A Survey on Cloud Sensor Integration

Chandrani Ray Chowdhury
Assistant Professor, Dept. of MCA., SDET-Brainware Group of Institution, Barasat, Kolkata, West Bengal, India

ABSTRACT: Now Wireless Sensor Network (WSN) has become an all pervading entity in human life. It reaches those areas where human cannot. The advancement and application of wireless sensor networks become an invincible trend into various industrial, environmental and commercial fields. However it has resource constraint and design constraint. Integrating cloud computing with WSN, we can overcome the weakness of WSN. This paper gives a brief survey of sensor cloud integration. Here mainly three areas of sensor cloud integration are considered namely Sensor Cloud Database, different cloud based sensor data sharing platform and cloud based sensor data processing approaches. A table is also given here to summarize the features supported by different cloud based sensor data sharing platforms.

KEYWORDS: WSN; Cloud Computing; Sensor Cloud; Cougar; TinyDB; CLEVER

I. INTRODUCTION

The advancement and application of wireless sensor networks become an invincible trend into various industrial, environmental and commercial fields. A typical sensor network consists of a number of sensor nodes acting together to monitor a region and fetch data about the surroundings. A WSN contains spatially distributed, self-regulated sensors that cooperatively monitor the environmental conditions like sound, temperature, pressure, motion, vibration, pollution, and so forth [1], [2]. Currently, WSNs are being utilized in several areas like healthcare, military target tracking and surveillance, natural disaster relief, hazardous environment exploration, seismic sensing and so forth. These sensors provide various useful data which can be used to monitor and control the environment in which they are deployed. The amount of data in a sensor network is huge, heterogeneous and multidimensional in nature. To store and process these data, high amount of storage and computation power is required. Unlike traditional networks, a WSN has its own design and resource constraints. Resource constraints include limited amount of energy, short communication range, low bandwidth, limited processing and storage in each sensor node. Design constraints are application dependent and are based on the monitored environment [6].

Cloud computing has evolved as the future generation computing paradigm [3]. The NIST (National Institute of Standards and Technology, US) defines the concept of cloud computing as follows: Cloud computing is a model for enabling convenient, on demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storages and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [4].

Cloud computing allows the systems and users to use Platform as a Service (PaaS) (e.g. Operating Systems (OSs)), Infrastructure as a Service (IaaS) (e.g. storages and servers) and Software as a Service (SAAS) (e.g. application level programs) and so forth at a very low cost which are being provided by several cloud providers (e.g., Amazon, Google and Microsoft) on the basis of pay per use services [5]. As Cloud provides a huge computing power and storage, we can integrate it with sensor network to resolve the problem of limited computation power and limited storage capacity. Sensor database system is integrated with cloud to provide database for cloud sensor network. To eliminate the computation constraint of WSN, cloud computing can be integrated. The deployment of WSN is provider specific i.e. providers define their own standard. Cloud is also integrated with WSN to provide a common data sharing platform. This paper survey the integration of cloud with WSN to provide database for cloud sensor, cloud based WSN data processing framework and cloud based data sharing platform for large scale sensor networks.

This paper organized as follows. Section I gives introduction. Section II, III and IV gives the survey of three approaches. Finally section V concludes the paper.
II. SENSOR CLOUD DATABASE APPROACH

A sensor network produces huge amount of multidimensional streaming data which are inherently noisy and ephemeral in nature. The data in sensor network have spatio temporal dimension which needs additional processing unlike traditional database system [36]. In sensor networks we must have both live data (sensor reading) and stored data (sensor metadata) processing i.e. sensor database must support continuous query processing and one-time query processing techniques [34]. A sensor network is spatially distributed and suffers from short range communication, small storage and limited computing power. So to manage such data we need lightweight distributed processing contrast to traditional network heavy-weight processing. As the data in sensor network is uncertain in nature [36], we must have pre-processing to manage missing value and calibration error. Sensor networks generate huge amount of data which needs data aggregation. For query processing, approximated query processing technique is used to manage the data volume [37]. We can use probabilistic approach in sensor networks for approximate query processing. Some sensor network uses RFID tag for identification purpose which needs special power [33]. Data in the sensor network is considered as continuous stream which leads to data structuring for processing. Generally sensor network data are represented as XML. Sensor modelling language (SML) [7] is used to represent any physical sensor’s metadata like their type, accuracy, physical location and so forth. The nodes in sensor network are highly failure prone, so we must have fault tolerant and data recovery mechanism to protect data. As sensor network is used to monitor and control the different environmental situations and to avoid different disaster, time critical real time database will be suitable for this network. Further security and reliability is important for this type of distributed network. Finally data in sensor network are heterogeneous in terms of data requirement, application needs, types of deployment and types of data generated by sensor network [24]. So to judge the sensor database the parameters are:

- Large set of multidimensional distributed heterogeneous streaming data processing.
- Structured data and unstructured data processing.
- Spatio-temporal data management.
- Continuous and one-time query processing.
- Single query and multi query optimization technique.
- Data pre-processing support to manage noise and ephemeral nature of data.
- Light-weight distributed processing considering resource constraint and energy efficiency.
- Inbuilt fault tolerant and recovery mechanism.
- Probabilistic approach support considering approximated query processing of sensor data.
- Security and reliability.

To save energy and reduction of unnecessary, irrelevant data transmission in-network processing is required. There are different types of databases are available on cloud which can be used for sensor network. This database may be key-value store or relational or combination of both. Cougar, TinyDB and Apache Hadoop cloud based database systems are used for sensor network. There are three types of cloud storage services namely Cloud Database Service, Cloud Blob Storage, Cloud Table Storage. Among the three clouds storage services, Cloud Table Storage and Cloud Database Service are more appropriate to store sensor data. Blob storage does not provide structured storage and query capabilities required for working with sensor data [13]. Different cloud based data management system frameworks are used to manage sensor data. RDB-KV CloudDB is a framework for managing massive heterogeneous sensor data stream. This framework combines the approach of relational database and key-value store. It supports global data distribution. For keyword based queries, a Global ObjectID B+-Tree (GOID B+-Tree) index is used. It focuses on spatio-temporal query processing and specify operator to process streaming data [9]. But this framework does not support data pre-processing, nested query processing, aggregated query processing and acquisitional query processing. Google Bigtable and apache Map-Reduce parallel processing model are combined to use as a sensor database. But Bigtable is not distributed outside Google and are only accessed within Google App Engine [10]. Both Bigtable and Map-Reduce are combined to support streaming multidimensional processing of huge amount of sensor data [11], [12]. Apache Hadoop HDFS, HBase and Map-Reduce are combined to provide a sensor data storage and parallel processing framework in cloud. This system can store large datasets suitable for sensor network using block oriented architecture and ensure high throughput of data. Feature replication strategy is used for fault tolerance of database [13]. But Hadoop supports batch processing style while sensor network supports real time interactive processing. In Managing Wearable
Sensor Data through Cloud Computing [14] authors propose an approach for data processing of sensors attached to patients. This approach uses java to develop an application that receives sensor data from patient body and stores them in Google public cloud. It also provides a visualization web application and APIs (Google charts). WSDL Web Services are used to access remote services and it helps different application to communicate with each other. This paper performs fault analysis, but does not consider QOS, data pre-processing and energy efficient data processing. It uses 3rd party Google public CLOUD SERVICE. Pachube [15] is a web service which provides a scalable infrastructure to manage real time sensor data. CLEVER [16] is a sensor data processing service which analyses and filters data from sensor networks and provides them as an input to other advanced services. It provides support for sensor provisioning. To provide cloud services it uses C-COMPUTING module for cloud data centre resource management, C-SENSOR module to add PAAS functionality for sensor network and C-STORAGE module to store necessary middleware data. CLEVER uses Sedna native XML database as distributed database and XMPP protocol to support interoperability. It uses distributed hierarchical clustering and provides special support of security management. In [17], authors propose a cloud based framework to collect ECG data from the sensors attached to the body of patients, store them and provides as input to different softwares which analyse and process this data. They use Amazon EC2 and Amazon S3 as IaaS. For PAAS they implement a new module called Dynamic Scalable Runtime with Aneka middleware to ensure QOS and provide different software to analyse, process and monitor patient health status as SAAS. This paper stresses the cost, scalability, dynamic load distribution and QOS of system. ThingSpeak is an open source platform available online to store and retrieve web based data. Users can store and create sensor-logging application in this platform. This platform is also used for sensor monitoring [18].

III. CLOUD BASED SENSOR DATA SHARING PLATFORM

When we design a sensor data sharing platform, we must ensure the delivery of information from WSN to middleware in time and as soon as possible. Sensor data sharing platform is required to access sensor data beyond the boundary of proprietary concern which implements this sensor network, sharing of data among several applications, integration of different sensor networks and to provide WSN data as information as a Service (IAS) in cloud.

A. Existing sensor data sharing platform:

In [19] authors propose a technique to integrate industrial sensor network with cloud to monitor and control the development process (e.g. Nuclear Plant Management System). This approach supports monitoring process parameters, threshold value violation checking, coordination of information among manufacturing sites, maintaining of the process quality, sensed parameters as a service, service searching and binding. But this approach does not consider distributed processing, user isolation, scaling of network, load balancing and faulty sensor analysis. Moreover a single EB-XML registry repository and single integration controller will be subjected to single point of failure. There is a system [20] which provides the smart device user, a cloud based wireless sensor network infrastructure which is used to access sensor data and hosting of computation tasks. This framework is designed to provide sensor information and applications as service. Authors use virtualization (virtual machine) to manage heterogeneity and hide underlying complexity. It uses SENSORML language for sensor data processing. It also provides support for automatic provisioning of sensors. For hosting services Google Public cloud (Google App Engine) is used. In [21] a publisher-subscriber based framework is proposed to integrate sensor network with cloud based community to provide data and applications as a SAAS. It provides the following services

- Sensor data delivery using content based pub-sub system.
- An event matching algorithm called Statistical Group Index Matching (SGIM) to match published events to appropriate consumers.
- Collaboration with other cloud provider by VO based dynamic collaboration to make the resource available.

It mainly considers data delivery. In addition to this it also provides multiple subscriber support, group key management and dynamic addition of publisher subscriber. GSN [22] is a middleware for dynamic integration and management of sensor network and produced data stream. It is mainly peer-to peer architecture and supports QOS aspects of system. VFSN: Virtual Federated Sensor Network [23] is used to share sensor data among multiple users and hiding actual resources. The main service is sensor data sharing. This framework is a peer to peer architecture using
PIAX where sink nodes are served as peer. As a whole this is an overlay network. It uses server virtualization and network virtualization to provide function expandability and resource security. In network processing of sensor data is used to reduce data transfer. This paper considers load distribution strategy. IEEE 1888 is used as data transmission protocol. It also specifies its operation. SenseWeb [24] is designed to provide common platform to share sensor data and build different application to manipulate data. But to control and manage all the activities it uses a centralized coordinator system which leads to single point of failure. In [25] authors propose a platform to develop a platform independent, adaptive, service centred application development for sensor networks. This platform helps end-users to develop their applications. Applications can be migrated. To support application migrability authors uses mobile script. Service binding is done by execution engine where authors implements most of the things as a service and include less functionality. So this paper mainly gives focus on application migration and addictiveness. It also considers multiple user support, adding new service and service provisioning. Though by run time binding we can achieve portability but execution time also increases which is not desirable in sensor network. Again, some execution are moved to sensor nodes which may increases power consumption. If multiple user access a single component then a strongly concurrency control system is required which is not considered in this paper.

IV. CLOUD BASED SENSOR DATA PROCESSING APPROACHES

Sensor networks have resource constraint. We cannot use heavyweight data processing approach inside a sensor network. But huge, heterogeneous streaming sensor data demands high level sophisticated processing. Integrating sensor networks with cloud can solve challenges involved in sensor networks data processing. A cloud based sensor data processing system can process data in two ways- First sensor network sense data and forward it to the cloud for processing [26], [31] and secondly cloud data processing system first process query, subdivide the query and forwards into sensor network. In response of query sensor nodes sends only required data [32].

A. Existing Cloud Based Platform for Sensor Data Processing:
Different authors use different cloud framework for data processing in sensor network. CLOUDVIEW is a framework for storage, processing and analysis of massive machine maintenance data collected from large number of sensors embedded in industrial machines in cloud computing environment [26]. This hybrid (offline and online) system uses distributed Hadoop and Map-Reduce for storage and parallel processing of streaming time series sensor data. System includes data collection (data aggregation, pre-processing, filtering, storing), case base creation, updating and maintenance, feature reduction, feature extraction and fault prediction using CBR (Case Base Reasoning) and rough set theory. But the problem is Hadoop supports batch processing style. So problem may arise for interactive real time processing. In [27] authors concentrate on maintaining the isolation and consistency property of sensor transaction. Some authors use Apache Hadoop and Map-Reduce to process semi structured and unstructured distributed large scale sensor data [28]. In [29] the authors propose a multilevel and flexible data processing model for sensor network. In this system, coordinator and gateway are used in each sensor network to collect and store sensor data. They provide local storage facility. Different sensor networks are connected to a local server which provides local storage backup and computing functionality. Local server provides uniform data access interface whose specification are stored at cloud and used by clients. This system uses different databases (SQLite, TinyDB, MySQL and MongoDB). In [30] authors integrate ERP and sensor network to minimize resource usage of sensor networks. They use processing and prediction model of cloud based ERP system to process data collected from sensor network, generate some decision and forwards this decision/control information back to sensor network to turn the sensors on/off and adjust their transmission frequency to reconfigure the sensor network. But the problem is sensor networks transmit all the data which increases power and energy consumption.

B. Sensor Network Query Processing Framework Consideration:
To design a cloud framework for sensor data processing we have to consider the following points

- The system has support for stored or historical sensor data processing as well as real time series streaming data processing. In this case we can consider relational database or Google Big-table approach for stored structured data and Apache/Google Hadoop for unstructured/semi structured streaming data processing.
- System must have support for one time, continuous, periodic query processing. As WSN is changed dynamically query optimization plan must be adaptive instead of static as in traditional database.
Support system defined and user defined grouped aggregation which reduces data transmission.
Explicity support of data pre-processing (especially uncertainty or discrepancy management).
Approximate query processing with certain level of error tolerance.
System must support acquitional query processing, on demand query processing and query on collected stored data.
System can support WSN resource management using processed data.
Explicit fault management, data consistency management and loads balancing.
Framework also includes power consumption monitoring and network legibility issues.

C. Comparison of different Cloud Processing Framework in Tabular Format:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Heterogeneity and abstraction</td>
<td>SOA based uniform interface</td>
<td>Virtual Machine</td>
<td>Virtualization</td>
<td>Well supported uniform interface</td>
<td>Yes (mobile script)</td>
<td>No standardized interface</td>
</tr>
<tr>
<td>Authentication</td>
<td>Check data forwarding request</td>
<td>NO</td>
<td>NO</td>
<td>Yes</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>Scalability</td>
<td>NO</td>
<td>Yes (virtualization)</td>
<td>Yes (virtualization)</td>
<td>both</td>
<td>Yes(main feature)</td>
<td>Yes(dynamic addition of subscriber/publisher)</td>
</tr>
<tr>
<td>Explicit load balancing</td>
<td>NO</td>
<td>NO</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>NO</td>
</tr>
<tr>
<td>QOS</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>partially</td>
<td>NO</td>
</tr>
<tr>
<td>Sensor mobility consideration</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>Mobile Proxy</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>Services</td>
<td>Data sharing</td>
<td>Raw Data acquisition Data Reformattin g Data aggregation</td>
<td>Data sharing</td>
<td>Sensor data and sensing resource sharing</td>
<td>Mallable, migratable application development</td>
<td>Sensor data sharing</td>
</tr>
</tbody>
</table>

TABLE I. COMPARISON OF SENSOR DATA SHARING APPROACHES
Integration of cloud computing with WSN can solve the resource constraint problems of WSN. In this paper we concentrate on three areas namely cloud sensor database, cloud based sensor network data sharing platform and cloud based sensor data processing approach. We have also give a comparison table of different cloud based sensor network data processing approaches and finally gives points to design cloud based sensor network data processing approaches. In future we will consider other areas where WSN and cloud computing can be integrated.

REFERENCES


BIOGRAPHY

Chandrani Ray Chowdhury is an Assistant Professor in the MCA Department of SDET-Brainware Group of Institution, Barasat, West Bengal. She received her M.Tech in Computer Science & Engineering degree in 2010 from KIIT University, Orissa, India. Her research interests are WSN, Cloud Computing, Information retrieval etc.