# Data Centric Security
# Issues and Challenges

Vijay Varadharajan

Global Innovation Chair Professor

Director: Advanced Cyber Security Engineering Research Centre

The University of Newcastle

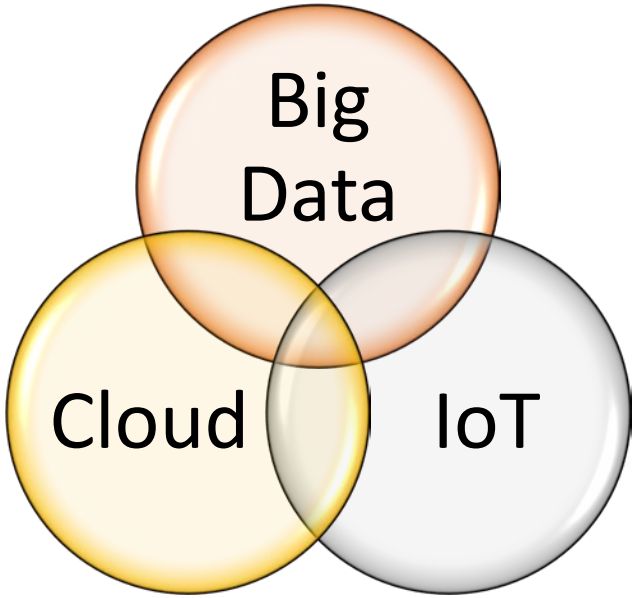
vijay.varadharajan@newcastle.edu.au

# Talk Overview

- Data Context and Related Challenges
- Data Centric Security
  - Controlling the Flow of Personal Data in Networks
  - Data Centric Security and Networks
  - Enforcement of Policies on Encrypted Data
- Concluding Remarks

# Technology Context

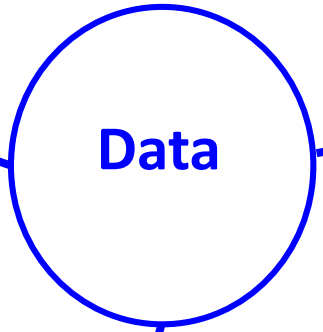
- Data Explosion → Big Data
- Systems of Systems → Cloud Computing
- Ubiquitous Computing → Internet of Things
- Pervasive Networking and Mobility
- Global Reach and Participation: Large Scale

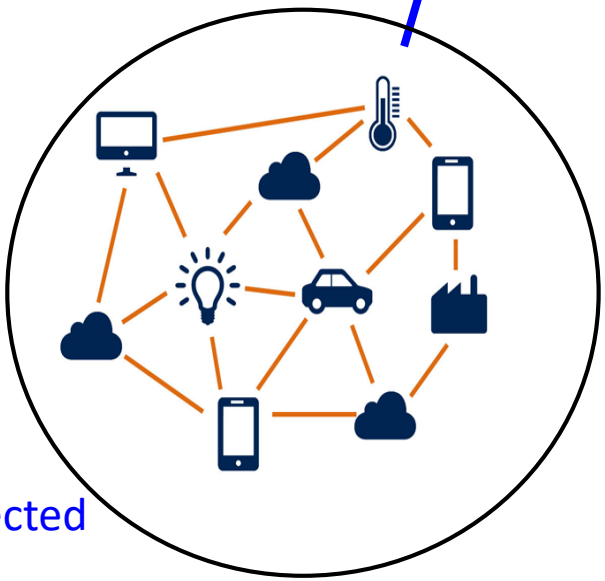


# Data Context

We are sharing data with many people

We are connected with more and more devices



**Data has become the new Currency in the Digital World**

→ New Innovations

→ Improve the quality of life

Devices are interconnected with each other

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

# Lots and Lots of Data

- Today
  - Over 3 billion online users added in last 10 years
  - Over 5 billion videos watched daily on Youtube alone
  - Some 300 hours of videos loaded every minute in Youtube
  - Dramatic Growth in Social Media – Facebook, Twitter, Instagram,…

❖ 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

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

# Data Context: 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
      - But 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
- So the presumption has changed
  - from controlling *opening* to controlling *closing*

# Data Context: Transformative Changes

- What is Public
  - Previously
    - one did somethings in public and some people saw them
    - they might or might 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 what Orwell imagined in 1984
    - In his case, not everyone in the population had a telescreen (only the Govt.)
      - Now, anyone can now use a mobile phone, video what is of interest, upload it to the Internet → pretty much available all over the world
    - May not be forgotten and can be recalled/aggregated
      - Can get a picture of someone else's life
      - (Potentially far more detailed than probably any secret police would have imagined some 30 years ago)

# Data Context

- Taking a data centric view
  - What rights should people have to control data about them
  - How to manage the data they create as to who gets access and for what purpose
  - We have always had this issue in security
    - Now compounded by the volume of data, different types of data, distributed, and dynamic changes
  - Not only security but also privacy
    - Much of privacy law assumes
      - when you entrust your private information to an entity (e.g. bank or hospital), that entity has an obligation to not to misuse that information and preserve the trust

# 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?
  - 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

- 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"
  - E.g. EU GDPR
- Industry
  - Usually prefer consistent rules to build customer relationships
  - Agreed rules to comply with regulations

# Data Related Challenges

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

# GDPR and Individual's Data

- 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 Centric Security

- Policy based Data Centric Approach
  - Secure Coupling Mechanism
    - Data coupled with policy
      - Policy explicitly attached to data
      - Data tagged with link to 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
  - Secure Processing Agents
    - Agents that process data must check and satisfy policies *before* using data
      - Trust in the Agents
      - Trusted Environment
    - Different types of policies
      - Audit and tracking policies
      - Access and usage policies
      - Obligation policies
  - Security Infrastructures
    - Policy to Data Mapping Infrastructure
    - Data to Policy Mapping Infrastructure

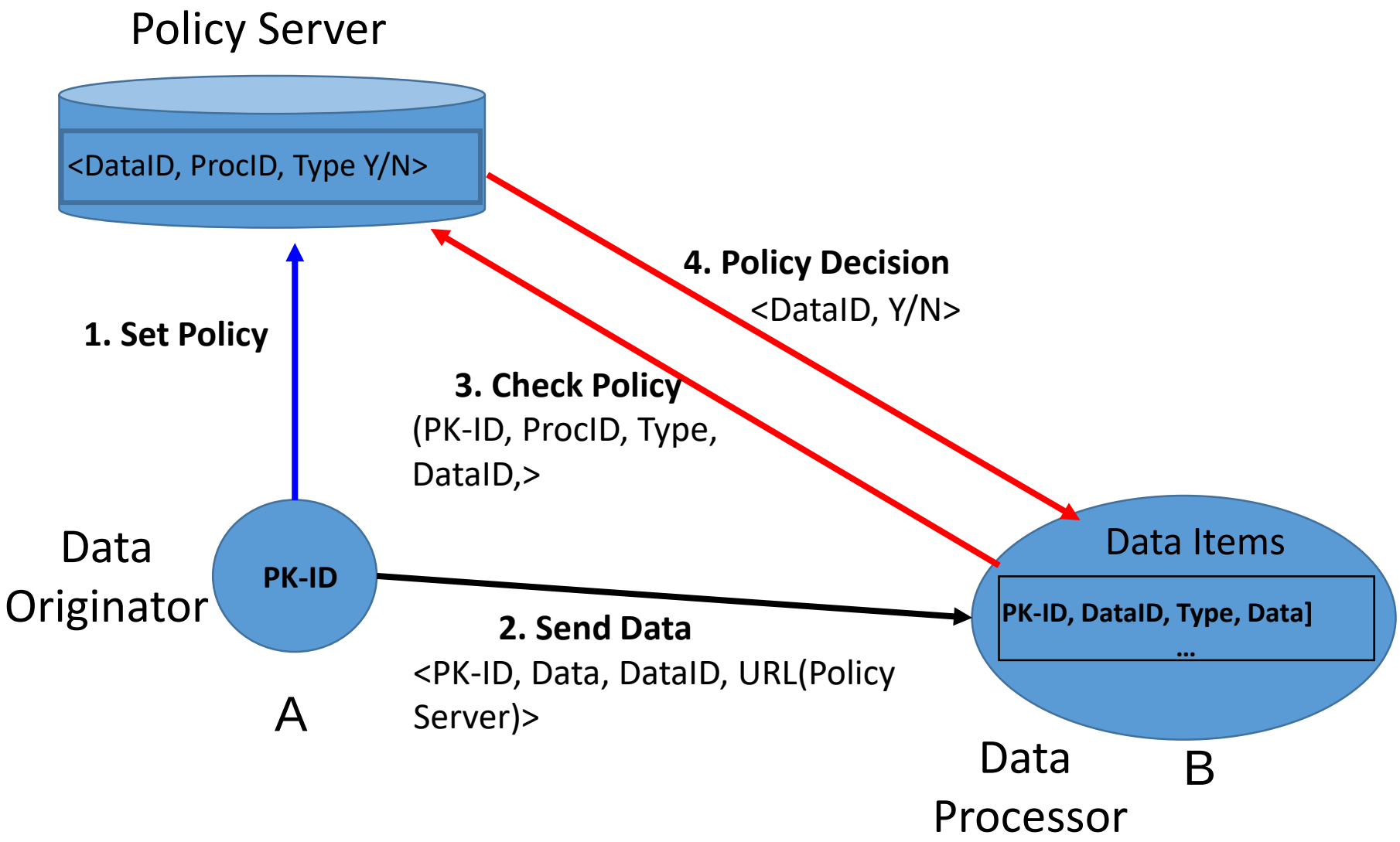
# Data Centric Security Project

- Initial Work: Data Tracking Protocol
  - A data tracking protocol that allows data originators/owners to track the flow of their data over Internet
  - Using existing infrastructures such as DNS and TLS for ease of adoption
- Further Work
  - Data originator/owner claiming the data
  - Data originator/owner providing proof of ownership
  - Data originator/owner enforcing data erasure
  - etc.
- Examples
  - Data from IoT devices transferred over the networks
  - Data in social media
  - Data sharing within enterprise and between enterprises
  - Data transfer over networks

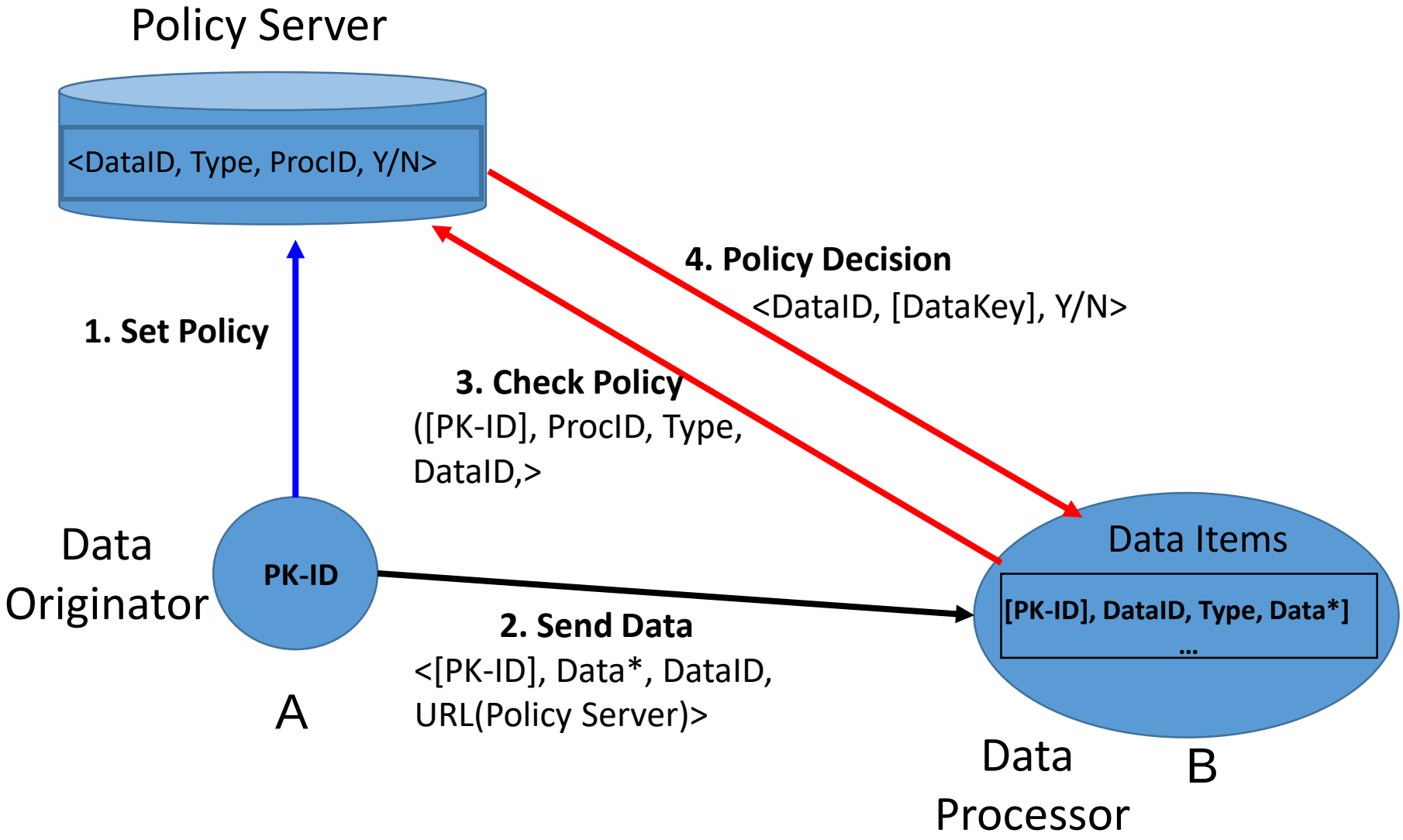
# Data Tracking Protocol (DTP)

- 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 services
  - Keeps track of data processors accessing user's data
  - Our Implementation
    - URL of policy service specified as part of DNS specification
    - Data protection using TLS

# Data Tracking Protocol (DTP)

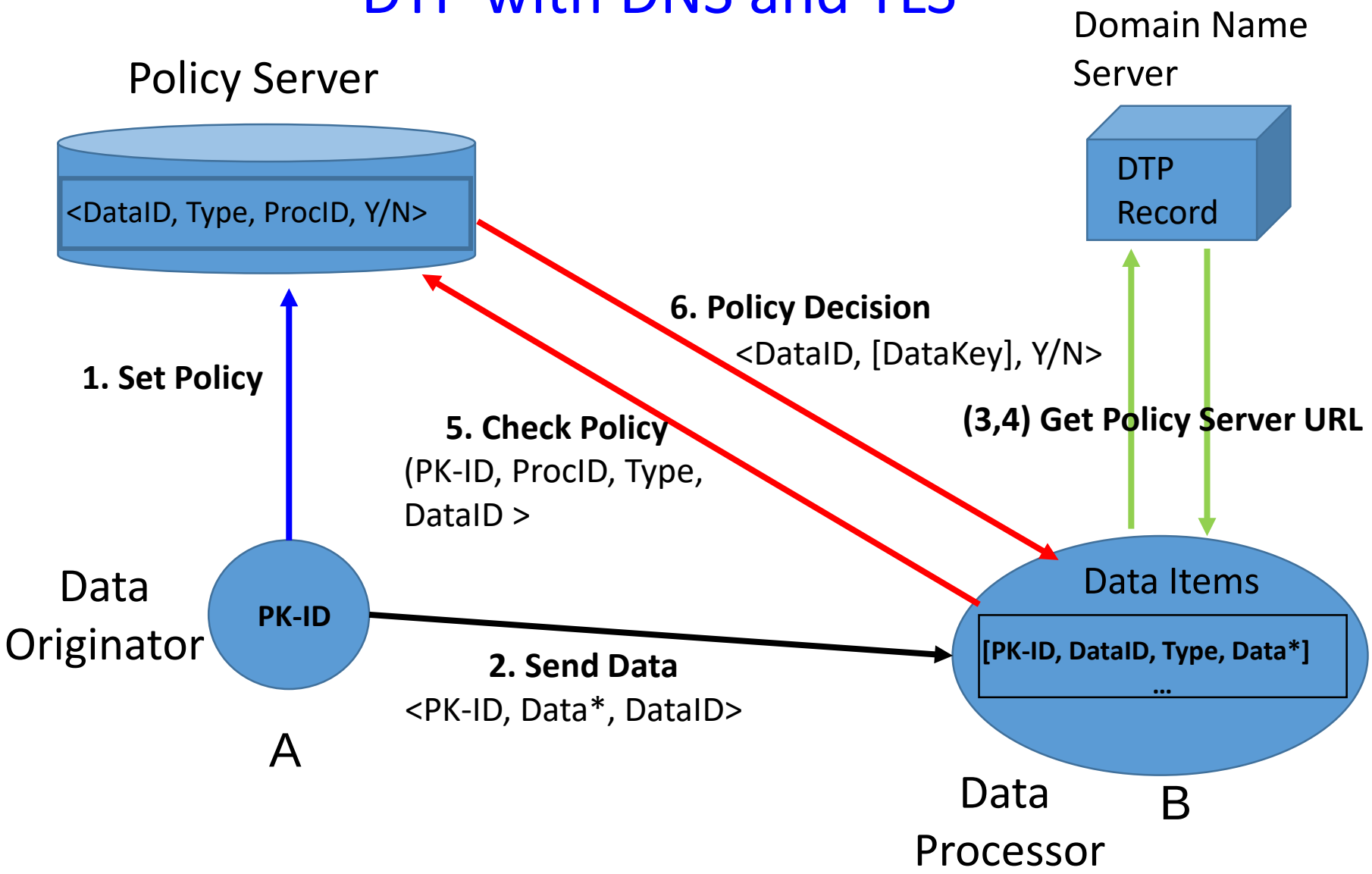


# DTP with Protected ID and Data (TLS)




- Policy Service can decrypt the DataID
- Policy Service provides the data key

# 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

# Data Centric Security

- 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, else on atomic policies
- Data Transfer Calculus
  - Compute allowed uses of data based on how it is derived
    - Data content dependent constraints
    - Data provenance based constraints
      - "a law-enforcement official may not act on improperly obtained evidence, but if the same information were redundantly obtained through lawful channels the official can act"
  - Transformation of data as it moves from one Processing Agent to another
  - Combination of data from multiple agents producing new data
  - Conflict Resolution

# Data Centric Security and Web Platforms

- Distributed Web Platforms
  - Third party applications running on a web platform
    - E.g. Facebook, Yammer
  - Users' control over what data third party apps can access
  - Many apps are more useful if they access users' sensitive data
  - Current solutions
    - Access control systems (often using some form of ACLs)
    - Security testing of app codes
    - Trust on software developers and producers
- Data Centric Approach to Web Applications
  - Developing web applications: Model-View-Controller pattern
    - Model (manages application data), Controller (interacts with users and model), View (model presentation)
  - Policy based data centric approach to accessing data
  - Also once data is accessed, policies control what apps can do with it

# Data Centric Security and Networks

- Network: Provide data (and services) to users
  - Communication model based on connections between devices rather than the actual data
  - IP Packets name communication endpoints
- Issue*: what data* one wants to access versus *where* it resides
  - Access to container where data resides
  - Access to the data in the container
- Security
  - Connection focussed rather than data focussed
  - Security mechanisms to identify, locate and communicate with the host
  - Security of data tied to the host where it is stored and how (over what kind of connection) it is obtained
- Data Centric Approach and Networks
  - Decouples the *what* from the *where*
  - Focus on securing the data itself, rather than the containers where it resides

# Data Centric Security and Networks

- Named Data Networks (NDN)
  - Request of data by *Name* rather than delivering data packets to receivers identified by IP addresses
  - If every data uniquely named, an NDN data packet is meaningful independent of where it comes from or where it may be forwarded to
  - If data changed in the packet, new packets generated with new *names*
- Securing Data in NDN
  - Naming data → Secure data directly by protecting every data packet
  - Securing the data
    - It is the *desired* data
    - From the *intended source*
    - *Not changed* in an unauthorized manner in transit
    - Sometimes *not disclosed* to unauthorized entities
  - E.g. Reading data from a host on the web
    - Host is the correct one
    - DNS giving a reliable indicator of where to find a host authorized to "speak for" the named data
    - Connection has been actually made to that host
    - Data retrieved over that connection is unaltered by any unauthorized intermediary
    - May be protected from eavesdropping

# Data Centric Security and Networks

- Securing NDN
  - Named, Secured Data Packets
    - Basic building block in NDN
    - Data packets in NDN need to be immutable
    - Each data packet carries a cryptographic signature (of the data producer) binding its name to the data content at the time of creation
    - Mechanisms for data packet to be encrypted if required
  - Key Management
    - All data producing entities to have cryptographic keys and facilities
    - Data consuming entities to verify the data packets
    - Efficient key management infrastructures required
  - Security Attacks
    - By default all data signed, including routing messages
      - Reduces spoofing attacks and malicious tampering easily detected
    - NDN messages are about data and not about addressing hosts
      - More difficult to send malicious packets to a particular target

# Data Centric Security and Networks

- Data Centric Access Control for NDN
  - Design Issues
    - Who is authorized to produce which type of data
    - Who is authorized to consume the data
    - Policies that specify privileges that enable a data producer to produce specific types of data
    - Policies that specify privileges that enable a data consumer to read specific types of data
    - Facility to deliver the credentials to the corresponding entities – producers, consumers
    - Facility to revoke the privileges of producers and consumers
    - If the data is encrypted, then we need techniques for enforcing policies on encrypted data

# Data Centric Security – Encrypted Data

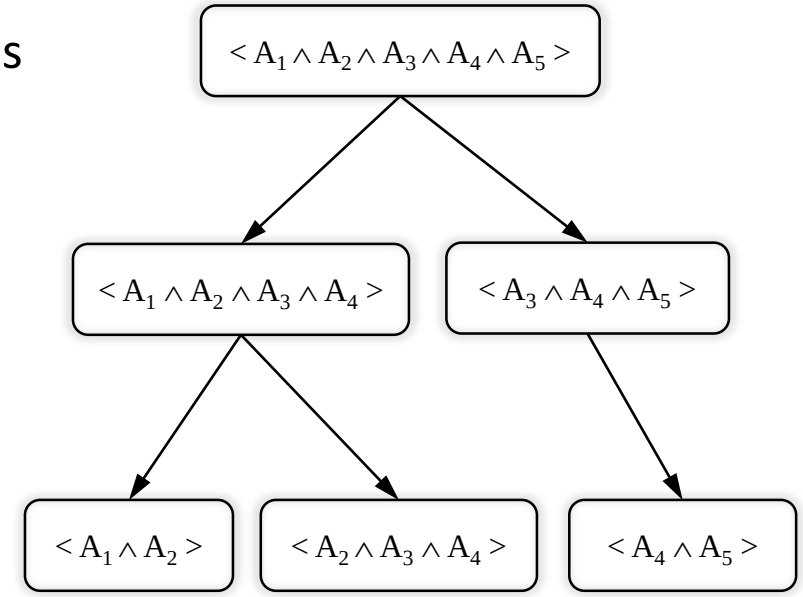
- Policy Enforcement on Encrypted Data
  - Assume now the Data Owners have encrypted their data
  - E.g. Cloud Data Storage System
  - Data Owners – Specify the Policies: Who can access their data
  - Cloud Providers – Enforce the Data Owners' Policies
- Access Control Solution
  - Only users that satisfy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

# Crypto based Access Control

- Some Existing Schemes
  - Hierarchical Identity-based Encryption Schemes (HIBE)
  - Attribute Based Encryption (ABE) and its variants
  - …

# Attribute Based Encryption (ABE)

- ABE: Goyal et al (2006)
- Each entity is associated to an attributes policy which is the "*and*" of a set of attributes
- Each entity is given the decryption keys for all the associated attributes
- Attributes set for each entity is a super set of the attributes for all its descendant entities



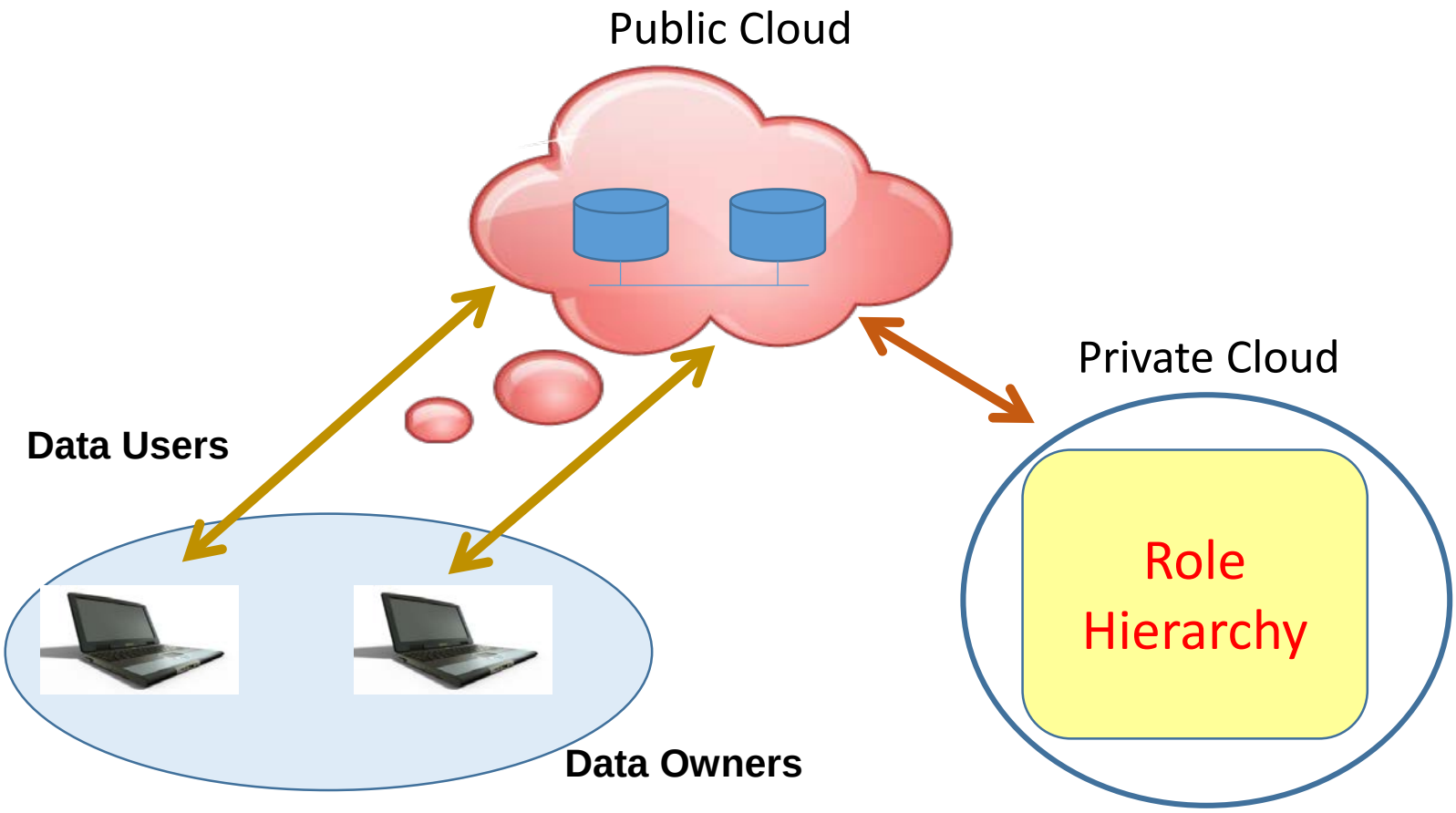
# Attribute Based Encryption (ABE)

- Data is encrypted to a set of attributes
- Users who have the private keys associated with these attributes can decrypt the data
- Disadvantages
  - Size of the user key is not constant
    - Changes with the number of attributes associated with each role and the role hierarchy
  - User revocation is inefficient (requires key update of all other non-revoked users in the same role)

# Role Based Encryption (RBE)

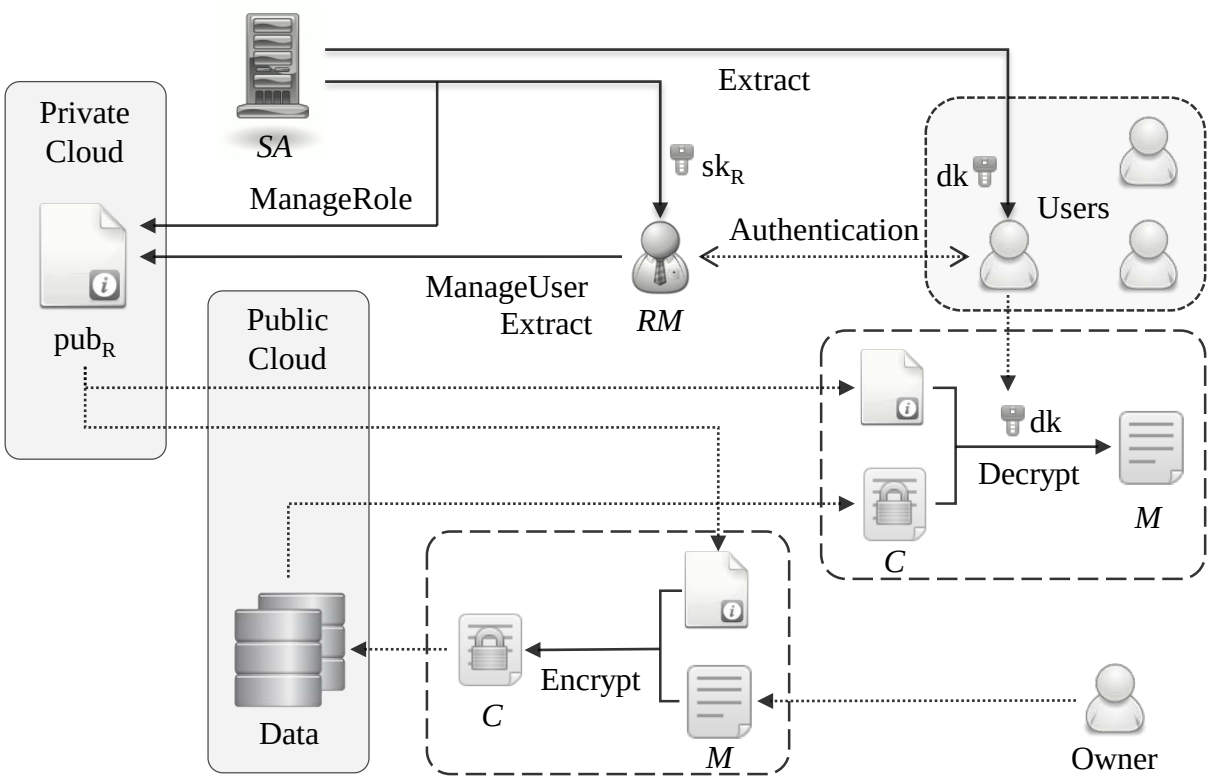
- New RBE Scheme for Secure Data Storage
  - Integrating Cryptographic Techniques with Role based Access Control
    - Data encrypted to role or roles
    - Data owner encrypt the data in such a way that only users who satisfy the role based access policies (specified by the owner) are able to decrypt 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.

# Policy Enforcement on Encrypted Data




THE UNIVERSITY OF NEWCASTLE AUSTRALIA

# Secure Hybrid Cloud Data Storage System

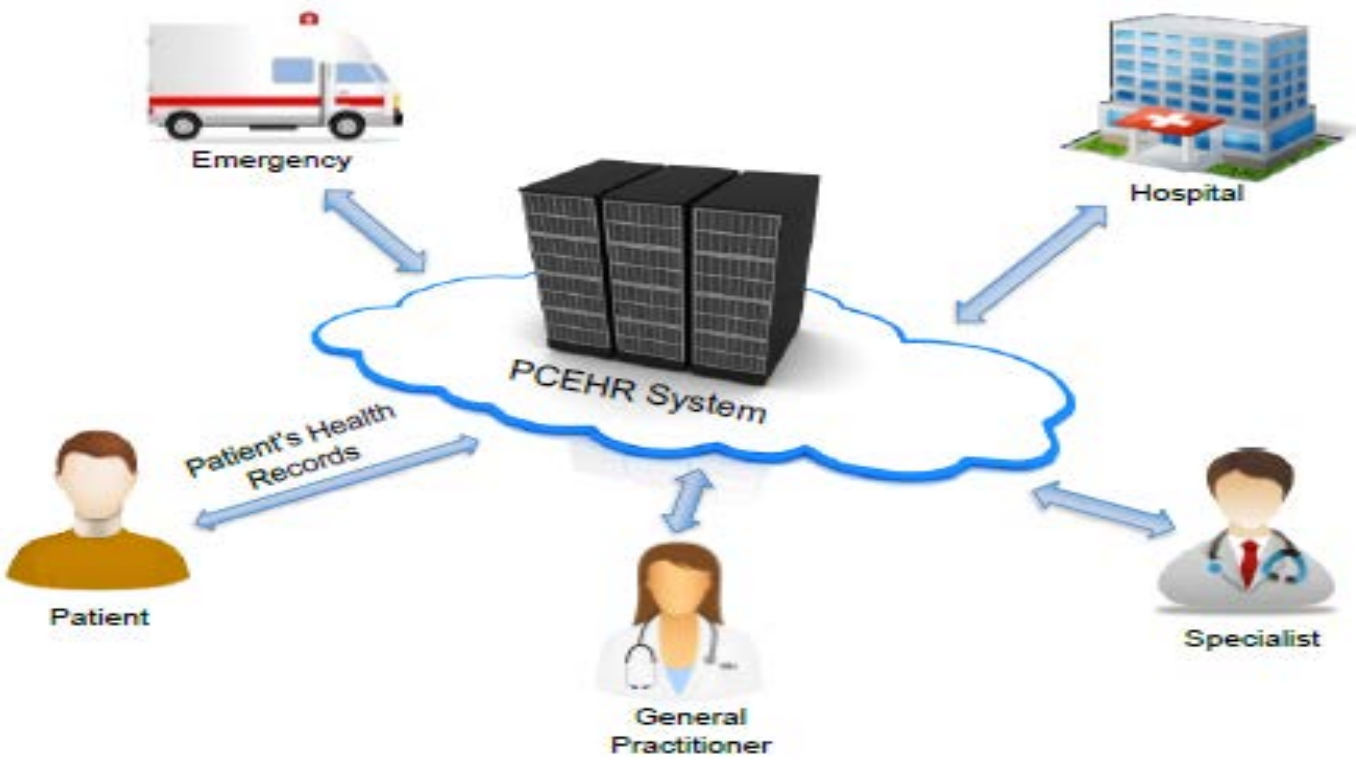



THE UNIVERSITY OF NEWCASTLE AUSTRALIA

# Role Based Encryption (RBE)

- Features
  - Constant size ciphertext
  - Constant size user secret keys
  - De-centralized role management
  - Efficient user revocation (no user secret key update)
  - Forward Secure

# Patient Centric Electronic Health Record (PCEHR)
## New Health Record System




THE UNIVERSITY OF NEWCASTLE AUSTRALIA

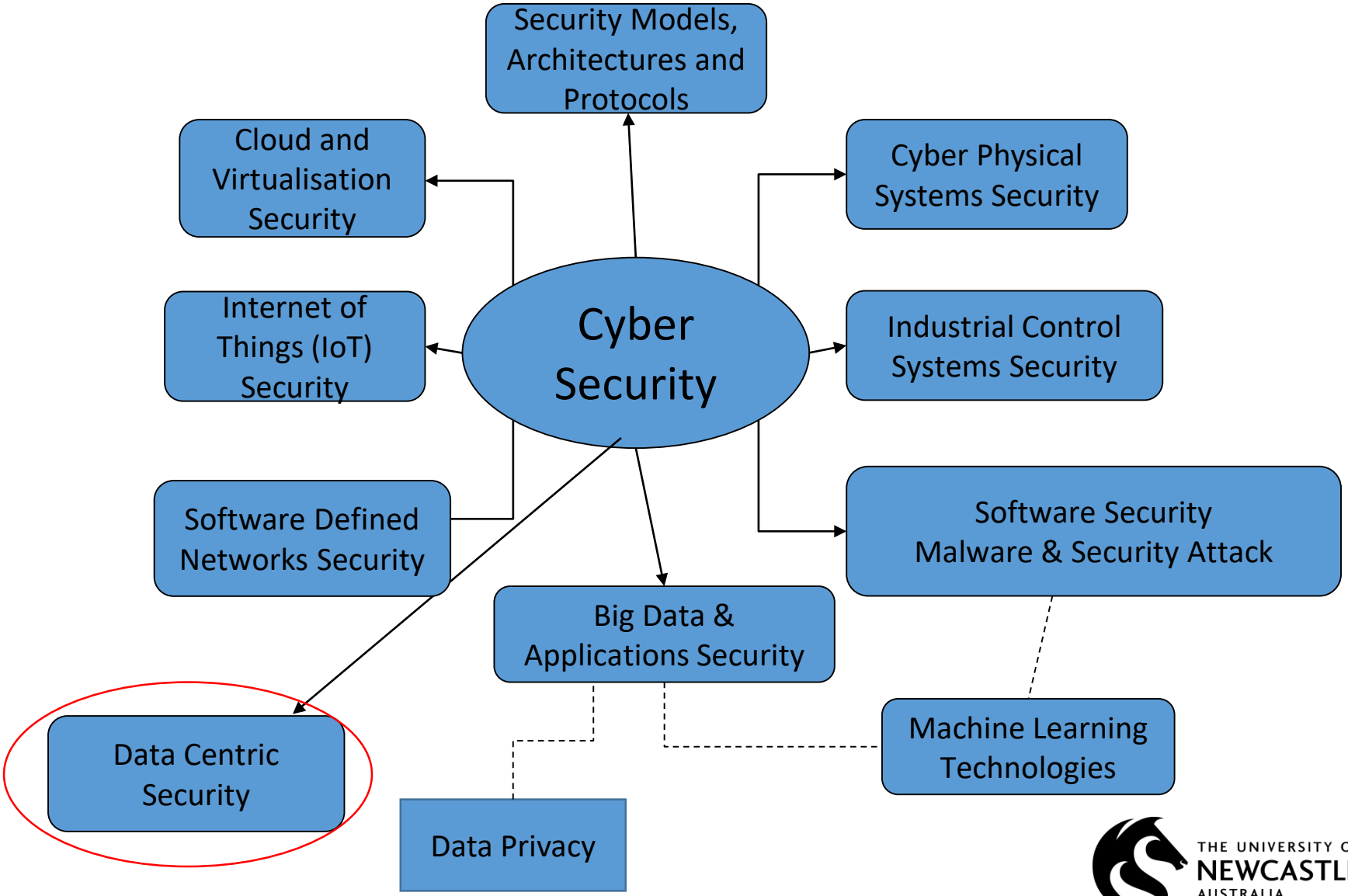
# Applications of RBE

- 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

- Challenges in Data Centric Security
- Secure Data Centric Approach
  - Policy based Control of Personal Data
    - Data Tracking Protocol with DNS and TLS
  - Data Centric Security and Networks
    - Named Data Networks Security
  - Enforcement of Policies on Encrypted Data
    - Role based Encryption (RBE) and its Applications

# Cyber Security Research Areas @ Newcastle




THE UNIVERSITY OF NEWCASTLE AUSTRALIA