ACCURACY OF DISTRIBUTED SYSTEMS TOWARDS INDUSTRY 4.0: SMART GRIDS AND URBAN DRAINAGE SYSTEMS CASE STUDIES

*Poonpakdee, P. and Koiwanit, J.

Department of Industrial Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

*Corresponding Author, Received: 16 June 2017, Revised: 17 Oct. 2017, Accepted: 1 Dec. 2017

ABSTRACT: With significantly improving the overall manufacturing operations commonly known in smart factory practices, vertical and horizontal integrations of various components are introduced throughout the entire value chain. The proliferation of smart factory practices introduces the fourth industrial revolution called Industry 4.0. Industry 4.0 creates new challenges and the application of networked manufacturing systems is one of the important features required to handle the systems in an efficient way by communicating and processing information to each other. Distributed systems provide coordination to allow the global information to be available for a better decision making and consequently achieving high efficiency. This connectivity helps for a better decision making and therefore achieve high efficiency. However, the evaluation on the accuracy of global information between different architectures of distributed systems (centralized systems and decentralized systems) has little work mentioning in the state of current manufacturing systems under the conception of Industry 4.0. As a result, this research will fill the gap by providing an accuracy of the decentralized system over the centralized system together with their sensitivity analysis. In this study, the decentralized system is built based on the concept of Epidemic protocols or Gossip-based protocols, while the centralized system is a simple client-server. The epidemic protocol is a bio-inspired communication paradigm that imitates the behavior of virus when the outbreak occurs in a community. The accuracy in both systems has been monitored by means of simulations. The effects of message loss to the accuracy of centralized and decentralized systems are studied. By comparing the system accuracy between both systems, it was found that the accuracy of the decentralized system is generally more accurate when the system is used for a long period. The accuracy tends to be lower down especially when the information is not completely distributed, while the accuracy of the centralized system receives an excessive suffer from message loss.

Keywords: Industry 4.0, Centralized systems, Decentralized system, Application accuracy

1. INTRODUCTION

The proliferation of smart factory practices introduces the fourth industrial revolution called Industry 4.0. Industry 4.0 creates new challenges and the application of networked manufacturing systems is one of the important features required to handle the systems in an efficient way by communicating and processing information to each other. Distributed systems provide coordination to allow the global information to be available for a better decision making and therefore achieve high efficiency.

The basic architecture of distributed systems, which is a part of smart factory systems, can be categorized into two major groups: centralized systems and decentralized systems. The most familiar form of topology is centralized systems, whose structure is typically similar to the clientserver. A server aims to centralize all function and information taken from clients by directly connecting to the clients. Clients share their resources by sending and receiving the information to a server [1]. In contrast, decentralized systems are similar to peer-to-peer systems where all peers communicate symmetrically and equally in roles. A well-known example of decentralized systems is Gnutella, the practical decentralized systems with only a centralized function for starting a new host [1]. In this study, Epidemic protocols or Gossipbased protocols, which bio-inspired is communication paradigm for information dissemination in extreme scale network systems, are used to construct the decentralized system.

In general, the convergence speed of centralized systems is faster than decentralized systems because the complete set of data is stored and managed within a hub that does not exist in decentralized systems. However, many circumstances such as message loss and node failure in centralized systems can cause data missing which can skew the results of the global aggregation. This introduces an inaccuracy in the decision making [2]. Unlike centralized systems, decentralized systems allow data from each node to be distributed to its neighbors [3]. This results in the networked infection over the whole system as the time goes by. Several works related to the accuracy of the global information in distributed systems have been reviewed. Reference [4] studied on the accuracy of failure detection in distributed system Reference [2] explored the accuracy analysis of data aggregation for distributed network monitoring. Reference [5] analyzed the convergence in decentralization in a specific problem. Reference [3] studied the accuracy of the epidemic spreading over a network. However, the literature review reveals that little work has been done under the concept of Industry 4.0 particularly focusing on the accuracy of global aggregation between both systems. As such, this paper presents how accuracy decentralized systems are evaluated against centralized systems in the context of manufacturing.

The rest of the paper is organized as follows. Section 2 gives the detail of Epidemic aggregation protocols and some of the related works. Application case studies are presented in Sections 3. The experimental scenario and their results are presented and analyzed in Section 4 and 5, respectively. Finally, Section 6 draws some conclusions and provides possible directions for future work.

2. EPIDEMIC AGGREGATION PROTOCOLS

An epidemic protocol or a gossip-based protocol aims to solve problems in an extreme-scale distributed system. The definition of an epidemic protocol is generally a bio-inspired communication and computation, which takes over the concepts of a virus-spreading mechanism and the gossip manners of a human. The characteristics of the protocol are robustness and scalability with respect to a global communication paradigm based on a deterministic interconnection network [6]. Epidemic protocols use randomized communication as the principal mechanism in order to achieve these characteristics.

A number of applications based on epidemic protocols have been proposed to serve different purposes in different environments. Examples of these applications are: Peer-to-Peer (P2P) overlay networks [7], distributed computing [8], mobile ad hoc networks (MANET) [9]-[11], failure detection [12,13], exascale high performance computing [14-16], data mining [17]-[18], data aggregation [19-22], and wireless sensor networks (WSN) [23-25].

Several epidemic protocols have been proposed to solve data aggregation problems because highly distributed systems such as P2P or large-scale network systems have currently become more popular and enabled for a broad range of applications [26,22]. Due to these systems, the information of data aggregation over the network is often more important than the information of an individual node [19], so tools that can compute significant system properties of a network are necessary [27]. A monitoring system is an example where the system administrator would be more interested in the aggregation information about the average of measured values from all sensors in an area than the individual measured value of a single sensor. More details about the utilizations of epidemic aggregation protocols are given in the next section.

3. APPLICATION CASE STUDIES

In this section, the detail of real case studies based on epidemic protocols is given due to the algorithm can be emerged in many fields. Examples of case studies are related to smart grids and urban drainage system. the detail of each example is shown below.

Firstly, the goal of Smart Grids (SG) is to change the structure of elect power system by enhancing the integration of renewables and promoting the generalized participation of different entities. The concept of SG allows the modern electric industry to ensure that high level of adaptability, scalability, security, economy, self-healing, robustness and protection in extremely dynamic systems by adopting the concept of Information and Communication Technologies (ICT) [28].

The objective of ICT in SG is the underlying service that supports the infrastructure in order to allow the information exchange between different entities.

In the context of communication paradigm, the epidemic protocol has been proposed as a communication framework to solve the problems of scalability and reliability in microgrids including information distribution and data aggregation in Automated Metering Infrastructure (AMI).

There are two scenarios illustrated the usage of proposed framework: propagation of information and retrieval of distributed metrics. The first scenario is for distribution of information, such as energy overproduction, to every entity which aims to stabilize energy consumption in the networks. As a result, the energy cost can be significantly reduced. On the second scenario, the information related to the energy requirements of intelligent electric devices is distributed across the network for a future power production plan.

Secondly, a case study of the monitoring system using epidemic aggregation protocol related to urban environmental aspects is studied to illustrate their use of the distributed systems in manufacturing. The interesting aspect is the urban flooding that occurs when an urban drainage system (UDS) becomes overloaded during an extreme rainy condition. As a result, the wastewater treatment plant (WWTP) is not able to treat the wastewater. This problem causes sewer flooding and combined sewer overflow in an urban environment, which is a potential risk to human life, economic asset, and the environment. In general, the UDS is managed by a real-time control (RTC) system that requires the network to be embedded with sensors and actuators permitting the network to be monitored real-time and regulated to adapt to the different rainfall events.

Many previous works of RTC are based on centralized systems that introduce several problems due to a large amount of data to be read, managed, and processed. The drawbacks of centralized systems include: 1) a complex mathematical model of the network is required, 2) the central unit needs to be connected with all physical parts, and 3) a single failure in any part of the system, especially the central unit, can compromise the entire system.

In order to solve these drawbacks, a decentralized system is considered Reference [28] developed an urban drainage network of the city of Cosenza in Italy, which adopts the decentralized real-time control (DRTC) system based on a distributed agent-based architecture and specifically an epidemic protocol, which exhibits good performance, and faults tolerance properties. The characteristics of the systems are shown as follows:

- **Autonomy**: Each agent is self-aware and has a self-behavior. It observes the environment, cooperates with others, and plans its execution autonomously.
- **Local views**: None of the agents have a full global view of the whole environment. However, its performance is solely based on the quality of local information.
- **Decentralization**: There are no master agents controlling the others, but the system is built from the interaction of peer agents.

The results are generated by using the Strom Water Management Model (SWMM), an opensource computer model for the simulation of hydrodynamic water and pollutant transport in sewer systems. The process to balance the water level throughout the conduits is simulated as the main contribution to their work. The water balancing process consists two tasks: 1) Computing the average water level in the generated network accomplished by epidemic protocol and, 2) Triggering the gate in order to bring the water level close to its average value.

Reference [29] confirms that epidemic protocol is able to support UDS to control water level successfully, ensuring a full utilization of the actual storage capacity of the system.

4. EXPERIMENTAL SCENARIO

The experiments in this study have adopted the scenario from the case studies mentioned in the previous section. The simulated scenarios are shown in the context of Industry 4.0, specifically in manufacturing systems under smart factory practices, where all the work units are connected. Machines, sensors, mobile devices, computers, laptops, and servers are some of the examples of the work units in smart factory practices. The simulation presents the activity in the manufacturing systems in terms of monitoring the system status in order to perform system maintenance. The goal of the experiments is to evaluate the accuracy of a global aggregation computation in different network environments: decentralized and centralized systems by mean of simulation. Fig. 1 presents two different architectures of simulated scenarios, which include centralized systems and decentralized systems.

In this study, the systems consist of two classes of work units: a base station and sensors. The base station's objective is to monitor the status of the system based on the data from the sensors. In centralized systems, a base station will compute the global aggregation by processing all of the local data sent by every sensor in the system. In general, a base station is a hub of data in centralized systems whereas decentralized systems do not have any hubs. Each of the sensors is distributing its local data to the system while processing the global aggregation locally. A base station aims to monitor and provide the global aggregation by obtaining the information from gateways, which are sensors that communicate with the base station.



b) Decentralized systems

Fig. 1 Architectures of simulated scenarios.

In comparison with the case studies, each sensor represents as an agent that maintains the measured value as local data. The differences between the case studies and the simulation are that the global aggregation in the case studies is the average of the measured values, whereas the global aggregation in the simulation is represented in terms of a maximum value.

5. EXPERIMENTAL RESULTS

The experimental results and their analysis are presented in this section. The simulations have been completed in PeerSim [30], a discrete-event based Peer-to-Peer simulator developed in Java. The simulations are based on an asynchronous network model with a network latency. Each sensor contains the local data initialized with a uniform distribution between 1 and 100,000, and the base station will monitor the system by computing the maximum value among the local data. Message loss is also introduced and included as one of the main parameters in the simulations. The value of the parameter indicates the possibility of message loss in percent.

For decentralized systems, the global aggregation is executed by an epidemic aggregation protocol. The protocol is based on push-only gossip-based communication modified from the SI model, one of the epidemic dissemination algorithms [31]. Algorithm 1 presents the mechanism of epidemic aggregation protocol where v is local data and j is a remote sensor. Every sensor uses the protocol to estimate and distribute the global aggregation. When the base station requires the global aggregation, it will send a request message to a gateway of the system, and then the gateway will reply its current estimated maximum value to the base station. In this work, the base station will make a request to a single fixed gateway.

For the centralized systems, the base station will collect the local data from every sensor to find the maximum value.

\mathbf{A}	gorithm 1 Epidemic Aggregation Protocol
1:	procedure SendPushMessage
2:	$j \leftarrow \text{get a random node}$
3:	send a <i>push</i> message to $j: m_{\rightarrow}(v)$
4:	
5:	procedure ReceivePushMessage (message m_{\rightarrow})
6:	if $m_{\rightarrow} v > v$ then
7:	$v \leftarrow m_{\rightarrow} . v$

The general configuration of a simulation is based on the evaluation criterion and the objectives of this study. The total number of work units using in this study is 1,000, which includes a single base station and 999 sensors. The initial topology depends on the models of simulated scenarios. For centralized systems, the initial topology is similar to a client-server architecture: a server (base station) contains a completed list of work units while clients (sensors) contain only a connection to the server. In contrast, the initial topology for decentralized systems is similar to a random graph that all sensors are randomly connected while a base station can only communicate with a gateway.

decentralized systems, In an epidemic membership protocol is required in this scenario. Epidemic membership protocols offer mechanisms intended to maintain the connectivity of the systems in terms of overlay topology [32]. Each sensor holds the local cache containing the connections to 20 random sensors. As a result, each sensor will recognize approximately 2% of other sensors. In each cycle, one-fourth of the connections in the local cache will be exchanged with other sensors due to the mechanism of an epidemic membership protocol. As a result, the structure of the system is continually rewired in every cycle. On the other hand, centralized systems are not required rewiring the structure, thus epidemic membership protocols are not taken into account.

The results can be categorized into two sets and presented in the following sections. First, the convergence speed of epidemic application in decentralized systems is examined, specifically the intention of the results showing how message loss affects the convergence speed. Second, the performance of base station is evaluated in both systems. The performance index is presented as the accuracy of the global aggregation.

5.1 Convergence Speed

This section shows how fast sensors can distribute the information across the system when different percentages of message loss occur. The results were collected from five trials with five different random seeds to avoid any experimental biases. The standard deviation of local estimated maximum value at each sensor is used to determine the convergence speed.



Fig. 2 Convergence speed with the different possibility of message loss.

The effect of message loss to the convergence speed can be observed from the results. Fig. 2 shows the average standard deviation of the local estimated maximum over five trials. The y-axis is the standard deviation and the x-axis is the number of cycles. A curve in the chart indicates a convergence to the target value.

At the beginning of the simulation, it is expected that the standard deviation is high because there are no sensors that are able to determine the maximum value. After the first cycle, the standard deviation is decreasing which shows that each sensor is starting to generate the maximum value. Different possibilities of message loss including 0%, 10%, 20%, and 30%, were simulated for the evaluation. Without message losses in the network, the Epidemic protocol provides the fastest convergence speed and it will be slower with a higher possibility of message loss.

5.2 System Accuracy

This set of simulations aims to evaluate the accuracy of the systems between decentralized and centralized systems when message loss occurs. Each simulation was run for 100, 300, and 700 cycles and was repeated five times with different random seeds in order to compute an average number of inaccuracies. The base station will compute the maximum for once at a cycle.

 Table 1 Average number of inaccuracies from the base station

Massaga	Centralized Systems			Decentralized Systems		
Loss	100 cycles	300 cycles	700 cycles	100 cycles	300 cycles	700 cycles
0%	0	0	0	7.6	7.6	7.6
10%	11	30.4	72.2	11.8	11.8	11.8
20%	19.6	61.8	146.6	12.4	12.4	12.4
30%	32.2	92.4	206	16.8	16.8	16.8

Table 1 shows that message loss has a correlation to the accuracy. The number of inaccurate estimations is increasing with the higher possibility of message loss. The best result comes from centralized systems which claim that the estimation of the base station is seamless regardless to message loss; however, the systems also receive a great suffer when message loss is considered. The system usage during a long period will lead to higher number of inaccuracies. In contrast, the long usage of decentralized systems does not affect an increase in the number of inaccuracies because they will only occur when the system is not completely distributed with the information. Nonetheless, a delay of information distribution is a necessary for decentralized systems. Figure 2 and table 1 show that there is a relationship between the convergence speed and the number of inaccuracies in decentralized systems. The systems with the fastest

convergence speed have the lowest number of inaccuracies.

6. CONCLUSION

Within the Industry 4.0's context, one of the smart manufacturing features is to focus on the establishment of the high-tech communication between not only humans but also machines in order to reach the goal of the smart factory practices. Industry 4.0 promises increased networking of these machines, which can be highly accurate and consequently customized to manufacture individual outcomes.

Centralized systems that are currently used in any industries require a single control center or hub to collect all required data to create a global information that can potentially suffer from being inaccurate, especially when messages are missing. This consequently results in poor decision making and most importantly, leading to an organization's future irreparably. Decentralized systems, however, redesign the networking system to allow each working unit to contain a global information without using any hubs.

This work has investigated the effect of systems architectures, centralized systems, and decentralized systems, to the accuracy of global aggregation. The characteristics of these architectures are also mentioned in this study. Epidemic aggregation protocols are used as the key for information distribution in the decentralized systems. The main task of these protocols is to provide a service to disseminate and generate the global information. The sets of results were produced under message loss conditions. By evaluating the accuracy of the decentralized systems over the centralized systems, the results show that message loss have a negative effect to the accuracy of the system and it is interesting to notice that the long period of system usage does cause higher inaccuracies in the centralized systems when message loss occurs, but not in decentralized systems. In addition, when there are no message losses, centralized systems run perfectly without any such inaccuracies while decentralized systems still have some error at the beginning of the simulations. These results can draw the conclusion that decentralized systems are suitable under the condition that the messages can be missing. On the other hand, centralized systems should be selected when the messages are guaranteed to be delivered.

In order to achieve the practical implications, further study on distributed systems in real-world manufacturing systems, e.g. node churn, information retrieval from multiple gateways and subbase stations, are required for future work. Another aspect to consider is how to improve the convergence speed of epidemic aggregation protocols under the context of Industry 4.0 with the variety of environmental conditions and case studies.

7. REFERENCES

- [1] O'Reilly Media, Inc., "Distributed Systems Topologies", O'Reilly Network: Distributed Systems Topologies, 2004.
- [2] Sofra N, He T, Zerfos P, Ko B J, Lee K, and Leung K K, "Accuracy Analysis of Data Aggregation for Network Monitoring", in Proc. 2nd Annual Conf. of the International Technology Alliance, London UK, September 2008.
- [3] Qu B, and Wang H, "The Accuracy of Mean-Field Approximation for Susceptible-Infected-Susceptible Epidemic Spreading", in Proc. of the 5th International Workshop on Complex Networks and their Applications, 2016.
- [4] Gupta I, Chandra D T, and Goldszmidt G S, "On scalable and efficient distributed failure detectors" in Proc. PODC'01, 2001, pp. 170-179.
- [5] Yuan K, Ling Q, and Yin W, "On the Convergence of Decentralized Gradient Descent", in SIAM Journal on Optimization, vol. 26, no. 3, 2016, pp. 1835-1854.
- [6] Poonpakdee P, and Di Fatta G, "Connectivity Recovery in Epidemic Membership Protocols", in IDCS, Springer, 2015, pp. 177 - 189.
- [7] Tang S, Jaho E, Stavrakakis I, Koukoutsidis I, and Van Mieghem P, Modeling gossip-based content dissemination and search in distributed networking, in Comput. Commun., vol. 34, 2011, pp. 765 - 779.
- [8] Di S, Wang C L, and Hu D H, "Gossip-based dynamic load balancing in an autonomous desktop grid," in Proc. of the 10th International Conference on High-Performance Computing in Asia-Pacific Region, 2009, pp. 85-92.
- [9] Vahdat A, and Becker D, "Epidemic Routing for Partially-Connected Ad Hoc Networks", Duke Tech Report CS-2000 – 06, 2000.
- [10] Ma Y, and Jamalipour A, "An epidemic P2P content search mechanism for intermittently connected mobile ad hoc networks", in IEEE GLOBECOM, 2009, pp.1-6.
- [11] Novotny P, "Fault Localization in Service-Based Systems hosted in Mobile Ad Hoc", Ph.D. thesis, Imperial College London, UK, 2013.
- [12] Van Renesse R, Minsky Y, and Hayden M, "A Gossip-Style Failure Detection Service", in Middleware'98 Proc. of the IFIP Int. Conf. on Distributed Systems Platforms and Open Distributed Processing, Springer-Verlaq, 2009, pp. 55-70.
- [13] Ahn J H, and Kim C Y, "Gossip-Pull Systembased Dissemination Protocol Satisfying

Message Causality Condition", in 2nd Int. Conf. on Advances in Computer Science and Engineering (CSE 2013), Atlantis press, 2013, pp. 323 325.

- [14] Strakov H, Niederbrucker G, and Gansterer W N, "Fault tolerance properties of gossip-based distributed orthogonal iteration methods", in Proc. Int. Conf. on Computational Science, vol. 18, 2013, pp. 189 - 198.
- [15] Soltero P, Bridges P, Arnold D, and Lang M, "A gossip-based approach to exascale system services", in Proc. of the 3rd Int. Workshop on Runtime and Operating Systems for Super computers, ser. ROSS 13. ACM, 2013.
- [16] Katti A, Di Fatta G, Naughton T, and Engelmann C, "Scalable and Fault-Tolerant Failure Detection and Consensus", in Proc. of the 22nd EuroMPI, 2015.
- [17] Di Fatta G, Blasa F, Cafiero S, and Fortino G, "Epidemic K-Means Clustering", in Proc. of the IEEE Int.l Conf. on Data Mining Workshops, 2011, pp. 151 158.
- [18] Mashayekhi H, Habibi J, Voulgaris S, and van Steen M, "GoSCAN: Decentralized scalable data clustering", in Computing, vol. 95, no. 9, 2013, pp. 759-784.
- [19] Kempe D, Dobra A, and Gehrke J, "Gossip-Based Computation of Aggregate Information", in 'FOCS', IEEE Computer Society, 2013, pp. 482-491.
- [20] Jelasity M, Montresor A, and Babaoglu O, "Gossip-based aggregation in large dynamic networks", in ACM Trans. Comput. Syst. vol. 23, no. 3, 2005, pp. 219-252.
- [21] Boyd S P, Ghosh A, Prabhakar B, and Shah D, "Randomized gossip algorithms", in IEEE Transactions on Information Theory 52 (6), 2006, pp. 2508-2530.
- [22] Blasa F, Cafiero S, Fortino G, and Di Fatta G, "Symmetric push-sum protocol for decentralized aggregation", in Proc. of the Int.l Conf. on Advances in P2P Systems, 2011, pp. 27-32.
- [23] Dimakis A, Kar S, Moura J, Rabbat M, and Scaglione A, "Gossip algorithms for distributed signal processing", in Proc. of the IEEE, vol. 98, no. 11, 2010, pp. 1847 - 1864.
- [24] Ragusa C, Liotta A, and Pavlou G, "An adaptive clustering approach for the management of dynamic systems," in Selected Areas in Communications, IEEE Journal on, vol. 23, no. 12, 2005, pp. 2223 2235.
- [25] Galzarano S, Savaglio C, Liotta A, and Fortino G, "Gossiping-based aodv for wireless sensor networks," in Systems, Man, and Cybernetics (SMC), 2013 IEEE Int. Conf., 2013, pp. 26-31.
- [26] Bawa M, Garcia-molina H, Gionis A, and Motwani R, "Estimating Aggregates on a Peer-

to-Peer Network", Technical Report. Stanford InfoLab, 2003.

- [27] Jesus P, Baquero C, and Almeida P S, "Dependability in Aggregation by Averaging", in CoRR abs/1011.6596, 2010
- [28] Campos F, Matos M, Pereira J, and Rua D, "A peer-to-peer service architecture for the Smart Grid", in 14th IEEE International Conference on Peer-to-Peer Comuting, London 2014, pp. 1-5
- [29] Garofalo G, Giordano A, Piro P, and Spezzano G, "A distributed real-time approach for mitigating CSO and flooding in urban drainage systems", in Journal of Network and Computer Applications, vol. 78, 2017, pp. 30-42
- [30] Montresor A, and Jelasity M, "PeerSim: A scalable P2P simulator," in Proc. of the 9th Int. Conference on Peer-to-Peer (P2P'09), 2009, pp. 99–100.
- [31] Jelasity M, "Self-Organising Software: From Natural to Artificial Adaptation", *Natural Computing Series*, Springer, 2011, pp. 139-162, DOI: 10.1007/978-3-642-17348-6_7
- [32] Poonpakdee P, Di Fatta G, "Robust and Efficient Membership Management in Large-Scale Dynamic Networks", in Future Generation Computer Systems, 2017, DOI:10.1016/j.future.2017.02.033

Copyright © Int. J. of GEOMATE. All rights reserved, including the making of copies unless permission is obtained from the copyright proprietors.