# Annihiliation of Redudant Data in Amalgamation of WSN's using Data Cleansing

Saleem Malik S Kanmani Vishwas S Yashashwini K.L M.H Sooraj Prasad
Department of Computer Science and Engineering
KVG College of Engineering, Sullia, 574327, India

## ABSTRACT
Nowadays manufacturing environments are very charismatic and inclement. Traditional management information systems (MISs) have mostly been implemented upon hierarchical architectures, which are inflexible to adapt changes and uncertainties promptly. Next-generation MISs must be agile and adaptable to accommodate changes without significant time delays. It is essential for an MIS to obtain real-time data from the distributed and vehement manufacturing environment for decision making. Wireless sensor networks (WSNs) and radio-frequency identification (RFID) systems provide an excellent infrastructure for data acquisition, distribution, and processing. In this paper, some key challenges related to the amalgamation of WSN and RFID technologies are discussed. A five-layer system architecture has been proposed to achieve synergistic performance. For the amalgamation of WSN and RFID, one of the critical issues is the low efficiency of communication due to redundant data as redundant data increases energy consumption and causes time delay. To address it, an improved data cleansing algorithm has been proposed; its feasibility and effectiveness have been verified via simulation and a comparison with a published algorithm. To illustrate the capacity of the developed architecture and new data cleaning algorithm, their application in relief supplies storage management has been discussed.

## Keywords
Management Information System, Data Cleansing, Wireless Sensor Networks(WSN).

## 1. INTRODUCTION
Todays manufacturing industries play a crucial role to play in economic development of any country, as it stimulates numerous economic sectors [1]. Manufacturing systems are very complex networks. consisting of numerous objects, decision-making units, materials, and information flows. Therefore, management information systems (MISs) are needed as technology platforms that enable the managements to integrate and coordinate their business processes at both intra-organizational and inter-organizational levels. The technological advances in MIS have provided a viable solution to the growing needs of information amalgamation in both manufacturing and service industries [2]. Traditionally, MISs have mostly been developed as centralized systems to ensure that information can be shared across all functional units and management hierarchies. However, today's manufacturing environment has been characterized by fierce global competition, narrow marketing window, and fluctuated and fractioned supplies and demands [3]–[5]. Next-generation MISs must support the global competitiveness, innovation, the introduction of new products, and strong market responsiveness. As a result, besides the cost and quality, manufacturing systems need tobecome more strongly time-driven and time-oriented. This evolution requires high flexibility to adapt changes [6]. Changes and uncertainties involved in a manufacturing environment should be accompanied by the changes of MIS architecture. The needs of developing next-generation MISs have been thoroughly discussed [7]–[9]; some critical requirements for MISs including *modularity*, *sustainability*, *adaptability*, *agility*, *autonomy*, and *scalability*. Performance of the MIS greatly depends on the availability of reliable and abundant data associated with dynamic changes. In this sense, the development of an integrated information system for the real-time data acquisition is crucial to the success of an MIS.

Decentralized and distributed MISs will allow a more flexible manufacturing system [10]. Wireless sensor networks (WSNs) and radio-frequency identification (RFID) systems provide excellent infrastructures to acquire, distribute, and process data in decentralized dynamic environments. Rapid advances in industrial information amalgamation methods have spurred tremendous growth in the use of WSNs and RFIDs for MISs [2]. With the applications of WSNs, event processing can fit well in EISs to improve the responsiveness [11]. Although WSN and RFID technologies have experienced great achievements in recent years, their applications in actual manufacturing environments are very limited. Bonivento *et al.* [12] stated that the lack of system-level design methodologies was one of the main obstacles to adopt WSNs or RFID systems in manufacturing industry.Vitturi *et al.* [13] drew a similar conclusion that the limited progress was made in the design of higher layer protocols for industrial applications of WSNs. To develop suitable WSNs for information amalgamation of MISs in a dynamic environment, this paper focuses on the adoption of RFID systems in the WSNs and the development of new data cleaning algorithm to eliminate redundancy data effectively.

## 2. RELATED WORK
### 2.1 Remonstrance in WSNs
RFID and WSNs represent two complementary technologies. RFID is widely used to identify, detect, or monitor objects. In comparison with other types of sensors, the low cost is the superior advantage of RFID; however, RFID is incapable of providing the detailed information about the conditions of objects. On other hand, a WSN can integrate logics into RFID nodes and allows an RFID system to operate in a multi-hop fashion and with the detailed information about the nodal conditions [9]. Lopez *et al.* [3] introduced a framework that operated on an integrated RFID and WSN network; similarly, Cho *et al.* [1] proposed to integrate RFID and WSN as an infrastructure for telecommunication. The tasks involved in the integration of WSNs and RFIDs are to design or select: 1) RFID tag memory; 2) WSN association protocols; 3) routing and addressing schemes; 4) RFID sensor-actuator data integration and management; 5) service definition and delivery; 6) context and service matching; and 7) distributed middleware [3]. However, the integration of RFIDs and WSNs is an emerging and immature technology; the identified

challenges in promoting such integration are *energy conservation*, *real-time performance*, *data cleansing and filtering*, *Localization*, *anti-collision*, and *authentication* [2]–[7]. Due to space constraints, only the first three challenges are discussed, and the focus is on data cleansing and filtering.

## 2.2 Data Cleansing Algorithms

Data cleaning is to eliminate redundant data mean while maintain the integrity of original data. Some progresses have been made in data cleaning or cleaning technologies. For example, Jeffrey [6] proposed an algorithm based on the pipeline framework. Different steps of cleaning are applied based on the characteristics of the raw data. This algorithm worked well for data leakage and repeated reading. In his sequential work [7], a data cleaning strategy based on the time correlation was proposed. This algorithm used a probability model and mainly developed to solve the problem of data leakage. Sarma [8] also introduced a pipeline algorithm to improve the quality of the data flow. All of the aforementioned algorithms were developed to address the problem of unreliability of RFID data caused by data leakage and repeated readings; the problem of data redundancy has not been tackled.

Carbunar [9] discussed the problem of redundant data; he suggested cleaning data by keeping inspection and silence of redundant readers. However, the proposed algorithm for detecting the device of redundant readers cannot avoid the fact that many readers have to work together at the same time. Based on the specified application, Khoussainova [5] developed a data cleaning method that relies on the application conditions and it needs restrictive rules. Barjesh [4] proposed a data cleaning, transformation and loading technique. It was implemented based on the probability theory. There is no further related research on data filtering of multiple readers. It has been found that all of the existing algorithms have their limitations in solving specific programs, but the solutions are not comprehensive and general. This paper will propose a comprehensive and efficient data cleaning algorithm using the correlations of time and space among readers and tags.

# 3. ARCHITECTURE FOR AMALGAMATION OF WSNs

## 3.1 Wireless Communication Protocols

The first task to integrate RFIDs with a WSN is to select a wireless communication standards and sensors. We choose one from two low-power wireless standards, i.e., Bluetooth and ZigBee. Bluetooth allows the creation and maintenance of a short-range personal area network (PAN). Bluetooth transfers data at the rate of 1 Mbps, the range of Bluetooth device is about 10 m. However, the main disadvantage of the Bluetooth is its relatively high energy consumption; usually, the Bluetooth cannot be used by sensors that are powered by a battery.

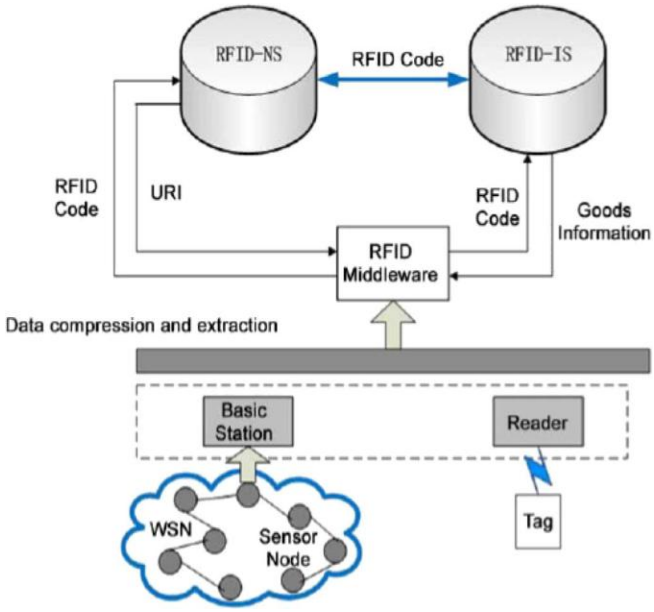


**Fig 1: Amalgamation of RFID and WSN**

On the other hand, a ZigBee-based WSN is a security network for a short distance communication based on the IEEE 802.15.4 standards. ZigBee has powerful networking capabilities. It is capable of supporting three types of self-organizing wireless networks: *star structure*, *network structure* and *cluster structure.* In addition, the ZigBee technology has the following characteristics: 1) it requires low power consumption; a sensor under a standby status can run more than six months; 2) it is cheap due to the opening standard; (3) it supports communication in a limited range from 10 to 100 m; 4) it provides a low rate and short latency; typically a response requires 15 30ms; 5) it has a high-capacity; each node can manage up to 254 child nodes; therefore, the ZigBee can have a maximized 64,516 nodes; 6) it possess high-security with the Advanced Encryption Standard AES-128 symmetric password and Access Control List (ACL). The ZigBee also has the anti-interference ability; moreover, the licensing of ZigBee network is free for industry. Due to the advantages of forming self-organized network systems and lower-cost, ZigBee technologies have been adopted for the WSN in this study.The integration of RFID and WSN is shown as Fig. 1. The integration of the ZigBee sensors and RFID technology make it possible to identify the targeted object globally and perceive the status of the object in a real time manner. The integration of WSNs and RFIDs can expand the scope of applications of the two technologies.

## 3.2 System Architecture

To take full advantages of WSNs and RFID technology, we propose a new robust system architecture which is shown in Fig. 2.. It contains five logical layers: *physical layer*, *data link layer*, *network layer*, *transport layer*, and *application layer*. The functions at each layer are explained below. 1) Physical layer: The functions in the physical layer include channel selection, signal monitoring, sending and receiving message. The design goal in this layer is to minimize the energy consumption and increase the link capacity. A working channel can be chosen from 16 channels with rate ranging from 250 KB/s at 2.4 GHz.2) Data link layer: This layer ensures that the physical data can be transmitted correctly; it also relates to the system spectrum efficiency and secure communication among system components. Based on the

IEEE802 standard, a data link layer consists of two sublayers: logical link control (LLC) and media access control (MAC). LLC is to ensure the security and reliability of transmission. MAC provides the service interface with the point-to-point communication for upper layer to access physical channels. To reduce power consumption and adopt the caching mechanism, MAC sub-layer in IEEE 802.15.4 is based on 802.11 wireless local area network (WLAN) standards for carrier-sense multiple access/collision avoidance (CSMA/CA) access.

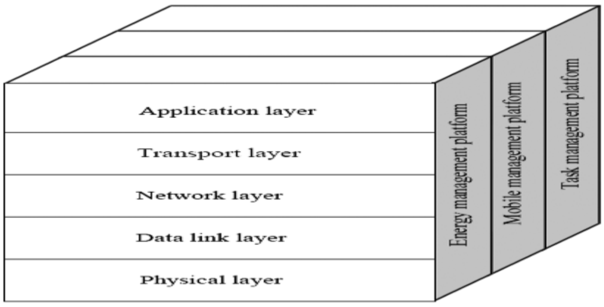


**Fig 2: System Architecture**

3) Network layer:. The functions at this layer include packet routing and congestion control. Packet routing answers which adjacent node to which the current node should send its packet so that the data can arrive in its eventual destination as quickly as possible. The selection of packet routing may be based on the shortest time, the minimized number of hops or the shortest distance. Congestion occurs when a node is carrying so much data that its quality of service deteriorates. Deteriorated quality may be caused by queuing delay, packet loss or the blocking of new connections. Congestion control is to avoid congestion collapse. 4) Transport layer: The functions of this layer are flow control, error control, and quality control for reliable transmission service. The network layer of ZigBee can be used as this layer. The routing protocols are based on *ad hoc* technology; it has similar characteristics of low energy consumption at the bottom layer of IEEE802.15.4 standard; meanwhile, it can help the organization and maintenance of the network. 5) Application layer.

The functions of this layer ensure safe data transmission and provide required services for the given application. An abstract interface is defined for the ZigBee network; the network platform also provides the application programming interface (API). APIs are used to define the rules such as how the application interfaces should be integrated into the ZigBee protocol stack. In our prototype system, the locations of readers are fixed. Therefore, the energy consumptions of sensor nodes will not be a concern since they can be continually loaded and data fusion and data management can be done in the readers. Readers transmit information that can be collected and initially fused to a middleware server, then to be delivered to an application server

## 3.3 Proposed Data Cleansing Algorithm

We assume there are m readers, m*n fixed reference tags whose attributes are known, and tags for arbitration. Each reader contains reference tags which can be detected and located in its working area of the reader. For the algorithm, the following concepts and definitions are defined.Steps involved in the proposed algorithm  are as follows as shown in fig 3.

1) Read the membership of configuration files that bind with reference tags and readers, create an initial CTL.

2) Update CTL in time from the RFID data stream for uninterrupted production.

3) Check if existing tags of waiting arbitration are nearby reference tags in CTL. If the answer is no, go to step 8; if there are some arbitration tags going to step 4.

4) Based on the sliding window, detect the tuple cache queue in tuple groups, and check if the tuple is redundant or not using the definition and inspection standard in definition 2. If it does not satisfy the condition, return to step 2; otherwise, go to step 5.

5) Compute the affiliation that tag responds to reader at the moment .

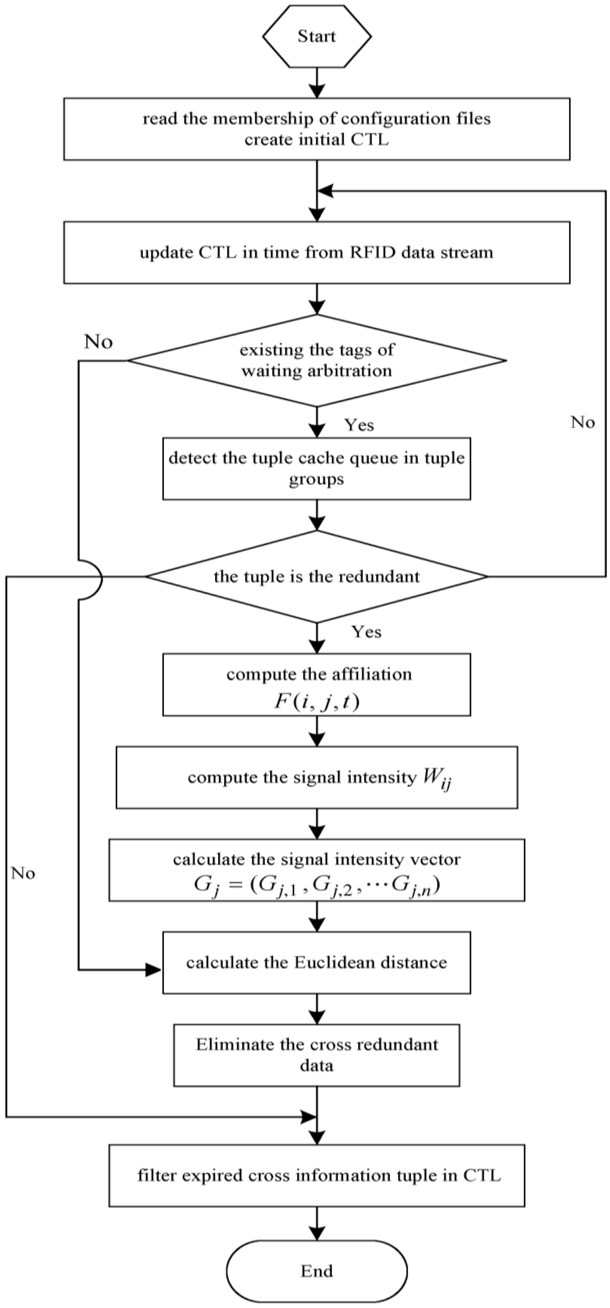


**Fig 3: Proposed data Cleansing Algorithm**

6) Compute the signal intensity to see that tag of waiting for arbitration is relative to the reader that already detected tag in sliding window .

7) Calculate the signal intensity vector; it denotes whether or not reference tags belongs to reader .

8) Calculate the Euclidean distance that tag of waiting for arbitration with respect to . The method of minimum relative position and to arbitrate the reader of cross redundant data are applied. Through an exclusive process, the cross redundant data can be eliminated.

9) Filter the expired cross information tuple in CTL to maintain a reasonable memory and then return to step .

The performance of our algorithm, especially compressibility and accuracy, has been taken into consideration and the results from the developed algorithm are compared with those obtained by a state-of-the-art data cleaning algorithm SMURF.

# 4. DESIGN AND IMPLEMENTATION

In this section, its application as a relief supply management system is discussed as a case study. One motivation to use a medical application as a case study is that RFID and WSN for medical applications have attracted numerous attentions. The recent progress in applying integrated information systems in healthcare industry has been reported in [4] and [5]. To rescue the affected population effectively, it is important to send relief supplies timely and manage the process wisely after some disasters or unexpected events occur [6]–[9]. is always difficult to monitor and manage the distributions of relief supplies on wide geographically distributed areas. With the rapid development of the WSN and RFID technologies, the implementation of intelligent logistic management and storage systems become possible; such systems will play a significant role in acquiring information regarding resources, scheduling. It collects real-time data, analyzes, assesses, and monitors the processes of distribution and transportation to ensure relief supplies are distributed appropriately in time and quality manner.

## 4.1 Design Framework

The proposed integrated system consists of four modules: relief supplies detection, relief supplies control, relief supplies warehousing, and relief supplies decision support. Relief supplies detection module is to detect all of the sensors and collect data including environmental data. The sources of data can be various sensors, attributes stored in RFID tag, and the locations of the relief supplies. All of the data is transmitted to the host data distribution center through the ZigBee-based WSN. Relief supplies controlmodule keeps the data from RFID tags and the environmental data received by sensors such as temperature, humidity, and pressure. This module also conducts comprehensive analysis and assessment. It is equipped with a software system for real-time monitoring and display. The software system can verify the attributes and control the distributing process. Relief supplies warehousing module is to manage the relief supplies to be put in specified storages. It monitors the transportation of relief supplies from suppliers to storage. If some mistakes are found during the monitoring process, this module suggests the solution for the correction. Decision support module is to analyze transmission data using the data mining technology. It focuses on the basic attributes of relief supplies; it is capable of rescheduling the distribution based on the latest needs of the various settlements. It verifies the locations of goods to determine whether or not the goods have been transported to destinations.

## 4.2. Implementation of Storage Management

The deployment of relief supplies management based on ZigBee and RFID technology is shown in Fig. 4. In order to manage the relief supplies more effectively, a number of readers and terminals must be set in the warehouse. The steps are given here.

1) For each object, an RFID tag or other type of tag is attached. Readers are installed at the entrance of the channel warehouse. It ensures the reader to obtain required information from tags when goods pass through the entrance.

2) In the warehouse, a certain number of RFID handheld/car terminals are set among shelves and exits. It is convenient to check inventory, track objects, retrieve accurate the information of relief supplies in the storage management system. It is very important that this network can ensure to acquire accurate data in the warehouse.

3) In the storage areas, based on the size of area, the distance signal transceiver module is set to form a backbone wireless network. Especially at the main drive roads, a number of ZigBee nodes are installed. When a vehicle with an integrated label passes by the coverage area, the monitoring and controlling can find the location of vehicle in real time.

# 5. CONCLUSION

The requirements of decentralization, high adaptability, and flexibility of next-generation EISs have been discussed. It has been observed that existing EISs are mostly developed based upon hierarchical architecture, which is not flexible to accommodate changes and uncertainties in today's turbulent and dynamic environments. EISs require sufficient and real-time data. It is critical to develop effective data acquisitionsystems for EISs. With the considerations of dynamic, distributed, and decentralized environments, RFIDs and WSNs become the better choices for many applications over traditional wired networks. In this paper, the literature on WSNs and RFIDs has been thoroughly reviewed. Most existing research is for specific applications which lack generality. In particular, there are no comprehensive algorithms to deal with the redundant data from multiple readers appropriately. In this paper, a five-layer system architecture is developed to integrate WSNs and the RFID systems. Bluetooth and ZigBee technologies are selected as the communication protocol of theWSNs to meet the requirements of large number of sensor nodes, wide areas, and low cost. To eliminate redundant data in the integrated WSN, an improved cross-redundant algorithm (ICRDC) is presented. Its effectiveness has been validated via simulation and comparison study: ICRDC can improve the data compressibility and accuracy more effectively in comparison to the SMURF. To illustrate the feasibility and effectiveness of integrated architecture and new algorithm, they have been proposed to be applied in relief supply management.

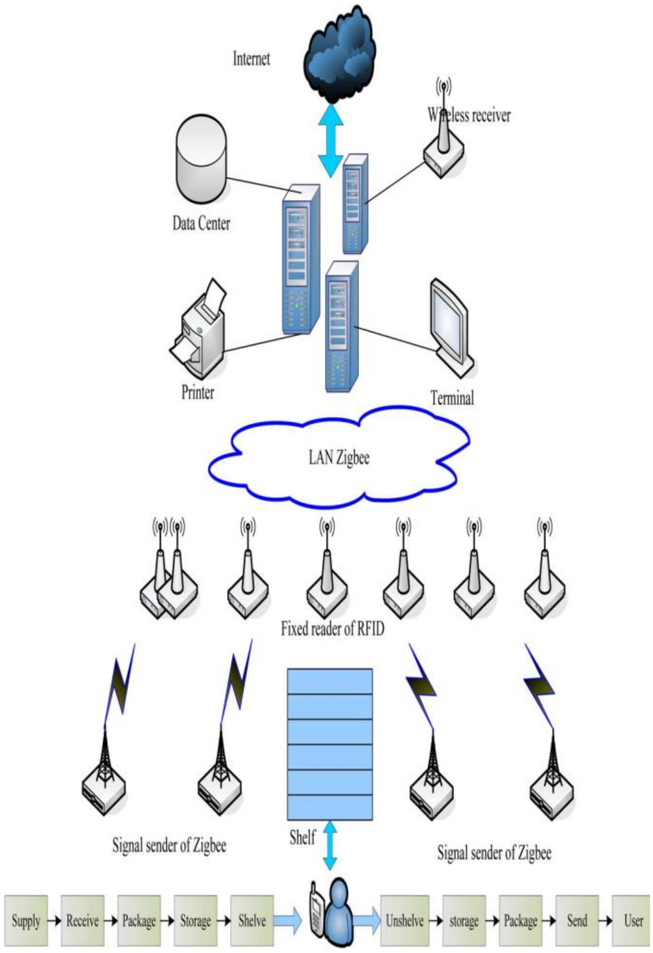


**Fig. 4. Deployment of relief supplies based on Zigbee and RFID**

## 6. REFERENCES

[1] Z. M. Bi, S. Y. T. Lang, W. M. Shen, and L.Wang, "Reconfigurable manufacturing systems: The state of the art," Int. J. Production Res., vol. 46, no. 4, pp. 967–992, 2011.

[2] L. D. Xu, "Enterprise systems: State-of-the-art and future trends," IEEE Trans. Ind. Inf., vol. 7, no. 4, pp. 630–640, Nov. 2012.

[3] A. Bratukhin and T. Sauter, "Functional analysis of manufacturing execution system distribution," IEEE Trans. Ind. Inf., vol. 7, no. 4, pp. 740–749, Nov. 2013.

[4] J. Zhu and N. Sun, "Research on integration of WSN and RFID technology or agricultural product inspection," Amer. J. Eng. Technol. Res,, vol. 11, no. 9, pp. 2299–2305, 2013.

[5] T.-H. Shin, S. Chin, S.-W. Yoon, and S.-W. Kwon, "A service-oriented integrated information framework for RFID/WSN-based intelligent construction supply chain management," Autom. Construction, vol. 20, pp. 706–715, 2013.

[6] P. Garcia-Ansola, A. Garcia, and J. de lasMorenas, "ZigID: Improving visibility in industrial environments by combiningWSN and RFID," J. Zhejiang Univ. Sci. A Appl. Phys. Eng., vol. 12, no. 11, pp. 849–859, 2013.

[7] S. Lee, B. Kim, H. Kim, R. Ha, and H. Cha, "Inertial sensor-based indoor pedestrian localization with minimum 802.15.4a configuration," IEEE Trans. Ind. Inf., vol. 7, no. 3, pp. 455–465, Aug. 2014.

[8] Y. Yin, Y. Fan, and L. Xu, "EMG and EPP integrated human-machine interface between the paralyzed and rehabilitation exoskeleton, IEEE Trans. Inf. Technol. Biomed., vol. 16, no. 4, pp. 542–549, Oct. 2014.

[9] K. Xie, J. Liu, G. Chen, P. Wang, and S. Chaudhry, "Group decisionmaking in an unconventional emergency situation using agile Delphi approach," Inf. Technol. Management, vol. 13, no. 4, pp. 351–361, 2014.

[10] L. Xu, C. Wang, Z. Bi, and J. Yu, "Object-oriented templates for automated assembly planning of complex products," IEEE Trans. Autom. Sci. Eng., vol. PP, no. 99, pp. 1–12, 2013.

[11] L. Li, "Effects of enterprise technology on supply chain collaboration: Analysis of China-linked supply chain," Enterprise Inf. Syst., vol. 6, no. 1, pp. 55–77, 2014.

[12] X. Shi, L. Li, L. Yang, Z. Li, and J. Choi, "Information flow in reverse logistics: An industrial information integration study," Inf. Technol. Management, vol. 13, no. 4, pp. 217–232, 2014.

[13] L. Zeng, L. Li, and L. Duan, "Business intelligence in enterprise computing environment," Inf. Technol. Management, vol. 13, no. 4, pp 297–310, 2014.