UNIT - 4

Resource Management

Dr. VIJAYALAKSHMI M N
Assoc. Professor
Dept. of AIML,RVCE

Resource Management

- Introduction
- Object Coding and Resolving
- Resolving Discussion for nID Objects
- Resource Naming
- Resource Addressing
- Resource Discovery
- Resource Allocation
- Resource Management Scheme in U2IoT

Introduction

- Internet of Things (IoT) aims to realize the interconnections among ubiquitous things with the considerations of cyber, physical, and social aspects.
- In order to address such interconnections in heterogeneous networks, resource management becomes an essential issue.
- In the Internet, traditional resources may exist as physical resources and cyber resources, which mainly refer to physical and cyber components of limited availability in computer systems, and can be accessed to support different applications.
- To a certain degree, external physical components (e.g., devices and equipment) connected to a computer system belong to physical resources, and internal cyber components (e.g., file data, network connections, computing capabilities, and memory storage) belong to cyber resources.
- In the Internet, cyber resources are essential components during network connections and data interactions; and physical resources are supplementary components to provide necessary hardware support.

Contd..

- In IoT, ubiquitous things are connected into networks, and it turns out that resources have more generalized aspects.
- In addition to the traditional resources, more physical things also belong to the physical resources, such as sensors, actuators, advanced metering infrastructure (AMI), and locators.
- Such physical objects and the attached abstract components are involved in IoT to ensure interactive ubiquitous resource cross-sharing and cross-utilization.
- Things include not only physical things (e.g., object, behavior, tendency, and physical event), but also cyber things (e.g., entity, cyber action, cyber event, and service).
- In IoT, things refer to both physical and cyber resources, and resources cover both physical things and cyber things.
- Accordingly, things and resources can be regarded as almost equivalent in IoT.

Contd..

- Generally, resources have two basic features: existence and availability.
- The former means that resources can exist or be created, such as information/service-related software infrastructures or supporting hardware infrastructures.
- The latter refers to resources that may be used by one or more users in an application.
- Resources may have the following additional features:
  - *Authority* - Resources are assigned different access authorities, and may be used, shared, and transferred by authorized users via intra networks and internetworks. Therefore, a distributed and hierarchical management mode can provide diverse resource utilization

Contd..

- *Functionality* - Resources should support particular functions by performing a single or multiple tasks
- For example, in radio frequency identification (RFID) systems, a tag is assigned or associated with sensitive information, and such information is regarded as a resource for further network access and data inquiry.
- During wireless sensing, communication channels are resources for distributed autonomous sensors to support data transmission.
- *Shareability* - Resources can be interoperated by different entities and applications to achieve enhanced resource sharing and utilization in the context of large distributed networks. For instance, in cloud computing and grid computing systems, the resources located on multiple sites are shared and accessed by multiple resource users.
- *Power demand-supply* - In specific applications, resources and power have subtle relationships. In one aspect, resources mainly consume power as an energy demand for functional operation; in another aspect, resources themselves can also provide power as an energy supply in some cases. For example, in the smart grid, distributed electric vehicles can be regarded as both power demand and power supply to achieve power balance, which may perform charging operations during the off-peak time, and feed power back into the power grid during the peak time.

Contd..

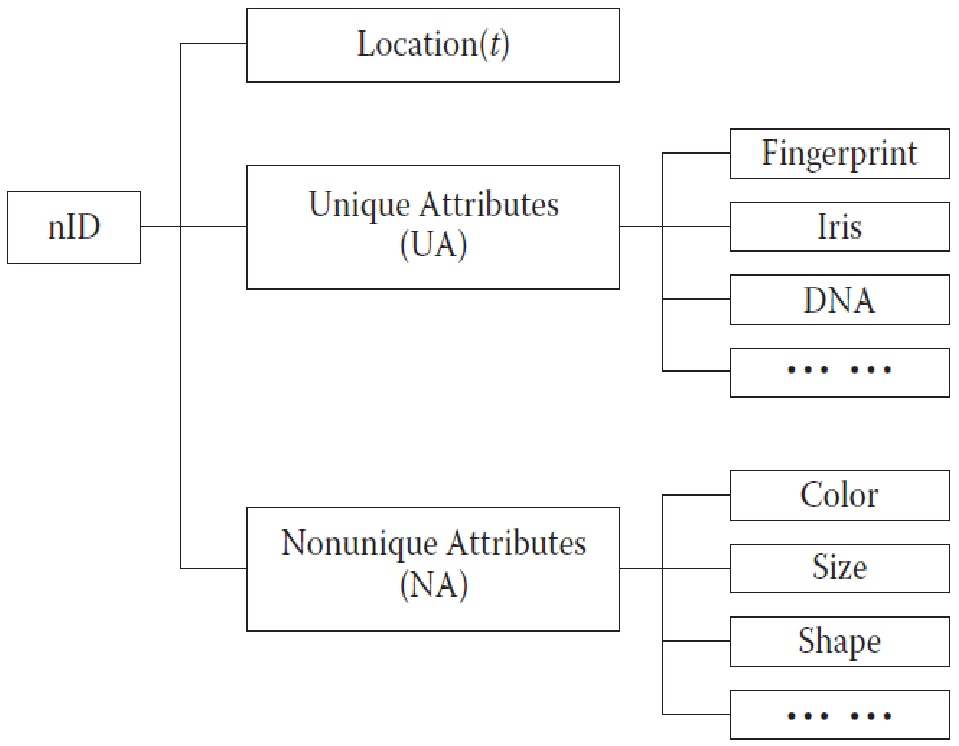
- *Cyber-physical collaboration* - The interrelationships are established among physical resources and cyber resources, which can be jointly applied to support a particular application according to the dynamic and hybrid environments.
- *Duty cycling* - Resources have the corresponding duty cycling, referring to the activity cycle and life cycle. It indicates that resources can be created, updated, released, and reloaded according to practical applications.
- Accordingly, resource management in IoT mainly includes two aspects: physical resource management and cyber resource management.
- Therefore, physical resources are usually mapped as cyber resources for further management, which makes cyber resource management become the main aspect of resource management in IoT.
- In this chapter, regarding ubiquitous resources, the main aspects of resource management are discussed according to resource naming, resource addressing, resource discovery, and resource allocation.
- Particularly, objects (i.e., identifier [ID] objects and non-ID [nID] objects) as a typical physical resource, coding and resolving are discussed in physical resource management.

Object Coding and Resolving

- Object Coding Discussion
  - In the process of objects' mapping from the physical world to the cyber world, objects' coding is significant to represent objects' information and reduce the data needed for transmission.
  - In the physical world, in addition to identifier-based objects, namely, ID objects, there are also nID objects, including objects unattached to any ID themselves, and objects attached to ID but that are unreadable or illegal. Currently, some kinds of mature object coding systems for ID objects exist. However, existing ID-based schemes may not be applicable for nID objects.
- Coding for ID Objects
  - For ID objects, there are two coding types according to the identifier: nonunique ID and unique ID.
  - For example, a bar code is a typical nonunique code, and an RFID tag is usually assigned with a unique ID (e.g., electronic product code [EPC]).
  - A bar code is a global common identification code, and is used in the EAN.
  - UCC uniform identification system. EPC is a product identification code and is stored in the RFID tag.
  - When ID objects are sensed and extra information needs to be added into the ID, coding should be further executed to create a new ID.
  - The assigned code should be able to be indexed to determine the code type.

Contd..

- Coding for nID Objects
  - Includes three main components: location(*t*), unique attributes (UA), and nonunique attributes (NA).



- *Location(t)*- Considering nID objects' existence, each object has a unique location attribute at a certain moment in a coordinate system. Location(t) means space-time information for one object. If the location is accurate and the detection interval is short enough to uniquely identify the object, objects' information at different times can be associated. In this case, their context information can also be associated.
- *UA* - UA refers to an object's unique attributes that can be used to identify the object uniquely. For instance, human biometric recognition attributes (e.g., fingerprint, iris, DNA, and face) belong to UA. Therefore, a fingerprint is a typical UA, and fingerprint identification is often employed in many important cases to identify a person uniquely.
- NA - NA refers to an object's non unique attributes that cannot be used alone to identify the object uniquely, including color, size, shape, and velocity. NA can be used along with other attributes to perform identification. For example, two cars that are the same color can also be differentiated by combing the information of location, size, or shape in some cases.

Contd..

- Combined Coding for Objects with ID and nID
- As objects may be attached to an ID when assigned a nID, codes for objects may consider the coexistence of ID and nID.
- The combination code should be composed by all or parts of the information, including its ID or nID.
- The object coding rule should facilitate resolving so that code can be resolved according to different requirements.
- The code structure should indicate the priority of resolving processes and be as simple as possible to reduce data transmission and storage.
- Compatibility is another significant factor that should also be seriously considered when constructing a combination code.
- Correspondingly, combined resolving is also needed according to corresponding coding approaches.

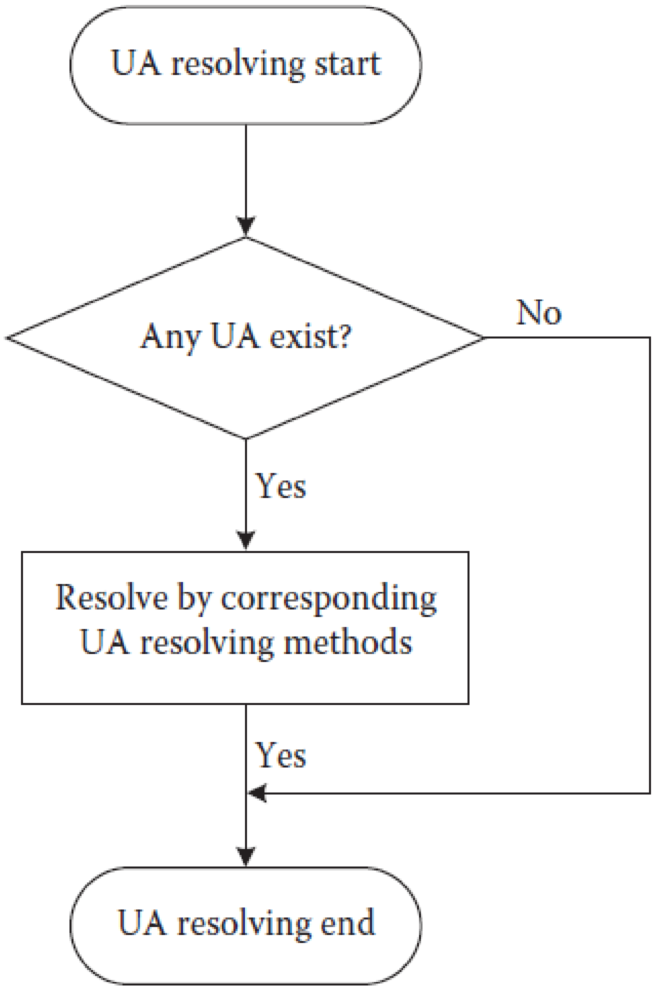
Resolving Discussion for nID Objects

- Object Name Service (ONS) is a typical object resolving method for ID objects,
- The resolving process refers to obtaining the corresponding readable information from the nID code so that physical objects can be matched to their corresponding identity and attributes in the cyber world.
- Since an object's different attributes exist separately, and multiple sensing technologies are applied, multiple resolving methods may be used to resolve a nID code.
- As a nID code consists of location(*t*), UA, and NA. Correspondingly, resolving for the three components is considered, including space-time information resolving for location(*t*), unique attribute resolving for UA, and nonunique attribute resolving for NA.

Contd..

Unique Attributes (UA) –

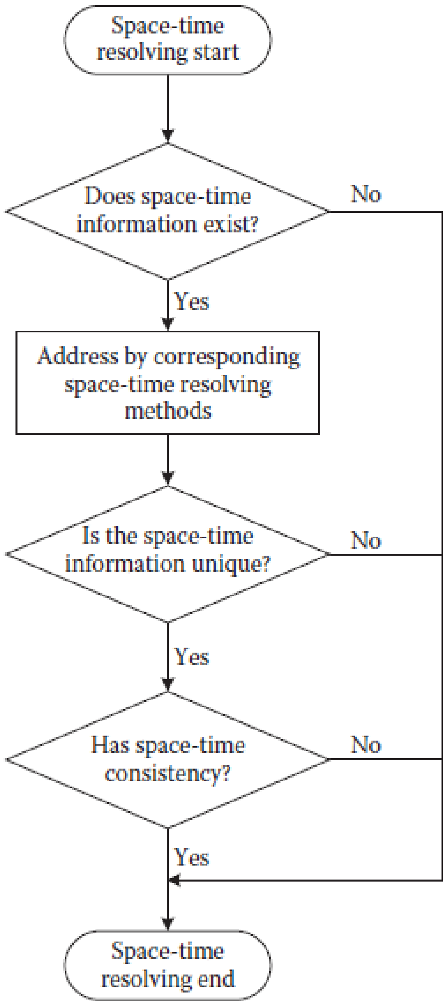
- Resolving When UA exists, UA resolving can be used—an example is shown in Figure
- Since different formats may be used for various UA, corresponding resolving methods shall be designed for UA.
- For the person's fingerprint, as an example, fingerprint resolving methods such as cluster analysis, principal component analysis (PCA), and nonlinear mapping (NLM) can be used.



Contd.

Go, change the world

- Space-Time Information Resolving A resolving process of its space-time information
- The goal of space-time information resolving can be reached when the space-time information exists uniquely and is unanimous and continuous.
- Different resolving methods are needed for different space-time formats.
- For instance, if the space-time information is in a global positioning system (GPS) format, the GPS resolving method should be applied.



Contd..

- Non unique Attributes (NA) Resolving - There are also various resolving methods for NA.
- NA can be used by combining other attributes for identification. The priority of the resolving processes depends on the concrete requirements, and may be combined and integrated for object identification.
- Resource Naming
- In addition to the above-mentioned physical objects, resources also have other existence forms.
- For example, the resource may be a back-end database in a distributed system, object information in a storage system, or the computing capabilities or memory storage in a cloud/grid computing system.
- Resource naming refers to the resource description based on formal languages, and it is important to achieve formal representation of such varied resources.
- In order to address the different types of resources, resource naming becomes important to achieve formal representation of such varied resources

Contd..

- In IoT, resource description refers to the physical and cyber resources in formal languages.
- An efficient resource description method can provide strong support for resource management.
- Heterogeneous resources may have different descriptive information with different formats and representations, and may also be indexed by different keywords referring to various attributes
- For instance, a uniform resource name (URN) is defined in the Internet to serve as a location-independent resource identifier, and to map the shared namespaces into URN-space.
- Here, ontology can be introduced to establish a reasonable resource description framework for both physical and cyber resources.

Contd..

- Ontology is a formal language to represent specific domain knowledge as a set of concepts within a domain and the corresponding relationships.
- The variants of ontology provide a compatible understanding among different applications.
- Meanwhile, ontology can describe the heterogeneous data in semantic contexts, and address the data integration with unambiguous conceptualizations.
- The typical ontology languages include Resource Description Framework (RDF), RDF Schema (RDFS), and Web Ontology Language (OWL), which provide the normative syntax to describe entity classes/concepts, properties, and relationships.
- For the ontology used in resource description, semantic conflict is an important obstacle to heterogeneous ontology languages.

Contd..

- Considering the importance of compatibility and interoperation in heterogeneous networks, semantic reconciliation becomes a major challenge for resource representation techniques.
- In the integration among different ontologies, metadata heterogeneity and instance heterogeneity are two major issues that should be considered .
- Metadata heterogeneity concerns the name- and structure-related information description conflicts, and instance heterogeneity refers to instance representation conflicts. Additionally, typical semantic conflicts, including naming conflicts and abstraction conflicts, should be considered

Contd..

- *Naming conflicts* refer to the relationships of “object, attribute, instance,” and include the synonyms and homonyms among concepts or properties of resources.
- Synonyms mean that similar or identical terms represent similar concepts, and homonyms refer to similar or identical terms that do not represent synonymous concepts.
- *Abstraction conflicts* may be caused by different attributes during modeling the concept from the physical world into the cyber world.
- Abstraction conflicts mainly consider the abstraction mapping relationship between two elements according to the aspects of class, generalization, aggregation, and computed function.

Contd..

- Based on ontology, RDF is introduced to provide functions of resource classification, indexing, keyword searching, and semantic analyses on metadata.
- RDF was originally developed by the World Wide Web Consortium (W3C), and this method can also be used to describe the things in future IoT.
- RDF is based on Extensible Markup Language (XML), which is a common syntax for the exchange and processing of metadata.
- XML syntax is a profile of Standard Generalized Markup Language (SGML), designed for the Web to alleviate the implementation of the parser.
- XML syntax also guarantees features, such as vendor independence, extensibility, validation, and the ability to represent complex structures. RDF further extends the general XML syntax and model to provide unambiguous semantic expression, and to enable consistent encoding and exchange of standardized metadata.
- Concretely, RDF applies the common conventions to facilitate modular interoperability among heterogeneous metadata sets, in which standard mechanisms for representing semantics should be established based on a simple, yet powerful and rigorous, data model.

Contd..

- Additionally, RDF provides both human-readable and machine-processable vocabularies to improve reuse and extension of metadata semantics among heterogeneous information.
- Vocabularies are the set of properties/attributes and metadata, and are defined by resource description communities to standardize the declaration of vocabularies.
- An *RDF triple* contains three components: subject, predicate, and object.
- The general format of an RDF statement is subject-predicate-object, in which the *subject* represents resources, the *predicate* represents all relative properties, and the *object* represents values.
- Note that the subject must be resources, the predicate must be properties, but the object may be both resources and literals.
- A set of triples is defined as an RDF graph, in which a subject is expressed as a *node*, a predicate is expressed as an *arc*, and an object is expressed as a node.

Contd..

- The set of *nodes* of an RDF graph is the set of *subjects* and *objects* of triples in the graph.
- RDF should consider human-friendly syntax, data integration/aggregation, nonexistent triples, bandwidth-efficient protocol, literal search, yes–no queries, addressable query results, and sorting results
- Physical Markup Language (PML) is another XML-based markup language used to describe physical resources and the corresponding relationships and interactions.
- PML can be applied to describe resources' inherent attributes of physical things, and can also present resources ascribed by humans.
- The main purpose of PML is to establish a general and standard method for formal modeling of objects, processes, and environments in the physical world.
- Based on PML, data structures and formats should be designed considering heterogeneous networks to provide efficient objects classification and generalization.
- Due to some degree of regularity and organization of physical objects, PML mainly uses a hierarchical organizational structure as a data configuration.
- Therefore, PML includes a special data structure name to contain an exact name of an entity, and the name element includes multiple attributes to distinguish the different name representations.

Resource Addressing

- Resource addressing is applied to locate the required resources by applying certain mechanisms .
- In IoT, resource addressing includes cyber resource addressing and physical resource addressing.
- Cyber resource addressing is similar to the addressing scheme in the Internet, and aims to access resources via domain names and an IP address–based scheme.
- Physical resource addressing includes two aspects: one is the physical objects' mapping from the physical world to the cyber world, and the other refers to the same as the cyber resource addressing.

Contd..

- The Domain Name System (DNS) is a typical resource addressing mechanism in the Internet, which is a typical hierarchical naming system to connect computers, services, or resources into the Internet or other private networks.
- Therefore, a domain name is applied as an identifier to define a realm of autonomy, authority, or control on the Internet.
- Domain names are formed by the rules and procedures of the DNS. Generally, a domain name represents an Internet Protocol (IP) resource, and can be applied in various networking contexts and application-aware naming and addressing purposes.
- Uniform resource locator (URL) is defined as a unique address for a document and other resources on the Internet, and applied to access a Web site.
- URL includes a protocol identifier and a resource name, in which the resource name indicates the IP address or domain name where the resource is located.

Contd..

- During resource addressing, a resource should be assigned with a unique available identifier—there are three typical addressing schemes in the Internet:
- *Universal Description, Discovery, and Integration* (UDDI)
  - It is a platform-independent, XML-based registry to realize self-presentation on the Internet.
  - UDDI is sponsored by the Organization for the Advancement of Structured Information Standards (OASIS), for enabling businesses to publish service listings and resource discovery.
  - The UDDI scheme applies a Universally Unique Identifier (UUID) for resource identification, and is based on the Transmission Control Protocol/Internet Protocol (TCP/IP), Hypertext Transfer Protocol (HTTP), XML, and Simple Object Access Protocol (SOAP) to establish a unified service description format and service discovery protocol.

Contd..

- *Electronic Numbering Mapping* (ENUM)
  - It aims to map the DNS E.164 Number to a uniform resource identifier (URI), and it includes two aspects:

    (1) creating a domain name from a telephone network (TN) and resolving it to an Internet address (a URI) using DNS, and

    (2) hosting ENUM domain names in the *E164.ARPA* domain.
- *Handle system*
  - It is based on a uniform resource name (URN), to achieve assigning, managing, and resolving persistent identifiers for cyber resources.
  - The scheme includes an open set of protocols, a namespace, and a reference implementation of the protocols.
  - The protocols enable a distributed system to store identifiers (e.g., names and handles) of arbitrary resources, and resolve the handles into the information that is necessary to locate, access, contact, authenticate, and use resources.

Contd..

- Additionally, Internet Protocol version 6 (IPv6) is developing to address the problem of IPv4 running out of addresses, and it is a revision of the Internet Protocol with extremely large address spaces.
- Such address expansion can accommodate more nodes, devices, and users, and is flexible for allocating addresses and routing traffic.
- For physical resource addressing, ONS is a typical approach to realize the interconnection between the physical world and cyber world.
- ONS, presented by EPCglobal, is applied to access an object's information and its related services by EPC.
- Here, EPC is a universal identifier that provides a unique identity for a tagged physical object.
- ONS as an automated networking service realizes the resource addressing in IoT, and also has similar functions to point each computer to Web sites.
- The main design idea is first to encode EPC into a fully qualified domain name (FQDN), and then to use DNS infrastructure to query for additional information.
- Concretely, in the case that a reader identifies a tag, the code is transmitted to an ONS on local networks or the Internet via middleware to find the stored information on the product.

Contd..

- During resource addressing, ONS points middleware to a server where a file about that product is stored.
- The middleware retrieves the file, and the information about the product in the file can be forwarded to an inventory or other applications.
- Furthermore, EPC Information Services (EPCIS) is designed to enable EPC-related data sharing within and across enterprises, and the EPCIS standard defines standard interfaces to enable EPC-related data to be identified and subsequently to be queried with service operations and an associated data model.
- Such data interaction is mainly based on persistent back-end databases. Meanwhile, application-to-application resource sharing may occur without needing a persistent database.

Resource Discovery

- Resource discovery is an essential component for searching and finding suitable resources (e.g., sensor nodes) to satisfy application requirements.
- The existing resource discovery mechanisms can be classified into two main categories: peer-to-peer (P2P) architectures and directory-based architectures
- P2P architectures are based on the distributed mechanisms, in which network entities negotiate one-to-one with each other to discover the available and required resources (e.g., services).
- There are two basic resource discovery mechanisms in P2P systems: pull mode and push mode.

Contd..

- *Pull mode*
  - In pull mode, users transmit a discovery message to networks by broadcast or multicast communications and require the matched resources for specific requirements.
  - Providers respond to the challenged query by transmitting a description of the service or resource.
  - For instance, the Service Discovery Protocol (SDP) is used in pull mode–based Bluetooth, as well as for on-demand *ad hoc* networks in the pull mode.
  - In the pull mode, resource information can be updated and notified along with the back-end resources.
- *Push mode.*
  - In push mode, providers advertise periodically by broadcasting or multicasting the location and attributes of resources and services, so that users can build a local database with all the resources available on the network.
  - The back-end resource information can be aggregated from resources in a periodic interval.

Contd..

Go, change the world

- Directory-based architectures mainly have a centralized or distributed database to aggregate and index the resource information.
- The resource providers register their resources/services in the directory, and resource users query the directory to obtain the required information.
- According to network topology, resource discovery mainly has three models: the centralized model, distributed model, and hierarchical model
- *Centralized model*
  - The centralization model refers to the allocation of all queries processing capability in a single node.
  - The main characteristics of a centralized approach include high controllability and efficiency.
  - All lookup and update queries are sent to a single entity in the system.
  - The centralized architecture is based on a central directory that aggregates information from every provider, and responds to queries from every user.
  - Central directory architecture is a simple solution, and easy to administrate, but the directory can represent a bottleneck and a single point of failure, which causes the whole system's failure.
  - Therefore, such a solution is suitable for small-scale networks, and a centralized model has inherent limitations for overloading the central server, and the model may not adapt well to dynamic network conditions.

Contd..

- *Distributed model*
  - There is no centralized control in the distributed model, in which complete autonomy, authority, and query processing capability are distributed over all resources in the system.
  - Note that no entities in the distributed mode are more important than others. In the case that any entity is compromised, the comprised entity will not disturb other entities in the system.
  - In the distributed model, several directories cooperate in a P2P mode to maintain a distributed database of information about resources and services.
  - Directories can exchange information with all other directories by multicasting communications so that each directory maintains a complete database about all resources and services.
  - Information exchange among directories in different clusters can be achieved using a P2P scheme, instead of using a lower advertising frequency than that within a cluster, although clustered solutions are more scalable and suitable for large-scale networks, in which complex algorithms should be designed to manage clusters and guarantee cluster stability.

Contd..

- *Hierarchical model*
  - Hierarchical organization has several advantages, including scalability, adaptability, and availability.
  - A tree-like structure is introduced to establish the resource discovery mode, in which only direct links in a hierarchy are from the parent nodes to their child nodes.
  - With the hierarchical architecture, the network is divided into domains with a hierarchical structure, and directories have a parent–child relationship.
  - This solution is fully scalable, but it enforces a rigid hierarchical network organization, which does not fit well in *ad hoc* environments.

Contd..

- During resource discovery, a distributed hash table (DHT) is usually applied as a distributed resource discovery strategy, and provides a lookup service similar to a hash table, in which the pairwise secrets are stored in a DHT, and any participating node can efficiently retrieve the value associated with a given key.
- The maintenance of the mapping from keys to values is performed by distributed nodes in such a way that a change in the set of participants causes a minimal amount of disruption.
- This allows a DHT to scale to an extremely large number of nodes and handle continual node arrivals, departures, and failures.
- Therefore, three main components (i.e., resource publication, resource lookup, and resource database) are included

Contd..

- The resource discovery challenges are as follows:
  - *Heterogeneous.* - The component discovery service should support discovery of heterogeneous components, which belong to different distributed computing models (i.e., Java RMI, NET, and CORBA), and apply different communication networks (IP network and wireless sensor network [WSN]).
  - *Activeness* - Activeness means that a user looking for a resource will initiate the request, for which the resource provider responds. The resource providers should be available and actively listen for the user’s requests.
  - *Scalability* - Scalability refers to the resource organizations being easily extended and able to enhance the system resource by adding new functionality at minimal effort.
  - *Multilevel matching* - There should be support for matching components at multiple levels, such as syntactic, semantic, protocol, and quality of service (QoS). This multiple-level matching helps to improve query matching.

Contd..

- *Reliability* - Reliability is defined as the ability of a system or component to perform its required functions under stated conditions for a specified period of time. The component discovery service should be reliable so that it consistently discovers the best components that match the given query. The discovery service should neither fail nor compromise the quality of the component matching process.
- *Security* - The component discovery architecture should be secure. This is mainly because the entities that form the component discovery architecture are in an open environment that suffers from severe threats.

Resource Allocation

- Resource allocation focuses on distributing resources to realize optimal allocation on the available resources.
- Resource allocation in the Internet aims at establishing a ubiquitous interconnection among the available resources in the networks, and focuses on realizing pervasive seamless and transparent access to heterogeneous resources and services.
- It is significant to present an adaptive resource reservation strategy according to the dynamical networks and application requirements, and an optimal resource allocation scheme is necessary to address underprovisioned or overprovisioned scenarios.
- There are several algorithms to be applied in resource allocation, mainly including artificial intelligence, theory of random graphs, and P2P-based approaches

Contd..

- *Artificial intelligence* (AI)
- Artificial intelligence–based resource allocation approaches include mainly two types: game theory and heuristic methods (e.g., ant colony optimization and genetic algorithms).
- Game theory is an effective strategy for resource allocation, bandwidth sharing, energy balance, and pricing models, in which the Nash equilibrium state is achieved among the competing users under the shared resources and services.
- The Nash equilibrium refers to that in which the optimal outcome of a game is one where no player has an incentive to deviate from the chosen strategy after considering an opponent's choices
- It turns out that an individual can receive no incremental benefit from changing actions, assuming other players remain constant in their strategies. Note that a game may have multiple Nash *equilibriums* or none.

Contd..

- Ant colony optimization (ACO)
  - is also an efficient algorithm inspired by the foraging behavior of ant colonies, and can be applied for dynamical resource allocation and optimal resource dispatching.
- Genetic algorithm (GA)
  - is another approach applied to address resource optimization.
  - It is noteworthy that for an online decision on optimal resource scheduling, the performance analysis of GA-based approaches needs further research.
  - GA is based on a search heuristic mode inspired by the mechanics of natural selection and natural genetics (e.g., inheritance, mutation, selection, and crossover), to realize optimization of resource allocation.
  - Additionally, particle swarm optimization (PSO) is a population-based stochastic optimization technique
  - It combines the social psychology principles in social cognition, human agents, and evolutionary computations.

Contd..

- *Theory of random graphs*
- Random graphs provide an important paradigm that may be used to realize the network connectivity, interoperation, and optimal resource distribution.
- The distributed network structure should be considered, and it generates regular resource allocation networks to achieve distributed and cooperative load trade-off.
- Another strategy is regarded as self-configuring and self-optimizing, which is inspired by Erdos-Renyi (ER) random graphs
- Such connectivity distribution of the resources satisfies a random graph.
- Thereby, the load distribution becomes equitable across all nodes with computing resources.
- Additionally, IoT network environments should consider the resource allocation with homogeneous nodes and heterogeneous nodes

Contd..

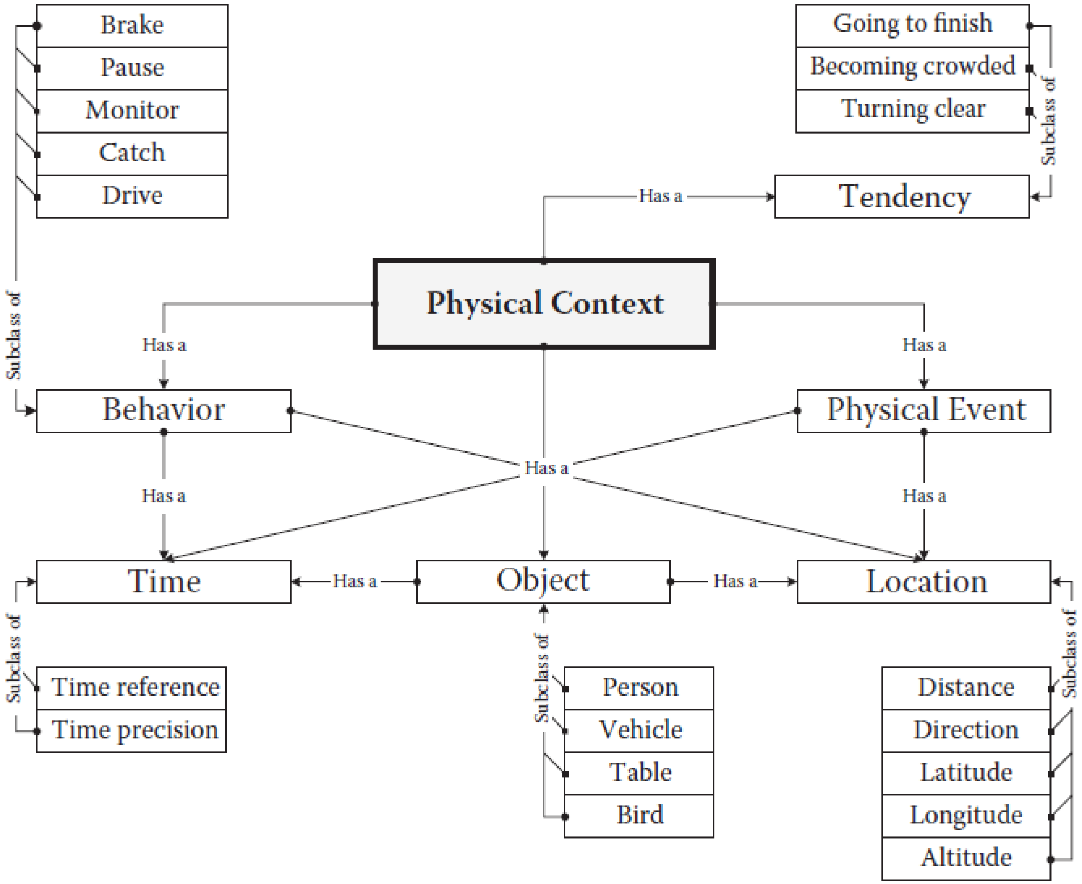
- *P2P-based approaches*
  - Interdomain resource scheduling is important for difficult provisioning of the same QoS requirements and strict threads/processes concurrency requirements across multiple domains.
  - P2P-based approaches address resource allocation in a single domain by using a typical broker.
  - The resource monitor may need to use the same information in all the domains even if the computing resources belong to different domains.
  - It is known that the topological structure of the underlying physical network has significant impacts on the resource utilization of the P2P overlay network
  - Meanwhile, P2P overlay routing may increase traffic congestion at the Internet backbone since the current traffic engineering algorithms for Internet service providers (ISPs) do not consider the traffic demands

Resource Management Scheme in U2IoT

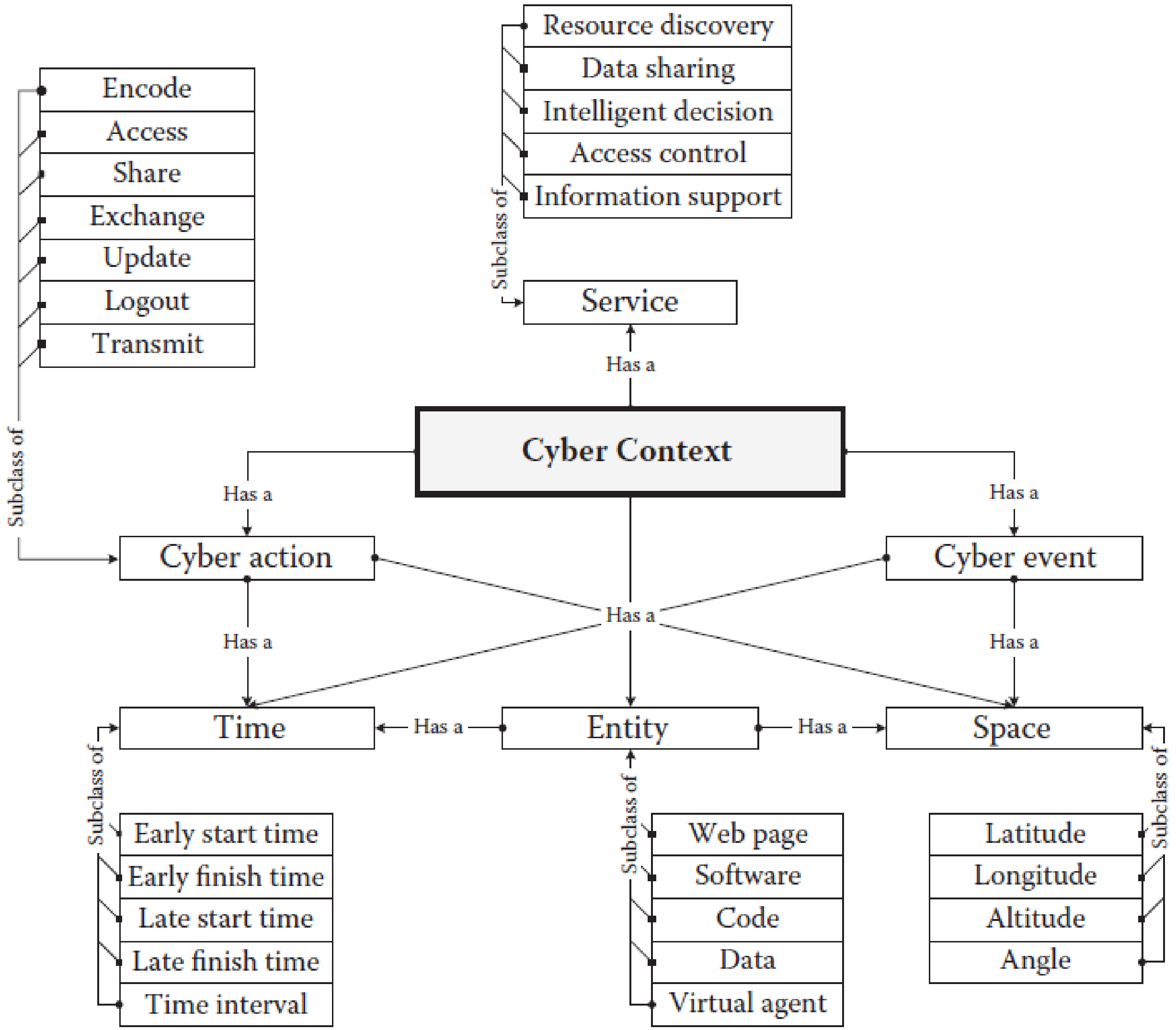
- In unit and ubiquitous IoT (U2IoT), resources are usually context sensitive, and therefore context information is relevant for resource management.
- Hereafter, context information is discussed according to physical- and cyber-related information, and a resource management scheme is established for U2IoT.
- Context Information
  - Context-aware computing has received great attention for acquiring various contexts from different context providers, interpreting contexts through context reasoning, and delivering contexts in both push and pull modes.
  - Initial efforts on context-aware service creation focused on an infrastructure-based solution.
  - A model-driven approach facilitates the creation of context modeling to tackle context awareness in the application layer.
  - The captured low-level static and dynamic contexts are aggregated by these devices to provide another set of deduced high-level contexts to build context-aware resource services

Contd..

- In U2IoT, the context mainly includes three aspects: physical context, cyber context, and physical-cyber context.
- Physical context mainly considers the context information during physical things mapping into cyber things.
- The physical context includes the subclasses of object, behavior, physical event, and tendency.
- Therefore, object, behavior, and physical event further have the mandatory subclasses of time and location.



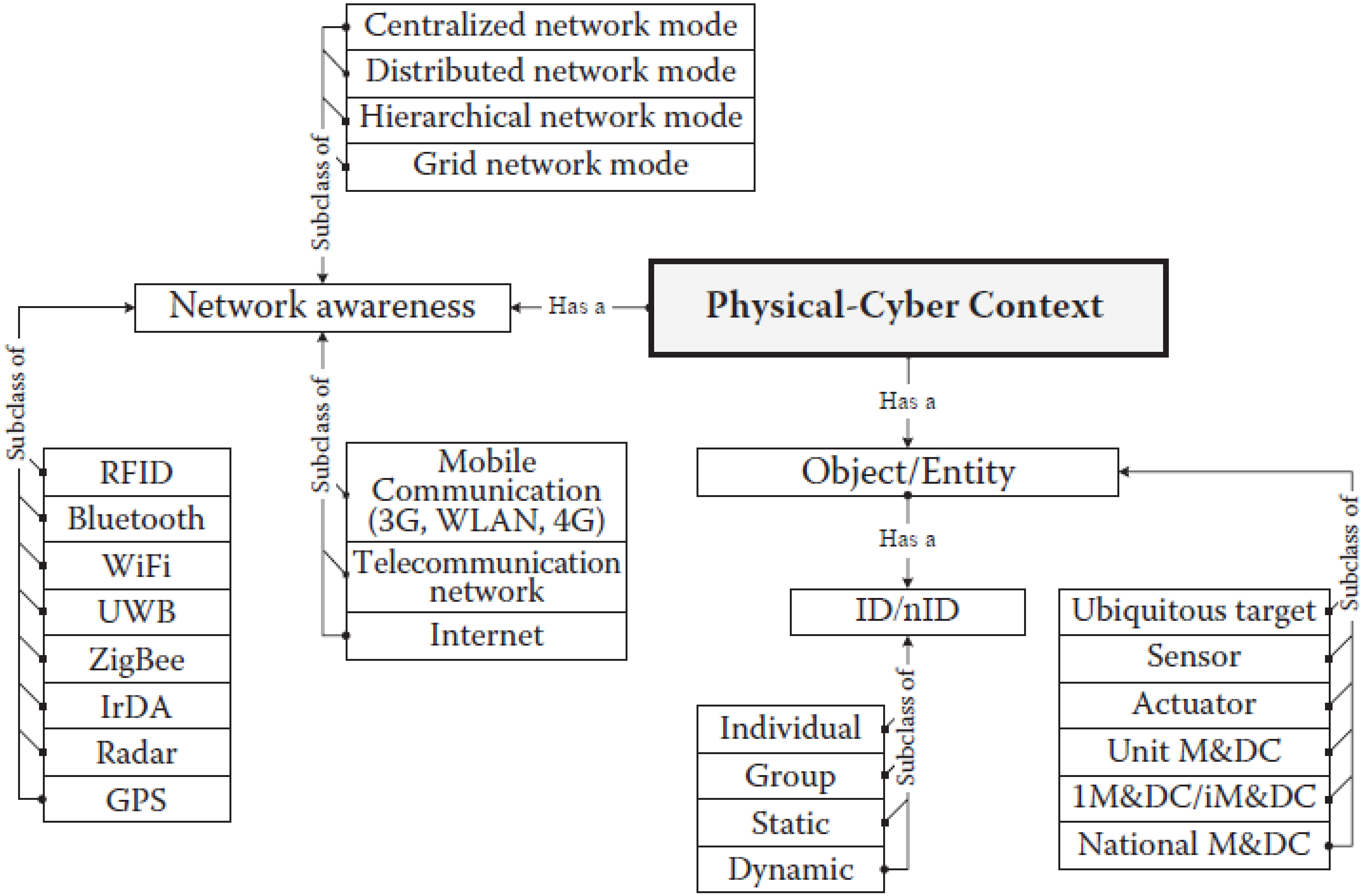
Contd..



The cyber context mainly considers the context information during cyber things mapping into physical things.

It includes cyber context descriptors, in which the cyber context includes the four main subclasses of entity, cyber action, cyber event, and service.

Contd..



The physical-cyber context mainly considers the context information during physical things mapping into cyber things, and then the cyber things remapping into the physical things. The physical-cyber context mainly considers the subclasses of network awareness and object/entity. .

Resource Management Scheme for U2IoT

Go, change the world

- A resource management scheme is established for U2IoT with the considerations of resource management in unit IoT and ubiquitous IoT.
- Concretely, resources in unit IoT (called unit resources) mainly include the resource-constrained sensors and actuators, unit management and data center (M&DC), and corresponding abstract resources, which provide basic functions and services for an application.
- Resources in ubiquitous IoT (called ubiquitous resources) mainly include physical and cyber forms of local, industrial, and national management and data centers (lM&DCs, iM&DCs, and nM&DCs) and other abstract resources in local IoT, industrial IoT, and national IoT.
- In the meantime, additional service selection and adaptation management are applied to support the resource management for the sensor and actuator resources

Resource Management in Unit IoT

Go, change the world

- Unit resources mainly refer to the available resources in unit IoT, in which resources are under intelligent management to achieve self-optimization functions for the sensors, actuators, and M&DCs.
- A universal management mode can be applied by resource components that implement service matching and binding services.
- When sensors receive a service request to access data, the sensors may challenge the service registry to obtain the available services.
- If the service is available, it will bind the service for internal resource supports.
- Otherwise, the service request will be transmitted to the upper management and data center for external service acquiring.
- In the case that the service selection component has selected a service provider, the service manager binds the available service to the appropriate request.
- Note that services in the sensors and actuators are implemented by the self-organized resource component, and the same resources may be concurrently used by multiple applications.
- The self-adaptive resource access mode is applied for unit resources.
- The self-adaptive mode has full-fledged service specifications and contextual information gathered by sensors, in which each service's functioning parameters can be autonomously changed according to specific network scenarios.
- The self-adaptive mode applies the adaptation control loop to adapt service behavior by polling the system state–related components.

Contd..

- Therefore, the resource components mainly include battery capability, process overhead, communication load, and available memory.
- During resource management, the system state–related components perform a comparison on the current system state against the existing adaptation policies, and the required adaptation is launched upon reaching a threshold.
- Furthermore, other abstract resources are managed by an upper service manager.
- Unit M&DC can also be applied to manage the distributed sensors and actuators in unit IoT, in which unit resources are managed by the service selection component.
- Unit resources provide services for the bottom sensors and actuators, and also request service from ubiquitous IoT.
- Unit IoT maintains a service directory for its resources, and matches the service request and available services for applications.

Resource Management in Ubiquitous IoT

- Ubiquitous resources refer to physical and cyber resources in local IoT, industrial IoT, national IoT, and global IoT applications–related resources.
- The local resources and industrial resources are relatively independent, and interact with each other to provide enhanced services and functions.
- Cloud management is introduced for IoT to provide anything as a service (XaaS), in which cloud computing may have three service delivery modes, including software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS).
- The combined use of these three modes is referred to as the service-platform-infrastructure (SPI) model

Contd..

- SaaS provides the service provider's applications running in a cloud infrastructure, which refers to the ability for a user to use on-demand software.
- The user has no need of management or control of the software infrastructure (e.g., storage, servers, and networks).
- PaaS provides the capability for an individual to use the applications without needing to maintain the hardware and software, which build service providers' platforms to create applications running in the cloud infrastructures.
- IaaS provides the service provider's hardware resources, including the storage (e.g., data replication, backup, and archiving), the server (e.g., computing requirements), and the connectivity domains (e.g., network load balancing and firewalls).

Contd..

- Additionally, XaaS also includes software as a service (SaaS), communications as a service (CaaS), network as a service (NaaS), and monitoring as a service (MaaS).
- Such cloud resource management realizes the resource optimal allocation and utilization.
- In U2IoT, three network-based interfaces are suitable for resource management, including network resource interface, network service interface, and network interaction interface .
- Concretely, network resource interface is defined to achieve configuration of the multiple physical resources, such as the wireless and wired infrastructures.
- The network service interface is used to provide service for applications by invoking network resources, in which network services can be measured by the capabilities, such as network delay, throughput, routing, and coverage areas.
- Network interaction interfaces are used to exchange information among different entities and heterogeneous networks.


Thank You!