.edu.au

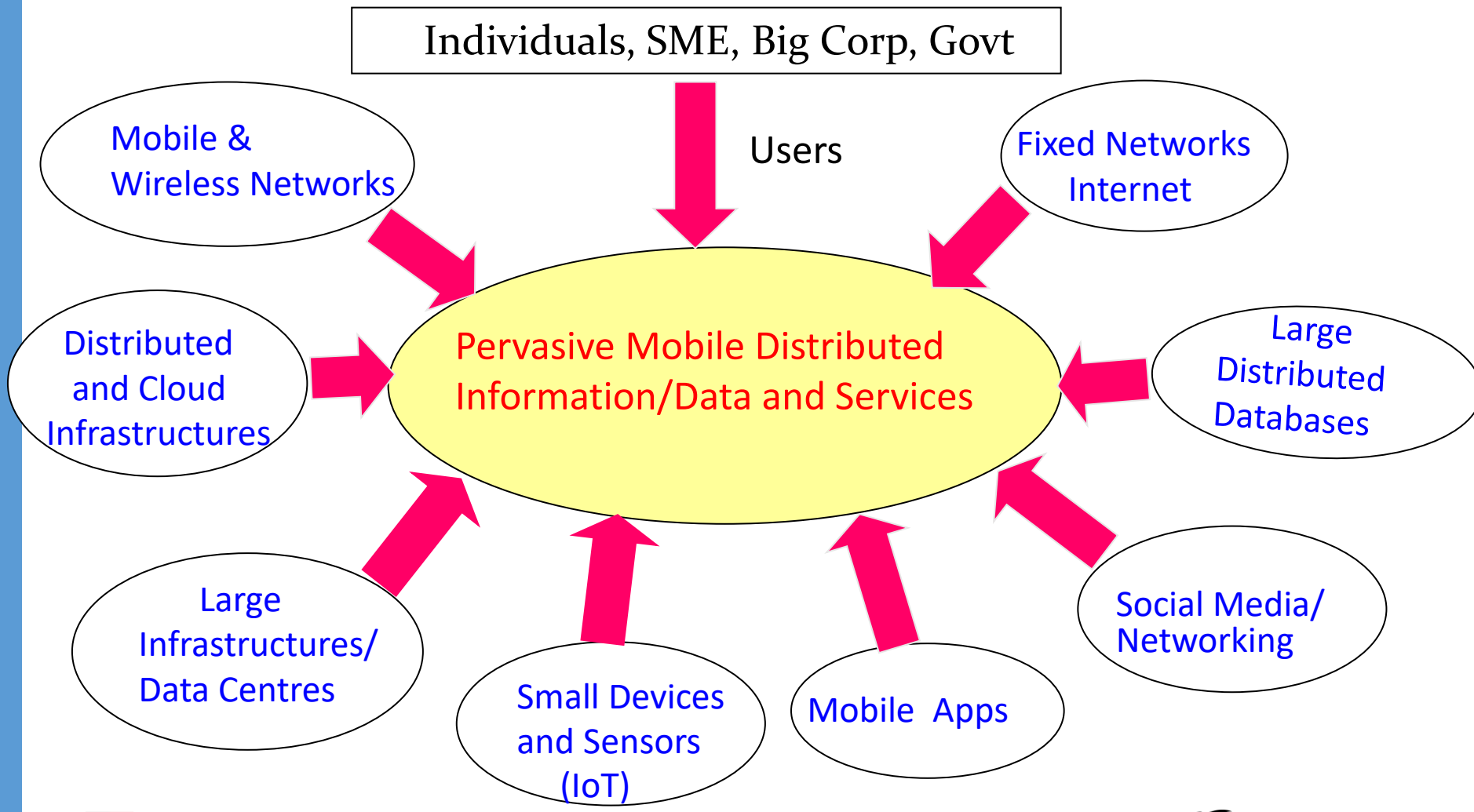
MACQUARIE University
SYDNEY·AUSTRALIA

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Talk Overview

- Technology Scenery and Cyber Security
- Data Context and Perspectives
- Data Centric Security
- Concluding Remarks

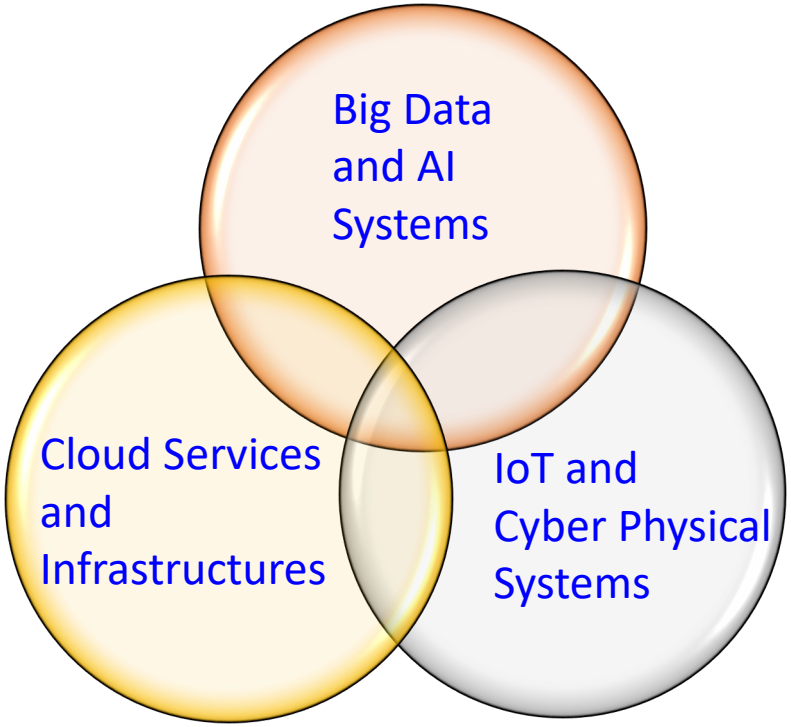
Technology Scenery



MACQUARIE University SYDNEY·AUSTRALIA

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

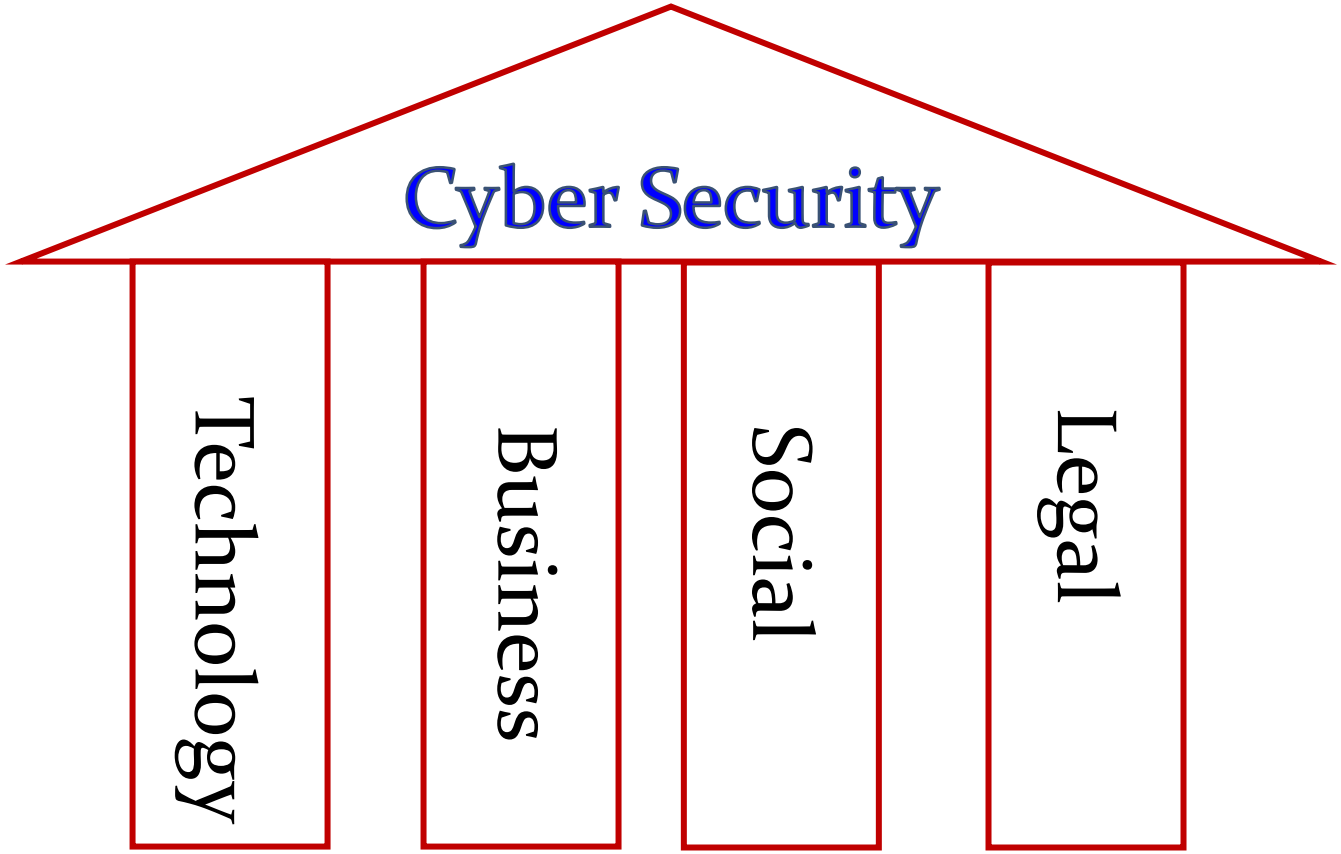
Cyber Security Challenges



- Information Explosion → Big Data
- Information Generation and Analytics → AI
- Systems of Systems → Cloud Computing
- Ubiquitous Computing → Internet of Things
- Cyber Physical Infrastructures → Industrial Control Systems

MACQUARIE University
SYDNEY·AUSTRALIA

THE UNIVERSITY OF NEWCASTLE AUSTRALIA


Cyber Security
Technology
Business
Social
Legal

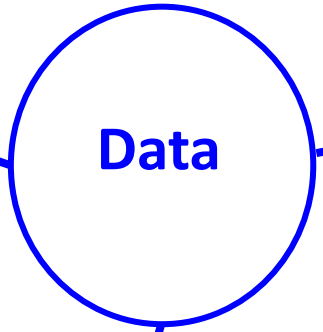
Security and Privacy

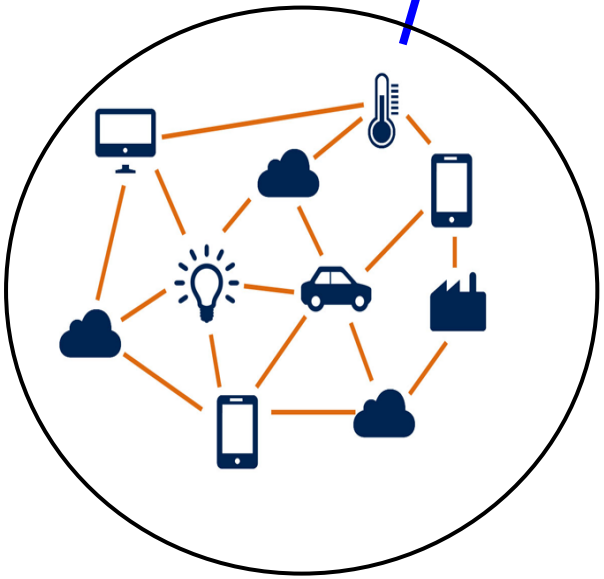
- Security
  - *Owner* of Information has control
  - Security is Not Privacy
- Privacy
  - *Subject* of Information has control
  - Privacy requires Security
- Anonymity
  - Has no subject
  - Requires Security and guarantees Privacy, but is neither

Data Context

We are sharing data with many people

We are connected with more and more devices





Devices are interconnected with each other

Data has become the new Currency in the Digital World

→ New Innovations

→ Improve the quality of life

MACQUARIE University SYDNEY·AUSTRALIA

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Data Context

- Today
  - Some 3 billion online users added in last 10 years
  - Over 5 billion videos watched daily on Youtube alone
  - Some 500 hours of videos loaded every minute in Youtube
  - Dramatic Growth in Social Media – Facebook, Twitter, Instagram,…
    - Approx 5 billion users of social media – over half of the world population

- We can probably store everything!
  - All movies made to-date : 1 petabyte or so
  - All music recorded to-date : 1 petabyte or so
  - 1 billion photos : 1 petabyte

- Capture everything you ever said from the time you are born to the time you die.
  - Less than a few percent of a petabyte

- Everything you ever did and experienced can be captured in living color
  - With only a few petabytes

Data Perspective: Transformative Changes

- What is Private
  - Previously
    - Default position was information was private until you opened up
      - E.g. Locked in safety deposit box, house, filing cabinet etc.
      - May be required by law to open it up, or choose voluntarily to open up
      - Basically, your personal information is locked and is private.
  - Now in the Cyber world
    - Private information is wide open
      - It may be in your laptop or mobile or some other device
      - As soon as it is connected to any form of network, it is not private any more
      - Default position is it is open
- The presumption has changed
  - From one of controlling the opening to controlling the closing

Data Perspective: Transformative Changes

- What is Public
  - Previously
    - One did somethings in public, and some people saw them
    - They may or may not discuss what they saw with other people, and
    - Over a period of time, people may forget what they had seen
  - Now in the Cyber world
    - What is public far exceeds even what Orwell had imagined in 1984
    - After all what Orwell only imagined, that we have government operated telescreens in our houses looking at everything we did.
    - He did not imagine that everybody in the population was a telescreen.
      - Anyone can now use a mobile phone, video/photo what is of interest to them, upload it onto the Internet
      - Then it becomes available pretty much all over the world
    - May not be forgotten and can be recalled/aggregated
      - To give somebody a picture of one's life for more granular than probably any secret police would have imagined some 30 years ago.

Data Related Challenges

- Just because data is accessible, it does not mean the data is trustworthy or reliable to make decisions, or even ethical to access and use it.
- There are several issues
  - Where does the data come from (data provenance)?
  - How trustworthy is it?
  - How do you know where your data is?
  - Do you know who can see the data and modify it without a trace?
  - Who can aggregate or summarize or embed your data for purposes other than what you specified?
  - How can data owners specify and enforce policies on the data as it moves over the Internet?
  - How to detect and prevent attacks on the services that operate on the data in a distributed environment?

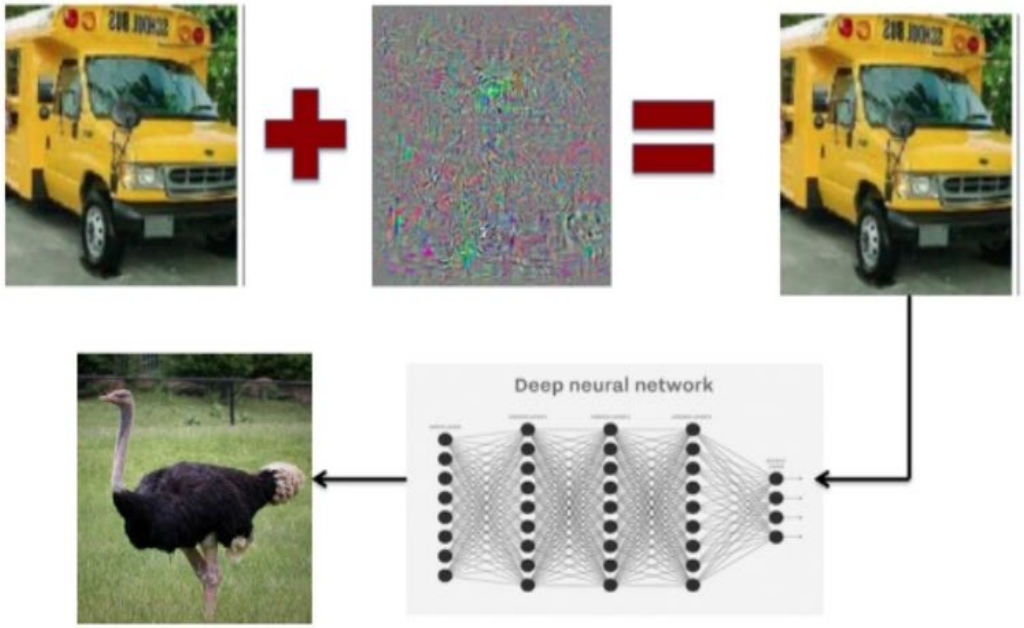
→ Enhance the trust and quality of decision making

Data Related Challenges

- Machine Learning at the core of many data analytic algorithms
  - Deep Learning
- Attacks on Big Data Analytics
  - Attackers can attack machine learning systems themselves
    - By injecting malicious data (at training time, at test time)
    - By exploiting algorithms' weaknesses
  - Attack on Data at training time
    - Poisoning attacks
  - Attack on Data at test time
    - Evasion attacks

Adversarial Machine Learning

Adversarial School Bus



Szegedy et al., Intriguing properties of neural networks, **ICLR 2014**
Biggio, Roli et al., Evasion attacks against machine learning at test time, **ECML-PKDD 2013**

MACQUARIE University
SYDNEY·AUSTRALIA

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Data Related Challenges

- Users
  - Typically want personal control of their data even if they don't want to exercise this control
  - Allow agents that they trust to access and process their data
- Regulators
  - Control of Data -- Fundamental human right
  - Mandatory Data Breach Regulation (Australia, Feb 2018)
  - EU GDPR (May 2018), California Consumer Privacy Act CCPA (Jan 2020)
- Industry
  - Usually prefer consistent rules to build customer relationships
  - Agreed rules to comply with regulations

General Data Protection Regulation (GDPR)

- Regulates the collection, storage, use, and sharing of "personal data."
- Personal Data
  - Any data that relates to an identified or identifiable individual.
  - Can include data such as
    - online identifiers (e.g., IP addresses),
    - employee information,
    - customer services data, customer feedback forms,
    - location data,
    - biometric data, CCTV footage,
    - loyalty scheme records,
    - health and financial information, …
  - Can even include information that does not "appear" to be personal
    - Such as a photo of a landscape without people – where that information is linked by an account number or unique code to an identifiable individual.
  - And even pseudonymized data can be personal data if the pseudonym can be linked to a particular individual.

GDPR and Organizations

- GDPR applies to any organization that is
  - Processing of anyone's personal data, *if the processing is done* in the context of the activities of *an organization established in the EU* (regardless of where the processing takes place)
  - Processing of personal data *of individuals who reside in the EU* by *an organization established outside the EU*, where that processing relates to the offering of goods or services to those individuals or to the monitoring of their behaviour

GDPR: Compliance and Data Breaches

- GDPR and Compliance
  - The maximum fine for serious infringements will be the greater of €20 million or four percent of an organization's annual global revenue.
  - In addition, the GDPR empowers consumers (and organizations acting on their behalf) to bring civil litigation against organizations that breach the GDPR.
- GDPR and Data Breaches
  - Personal data breach
    - A breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed."
  - Provide notice to regulators within 72 hours of detecting the breach.
  - You may also need to notify affected individuals if there is a significant risk of harm due to the breach.

GDPR and Individual's Rights

- GDPR is oriented towards individuals' rights
  - the right to know how data about you is processed (collected, analyzed, and used)
  - the right to object to such processing
  - the right to see the data that is stored about you
  - the right to a meaningful explanation about automatic data processing
  - the right to withdraw consent to processing
  - the right to have your data erased under certain conditions
  - the right be able to easily move your data from one provider to a different one

Data Related Challenges

- Find where the data is and limit its use
  - E.g. Tracking Sensitive Data: Credit Card Number, Social Security Number, Aadhaar ID
- Across the whole Internet
  - Lots of Transactions over the Internet.
  - E.g. opening bank accounts, booking hotels, air and train travels etc.
- Ability to track data anytime (data provenance) and not just at the time of collection
- Across different agents and devices that handled the data
  - Users share sensitive data with several organizations
  - Government entities, private organizations

Data Centric Security Approach

- Data tagged with *metadata* that links to Policy
  - Secure coupling mechanism between Data and Policy
  - Should not be able to decouple data and policy
    - Security mechanisms enforcing this coupling
    - Policy stays with the data when the data is copied
- Data Processing Agents
  - Agents that process data must check and satisfy policies *before* using data
- Policies
  - Ideally simple, coarse-grained policy and good defaults
  - Different types of policies
    - Audit and tracking policies
    - Access and usage policies
    - Obligation policies

Data Centric Security Scenario

- Simple Scenario: Data Tracking Protocol
  - Allowing data owners to track the flow of their data
- Examples
  - Data from IoT devices transferred over the networks
  - Data in social media
  - Data sharing within enterprise and between enterprises
  - Data transfer over networks

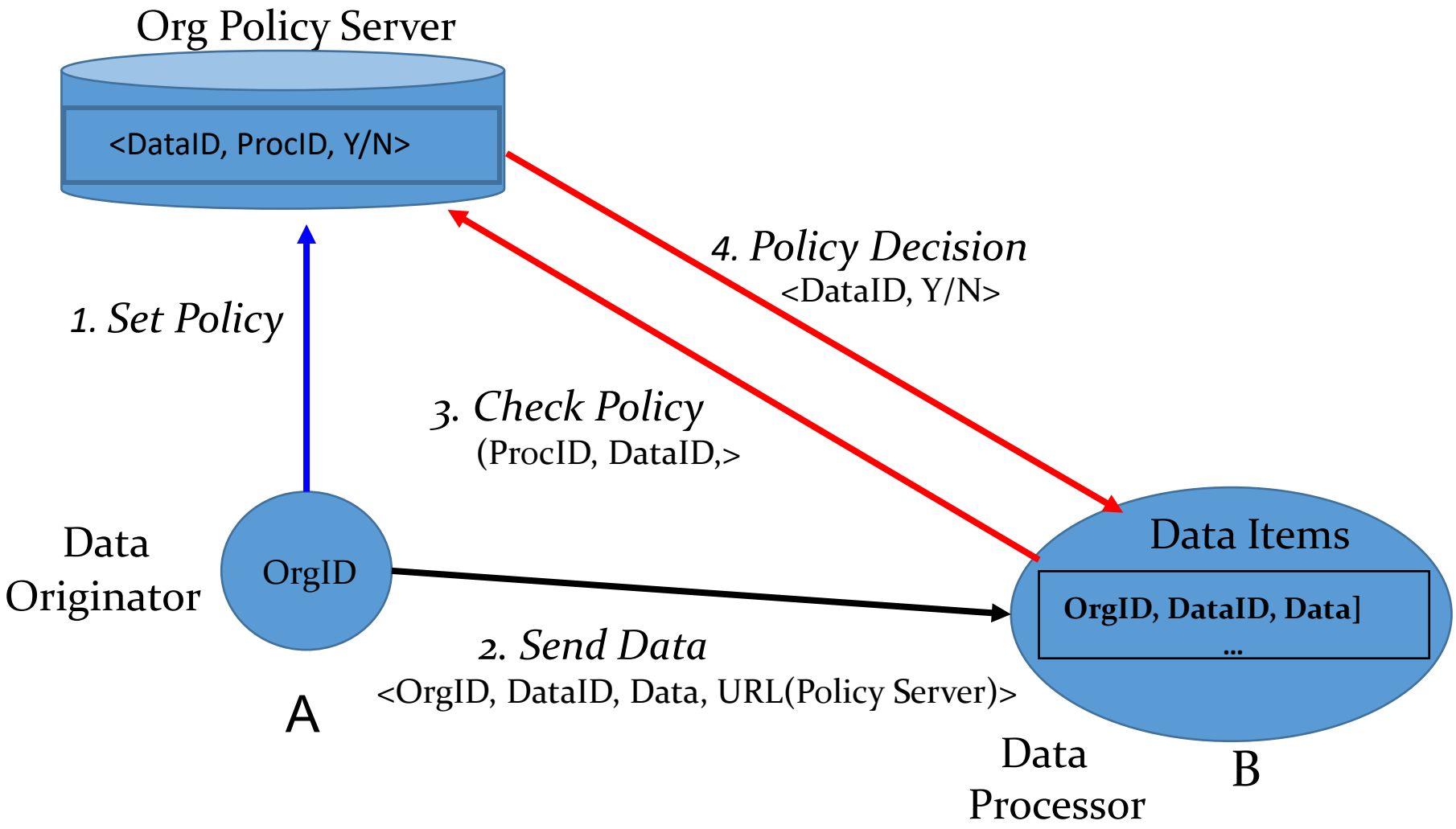
Personal Control of Data

- Personal Agent
  - Handles Personas
  - Several Personas to manage your different identities and claims
  - Can be offline
- Your Policy Service
  - Tells Processing Agents your policy
  - Should be online
- Data Processing Agents
  - Subject to regulation
  - Anyone who stores or processes your data and is following the rules

Data Tracking Protocol (DTP)

- Data tagged with metadata that links to policy
- Data Processing Agents process and store the data
  - Trusted – willing to follow the rules and subject to regulation
- User chosen Policy Service
  - Online service that stores policies for user's data items
  - User specifies policies for his/her data
  - Tells data processing agents the policy associated with the data
  - Different data items and different data processors can have different policy service
  - Keeps track of data processors accessing user's data
  - Simple Implementation
    - URL of policy service specified as part of DNS specification

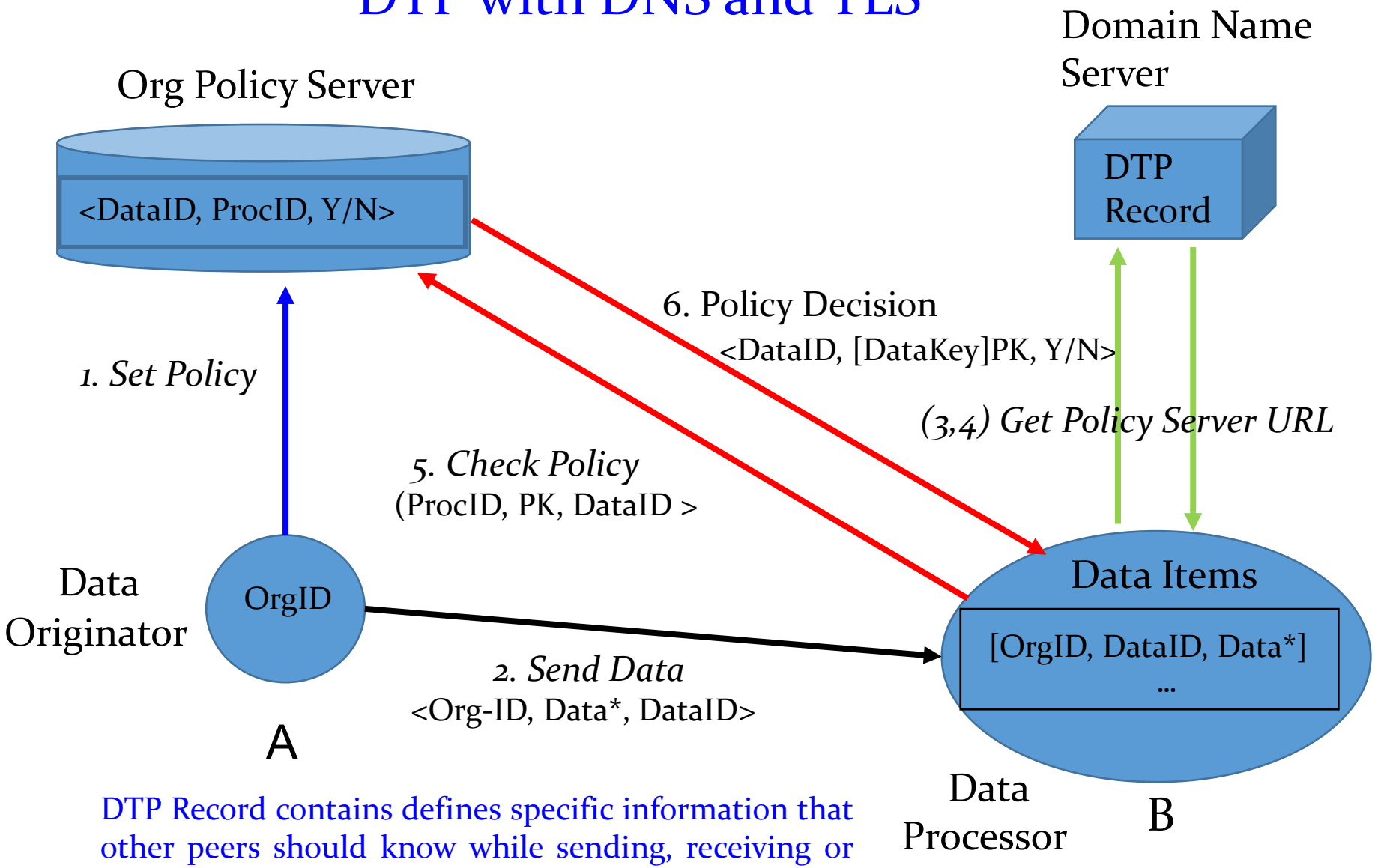
Data Tracking Protocol (DTP)



MACQUARIE University
SYDNEY·AUSTRALIA

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

DTP with DNS and TLS



DTP Record contains defines specific information that other peers should know while sending, receiving or sharing data under the purview of DTP

MACQUARIE University SYDNEY·AUSTRALIA

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Data Centric Security

- Policy based Data Centric Approach
  - Secure Coupling Mechanism
    - Data coupled with policy
  - Secure Processing Agents
    - Trust in the Agents that process data will check and satisfy policy before using data
    - Trusted Environment
  - Security Infrastructures
    - Policy to Data Mapping Infrastructure
    - Data to Policy Mapping Infrastructure

Data Centric Security Polices

- Policies
  - Metadata pointing to policies associated with data
    - Not device or service centric
  - Basic Policies can be simple
    - Data Processor *ProcID* can use Data *d* Type *t*
  - Composition of Policies
    - Boolean operators, such as *and*, *or*, on atomic policies
  - Content Dependent Policies
    - Data content dependent constraints
    - Data provenance-based constraints
      - "a law-enforcement official may not act on improperly obtained evidence, but if the same information was obtained through lawful channels the official can act"
  - Different types of policies
    - Audit and tracking policies
    - Access and usage policies
    - Obligation policies

Data Centric Security – Access Control

- Access Control
  - Typically address access to data by principals, generally no further control is applied to ensure the data is handled properly
  - That is once access to data is granted, application is **trusted** not to leak the data.
  - That is, they are principal-centric (often principal – user)
  - As data flows through a complex multi-component system, it may fall under different access control regimes, with varying granularity
- Data Centric Access Control
  - Access control for the cloud should no longer solely be principal or application-specific but *should be data-centric, controlling data flows between applications.*
  - Useful work
    - Require data sharing between applications (and services) across and outside isolation boundaries.

Data Centric Access Control

- Consider a scenario where personal medical data gathered by sensors, say from monitoring a patient at home, are used by cloud services and databases.
- Patient's policy should specify on how the data can be used throughout its lifetime
- For instance, if we have a *tag* associated with the data, it can specify that this is medical-research data. This can in turn be used to ensure that it can flow only to those conducting medical research, who also have this tag.

Data Centric Access Control

- Web Platforms
  - Extensible: Allow third party applications to integrate with the platform
    - Facebook popularized this extension model for social networking.
    - Yammer provides a similar social platform for enterprises
  - Functionality users experience on these sites is no longer the product of a single entity
    - It is a combination of a core trusted platform, and apps written by less-trusted third-parties.
- Many apps are only useful when they are able to manipulate user data
  - Including sensitive and personal information such as financial or medical details, or non-public social relationships
  - Again, once access to this data has been granted, often there is no mechanism to constrain what the app may do with it on many platforms.
  - Furthermore, 3rd party apps run on servers outside the control of the trusted platform
    - This means that all data the app accesses is exfiltrated.
  - As apps only function if all their access requests are granted
    - Forces the users to choose between privacy or functionality.

Data Centric Security: Access Control

- Data Centric Approach to Access Control
  - Mandatory data access policies
  - Policies follow data throughout the system and enforced even *after* apps given access to data
  - That is, policies to control what apps can with it
- Information Flow Policies
  - Data confidentiality (Bell La Padula Model)
  - Data Integrity (Biba Model)
- Entities: Active (e.g. processes), Passive (e.g. data)
- Each entity A has two security labels
  - Confidentiality Label C(A), Integrity Labe I(A)
- Security Context is determined by the state of the two labels, State(C, I).
- Information Flow Policy
  - Information can flow from A to B: A → B
    - If an only if {C(A) is subset of C(B)} AND {I(B) is a subset of I(A)}
    - That is, lower confidentiality (A) to higher confidentiality (B) AND higher integrity (A) to lower integrity B

Data Centric Access Policies

- If an entity creates an entity (active or passive), then the created entity inherits the labels of its parent.
- In addition, we will need certain active entities to have privileges to add and/or remove tags from these labels.
- An active entity can have up to four privilege sets in addition to its security context.
  - ***Add confidentiality label, Remove confidentiality label***
  - ***Add integrity label, Remove integrity label***
- No inheritance when it comes to these privileges

Data Centric Access Policies

- Security Context Domain (SCD)
  - Security Context Domain comprises entities with the same sets of labels.
- Information Flow Policies
  - Flow of data is therefore allowed within a security context domain
  - Flow of data allowed into a more constrained domain.
  - Once data has flowed into a more constrained domain further flows are to that domain or into increasingly constrained domains.
  - Problem of “label creep”

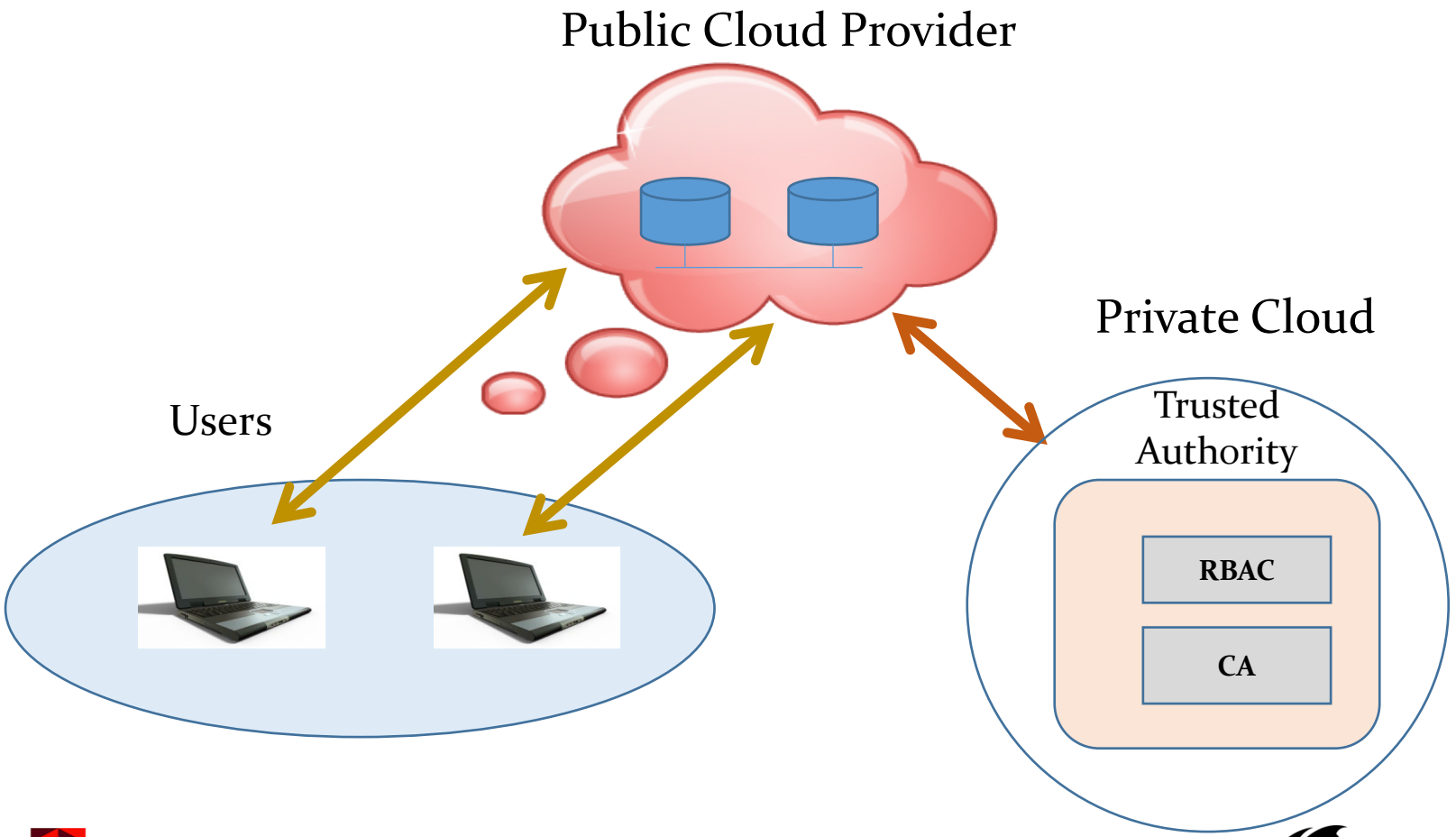
Data Centric Security Policies

- Meta Policies
  - Certain entities are given the capability to ***modify*** their labels in order to transfer information across security contexts
  - ***De-classifier***
    - An entity that **modifies confidentiality labels** in security context
  - ***Endorser***
    - An entity that **modifies integrity labels** in security context
  - ***Endorsers and De-classifiers***
    - Trusted gateways between security context domains

Data Centric Security Policies

- Examples
  - *Requirement: Medical data must only be stored in encrypted form*
    - Database labels set up so that only appropriately labelled data, e.g. encrypted, can flow into it.
    - An encryption function must be applied to the data together with an *Endorser* to add encrypted to its integrity label.
  - *Medical data can be used for research purposes only if the consent of the owner is obtained, and the data is anonymised*
    - Data owner's consent must be established and indicated with the data
    - An approved anonymising function must be applied to the data and a *De-classifier* must transform its labels.
    - The data is therefore constrained to flow between related applications in a medical domain
- Ability to express and enforce such policies makes deployment of cloud application domains handling sensitive data feasible

Data Centric Security: Cloud Data Storage



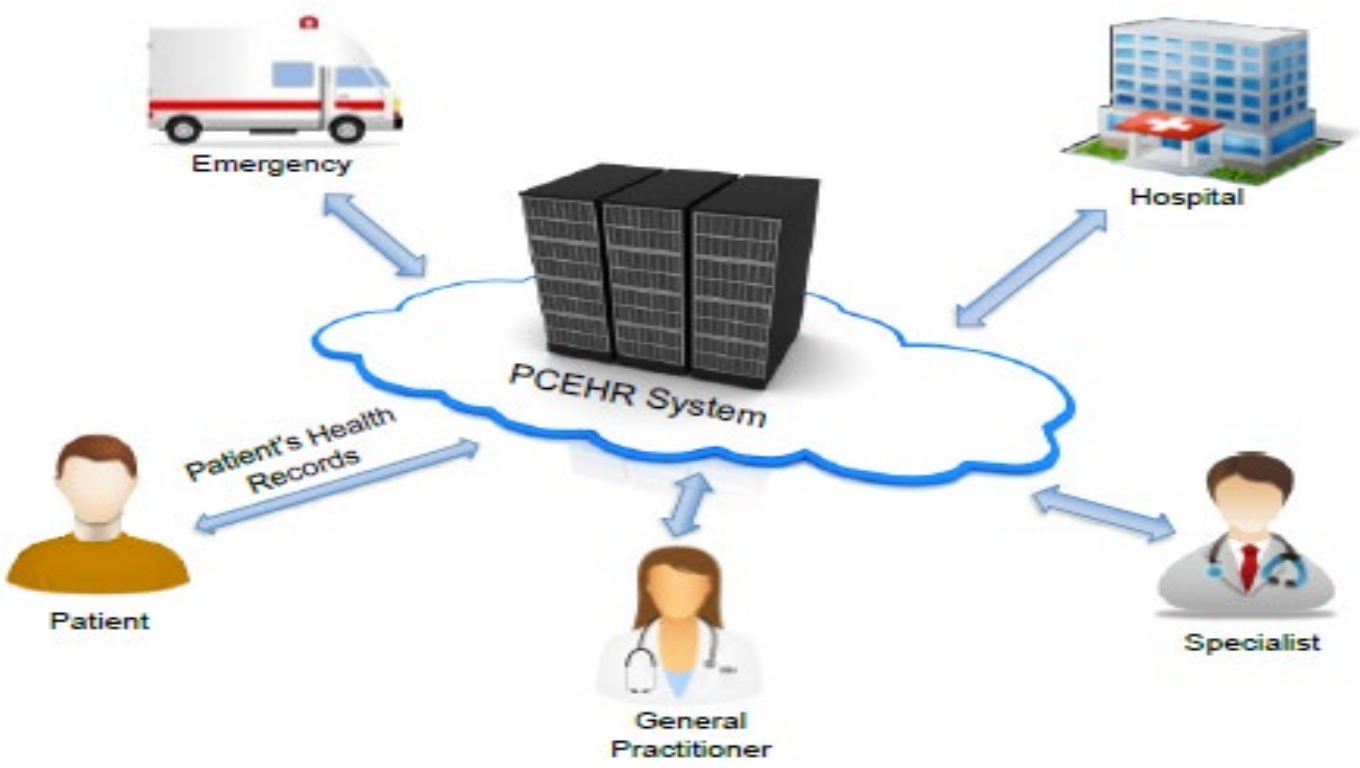
Data Centric Security: Cloud Data Storage

- Policy based decision making for stored data
  - Assume now the data owners have encrypted their data
- E.g. Cloud Data Storage System
  - Data Owners – Specify the Policies: Who can access their data
  - Cloud Providers – Enforce the Data Owners' Policies
- Access Control Solution
  - Only users that satisfy the access policies specified by the data owner is able to decrypt the data
  - Approach: Integrating Cryptographic Techniques with Access Control

Role Based Encryption (RBE)

- New RBE Scheme for Secure Data Storage
  - Integrating Cryptographic Techniques with Role based Access Control
    - Data encrypted to role or roles
    - Data owner encrypt the data in such a way that only users who satisfy the role based access policies (specified by the owner) are able to decrypt the data.
    - If data stored in cloud, if the cloud provider does not have the appropriate role(s), will not be able to decrypt the data
  - Characteristics
    - A user is able to join a role after the owner has encrypted the data for that role.
    - The user will be able to access the data from then on, and the owner does not need to re-encrypt the data
    - A user can be revoked at anytime, and the revoked user will not have access to any future encrypted data for that role
    - Revocation of a user from a role does not affect other users or roles in the system
    - Our scheme caters for role hierarchies and inheritance, thereby enabling roles to inherit permissions from other roles.
    - Reduces the level of trust on the cloud provider

Secure New Health Record System



MACQUARIE University
SYDNEY·AUSTRALIA

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

Role Based Encryption (RBE)

- Features
  - Constant size ciphertext
  - Constant size user secret keys
  - User-role assignment managed by individual role managers
  - Efficient user revocation (no user secret key update)
  - Forward Secure

Applications of RBE

- Secure Cloud Data Storage (e.g. New Health Records)
- Secure Data Sharing within a Large Organization
- Secure Data Sharing in a Multi Organization Context
  - Secure Data Sharing in a Consortium
- Large Scale Identity Management System - Aadhaar
- …

Concluding Remarks

- Secure Data Centric Approach
  - Examples: Cloud Platforms, Networks, Cloud Data Storage
- Challenges in Data Centric Security
  - Data Visibility
  - Data Categorization
  - Dynamic Data Evolution
  - Legacy System Integration
- Data Centric Security and Regulatory Issues
  - Evolving Regulations
  - Cross Border Complexity
  - Demonstrating Compliance