Hybrid Bandwidth Scheduling for CAN-based Networked Control Systems

BAI Tao[1, 2] WU Zhi-Ming[1]

Abstract A hybrid bandwidth scheduling scheme is proposed to improve the quality of service and the bandwidth utilization for the CAN-based networked control systems. It combines rate monotonic and improved round-robin scheme for both the real-time and non-real-time data. Moreover, considering the constraints of control performance and network schedulability, a heuristic branch and bound & genetic algorithm (GA) algorithm is presented for the control data to minimize their bandwidth occupancy and the jitter caused by improper scheduling. The residual bandwidth is allocated to non-real-time data by the proposed scale round-robin scheme such that their network loads are balanced.

Key words Networked control systems, QoS, network-induced delay, loop delay

1 Introduction

Because of the low implementation costs, optimum responsiveness, reconfiguration flexibility, and high reliability of the CAN protocol[1], the CAN-based networked control systems (NCSs) are widely used in the automotive industry, industrial automation, and process control areas these years. However, because of the limited bandwidth shared by real-time and non-real-time data, networked-induced delay and jitter, the transmission fluctuations of successive instances of the data degrade the dynamic performance of the NCSs inevitably and even cause instability (see [2∼4] and references therein). Therefore, many comprehensive studies [5∼12] have been focussed on the bandwidth utilization improvement and the control performance optimization.

A mixed traffic scheduler (MTS), combined with earliest deadline first (EDF) scheme with linear encoding and deadline monotonic (DM) scheme, was proposed in [5] to schedule network traffic with different real-time requirements. A relative deadline logarithmic encoding method was presented in [6] to realize EDF scheduling, which achieved much higher schedulability than MTS. A round-robin access scheme was realized in [7] to obtain the same QoS without limitation on the permitted serving number. In [8], a served-based hierarchical scheduling was presented for flexible time-triggered CAN protocol (FTT-CAN) to isolate the bandwidth among the different data streams and improve flexibility. A closed-loop fuzzy priority scheduling was proposed in [9] to widen the service range of the CAN without increasing overhead. Jitter control was mainly considered in [4, 10] by optimizing the task temporal parameters. The researches mentioned above emphaszed particularly on the improvement of the real-time scheduling performance, but failed to consider the network effects on the control system performance. Only a few papers were based on the relationship between the control performance and the bandwidth utilization. In [3], an optimal scheduling problem was formulated with both constraints of the rate monotonic (RM) utilization-based schedulability and the control stability. Hong[11] developed a non-jitter scheduling algorithm for multiple control loops with cyclic service discipline to satisfy the control performance. However, this is not always economical because excessive bandwidth is wasted to achieve the same control performance. The relationship between the sampling rates of control system and the network transmission rates has been analyzed in [11, 12]. These studies mainly focused on the control performance, scarcely paying attention to the network. Because only the sufficient conditions for network schedulablity were used, the results in [3, 12] were too conservative.

In this paper, a hybrid bandwidth scheduling scheme is proposed for the CAN-based NCSs to satisfy the different QoS requirements of data and to improve the bandwidth utilization. An RM-based scheduling scheme is used for real-time data, and the response-time based network schedulability is tested to guarantee their real-time requirement. Under both constraints of the control performance and the network schedulability, a progressive heuristic branch and bound & GA algorithm is proposed to optimize the sampling periods and the initial phases for the control loops. Therefore, the bandwidth occupancy of control data and the jitter caused by improper periodic data scheduling are minimized. And a scale round-robin scheme is presented to allocate the residual bandwidth to non-real-time data proportionally such that different non-real-time data nodes have the same QoS and the network loads are balanced.

2 Hybrid bandwidth scheduling scheme

In the CAN-based NCSs, the bandwidth is shared by real-time and non-real-time data. Real-time data usually subdivided into control and event data, have a stringent real-time transmission requirement. Control data are generated from the control loops. Each control loop has two data transmitting nodes namely, the sensor node and the controller node, whereas the actuator node is not included because it does not transmit its data through the medium. The transmitting nodes sample the data on a given periodic basis. If the network-induced delay is longer than the sampling interval, the control loop will experience such undesirable conditions as message rejection and vacant sampling, which degrade the control performance and distort the controller signals. The distortion will cause the high frequency noise in the actuator leading to excessive wear[2,11]. There-fore, control data should be transmitted within one sampling interval. Event data, including the monitoring, alarms, and diagnostic information, are generated aperiodically or periodically. They are generally very short in length and must be transmitted within a very short time interval. Non-real-time data are mainly of the numeric control programs, the data or graphic files, and the database

---

Received June 23, 2006; in revised form February 5, 2007
Supported by Natural Science Foundation of Shanxi Province (20051020)
1. Department of Automation, Shanghai Jiao Tong University, Shanghai 200030, P. R. China 2. Department of Automation, Taiyuan University of Science and Technology, Taiyuan 030024, P. R. China
DOI: 10.1360/aas-007-0963

management information. They are large in amount and their delay is not significant compared with that of real-time data. Therefore, the bandwidth for real-time data must be allocated exclusively to satisfy their real-time requirement, and the residual is then used to non-real-time data. Because of the priority-based medium access mechanism of CAN bus, the priorities are divided into three grades as a whole in this paper, with the highest priority for event data, the medium for control data, and the lowest for non-real-time data. In the following, it is assumed that there are N_e event data nodes, N_n non-real-time data nodes, and N_c control loops in the CAN-based NCS.

2.1 RM-based bandwidth scheduling for real-time data

When RM scheduling scheme is used, the shorter the period or the minimum interarrival time is, the higher the priority of data will be. In order to guarantee the real-time requirement, the schedulability should be tested. The schedulability for event data will be neglected because the same one for control data will be discussed below. The amount of control data generated by control loops is related directly to their sampling rates and affects the bandwidth occupancy consequently. The shorter the sampling period is, the better the control performance is obtained while the more the bandwidth is occupied. Therefore, considering both constraints of the control performance and the network schedulability, the sampling period for each control loop should be chosen properly to minimize the bandwidth occupancy.

2.1.1 Analysis of the control data model

The controller for each control loop is assumed to be designed in advance without considering the effect of the network. Sensor and controller node in the same control loop sample the data with an identical sampling rate. The performance of a feedback control loop directly depends on the loop delay[2], which is defined as the interval between the instant when the sensor node samples a datum from the plant and the instant when the actuator command generated based on the same data acts on the plant. Considering a control loop i with a sampling period T_i, the loop delay is expressed as

$$D_i = \left\lceil \frac{\tau_{i1}}{T_i} \right\rceil T_i + \tau_{i2} \tag{1}$$

with the network-induced delay τ_{i1} from sensor to controller and τ_{i2} from controller to actuator. Control data should be transmitted within one sampling interval, *i.e.*, the data deadline for each node is $\tau_{im} \le T_i$ $(m = 1, 2)$. The loop delay D_i is time-varying because of the time-variant τ_{i1}. The control loop is still kept stable when D_i is replaced by its constant supremum D'_i[11]

$$D'_i = 2T_i + (\max \tau_{i2} - \min \tau_{i1}) \tag{2}$$

Let ϕ_i $(\phi_i \le \phi_{i+1},\ i = 1, 2, \cdots, N_c)$ be the predetermined maximum allowable loop delay of control loop i, which can be obtained from the conventional stability criterion and/or considering the response smoothness (see [12] and references therein). Therefore, from the viewpoint of control, $D'_i \le \phi_i$ should be guaranteed so as to maintain the required control performance.

On the other hand, control data are scheduled by RM, so for node im with priority $pri(im)$, its minimum network-induced delay is

$$\min \tau_{im} = C_{im}, \quad (m = 1, 2) \tag{3}$$

where C_{im} is the maximum transmission time of its frame. The maximum network-induced delay is

$$\begin{aligned} \max \tau_{im} &= C_{im} + I_{im},\ (m = 1, 2) \\ I_{im} &= B_{im} + \sum_{pri(k) \in hp(pri(im))} \left(\left\lceil \frac{I_{im}}{T_k} \right\rceil \times C_k\right) \end{aligned} \tag{4}$$

where I_{im} is the interference time, including the longest blocking time all higher-priority data can occupy the bus and the blocking time B_{im} caused by the transmission of some lower-priority data, *i.e.*, $B_{im} = \max(C_k, C^n_j),\ \forall\ pri(k) \in lp(pri(im)), j = 1, 2, \cdots, N_n$. C_k and C^n_j are the maximum transmission time of a frame for node k with the priority of $pri(k)$ and for non-real-time node j, respectively. Noted that the set with the higher priority than $pri(im)$ also includes event data. T_k in (4) represents the period of periodic event data or the minimum interarrival time of aperiodic event data, too. If the maximum network-induced delay calculated by (4) is bounded to the permitted value, the real-time requirement is satisfied and the network is schedulable.

The network-induced delay experienced by the lower-priority control data is directly affected by the selected sampling periods of the higher-priority control data. In order to formulate the optimal sampling periods scheduling problem, an auxiliary variable of $\beta_{im}C$ is introduced where $C = \max(C^e_k, C_{im}, C^n_j)$ $(k = 1, 2, \cdots, N_e; im = 1, 2, \cdots, 2N_c; j = 1, 2, \cdots, N_n)$ is the time granularity. The maximum network-induced delay $\max \tau_{im}$ for node im is bounded by $\max \tau_{im} \le \beta_{im}C$. The sampling period could be determined as

$$T_i \le \left\lfloor \frac{\phi_i - \beta_{im}C + C_{im}}{2} \right\rfloor \tag{5}$$

Assume that $T_i = k_iC$, where $k_i \le \lfloor \frac{\phi_i - \beta_{im}C + C_{im}}{2C} \rfloor$ is an integer. With the increase of $\beta_{im}C$, T_i will decrease; thus in order to guarantee $\tau_{im} \le T_i$, $\beta_{im}C \le T_i$ should be kept. Moreover, $T_{i-1} \le T_i$ $(\forall\, i)$ is maintained for the NCS with $\phi_{i-1} \le \phi_i$ so as to guarantee the RM rule. Therefore, the minimum bandwidth occupancy and non-jitter scheduling are obtained by finding the optimal sampling periods $(T_i, \forall i)$ and the optimal initial phases $(\varphi_i, \forall i)$ for each control loop, which is formulated as follows

$$\begin{aligned} &\min J = \alpha_1 \sum_{\forall im} \frac{C_{im}}{T_i} + \alpha_2 \sum_{\forall i, im} \sum_{k=1}^{T/T_i} \frac{ts_{im,k} - KT_i - \varphi_i}{T_i} \\ &\text{s.t.:}\quad \max \tau_{im} < \beta_{im}C \le T_i \\ &\qquad T_i \le \left\lfloor \frac{\phi_i - \beta_{im}C + C_{im}}{2C} \right\rfloor C \\ &\qquad T_i \le T_{i+1}\ \ \forall\, i = 1, 2, \ldots, N_c, \ \ \forall\, im = 1, 2, \ldots, 2N_c \end{aligned} \tag{6}$$

In the performance measure, the first term is the network bandwidth occupancy and the second is the overall system jitter (OSJ) with the penalty factors $\alpha_1 > 0$, and $\alpha_2 > 0$, respectively. The OSJ can be obtained by integrating the individual jitter for each instance of control data over the system's macro-cycle $T = LCM[T_i, i = 1, 2, \cdots, N_c]$. $\frac{ts_{im,k} - KT_i - \varphi_i}{T_i}$ is the jitter for the kth instance of the data for the node im, where $ts_{im,k}$ is the actual transmission time of this instance. The first constraint in (6) is the RM response time-based schedulable condition, which is the sufficient and necessary condition, so it may give less conservative results than [3, 12].

2.1.2 Progressive heuristic branch and bound & GA algorithm for control data

A progressive heuristic branch and bound & GA algorithm is proposed to solve this optimal problem by the following steps:

Step 1. Determine the feasible domain of the sampling period T_1 for the first control loop and choose its optimal values that minimize the jitter.

Step 2. For each of the optimized branch, add the next ith control loop with the highest priority to the remainder undetermined set, determine the feasible domain of its sampling period T_i, and search for its optimal values with the previous optimized set that minimize the jitter using GA.

Step 3. Repeat Step 2 until the complete set of the control loops are determined and then compare all the optimal branches of performance measure and obtain the global optimal solution.

The feasible range of T_i for each control loop can be determined when some $\beta_{im}C$ is chosen with its initial value as $C + \sum_{pri(k)\in hp(pri(im))} C_k + C_{im}$. For a given $\beta_{im}C$, if the feasible range $\Gamma_{\beta_{im}}$ exists, it is needless to search for the feasible range $\Gamma_{>\beta_{im}}$ with the given $> \beta_{im}$, because of $\Gamma_{>\beta_{im}} \subseteq \Gamma_{\beta_{im}}$. When $\beta_{im}C = T_i$, if no feasible range $\Gamma_{\beta_{im}}$ exist, it means that this branch is not feasible and the network is not scheduled in this branch.

Jitter minimization can be achieved by GA. A similar method was used in [4]. The genome is a vector that holds the values of $(\varphi_i, i = 1, 2, \cdots, N)$ for all control loops. The allele represents the initial phase of control loop i, which is a value integer times the granularity within the range [0, T_i]. The fitness of each individual is computed as $\frac{1}{OSJ+1}$, so the best solutions generate the smallest OSJ.

This algorithm works iteratively and only the optimal branches are considered for the next iteration. Thus, the amount of the branches decreases considerably and the process of jitter minimization is also speeded up because of the obtained partial optimal solutions.

2.2 Scale round-robin scheduling for non-real-time data

The priority-based medium access mechanism of CAN protocol makes the higher priority data occupy the bandwidth preferentially , thus giving rise to better QoS. It is important to guarantee the real-time requirement for real-time data excluding the non-real-time data via accessing the bus. However, it dose not guarantee the same QoS between non-real-time data because the higher priority data can hog all available residual bandwidth such that the lower priority ones are seldom served under the heavy traffic condition. Therefore, the CAN protocol cannot guarantee the fair competition between different non-real-time data nodes and causes unbalanced workload.

In virtue of the broadcast characteristics of CAN protocol, a round-robin access scheme was presented in [7] by remembering the priority information of the last transmitted frame, where the state transition of node i is shown as Fig. 1. M is the priority mark, which indicates that all nodes with priority higher than M will be forbidden to compete with the bandwidth. As a consequence, each node is served at most once in each transmitting cycle, as in the case of token-based network. However, the scheme does not consider the actual QoS requirements for non-real-time nodes, which may make the network load unbalanced.

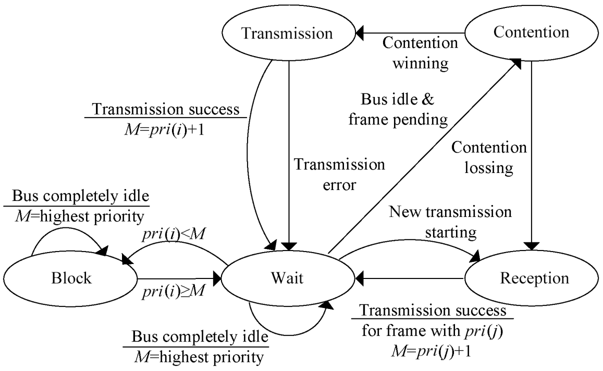

Fig. 1 State transition for non-real-time node i under round-robin scheme

An improved scheme, the scale round-robin access scheme, is proposed in this paper. The state transition of node i is shown as Fig. 2. $NC(i)$ is its maximum served number in one transmitting cycle, which is related to the actual QoS requirement of node i. $CR(i)$ is the counter to record the served number for node i in the present cycle. Different from the round-robin scheme, the priority mark M is only updated after node i is served for the $NC(i)$ times, *i.e.*, when $CR(i) = NC(i)$, $M = pri(i) + 1$, and $CR(i) = 0$.

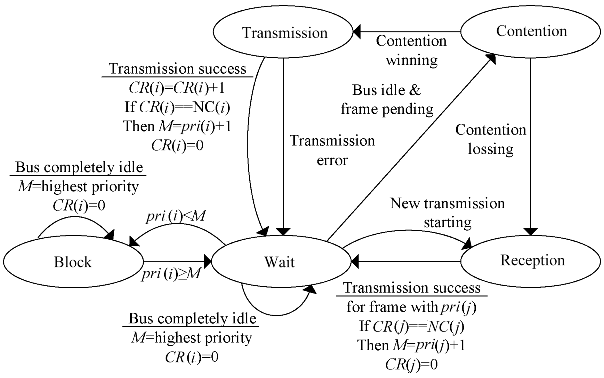

Fig. 2 State transition for non-real-time node i under scale round-robin scheme

Usually non-real-time data are long and can vary in length, so they are segmented into several frames and added to the transmission queue, then transmitted one by one and reassembled at the destination node. Assume that the arrival of non-real-time data at node i is a poisson distribution with a parameter of Λ_i^n, and its average length of the non-real-time data is M_i^n. If it is segmented into $f_i^n = \lceil \frac{M_i^n}{C_i^n} \rceil$ frames, then the average arrival rate of a non-real-time data frame generated unit time at node i is $\lambda_i^n = f_i^n \Lambda_i^n$. Therefore, in order to balance the network load among different non-real-time nodes, $NC(i) = \lceil \frac{\lambda_i^n}{\min(\lambda_i^n, i = 1, 2, \cdots, N_n)} \rceil$ can be set.

2.3 Network stability

The network is stable if all kinds of data are served without any overflow of the corresponding transmission queues. For real-time data, the schedulability is also their network stability. In this paper, the RM-based schedulability condition for real-time data, the first constraint in (6), is their network stable condition. For non-real-time data, the network stability means that all data generated during a long period of time t must be transmitted in less than t. Therefore, the network stable condition for the non-real-time data scheduled by the scale round-robin scheme

$$\sum_{im}^{2N_c} C_{im}\left\lceil \frac{t}{T_i} \right\rceil + \sum_{j}^{p} C_j^e \left\lceil \frac{t}{T_j^e} \right\rceil + \sum_{j=p+1}^{N_e} C_j^e(\lambda_j^e t) + \sum_{k}^{N_n} C_n^k(\lambda_k^n t) \leq t, \quad 1 \leq p \leq N_e \tag{7}$$

3 Simulation tests

The hybrid bandwidth scheduling scheme is tested when the network load is approximately 95%, assuming that the CAN-based NCS has $N_c = 5$ control loops, $N_e = 2$ event data nodes, and $N_n = 5$ non-real-time nodes with the transmission rate of 125 KB/s. The maximum allowable loop delays for the control loops are $[\phi_1, \phi_2, \phi_3, \phi_4, \phi_5] = [24, 40, 65, 100, 150]$ ms. The control data length is 130 bits, including the effective data with 8 bytes, overhead, worst-case stuffing bits and inter-frame space (3 bits), and the frame transmission time $C_{im} = 1$ ms. Event data are generated randomly with the minimum interarrival time 4 ms and 10 ms, respectively, and their deadlines are both 5 ms. The event data length is 2 bytes with the transmission time of $C_i^e = 0.6$ ms. The average arrival rate λ_i^n of non-real-time data frame generated unit time is 0.167, 0.067, 0.048, 0.042, and 0.033 ms, respectively, and the non-real-time data length is 8 bytes, so $C_j^n = 1$ ms. By the progressive heuristic branch and bound & GA algorithm, the optimal sampling periods of control loops are determined as $[T_1, T_2, T_3, T_4, T_5] = [8, 16, 24, 40, 64]$ ms and the corresponding optimal initial phases are $[\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5] = [0, 2, 4, 14, 6]$ ms with the minimum bandwidth occupancy 53.96%. By Hong′s scheme[11], $[T_1, T_2, T_3, T_4, T_5,] = [8, 16, 16, 32, 64]$ ms are determined for the same control loops with the bandwidth occupancy 59.38%. There will be 5.42% more bandwidth occupied by the control data than ours while both scheme satisfy the same control performance requirement. It means that too much bandwidth is wasted and the network serving capability is cut down. Therefore, our scheme is better than Hong′s.

The QoS for real-time are listed in Table 1. The real-time requirements are all satisfied because the maximum network-induced delays for real-time data are bounded by their respective deadlines. Moreover, the maximum loop delays for the control loops are 10.987, 19.1, 27.635, 47.05, 73.106 ms, respectively, and they are all smaller than their corresponding maximum allowable loop delays. Therefore, the required control performance for each control loop is guaranteed.

The QoS for non-real-time data by three different schemes are shown in Fig. 3. Compared with the other two schemes, the residual bandwidth allocated to each non-real-time node is proportional to the average frame arrival rate of λ_i^n by the scale round-robin scheduling. Therefore, the proposed scheme not only provides fair service and similar QoS to the non-real-time data nodes, but also balances the network load of the non-real-time data nodes.

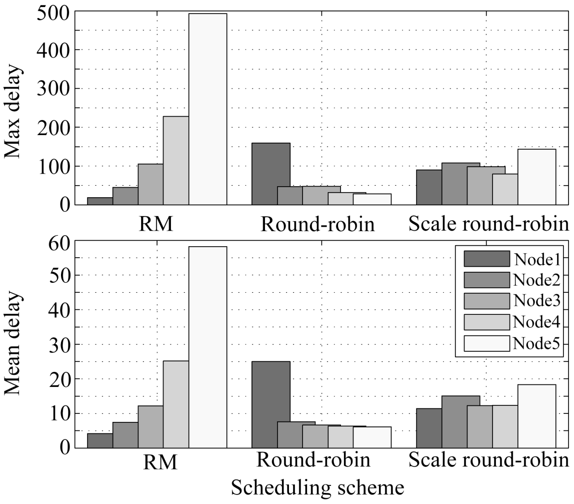

Fig. 3 QoS comparison for non-real-time nodes by different scheduling schemes

Table 1 QoS for the real-time data by hybrid bandwidth scheduling

	Deadline (ms)	Min delay (ms)		Max delay (ms)		Avg delay (ms)	
Event data node1	5	0.6		1.600		1.074	
Event data node2	5	0.6		2.195		1.116	
		Controller	Sensor	Controller	Sensor	Controller	Sensor
Control loop1	8	1	2	2.987	4.178	1.460	2.520
Control loop2	16	1	2	3.100	4.170	1.564	2.634
Control loop3	24	1	2	3.635	4.635	1.575	2.635
Control loop4	64	1	2	9.106	10.657	2.457	3.874

4 Conclusion

It is important to assign the limited bandwidth resource among different kinds of data properly for the NCSs such that the network influence on the controlled plants is as little as possible and the different quality of service (QoS) requirements of these data are guaranteed. In this paper, a hybrid bandwidth scheme is proposed for the CAN-based NCSs to satisfy the different quality of service (QoS) requirements and improve the bandwidth utilization. Based on RM scheduling, the optimal sampling periods with their optimal initial phase of control data are determined by the progressive heuristic branch and bound & GA algorithm such that the bandwidth occupancy and jitter caused by improper periodic data scheduling are minimized. The scale round-robin scheme is proposed to allocate the residual bandwidth to non-real-time data proportionally such that different non-real-time data nodes are served fairly and their network loads are balanced. Based on the broadcast nature and priority-based scheduling mechanism, the proposed hybrid scheme can also be extended to other networks if they have the same features, such as IEEE802.3p.

References

1 Torngren M. A perspective to the design of distributed real-time control applications based on CAN [Online], available: http:// www.md.kth.se, Octobor 6, 2003

2 Ray A, Halevi Y. Integrated communication and control systems: part I - analysis and part II - design consideration. *ASME Journal of Dynamic System Measurement & Control*, 1998, **110**: 367∼381

3 Zhang W. Stability analysis of networked control systems [Online], available: http://dora.cwru.edu/msb/students.html, December 28, 2005

4 Coutinho F, Fonseca J, Barreiros J. Using genetic algorithms to reduce jitter in control variables transmitted over CAN [Online], available: http://www.ieeta.pt/%7Ejaf/, September 25, 2003

5 Zuberi K M, Shin K G. Design and implementation of efficient message scheduling for controller area network. *IEEE Transactions on Computers*, 2000, **49**(2): 182∼188

6 Natale M D. Scheduling the CAN bus with earliest deadline techniques. In: Proceedings on Real-time Systems Symposium. IEEE, 2000. 259∼268

7 Cena G, Valenzano A. Achieving round-robin access in controller area networks. *IEEE Transactions on Industrial Electronics*, 2002, **9**(6): 1202∼1213

8 Nolte T, Sjodin M, Hansson H. Hierarchical scheduling of CAN using server-based techniques. In: Proceedings of 3rd International Workshop Real-time Networks (RTN′04) & 16th Euromicro International Conference Real-time System (ECRTS2004). Catania, Italy, 2004

9 Bai T, Hu L S, Wu Z M, Yang G K. Flexible fuzzy priority scheduling of the CAN bus. *Asian Journal of Control*, 2005, **7**(4): 401∼413

10 David L, Cottet F, Nissanke N. Jitter control in on-line scheduling of dependent real-time tasks. In: Proceedings on Real-time Systems Symposium. IEEE, 2001. 49∼58

11 Hong S H, Kim W H. Bandwidth allocation scheme in CAN protocol. *IEE Proceedings of Control Theory and Applications*, 2000, **147**(1): 37∼44

12 Lian F L, Yook J K, Tilbury D M, Moyne J R. Network architecture and communication modules for guaranteeing acceptable control and communication performance for networked multi-agent systems. *IEEE Transactions on Industrial Informatics*, 2006, **2**(1): 13∼24

BAI Tao Postdoctor at Department of Computer Science, Shanghai Jiao Tong University. Her research interest covers industrial networked control system modeling, control, and optimization. Corresponding author of this paper.
E-mail: Bettertao.bai@gmail.com

WU Zhi-Ming Professor at Department of Automation, Shanghai Jiao Tong University. His research interest covers hybrid dynamic systems, distributed realtime embedded systems, and information integration. E-mail: ziminwu@sjtu.edu.cn