Deadline-Aware Routing: Quality of Service Enhancement in Cyber-Physical Systems

Sunghwa Son[†] · Byeong-Hoon Jang[††] · Kyung-Joon Park[†††]

ABSTRACT

Guaranteeing the end-to-end delay deadline is an important issue for quality of service (QoS) of delay sensitive systems, such as real-time system, networked control system (NCS), and cyber-physical system (CPS). Most routing algorithms typically use the mean end-to-end delay as a performance metric and select a routing path that minimizes it to improve average performance. However, minimum mean delay is an insufficient routing metric to reflect the characteristics of the unpredictable wireless channel condition because it only represents average value. In this paper, we proposes a deadline-aware routing algorithm that maximizes the probability of packet arrival within a pre-specified deadline for CPS by considering the delay distribution rather than the mean delay. The proposed routing algorithm constructs the end-to-end delay distribution in a given network topology under the assumption of the single hop delay follows an exponential distribution. The simulation results show that the proposed routing algorithm can enhance QoS and improve networked control performance in CPS by providing a routing path which maximizes the probability of meeting the deadline.

Keywords : Deadline-Aware Routing, Quality of Service, Cyber-Physical Systems

사이버물리시스템 서비스 품질 향상을 위한 데드라인 인지 라우팅

손 성 화[†] · 장 병 훈[††] · 박 경 준[†††]

요　　약

실시간 시스템, 네트워크 제어 시스템, 사이버물리시스템과 같이 지연에 민감한 시스템의 서비스 품질을 위해 종단 간 지연 데드라인을 보장하는 것은 중요하다. 대부분의 라우팅 알고리즘은 일반적으로 종단 간 평균 지연을 성능 메트릭으로 사용하고 평균 성능 향상을 위해 이를 최소화하는 라우팅 경로를 선택한다. 하지만 최소 평균 지연은 평균값만을 나타내기 때문에 예측할 수 없는 무선 채널의 특성을 반영하기에 불충분한 라우팅 메트릭이다. 본 논문에서는 평균 지연보다는 평균 분포를 고려하여 사이버물리시스템의 주어진 데드라인 내에 패킷이 도착할 확률을 최대화하는 데드라인 인지 라우팅 알고리즘을 제안한다. 제안한 라우팅 알고리즘은 단일 홉 지연이 지수 분포를 따른다는 가정 하에 주어진 네트워크 토폴로지에서 종단 간 지연 분포를 구성한다. 시뮬레이션 결과는 제안한 라우팅 알고리즘이 데드라인을 만족할 확률을 최대화 하는 라우팅 경로를 제공하여 사이버물리시스템의 서비스 품질과 네트워크 제어 성능을 향상시킬 수 있음을 보여준다.

키워드 : 데드라인 인지 라우팅, 서비스 품질, 사이버물리시스템

1. Introduction

Cyber-physical systems (CPS) have been recently developed and widely researched[1]. Considering various

systems are connected through network in CPS, network delay is one of the most significant factor that influences quality of service (QoS). In particular, in case of delay sensitive systems, such as real time system and networked control system (NCS)[2], a certain level of quality of control (QoC) should be guaranteed to provide reliable control performance.

The time taken for a packet traverses a network through routing path from source to destination is called network delay, which includes deterministic and non-deterministic

delays. Deterministic delays are caused by hardware performance issues, e.g. routers, whereas non-deterministic delays depend on software performance, such as routing algorithms, etc. Therefore, routing algorithm have to provide the optimal path with respect to the system's objectives.

Generally, most routing algorithms use the minimum mean delay as a routing metric to reduce the end-to-end delay for enhancing QoS[3]. However, minimizing the mean delay is not sufficient enough to achieve the performance of CPS since network delay includes non-deterministic delay which is unpredictable factor. Consequently, we need to introduce a performance metric that can properly reflect the stochastic nature of network delay, such as the probability of packet arrival within given deadline.

In this paper, we propose a deadline-aware routing algorithm that considers the probability of packet arrival within a given deadline as the major metric. Minimum mean delay does not maximize the probability of packet arrival within the deadline. For example, suppose a routing path, R1, has the same minimum mean delay but higher delay variance compared to another path, R2. Then R1 may not be able to deliver a sufficient number of packets to the destination within a given deadline. Therefore, we focus on the QoC of CPS, and propose a routing algorithm that can improve control performance over networks.

The rest of the paper is organized as follows. We introduce background fundamental details in Section 2, to illustrate how the proposed solution contributes to improve network performance. Section 3 reviews current research related to network packet transport and development of the proposed routing algorithm. In Section 4, we explain the key concepts and motivations, and presents the proposed deadline-aware routing algorithm. Section 5 presents the parameters and structures for a simulation model to test the proposed algorithm, and compares the performance with respect to the conventional shortest path routing algorithm. In particular, we show the effect of routing on networked control performance. Finally, we summarize the outcomes and present our conclusions in Section 6.

2. Background

2.1 Network Delay

Network delay represents the elapsed time from when a packet leaves from a source until it reaches its final destination, passing through the various nodes and other elements of the network. The packet experiences various processes as it is delivered to various network devices in the route. Fig. 1 shows the types of delay that can occur during

packet delivery, including processing (D_{proc}), transmission (D_{trans}), propagation (D_{prop}), and queueing (D_{queue}) delay. Then, the sum of delays occurring from one router to the next is nodal delay (D_{nodal}). End-to-end delay ($D_{end-end}$) is the sum of D_{nodal} from source to destination, but is not the product of all D_{nodal} values, due to the non-deterministic term arising from D_{queue}. Since $D_{end-end}$ is the critical delay that affects end users, minimizing $D_{end-end}$ is the main issue for routing algorithms[4].
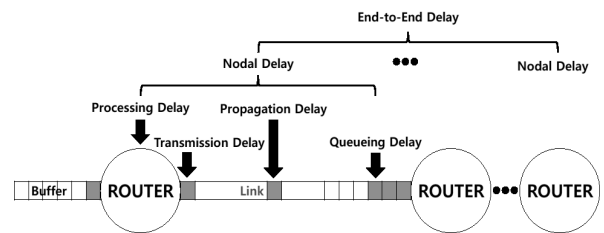

Fig. 1. Types of Network Delay

2.2 Routing Protocols

The concept of routing is to allocate the optimal path from source to destination, which applies not only to computer networks, but also to roadways, etc. The optimal path varies depending on the metric employed. The metric is a calculated factor expressing the "cost" of a given route, and incorporates hop count, delay, bandwidth, reliability, and load. Thus, we may choose different optimal paths depending on the specific metric chosen, which may vary for different purposes.

Routing protocols are classified differently depending on their table management and information exchange methods. Table management methods include static, dynamic, and default routing. The network manager directly designates the path for static routing. This routing method is usually employed only when the network environment is static and relatively small, since the routing table is not changed unless the network manager intervenes. Dynamic routing updates modified information among routers automatically. Although this consumes more resource than static routing, unexpected malfunctions in any router or other network devices are actively resolved.

Information exchange methods are distance vector and link state routing protocols, as shown diagrammatically in Fig. 2. Distance vector protocol updates router information across the whole network specific, periodic, times. Hop count and vector to destination require relatively little effort to update. However, they should be updated periodically regardless of network change, which wastes network traffic. Moreover, it takes longer to update routers when the convergence time extends due to router malfunctions, because the update is

executed by broadcast methods. Bellman-Ford discuss a representative updating algorithm[5]. In contrast to the distance vector protocol, link state protocol knows all the routing information to the packet destination, which provides short convergence time and infrequent information exchange. However, maintaining the entire routing information consumes significant memory. Dijkstra presents and discusses a representative and popular algorithm[6].

Fig. 2a shows that in distance vector routing, the optimal routing path is set as A-B, since the router only stores hop count and direction to destination. On the other hand, as shown in Fig. 2b, the link state router knows the entire network information to the destination, and is able to optimize the routing path as A-C-D-B. Link state protocol is generally used for larger networks, and the proposed scheme follows this protocol.

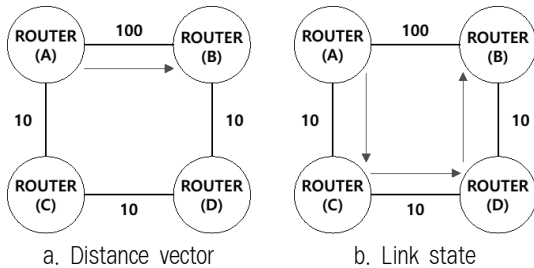a. Distance vector   b. Link state
Fig. 2. Comparison of Distance Vector Routing and Link State Routing

2.3 Networked Control System

Fig. 3 shows how the network control system (NCS) connects various devices in different locations via the network and exchanges control and input/output signals. The NCS is itself connected via the network, which reduces system maintenance costs by minimizing wire connections among related devices, and assists with system expansion and management due to network flexibility.
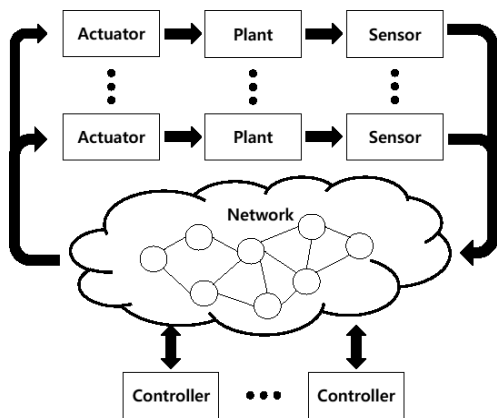

Fig. 3. Networked Control System Overview

CPS is one of the most popular NCS models. It provides a feedback control system that affects the physical system based on observation from network connected systems. Since CPS is a real time system, it requires immediate responsiveness, and to guarantee this, it is essential to minimize network delay. QoS and QoC in CPS environments have been widely investigated.

3. Related Work

Routing algorithms generally use packet transmission times for the network and builds a path decision using vehicle and plane concepts. To select a path, the algorithms choose the desired metric, such as delay, bandwidth, packet loss, stability, or hop count, and calculates the optimal path by comparing the calculated metric for candidate paths. Most networks use the conventional shortest path routing algorithm with the minimum mean delay metric. This section discusses previous routing algorithm studies considering metric options.

3.1 Quality of Service Routing

Networked control has become increasingly powerful and popular, and hence, QoS routing has been extensively studied. Systems employing NCS are very broad, including healthcare, CPS, and industry. Multiple performance metrics were considered in [3] and [7], and showed that although employing multiple metrics makes it more complex to calculate the optimal path, system performance can be significantly improved, with guaranteed QoS by considering bandwidth, delay, jitter, packet loss, and other metrics. In terms of robustness, path diversity routing for QoS was proposed[8]. QoS routing has also been widely studied in the context of wireless multimedia sensor networks[9], and many topological control algorithms that minimize interference among nodes have been proposed[10].

3.2 Road Network

In transportation literature, a stochastic vehicle routing algorithm is proposed[11]. They considered the delay distribution of real road network. This paper assumed that road network delay follows Gaussian distribution by gathering the real delay. In order to select the best path, this algorithm considered maximum probability reached within deadline based on [12-13]. The algorithm then compare the distribution of pre-stored delay data in a database, and since the delay distribution is Gaussian, calculates the optimal path. We exploit this method for our proposed algorithm, migrating the broad principles of these algorithms into the computing network environment.

4. Deadline-Aware Routing Algorithm

4.1 Key Idea and Motivation

As already mentioned, networked control in CPS requires timely delivery of each packet rather than average performance. The most important aspect is that a typical digital control periodically receives data from sensors and sends control inputs to the physical system. Thus, the probability of successful packet delivery within a given deadline is critical for system performance and physical system stability. This requirement is fundamentally different from average performance requirements, such as average delay and throughput for best-effort traffic. However, due to the stochastic nature of network delay, we need to consider a routing metric that incorporates the probability that each packet is delivered within a given deadline.

Fig. 4 shows two typical probability density functions (PDF), one with larger mean and smaller variance (red line), and the other with smaller mean and larger variance (blue line). In Fig. 4, Prob(delay > deadline) denotes probability when delay is bigger than given deadline; P_1 and P_2 denotes red and blue line, respectively. Although P_2 has less average delay than P_1, due to delay variance, P_1 has less probability of packet delivery within the deadline than P_2 (see Fig. 4). Therefore, the routing path with minimum mean delay may not maximize the probability of packet delivery within the deadline. Thus, rather than minimizing mean delay, we focus on minimizing the probability the network delay exceeds the given deadline.
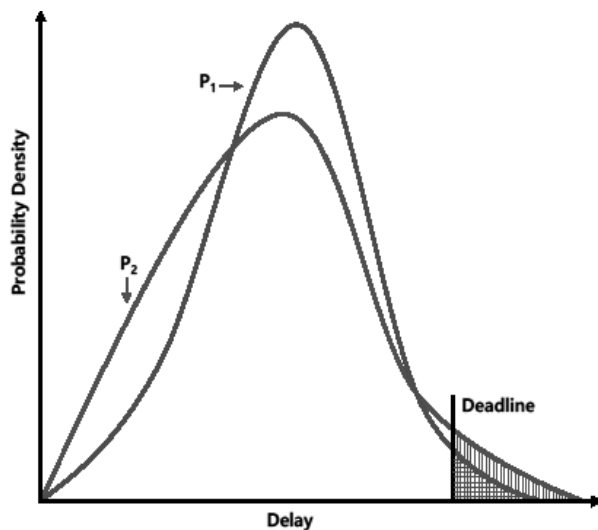

Fig. 4. Comparison of Outage Probability of Two Different Density Functions: Larger Mean with Smaller Variance (red) vs. Smaller Mean with Larger Variance (blue)

4.2 Deadline-Aware Route Selection

D_{queue} is non-deterministic, whereas all other network delays, D_{trans}, D_{prop}, and D_{proc} are deterministic, and is dependent on the router buffer statuses, which depend on network traffic status, i.e., busy or idle. Therefore, this study proposes a routing algorithm focusing on D_{queue}.

We assume the distribution of D_{queue} for a single hop link follows an exponential distribution. This is a reasonable assumption since single hop delay is measured from the backbone network, and D_{queue} has been shown to be at least approximately exponentially distributed over several data sets gathering packets passing a router[14]. It has also been shown that link-level D_{queue} distribution is exponentially distributed [15]. In summary, the assumptions of our study for network delay are:

• The distribution of single hop queueing delay follows an exponential distribution.

• D_{queue} is dominant term thus other network delays except for queueing delay are ignored.

• Conditions such as interference, bandwidth, and packet drop are not considered.

For the next step of modeling the distribution of end-to-end network delay, we need to aggregate each single hop delay distribution. We use hypo-exponential distribution which is a sum of exponential distribution. First, i of each link is exponentially distributed with its own rate of λ_i. Then, the $D_{end-end}$ distribution is expressed as a sum of independent exponential distributions as follows:

$$X = \sum_{i=1}^{n} X_i, \qquad (1)$$

where X is the hypo-exponential random variable, and X_i is the exponential random variable for the ith link, with rate λ_i.

Fig. 5 shows example PDFs that sum up 2-4 exponential distribution which has rate as 10, 10.1, 10.2, and 10.3, respectively. The distribution of hypo-exponential $D_{end-end}$ follows long-tailed distribution due to D_{queue}[16]. Like that, although network delay happens randomly during packet delivery, it shows certain distribution.

Our objective is maximizing a probability that $D_{end-end}$ is smaller than the given deadline. Delay distribution of each link is summed up to obtain $D_{end-end}$. As mentioned before, exponential distribution is sum up since all delay except D_{queue} is deterministic.

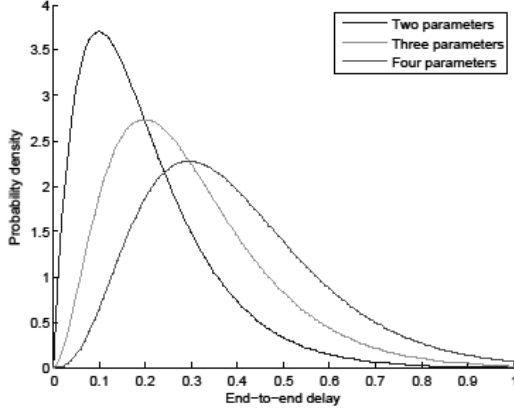The mean and variance of the end to end path delay in (1) can be expressed as

Fig. 5. PDF of Hypo-Exponential Distributions: Each Rate of Exponential Distributions is 10, 10.1, 10.2, and 10.3

$$X_{mean} = \sum_{i=1}^{n} \lambda_i, X_{variance} = \sum_{i=1}^{n} \frac{1}{\lambda_i^2}, \quad (2)$$

and the probability that $D_{end-end}$ delay is smaller than the given deadline may be calculated using the hypo-exponential distribution and cumulative distribution function (CDF) as,

$$\Pr ob(X \le x) = F(x) = \sum_{i=1}^{n} \frac{e^{-\lambda_i x} \prod_{j=1, j \ne i}^{n} \lambda_j}{\prod_{j=1, j \ne i}^{n} (\lambda_j - \lambda_i)}, \quad (3)$$

where λ_i and λ_j denotes rate of link i and link j, respectively. Using the probability calculated from (3), an optimal routing path can be chosen among possible paths from source to destination.

5. Simulation and Performance Evaluation

5.1 Simulation Environment

This section evaluates the proposed algorithm performance using MATLAB and Simulink[17]. Fig. 6 shows the network topology considered for the simulation, where the number on each link denotes the rate, λ, of the exponential distribution, and the mean delay is $1/\lambda$. Path 1 (Fig. 6, red line) and 2 (Fig. 6, blue line) denote end-to-end routing paths from source to destination chosen by deadline-aware and shortest path routing algorithms, respectively.

Network delay was explained in Section 2, and the parameters used in the simulation are described in Table 1. All parameters except D_{queue} are assumed to be constant for the convenience of calculation, and to focus on the non-deterministic nature of D_{queue}.
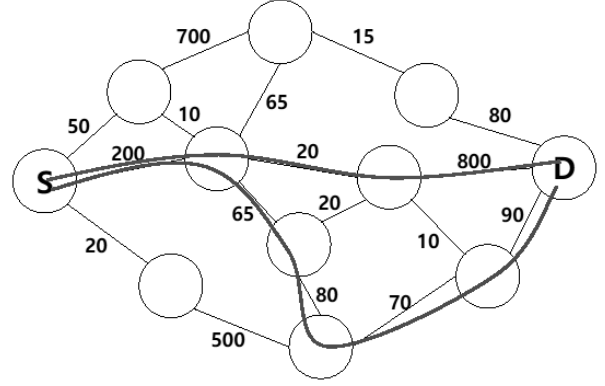

Fig. 6. A Network Topology for Performance Evaluation: The Number on Each Link Denotes the Rate λ and Hence Mean Delay of Each Link is $1/\lambda$

Table 1. Delay Parameters Used in the Simulation

Parameter type	Value
Distance (D)	1000 meters
Packet length (N)	128 bytes
Data bit rate (R)	10 Mbps
Speed (S)	3.0×10^8 m/s
D_{proc}	50 us
D_{trans}	1.024 ms
D_{prop}	3.33 us
D_{queue}	Randomly generated

As discussed above, CPS is a type of NCS that controls physical systems using feedback via the network. Hence, CPS performance is significantly affected by network delay. A typical Simulink model is considered, as shown in Fig. 7, incorporating an integrator plant, generally used in industrial applications; and proportional integral controller for system stability, where proportional gain and integral gain were set as shown in Table 2, along with the sine wave and constant reference values. In addition to this, we add a zero mean Gaussian noise with variance 0.01 to the feedback loop as a disturbance.
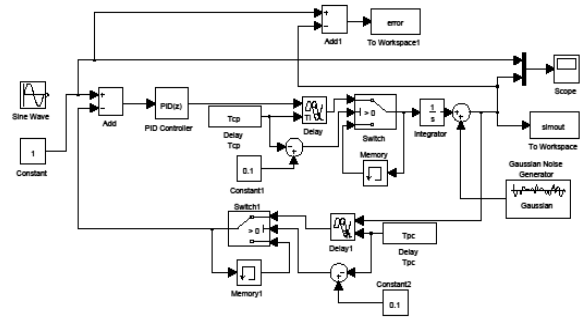

Fig. 7. Typical Simulink Model of Networked Control

Table 2. Simulink Parameters for Simulation

Sine wave reference	
Amplitude	1
Frequency	2
Constant reference	
Value	1
Gaussian noise	
mean	0
variance	0.01
Controller	
Sampling time	200 ms
Proportional gain	3
Integral gain	1

The sine wave reference shows how well the system is tracking, and the constant reference is utilized to check maximum overshoot of the plant. $D_{end-end}$ for each path is added between the controller and plant. Overall NCS delay includes delay from controller to plant, T_{cp}, and from plant to controller, T_{pc} which are expressed as delay block in Fig. 7. After each delay block, we added switch and memory block to determine delivering current or previous value of control and sensor signal with respect to $D_{end-end}$. As described in Table 2, the sampling time of the controller is set to 200 ms thus one way $D_{end-end}$ should be under the 100 ms. However, if T_{cp} or T_{pc} exceeds 0.1 second then switch block switches and transmits previous control/sensor signal which is stored in memory block.

5.2 Simulation and Performance Evaluation

We calculated the mean, variance, and CDF for paths 1 and 2 shown in Fig. 6 from (1) and (2). The shortest path routing algorithm chose path 2 as the optimal routing path because it had smaller mean delay than path 1, whereas our proposed algorithm chose path 1 as this had higher probability of reaching the destination within the given deadline, even though the mean delay was larger than path 2.

Fig. 8 shows the CDF for paths 1 and path 2 defined in Fig. 6. The network delay deadline was set to 100 ms. From equation (2) and (3), the mean delay of paths 1 and 2 were 58.3 and 56.3 ms, and the probability of arrival within the deadline was 0.9216 and 0.8458, respectively. Although mean delays were similar, the probability of packet arrival within the deadline was significantly influenced by the variance, as shown in Fig. 8.

Fig. 9 show the mean square error (MSE) between the sine wave reference and simulated output for paths 1 and 2, respectively. The run-time of the simulation was 20 seconds and repeated 1000 times. The average MSE for paths 1 and

2 seem to be adequately low as 0.2721 and 0.8090, respectively. However, in terms of distribution of 1000 times result, path 2 has several irregular MSE points even up to 29.4293. These peak MSE values signify a failure of networked control system which induced by large network delay for path 2.
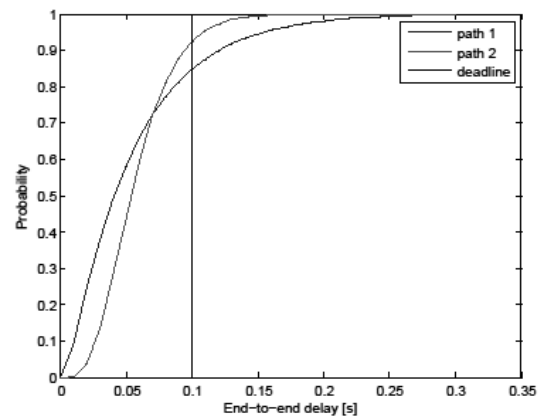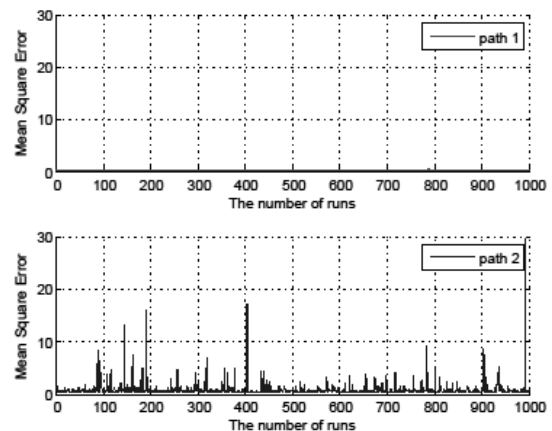

Fig. 8. The CDF of Path 1 and Path 2 from Fig. 6

Fig. 9. The Mean Square Error of Networked Control with Sine Wave Source for Path 1 and Path 2

Fig. 10 and 11 show the MSE and peak overshoot, respectively, relative to the constant reference for paths 1 and 2. The simulation environment was the same as for Fig. 9. The result of MSE has similar trend with average overshoot for paths 1 and 2 were 0.0321 and 0.0633, respectively. In case of measuring overshoot, path 2 has 7.3218 with maximum value which correspond to 732.18% of the reference value. In summary, simulation results show though path 2 has smaller mean network delay than path 1, it has a high risk that would make control system to unstable.

The proposed algorithm performance was simulated for several conditions, MSE and peak overshoot with sine and constant reference, important factors in assessing control

system performance. Control performance was significantly affected by delay variance, and the proposed algorithm showed significantly better control performance compared to conventional shortest path routing algorithm in CPS. Thus, to improve networked control performance, not only the mean network delay, but also the variance must be considered, and the optimal routing pathway should be derived using a stochastic algorithm, as per that proposed.
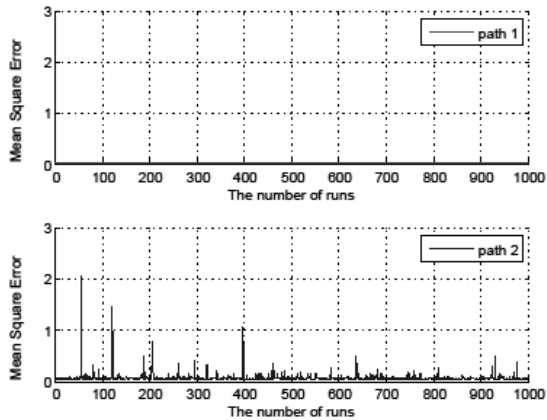

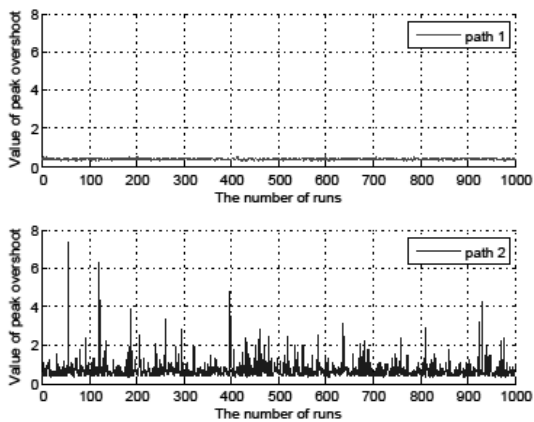Fig. 10. The Mean Square Error of Networked Control with Constant Source for Path 1 and Path 2

Fig. 11. The Peak Overshoot of Networked Control with Constant Source for Path 1 and Path 2

6. Conclusions

In this paper, we have proposed a deadline-aware routing algorithm to satisfy a probabilistic delay constraint in CPS. Since CPS requires timely packet delivery for QoS, the probability of packet arrival within the deadline is a critical factor. In order to maximize the probability of packet arrival within a given deadline, we consider the mean and variance of the overall delay, $D_{end-end}$. We have modeled the end-to-end delay distribution to hypo-exponential distribution

under the assumption of the link-level queueing delay is exponentially distributed. Through the simulations, we compare the proposed routing algorithm with the conventional shortest path routing algorithm in terms of MSE, and peak overshoot. The simulation results show that the proposed algorithm significantly improves the networked control performance as well as the system stability.

References

[1] K.-J. Park, R. Zheng, and X. Liu "Cyber-physical systems: Milestones and research challenges," *Elsevier Computer Communications*, Vol.36, No.1, pp.1-7, 2012.

[2] L. G. Bushnell, "Networks and control," *IEEE Control Systems Magazine,* Vol.21, No.1, pp.22-23, 2001.

[3] D. Medhi and K. Ramasamy, "Network routing: algorithms, protocols, and architectures," Morgan Kaufmann, 2010.

[4] H. Li, Y. Cheng, C. Zhou, and W. Zhuang, "Minimizing end-to-end delay: A novel routing metric for multi-radio wireless mesh networks," in *Proceedings of INFOCOM*, pp.46-54, 2009.

[5] R. Bellman, "On a routing problem," *Quarterly of applied mathematics*, Vol.16, No.1, pp.87-90, 1958.

[6] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische mathematik*, Vol.1, No.1, pp.269-271, 1959.

[7] A. Malik, J. Qadir, B. Ahmad, K.-L. A. Yau, and U. Ullah, "QoS in IEEE 802.11-based wireless networks: A comtemporary review," *Journal of Network and Computer Applications*, Vol.55, pp.24-46, 2015.

[8] K.-J. Park, J. Kim, H. Lim, and Y. Eun, "Robust path diversity for network quality of service in cyber-physical systems," *IEEE Transactions on Industrial Informatics,* Vol.10, No.4, pp.2204-2215, 2014.

[9] A. Alanazi and K. Elleithy, "Real-time QoS routing protocols in wireless multimedia sensor networks: Study and analysis," *Sensors*, Vol. 15, No. 9, pp. 22209-22233, 2015.

[10] X. M. Zhang, Y. Zhang, F. Yan, and A. V. Vasilakos, "Interference-based topology control algorithm for delay-constrained mobile ad hoc networks," *IEEE Transactions on Mobile Computing*, Vol.14, No.4, pp.742-754, 2015.

[11] S. Lim, H. Balakrishnan, D. Gifford, S. Madden, and D. Rus, "Stochastic motion planning and applications to traffic," *The International Journal of Robotics Research*, Vol.30, No.6, pp.699-712, 2011.

[12] E. Nikolova, J. A. Kelner, M. Brand, and M. Mitzenmacher, "Stochastic shortest paths via quasi-convex maximization," in *European Symposium on Algorithms*, Springer, pp.552-563, 2006.

[13] E. Nikolova, M. Brand, and D. R. Karger, "Optimal route planning under uncertainty," *ICAPS*, Vol.6, pp.131-141, 2006.

[14] K. Papagiannaki, S. Moon, C. Fraleigh, P. Thiran, and C. Diot, "Measurement and analysis of single-hop delay on an IP backbone network," *IEEE Journal on Selected Areas in Communications*, Vol.21, No.6, pp.908-921, 2003.

[15] A. Di Pietro, D. Ficara, S. Giordano, F. Oppedisano, and G. Procissi, "End-to-end inference of link level queueing delay distribution and variance," in *Proceedings of IEEE International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS)*, pp.503-510, 2008.

[16] P. Fransson and A. Jonsson, "End-to-end measurements on performance penalties of IPv4 options," in *Proceedings of Global Telecommunications Conference (GLOBECOM)*, Vol.3, pp.1441-1447, 2004.

[17] MATLAB version 8.1.0.604 (R2013a), The Mathworks, Inc., Natick, Massachusetts, 2013.

Byeong-Hoon Jang

https://orcid.org/0000-0002-0334-4250
e-mail : bhj@1011.co.kr

He received the B.S. degree in Computer Engineering from Yeungnam University, Gyeongbuk, Korea, in 2015 and M.S. degree in Information and Communcation Engineering from Daegu Gyeongbuk Institute of Science and Technology (DGIST) in 2017. He has been a Software Engineer at Teneleven Inc., Seoul, Korea since 2017. His research interests include cyber-physical systems and network security.

Sunghwa Son

https://orcid.org/0000-0002-5300-967X
e-mail : ssh@dgist.ac.kr

He received the B.S. degree in Electronic Engineering from Kyungpook National University, Daegu, Korea, in 2010. He is currently pursuing his Integrated M.S. and Ph.D. degree at Daegu Gyeongbuk Institute of Science and Technology (DGIST). His research interests include QoS in wireless networks and cyber-physical systems.

Kyung-Joon Park

https://orcid.org/0000-0003-4807-6461
e-mail : kjp@dgist.ac.kr

Kyung-Joon Park (M'05) received the B.S. and M.S. degrees in electrical engineering from the School of Electrical Engineering, and the Ph.D. degree in electrical engineering and computer science from Seoul National University, Seoul, Korea, in 1998, 2000, and 2005, respectively. From 2005 to 2006, he was a Senior Engineer with the Samsung Electronics, Suwon, Korea. From 2006 to 2010, he was a Postdoctoral Research Associate with the Department of Computer Science, University of Illinois at Urbana-Champaign (UIUC), Champaign, IL, USA. He is currently an Associate Professor with the Department of Information and Communication Engineering, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu, Korea. His research interests include resilient cyber-physical systems and smart production systems.