Control Engineering Practice ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Control Engineering Practice

journal homepage: www.elsevier.com/locate/conengprac

Bandwidth allocation and scheduling of networked control systems with exponential and quadratic approximations

Jiawei Dong*, Won-Jong Kim

Department of Mechanical Engineering, Texas A&M University, College Station, TX, USA

A R T I C L E I N F O

Article history:
Received 6 November 2012
Accepted 2 January 2014

Keywords:
Networked control system
Bandwidth allocation
Scheduling
Exponential approximation
Quadratic approximation

A B S T R A C T

This paper investigates bandwidth allocation and scheduling of networked control systems (NCSs) with nonlinear-programming techniques. The bandwidth utilization (BU) is defined in terms of sampling frequency. An exponential and a quadratic approximation are formulated to describe system performance versus the sampling frequencies. The optimal sampling frequencies are obtained by solving the approximations with Karush–Kuhn–Tucker (KKT) conditions. Experimental results verify the effectiveness of the proposed approximations and scheduling algorithms. The two approximations could find an optimal BU of an NCS with a given sequence of plants and maximize the total BU up to 98% of the total available bandwidth.

1. Introduction

Networked control systems attract significant attentions recently due to their advantages of easy maintenance, architecture flexibility, reduced wiring cost, etc. However, the design of an NCS requires not only participation of controller designers but also real-time operating-system (RTOS) designers because of the introduction of the networks. Traditionally, a controller design problem is separate from software design and implementation. This separation allows controller designers to disregard the characteristics of the computational and communication resources, but mainly focus on the stability and performance of the controllers and the systems. On the other hand, the RTOS designers consider the control loops as periodic tasks with hard deadlines and focus more on how to schedule all the tasks and guarantee that the tasks do not miss their deadlines (Arzen, Cervin, Eker, & Sha, 2000). In the NCSs, however, these two fields are correlated in a closer way so that their separation will lead to poor system performances. The ideal linear relation between the system performance and the sampling frequency is no longer the case for the NCS design because of the existence of the network.

A representative framework of an NCS is shown in Fig. 1. In this framework, the NCS includes several operation scenarios—(i) a single controller controls a single plant, (ii) a single controller controls multiple plants, and (iii) multiple controllers collaboratively control a single plant, etc. In this framework, all the controllers and the plants will compete for the limited resources, such as the central processing unit (CPU) time, network bandwidth, and battery, in the NCS to maintain the stability and performance. More often, one could expect the global information sharing and resource allocation could dynamically adjust the performance of each plant so that the entire NCS could be maintained at a desirable level. Therefore, the communicational and computational resource allocation and scheduling plays a crucial role in the design of an NCS.

Traditionally, digital control ideally assumes that the system performance index can be reflected by a monotonically decreasing linear or exponential function of the sampling frequency. In other words, a higher sampling frequency yields better performance. In practice, this is not always the case as noise, numerical errors, and hardware limits exist in reality. NCS is one of the exceptions because networks have bottlenecks as hardware limits such that the ideal monotonic linear performance function no longer holds for the design of an NCS (Lian, 2001). A higher sampling frequency will increase the number of data packets in the network, which will cause longer time delays and might even overload and destabilize the network. Therefore, the linear models of the system performance proposed in the aforementioned literature could not completely represent the system dynamics in an NCS. The effects on the system performance with possibly longer time delays brought by a high sampling frequency should be considered when formulating the performance index function (PIF) of an NCS. Fig. 2 gives intuitive trends of the system performance of an NCS with respect to the sampling frequency. In Fig. 2, f_γ is the optimal sampling frequency that yields the optimal system performance of an NCS. Sampling frequencies f_α and f_β are the boundaries of the acceptable performance range.

* Corresponding author. Tel.: +1 979 842 2649.
E-mail address: jwdong@aggienetwork.com (J. Dong).

0967-0661/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.conengprac.2014.01.001

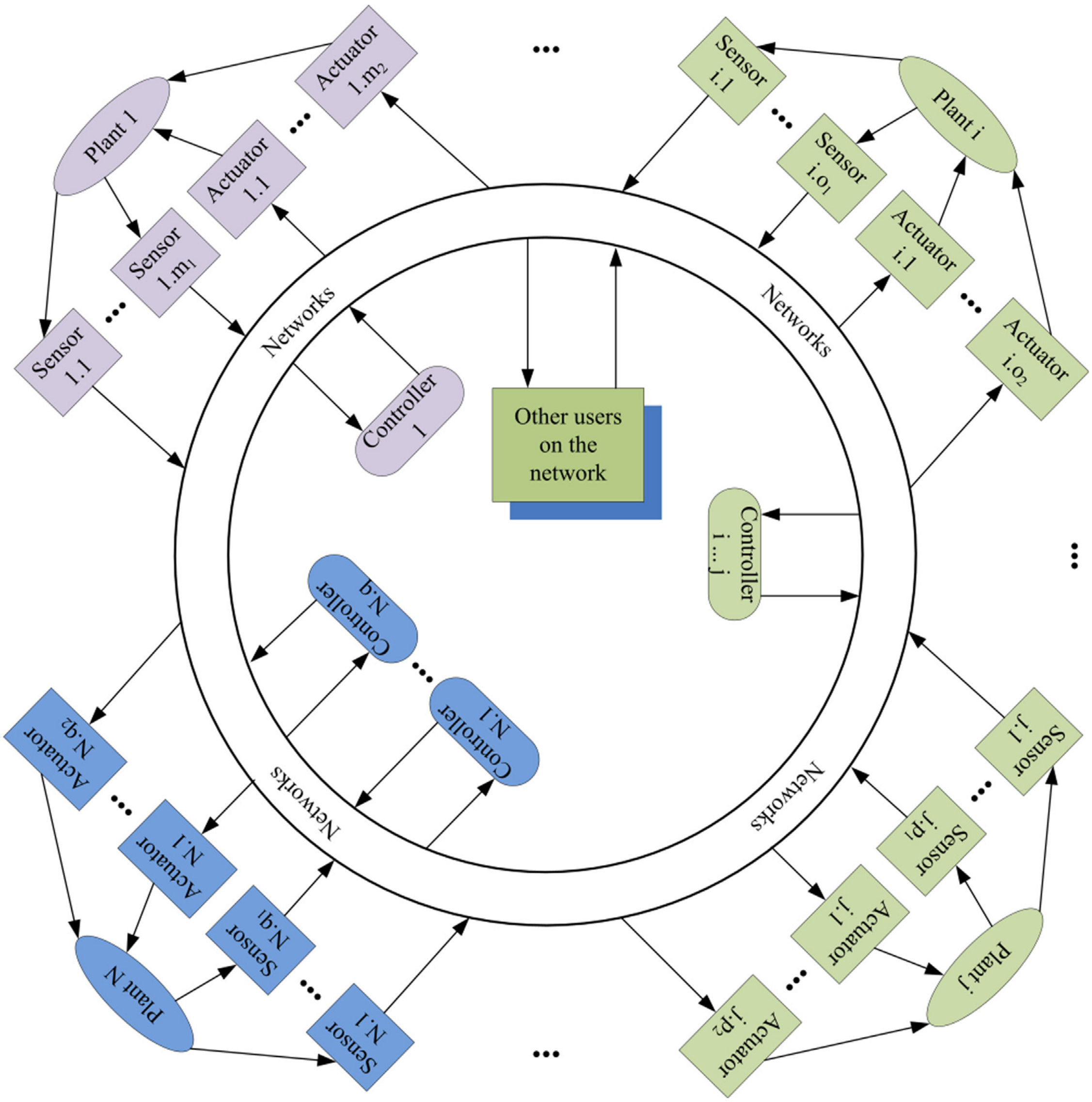

Fig. 1. A representative framework of an NCS.

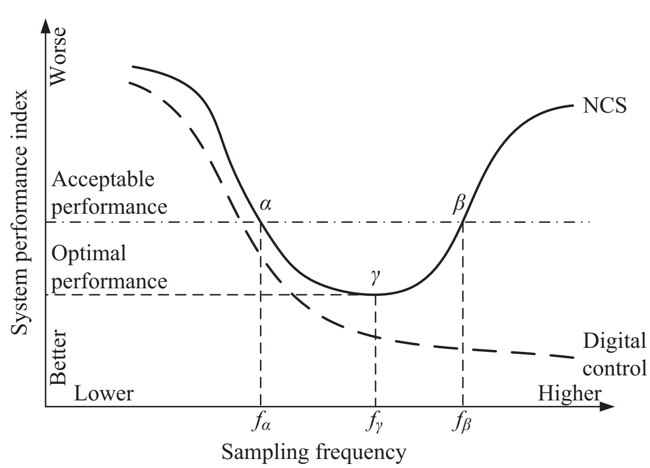

Fig. 2. Performance index versus sampling frequency (Lian, 2001).

Several research projects have been conducted in this area. Zhang, Gao, and Kaynak (2013) gave a research survey of network-induced constraints of the NCS including time delays, packet losses, resource competition, data quantization, etc. The resource competition should be solved from the control perspective to equip the NCS with calculating optimal sampling frequencies and a dynamic scheduler based on Zhang's survey. The convex optimization applied in the paper is one of the most powerful and popular tools that could solve the resource-allocation issue effectively. A dynamic bandwidth allocation algorithm based on captured visual content information was presented to raise the BU of an NCS (Lin & Lian, 2012). The algorithm was to evenly distribute resources to each node in the NCS and then dynamically revise their allocation based on a linear performance evaluation whereas our proposed methods dynamically reschedule each client based on their performance changing rate with respect to the sampling frequency. A network bandwidth allocation with time reservations was studied (Belzarena, Ferragut, & Paganini, 2009). The NCS involved fully distributed solutions over an arbitrary network topology in terms of revenue that was computed via distributed convex optimization. Effects of the sampling period to an NCS were discussed, and an optimized model and optimal sampling period selection algorithm was proposed based on the control performance optimization and network scheduling condition (Wang & Liu, 2011). A bandwidth-allocation scheme formulated as a convex optimization problem for NCSs was proposed (Al-Hammouri, Branicky, Liberatore, & Phillips, 2006). A co-design approach was proposed to treat communication protocols and interact controlled plants as a coupled system (Branicky, Phillips, & Zhang, 2002). These papers all considered that the NCS performance could be modeled by convex functions, and the resource allocation could be achieved by solving the convex functions with existing optimization algorithms. However, the above studies neglected the network-induced time delays, which could significantly affect the system performance. This paper strives to include the impacts of the time delays in an explicit way when resolving the NCS resource allocation problem.

The NCS system performance was approximated as an exponential function, and the resource allocations were solved by convex optimization (Lin, 2004; Peng, Yue, Gu, & Xia, 2009; Seto, Lehoczky, Sha, & Shin, 2001). However, the negligence of the time delays in the approximations would not reveal the characteristics of an NCS. Those studies motivated us to find a proper performance approximation that considers the effects of time delays. Compared to those existing methods, our proposed methods consider not only the effect of time delays brought by high sampling frequencies as an essential part when setting the system PIFs of an NCS but also the scheduling sequences of the controlled plants. The proposed approximations and the optimal solutions are expected to exhaust the entire network bandwidth available to the NCS to maximize the BU and the system performance. On the other hand, the consideration of the time delays also introduces certain complexity into the optimization compared to the ones in the cited papers. As can be seen in Section 3.1, a transcendental equation was generated as the derivative of the Lagrange equation. This causes the on-line calculation of the optimization to be infeasible as it would bring more delays in control-law generation. Therefore, the proposed methods should calculated off-line and tabulated in the NCS controller.

Note that although the proposed approximations and scheduling algorithms are mainly for an NCS with a single controller and multiple plants, they can be applied to an NCS with multiple controllers and multiple plants easily with proper adjustments.

The rest of the paper is organized as follows: in Section 2, the system PIFs are formulated by an exponential approximation and a quadratic approximation. In Section 3, the optimal solutions of both approximations are given, and a scheduling algorithm is proposed. In Section 4, experimental results are provided to illustrate the effectiveness of the proposed approximations and scheduling algorithm. In Section 5, our conclusions are presented.

2. NCS performance approximations

Consider an NCS with the framework that contains one controller and multiple controlled plants. To guarantee the stability and enhance the system performance, all the controlled plants are assumed to compete for the CPU time and the network bandwidth to calculate control inputs and transmit data packets. Accordingly, the most common objective in the resource allocation of an NCS is to optimize the overall quality of control subjected to certain resource limitations. A representative diagram of an NCS with the framework of single controller and multiple plants is shown in Fig. 3, where $\mathbf{e}_i(k)$, $\mathbf{u}_i(k)$, and $\mathbf{y}_i(k)$ are error, control-input, and plant-output vectors of Plant i, respectively. As in the dashed box, the network nodes are indicated as Ethernet. A central controller that contains all the corresponding control laws for N plants reside at one end of the communication network whereas the N plants including sensors and actuators, at the other end. From Fig. 2, one can see that all these N plants share the same communication networks and their corresponding control laws share the CPU and other resources on the controller.

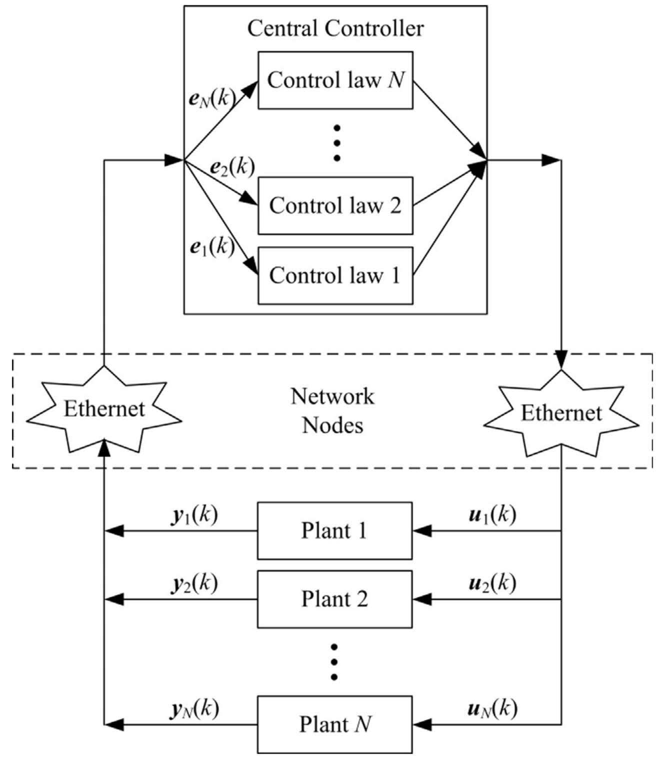

Fig. 3. A representative NCS block-diagram with a single controller and multiple plants.

2.1. Network bandwidth of NCS

To achieve the optimal resource-allocation objective, a system PIF in terms of various resources is set up. The relation between the sampling period or frequency and the BU can be indicated by the following equation (Park, Kim, Kim, & Kwon, 2002):

$$b_i^k = \tau_i^k / h_i^k \quad \text{or} \quad b_i^k = \tau_i^k f_i^k \tag{1}$$

where b_i^k is the BU, h_i^k is the sampling period, f_i^k is the sampling frequency, and τ_i^k is the total time delay in the NCS that includes the propagation delay from the network and the data processing time. The subscript i indicates the index of the plants in the NCS, and the superscript k indicates the number of control iterations so far. That is, the controller generates the k-th control signal in the k-th sampling period. Note that given a certain amount of time delays, Eq. (1) gives a means to evaluate the plants′ sampling frequencies and represents the portion of network bandwidth assigned to each plant. Since τ_i^k includes the data processing time on Server′s CPU, this bandwidth definition also implicitly indicates the CPU resource allocation on Server. In control system design, the sampling frequency directly relates to the system stability and performance. Therefore, Eq. (1) also gives an implicit means to measure the plant′s stability and performance. A large BU implies a high sampling frequency as indicated by Eq. (1) if given a certain amount of time delays so that a plant can have a better performance. However, an upper bound exists on the NCS bandwidth. If the BU reaches the network bandwidth saturation threshold, the network will be overloaded and induce more time delays or packet losses, and the performance of an NCS will be degraded.

2.2. Performance index function

In this paper, discrete integral absolute error (DIAE) is adopted to be the performance index formulated as follows (Franklin, Powell, & Workman, 2010):

$$\mathrm{DIAE}_i = \sum_{k=k_0}^{k_f} |e_i^k|, \tag{2}$$

where k_0 and k_f are the initial and final times of the interval of interests, and e_i^k is the error of Plant i. For each individual plant, at various sampling frequencies, the DIAE will take a different value. Hence, a set of accumulated DIAEs of a plant over a stability range of sampling frequencies will imply the performance of an NCS and can be applied to find the optimal sampling frequency of the plant. Hereafter, the practical PIF will be defined as a piecewise function of the sampling frequency follows,

$$\bar{J}_i(f_i) = \sum_{k=k_0}^{k_f} |e_i^k(f_i)|. \tag{3}$$

Two approximations will be proposed to capture the trends of the practical PIF as in Eq. (3) so that the analytical optimal bandwidth allocation can be achieved.

2.3. Exponential approximation of PIF

The time delays depend on many aspects such as the data-packet size, number of packets in the network, network conditions, router's capacity, or unpredictable uncertainties, etc. Although the time delays affect the system performance, they are not directly controllable variables during a design of an NCS. However, by controlling the sampling frequency of each plant in an NCS, the number of data packets in the network can be maintained at a certain level so that the average of time delays can be controlled within a certain range. Therefore, for simplicity, one can assume that the effects of the time delays can be reflected by an increasing function of the sampling frequencies of the plants in an NCS. The details of the system performance versus the time delays can be referred to (Kim, 1998). Hence, from an NCS design perspective, the PIF that reveals the effects from a high sampling frequency can be revised as an increasing exponential function of the sampling frequency. Hereafter, from a traditional digital design perspective, $E_i(f_i^k, t)$ defines an approximated PIF of Plant i as a decreasing exponential function of the sampling frequency. Similarly, from an NCS design perspective, $F_i(f_i^k, t)$ defines an approximated PIF of Plant i as an increasing exponential function of the sampling frequency. Therefore,

$$E_i(f_i^k, t) = e^{-\beta_i {f_i}^k}, \tag{4}$$

and

$$F_i(f_i^k, t) = e^{\delta_i f_i^k}, \tag{5}$$

where β_i and δ_i are the approximation coefficients. These parameters can be obtained from simulation or experiments by a least-square fitting approach. Refer to Eqs. (36) and (38) in Section 4 as examples. Therefore, for each individual plant, the PIF can be defined as

$$\bar{J}_i \cong J_i = \sum_{k=0}^{M-1} (\alpha_i E_i(f_i^k, t) + \gamma_i F_i(f_i^k, t)) \tag{6}$$

where M is the maximum control iteration. The coefficients α_i and γ_i balance the impacts of the errors and time delays in the PIF of the corresponding plant.

For the entire NCS, the purpose of optimal bandwidth allocation is to minimize the PIF as

$$\begin{aligned} \min_{f \in \Omega} J &= \min_{f \in \Omega} \sum_{i=1}^{N} \omega_i^k J_i = \min_{f \in \Omega} \sum_{i=1}^{N} \sum_{k=0}^{M-1} \omega_i^k (\alpha_i E_i(f_i^k, t) + \gamma_i F_i(f_i^k, t)) \\ &= \min_{f \in \Omega} \sum_{i=1}^{N} \sum_{k=0}^{M-1} \omega_i^k (\alpha_i e^{-\beta_i {f_i}^k} + \gamma_i e^{\delta_i {f_i}^k}), \end{aligned} \tag{7}$$

$$\text{subject to} \quad \sum_{i=1}^{N} \tau_i^k {f_i}^k \le B, \quad \forall k = 1, \cdots, M, \tag{8}$$

where $0 \le B \le 1$ is the network bandwidth saturation threshold in the NCS, N is the number of plants, ω_i^k is the weight for the i-th plant at the k-th control iteration, and Ω is the set of sampling frequencies that maintains the stability of the plants. The selection of ω_i^k can be based on the system requirements. For example, the plant with the largest sampling frequency may indicate the difficulties in maintaining the stability and system performance and wins the largest weight. Furthermore, the PIF is a convex function of the sampling frequencies, and it is this convexity that allows for the optimal sampling frequency for a set of plants with appropriate convex optimization methodologies. The details of the optimal solution of the given exponential approximation will be illustrated in Section 3.1.

2.4. Quadratic approximation of PIF

In Section 4, one will see that the exponential approximation can closely approximate the practical system performance, but a closed-form optimal solution of Eqs. (7) and (8) is not easy to obtain in real time. Therefore, a quadratic approximation is proposed as a replacement of the exponential approximation. Even though a quadratic approximation can hardly capture all aspects of an NCS, it can still be applied when one wants to quickly evaluate the performance of a given NCS. The quadratic approximation has a simple closed-form optimal solution to Eqs. (10) and (11). For each individual plant, the PIF can be defined as

$$\bar{J}_i \cong J_i = \sum_{k=0}^{M-1} (a_i (f_i^k)^2 + b_i f_i^k + c_i), \tag{9}$$

where a_i, b_i, and c_i are the approximation coefficients. They can be obtained from simulation or experiments by a least-square fitting approach. Refer to Eqs. (37) and (39) in Section 4 as examples.

For the entire NCS, the objective function and the constraints could be formulated as

$$\min_{f \in \Omega} J = \min_{f \in \Omega} \sum_{i=1}^{N} \omega_i^k J_i = \min_{f \in \Omega} \sum_{i=1}^{N} \sum_{k=0}^{M-1} \omega_i^k (a_i (f_i^k)^2 + b_i f_i^k + c_i), \tag{10}$$

$$\text{subject to} \quad \sum_{i=1}^{N} \tau_i^k {f_i}^k \le B, \quad \forall k = 1, \cdots, M. \tag{11}$$

Note that this quadratic PIF is also a convex function of the sampling frequencies.

3. Optimal bandwidth allocation and scheduling

In this section, the optimal solution of the proposed exponential and quadratic approximations and the scheduling of the bandwidth assignment sequence of the plants are given. To facilitate the development, the following assumptions are made.

Assumption 1. The total time delay τ_i^k in Eq. (1) is a random variable by the nature of the network. For the simplicity of analysis and optimization, however, it is assumed to be a constant at $\bar{\tau}_i$ each sampling frequency f_i of Plant i, and an average value is used in this paper. Then the superscript k in all the approximations can be dropped, and $\bar{\tau}_i$ is the average time delay for Plant i.

Note that the average value of time delays is not assumed for the entire sampling frequency set but for each specific sampling frequency. For instance, we assumed average values of time delay τ_1 for f_1, τ_2 for f_2, and so on. Refer to (Heemels, Teel, Wouw, & Nešić, 2010) for details of the reason that an average value of time delays can be taken.

Assumption 2. All the plants can be scheduled at their minimum sampling frequency. That is, when $f_i = f_i^{\min}$, one can have $\sum_{i=1}^{N} \overline{\tau}_i f_i^{\min} < B$, where $f_i^{\min}$ is the minimum sampling frequency of Plant i. When all the plants are at their maximum BU (or maximum sampling frequency), the total BU of the entire NCS may or may not exceed the network bandwidth saturation threshold B.

An NCS could contain various plants that have different system specifications and requirements. These plants can be categorized into two groups, the one with variant sampling frequencies, and the one with fixed sampling frequencies. If an NCS includes both groups of plants, the bandwidth threshold B needs to be modified as $\hat{B} = B - \sum_{j \in \mathbb{J}} \overline{\tau}_j f_j$, where $\mathbb{J}$ is the set of the indices of the plants with fixed sampling frequencies. Then, for the rest of the controlled plants in the NCS with variant sampling frequencies, the new bandwidth threshold $\hat{B}$ will be used for the optimization purpose so that the objective function of Eqs. (7) or (10) can still be applied. Or if a certain percentage of the network bandwidth should be reserved for other functionalities, the newly defined $\hat{B}$ can also be applied such that $\hat{B} = B - \tilde{B}$, where $\tilde{B}$ is the reserved network bandwidth.

3.1. Optimal solution of exponential approximation

For the exponential approximation in Eqs. (7) and (8), the optimal solution will be given by the following theorem.

Theorem 1. *Given an NCS with N plants, and with the PIF in Eq. (7), an optimal solution, is given by*

$$f_i^* = f_i^{\min}, \quad i = 1, \cdots, p \tag{12}$$

$$f_j^* = g_j(\lambda), \quad j = p+1, \cdots, N \tag{13}$$

where p is the smallest index such that

$$\sum_{i=1}^{p} \overline{\tau}_i f_i^{\min} + \sum_{j=p+1}^{N} \overline{\tau}_j f_j^* \geq B, \tag{14}$$

and $g_j(\lambda)$ is the solution to

$$\Gamma_i e^{-\beta_i f_i} + \Phi_i e^{\delta_i f_i} + \lambda \overline{\tau}_i = 0, \tag{15}$$

where $\Gamma_i = -\omega_i \alpha_i \beta_i$ and $\Phi_i = \omega_i \gamma_i \delta_i$.

Proof. The KKT conditions (Griva, Nash, & Sofer, 2008) and the Lagrange multipliers λ, λ_{i1}, and λ_{i2} will be introduced. Then define the Lagrange equation as

$$L = \sum_{i=1}^{N} \omega_i (\alpha_i e^{-\beta_i f_i} + \gamma_i e^{\delta_i f_i}) + \lambda \left(\sum_{i=1}^{N} \overline{\tau}_i f_i - B\right) + \sum_{i=1}^{N} \lambda_{i1} (f_i^{\min} - f_i) + \sum_{i=1}^{N} \lambda_{i2} (f_i - f_i^{\max}). \tag{16}$$

Then from the KKT conditions, the dual feasibility of each plant is

$$\frac{\partial L}{\partial f_i} = \omega_i (-\alpha_i \beta_i e^{-\beta_i f_i} + \gamma_i \delta_i e^{\delta_i f_i}) + \lambda \overline{\tau}_i - \lambda_{i1} f_i + \lambda_{i2} f_i = 0, \tag{17}$$

the primal feasibilities are

$$\lambda \left(\sum_{i=1}^{N} \overline{\tau}_i f_i - B \right) = 0 \tag{18}$$

$$\lambda_{i1} (f_i^{\min} - f_i) = 0 \tag{19}$$

$$\lambda_{i2} (f_i - f_i^{\max}) = 0, \tag{20}$$

and the complementary slacknesses are

$$\lambda \geq 0 \tag{21}$$

$$\lambda_{i1} \geq 0 \tag{22}$$

$$\lambda_{i2} \geq 0, \tag{23}$$

where $i = 1, 2, \cdots, N$.

Based on Assumption 2, all the plants are given initially the minimum frequencies, $f_i = f_i^{\min}$, and there will be an idle network bandwidth available. For those plants, $\lambda_{i1} = 0$ and $\lambda_{i2} = 0$ based on the KKT conditions. Therefore from Eq. (17),

$$\Gamma_i e^{-\beta_i f_i} + \Phi_i e^{\delta_i f_i} + \lambda \tau_i = 0, \quad i = p+1, \cdots, N \tag{24}$$

where $\Gamma_i = -\omega_i \alpha_i \beta_i$ and $\Phi_i = \omega_i \gamma_i \delta_i$.

Solve f_i from Eq. (24) in terms of λ, defined as $f_j^* = g_j(\lambda)$ and then substitute f_i^* into Eq. (14) as

$$\sum_{i=1}^{p} \overline{\tau}_i f_i^{\min} + \sum_{j=p+1}^{N} \overline{\tau}_j g_j(\lambda) \geq B. \tag{25}$$

And solve for λ from Eq. (25), which yields the optimal solution of the exponential approximation.

Note that Eq. (24) is a transcendental equation, and a closed-form solution for f_i may not be easily obtained. □

3.2. Optimal solution of quadratic approximation

For the quadratic approximation in Eqs. (10) and (11), the optimal solution will be given as follows.

Theorem 2. *Given an NCS with N plants, and with the PIF in Eqs. (10) and (11), an optimal solution, is given by*

$$f_i^* = f_i^{\min}, \quad i = 1, \cdots, p \tag{26}$$

$$f_j^* = \frac{-\lambda \overline{\tau}_i - \omega_i b_i}{2\omega_i \alpha_i}, \quad j = p+1, \cdots, N \tag{27}$$

where p is the same as in Theorem 1 and

$$\lambda = \frac{\sum_{i=1}^{p} \overline{\tau}_i f_i^{\min} - \sum_{j=p+1}^{N} \overline{\tau}_j b_j / 2a_j - B}{\sum_{j=p+1}^{N} \overline{\tau}_j^2 / 2\omega_j a_j}. \tag{28}$$

Proof. The proof follows that of Theorem 1 with Eqs. (17) and (24) replaced by Eqs. (29) and (30), respectively.

$$\frac{\partial L}{\partial f_i} = \omega_i (2a_i f_i + b_i) + \lambda \tau_i - \lambda_{i1} f_i + \lambda_{i2} f_i = 0, \tag{29}$$

such that when $f_i^{\min} < f_i < f_i^{\max}$ for those plants, $\lambda_{i1} = 0$ and $\lambda_{i2} = 0$ based on the KKT conditions. Therefore from Eq. (29),

$$2\omega_i a_i f_i + \omega_i b_i + \lambda \tau_i = 0, \quad i = p+1, \cdots, N \tag{30}$$

such that

$$f_i = \frac{-\lambda \overline{\tau}_i - \omega_i b_i}{2\omega_i a_i}, \quad i = p+1, \cdots, N \tag{31}$$

Substitute Eq. (31) into Eq. (29), and solve for λ,

$$\sum_{i=1}^{p} \overline{\tau}_i f_i^{\min} + \sum_{j=p+1}^{N} \overline{\tau}_j \frac{-\lambda \overline{\tau}_j - \omega_j b_j}{2\omega_j a_j} = B, \tag{32}$$

and

$$\lambda = \frac{\sum_{i=1}^{p} \overline{\tau}_i f_i^{\min} - \sum_{j=p+1}^{N} \overline{\tau}_j b_j / 2a_j - B}{\sum_{j=p+1}^{N} \overline{\tau}_j^2 / 2\omega_j a_j}. \tag{33}$$

Note that Eqs. (12) and (13) and Eqs. (26) and (27) may vary depending on the selection of the weights ω_i^k and the approximation coefficients. □

3.3. Unique global optimal solution

Note that the exponential and quadratic functions are convex functions. The additional operation preserves the convexity of the functions. Hence, the two proposed approximations defined in Eqs. (7), (8), (10) and (11) are convex approximations. Also note that the feasibility of convex optimizations only depends on the constraints, not the objective function (Griva et al., 2008). If the constraints of a convex optimization are feasible, then there at least exists an optimum. Since the constraints of the proposed PIFs are stable sampling frequency sets, they are non-empty feasible sets. That being said, the approximations proposed in this paper guarantee the existences of optimal solutions. With the convexity of the approximations, the following theorem holds.

Theorem 3. *Given the two approximations in Eqs. (7), (8), (10) and (11), the optimal solutions in Eqs. (12), (13), (26) and (27) will be the unique global optimal solutions.*

Proof. Consider a convex optimization problem as follows:

$$\begin{aligned} &\text{minimize} \quad f(\mathbf{x}) \\ &\text{subject to} \quad \mathbf{h}(\mathbf{x}) = \mathbf{0} \\ &\qquad\qquad\quad \mathbf{g}(\mathbf{x}) \le \mathbf{0} \end{aligned}$$

and let $\mathbf{x}^* \in R^n$ be an existing local optimal solution. Assuming that the given convex optimization problem $f(\mathbf{x})$ is feasible, then there exists ε such that

$$f(\mathbf{x}^*) = \inf\{f(\mathbf{x}) : g_i(\mathbf{x}) \le 0,\ \ i = 1,\ \cdots,\ m; \quad h_j(\mathbf{x}) = 0,\ \ j = 1,\ \cdots,\ p; \\ ||\mathbf{x} - \mathbf{x}^*|| \le \varepsilon\} \tag{34}$$

Suppose that $\mathbf{x}^*$ is not globally optimal. Then there exists a feasible $\mathbf{y}$ so that $f(\mathbf{y}) < f(\mathbf{x}^*)$, which implies that $||\mathbf{y} - \mathbf{x}^*|| > \varepsilon$. Consider that a point $\mathbf{z}$ is given by

$$\mathbf{z} = (1-\theta)\mathbf{x}^* + \theta\mathbf{y}, \quad 0 < \theta = \frac{\varepsilon}{2||\mathbf{y} - \mathbf{x}^*||} < 1$$

Then $||\mathbf{z} - \mathbf{x}^*|| \le \varepsilon/2 < \varepsilon$ and by convexity of the objective function $f(\mathbf{x})$,

$$f(\mathbf{z}) \le (1-\theta)f(\mathbf{x}^*) + \theta f(\mathbf{y}) \le f(\mathbf{x}^*),$$

which contradicts Eq. (34). Therefore, if local optimal solutions of Eqs. (12), (13), (26) and (27) exist, they are also the unique global optimal solutions of the optimization problems, respectively. □

3.4. Time complexity of the optimizations

Note that the solution to the optimization problems in Eqs. (7), (8), (10) and (11) would eventually be converted to the solution to a transcendental Eq. (24) and a linear Eq. (30). This conversion was done by the derivative of the Lagrange equation of the KKT conditions. Therefore, the time complexity of those optimization problems will be bounded by the time complexity of the transcendental equation and the linear equation. For the linear equation, its time complexity is simply $O(n)$ (Spiser, 2012). For the transcendental equation, the following theorem gives its time complexity.

Theorem 4. *(Karatsuba, 1991): Let $y = f(x)$ be an elementary transcendental function that is an exponential function, a trigonometric function, or an elementary algebraic function, or their superposition, or their inverse, or a superposition of the inverses. Then*

$$s_f(n) = O(M(n)\log^2 n).$$

where $s_f(n)$ is the complexity of computation of the function $f(x)$ with accuracy up to n digits, $M(n)$ is the complexity of multiplication of two n-digit integers.

3.5. Scheduling algorithm

Scheduling of the NCSs consists of two parts: (i) priority assignment and plant arrangement and (ii) scheduling algorithm implementation. In general, the second part can be achieved by introducing the existing scheduling algorithms in the RTOS to the NCSs.

To be beneficial to the scheduling of an NCS, the changing rate of the system PIF in terms of the sampling frequency is defined as $U_i(f_i) = \partial J_i / \partial(\overline{\tau}_i f_i)$. Initially, the plants will be arranged by the following sequence (Seto et al., 2001) assuming that the NCS has sufficient bandwidth available for the initial bandwidth allocation under Assumption 2:

$$U_1(f_1^{\min}) \le U_2(f_2^{\min}) \le \cdots \le U_N(f_N^{\min}). \tag{35}$$

In this preferred sequence, by changing the same amount of the sampling frequency of each plant, Plant N will yield the largest change in PIF so that the performance of the NCS can be improved in the fastest rate. Consequently, Plant N should be first given the idle network bandwidth of the NCS if available. As long as sufficient bandwidth is available, the increment of BU for Plant N will continue until the moment that $U_N(f_N) = U_{N-1}(f_{N-1}^{\min})$. When $U_N(f_N) = U_{N-1}(f_{N-1}^{\min})$, the benefits from Plants N and $N-1$ will the same as the sampling frequency changes. Then BU of Plants N and $N-1$ will increase by maintaining $U_N(f_N) = U_{N-1}(f_{N-1})$ until the moment that $U_N(f_N) = U_{N-1}(f_{N-1}) = U_{N-2}(f_{N-2}^{\min})$. This process will continue until either the total available network bandwidth in the NCS is exhausted, or each plant reaches its maximum sampling frequency,$f_i = f_i^{\max}$, or its calculated optimal sampling frequency, $f_i = f_i^*$. If p clients are set to their minimum frequencies, where $1 \le p \le N-1$, then total available BU of $B - \sum_{i=1}^{p} b_i f_i^{\min}$ will be available for clients $p+1, \cdots, N$. A detailed definition of p can be referred to (Peng et al., 2009; Seto et al., 2001). Initially, $p = N$. The optimization program solves the first optimal sampling frequency and decreases p by 1. Then, the program checks Eq. (14) or Eq. (25). If the conditions are not satisfied, the program solves the second optimal sampling frequency, and so on. The optimization program stops until Eq. (14) or Eq. (25) is satisfied and returns all the optimal sampling frequencies.

The total available network bandwidth B of an NCS depends on the scheduling algorithms. There exist several scheduling algorithms for the RTOS that could also be implemented in the NCSs. However, there is a significant difference between RTOS scheduling and NCS scheduling. RTOS scheduling is able to put the lower-priority tasks into a preemptive status to guarantee the time performance of the higher-priority tasks. However, when a data packet is transmitted in the network, the controller will be unable to suspend the data packet into a pre-emptive status although there might be higher-priority tasks in the NCS. Therefore, only real-time non-preemptive scheduling algorithms can be applied to the NCSs such as non-preemptive Rate Monotonic (RM) and Earliest Deadline First (EDF) algorithms (Liu, 2000).

The network bandwidth saturation threshold B of a non-preemptive RM algorithm is defined as the ratio of the smallest task period over the largest task period in the system. In an NCS, the task period can be assumed to be equal to the sampling period of the controlled plant. For instance, arranging the sampling period of each plant in an ascending order, $h_1 \le h_2 \le \cdots \le h_N$, then $B = h_1/h_N$. The network bandwidth saturation threshold B of a non-preemptive EDF algorithm, by definition, is 1. The EDF scheduling algorithm was chosen since it guarantees a 100% resource availability. The non-preemptive RM may guarantee a 100% resource availability only under the circumstance of the universal sampling period of each plant (Park, 2007). Meanwhile, EDF itself is a dynamic scheduling algorithm that gives the tasks with the earliest deadline the highest priority to guarantee the entire performance of the system.

The algorithm of the proposed approximations of an NCS is illustrated in Fig. 4. In the beginning, the user will choose which system PIF and scheduling algorithm the NCS will be implementing. The NCS will then arrange all the plants by the preferred sequence Eq. (35). The controller will then either look up or calculate the optimal sampling frequency for the exponential approximation or the quadratic approximation, respectively. As discussed in Section 3.1, the closed-form analytic solution of the exponential approximation is difficult to calculate on-line. Hence, the optimal sampling frequencies will be calculated off-line and be tabulated in the controller for both the non-preemptive RM and EDF algorithms.

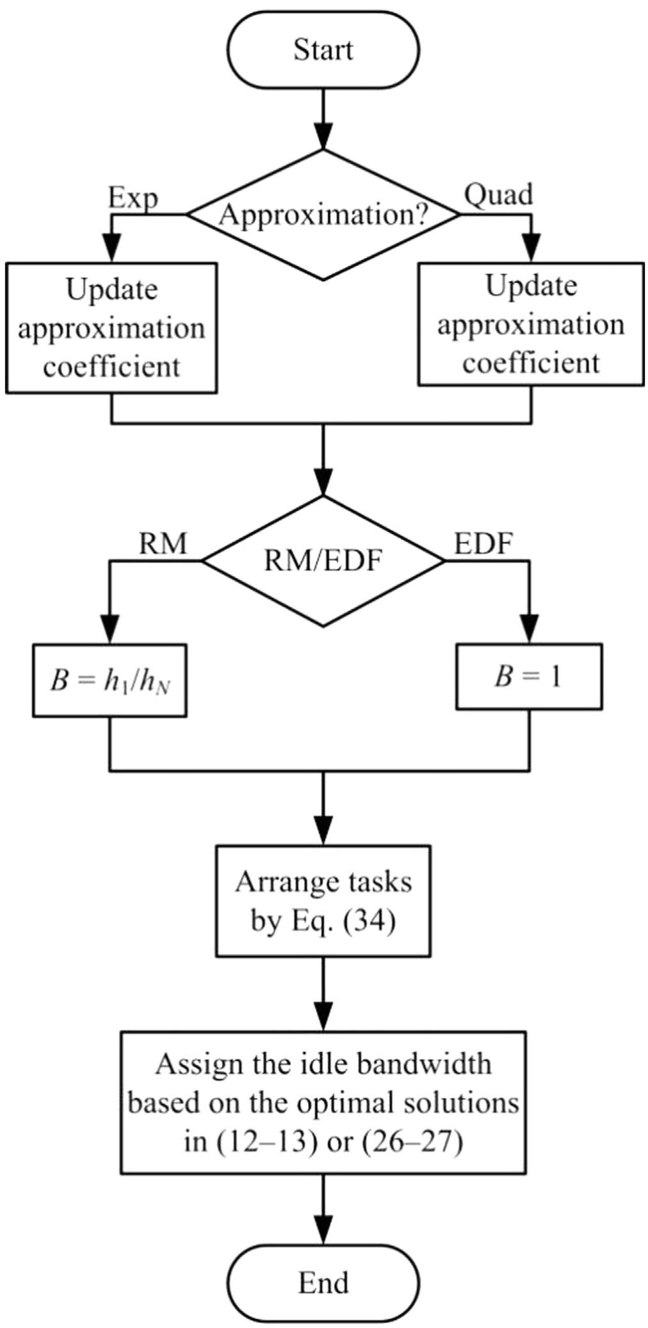

Fig. 4. Scheduling algorithm with the proposed approximations of the NCS.

4. Experimental results

In this section, the experimental results are presented to demonstrate the effectiveness of the proposed methods and their scheduling performances. Four DC-motor speed-control systems were set up as the test bed (Lee, 2009) for experimental verification as shown in Fig. 5.

The objective of this experiment is to control the speed of a DC motor over the Ethernet local area network (LAN) in our lab. The transfer function of the DC motor is

$$G(s) = \frac{20.2}{9.92s + 2.57}. \tag{36}$$

A proportional-integral (PI) controller was applied to control the DC motor as

$$D(s) = \frac{1.5s + 5}{s}. \tag{37}$$

The reference speed in the experiment was set to be 10 revolutions per second (rps). The minimum sampling frequency without causing instability of each DC motor is 65 Hz. The average of total time delays for the DC motor system is measured to be 1.360 ms.

4.1. Experimental setup

Linux Redhat 7.3 with Real-time Application Interface 3.4 (RTAI 3.4) is the OS running on the controller (Mantegazza, 2012), and Linux Ubuntu 6.10 with RTAI 3.4, on the plants. The control and measurement device interface (Comedi) is used as the drivers and libraries of data acquisition on the plants (Schleef, 2012). A PCI-6221 data-acquisition card by National Instruments enables the DC motor test bed to send out sensor data and receive control data through the LAN. The speed control is achieved by controlling the output voltage of a pulse-width modulation (PWM) amplifier. Fig. 6 shows the block diagram of the entire experimental setup. The communication network in the experiment is a 100-Mbps Ethernet with unblocked User Datagram Protocol (UDP) sockets.

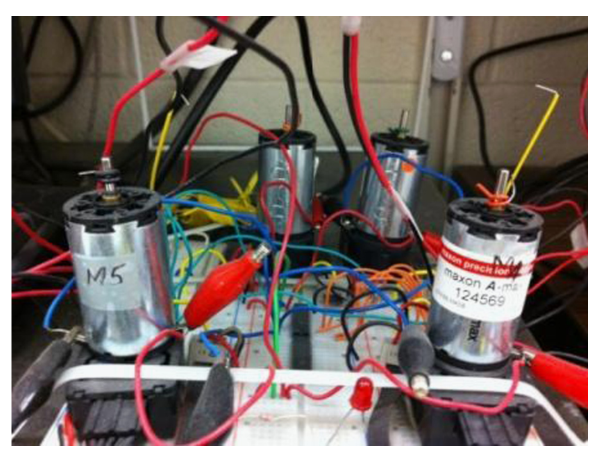

Fig. 5. DC motor speed-control systems.

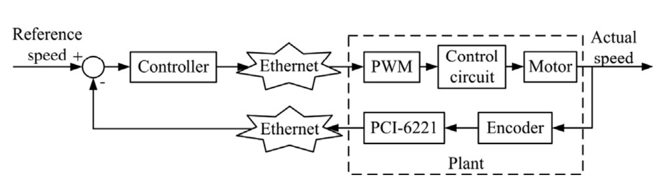

Fig. 6. Block diagram of DC motor system.

Please cite this article as: Dong, J., & Kim, W.-J. Bandwidth allocation and scheduling of networked control systems with exponential and quadratic approximations. *Control Engineering Practice* (2014), http://dx.doi.org/10.1016/j.conengprac.2014.01.001

4.2. Experimental results without reserved bandwidth

The relation of the DIAE versus the sampling frequencies of the experiments and exponential/quadratic approximations is given in Fig. 7. Five sets of experiments were conducted under the same network conditions. In each individual sampling period, the experiments ran for 20,000 control iterations. Fig. 7 shows the experimental results of a single DC motor and its exponential and quadratic approximations. From Fig. 7, the DIAE of the DC motor was quite large in the lower sampling-frequency range because the DC motor could not have adequately frequent control inputs from the controller to maintain the system performance. As the sampling frequency increased, the DIAE of the DC motor decreased. In the higher sampling-frequency range, the DIAE of the DC motor increased again because the number of data packets in the network increased as the sampling frequency kept increasing. The large number of the data packets would bring longer time delays into the NCS or even packet losses so that the performance of the DC motor could be degraded. Note that there are a few discrepancies among the five sets of experiments in the high sampling-frequency range. Because of the large number of the data packets in the network in the high sampling-frequency range, any possible disturbances or irrelevant data-packet transmissions from other network users may cause the NCS performance to be degraded. Therefore, the NCS is more sensitive at high sampling frequencies. Both the transient and steady-state system responses of the NCS were considered when calculating the integral absolute errors.

The PIFs with exponential and quadratic approximations are

$$J_i = 8781e^{-0.05f_i} + 0.001e^{0.0556f_i}, \tag{38}$$

and

$$J_i = 0.1316f_i^2 - 67.1258f_i + 7941.49, \tag{39}$$

respectively.

From the optimization programming, $p=1$ for both the exponential and quadratic approximations, respectively. Therefore, based on Theorems 1 and 2, the optimal sampling frequency of each DC motor is given in Table 1.

Fig. 8 shows the profile of the sampling-frequency and the BU changes for each DC motor during the experiments. Here, a linear changing rate was adopted for the sampling frequencies. From Table 2 and Eq. (1), the total BUs of the exponential and quadratic approximations are 98.98% and 95.66%, respectively. The sampling frequency were adjusted automatically by $f = cf$, where c is a constant rate of change. In Fig. 8, as the control iteration increased, the sampling frequency of each DC motor approached their optimal values listed in Table 2. The total BU of the experiments may not be exactly the same as the ones calculated from Table 1. However, it will be close to the optimal value eventually. From Fig. 8, the final total BU is 98.26%.

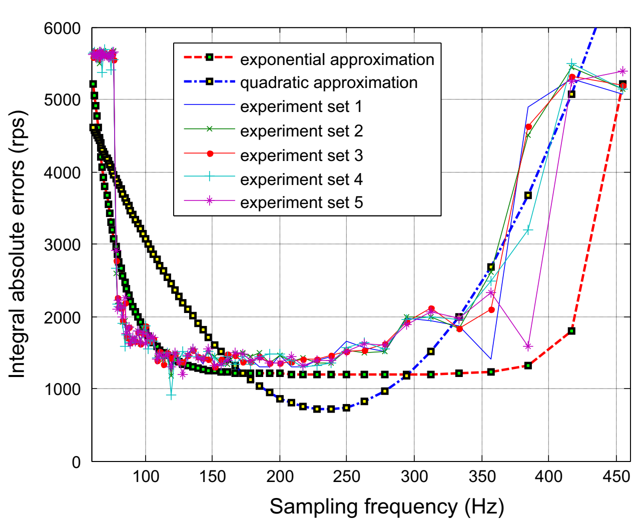

Fig. 7. DIAE versus sampling frequencies of the experiments and exponential and quadratic approximations.

Table 1
Optimal sampling frequencies.

Number of DC motors	ω_i	Exponential approximation [Hz]	Quadratic approximation [Hz]
1	1	65	65
2	2	173.5	162.7
3	4	237.1	225.8
4	5	252.2	249.9

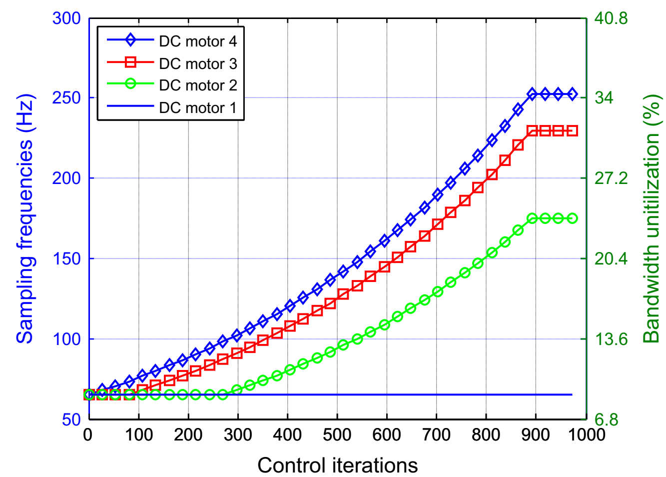

Fig. 8. Profile of the sampling-frequency and BU changes for each DC motor during the experiments.

Table 2
Optimal sampling frequencies.

Number of DC motors	ω_i	Exponential approximation [Hz]	Quadratic approximation [Hz]
1	1	65	65
2	2	65	65
3	4	117.68	119.35
4	5	151.35	148.02

4.3. Experimental results with reserved bandwidth

In the next experiment, a ball magnetic-levitation (maglev) system was introduced in to the NCS as Plant 5. It has a fixed sampling frequency of 333 Hz as shown in Fig. 9. The average of total time delays τ_5 is measured to be 1.350 ms. From Eq. (1), the BU of the ball maglev system is $b_5 = \overline{\tau}_5 f_5 = (1.350 \times 10^{-3}\ \text{s}) \times (333\ \text{Hz}) = 44.96\%$. The total time delay of the ball maglev system and the DC motor are very close because they are tested under the same network conditions and have the same size of data packets. With $\hat{B} = B - \sum_{j \in \mathbb{J}} \overline{\tau}_j f_j = 1 - 0.4496 = 55.04\%$, where $\mathbb{J} = \{5\}$, $p=2$ for the exponential approximation, and the quadratic approximation, respectively. The optimal sampling frequency of each DC motor is given in Table 2.

Fig. 10 shows the profile of the sampling-frequency and the BU changes for each DC motor during the experiments with the ball

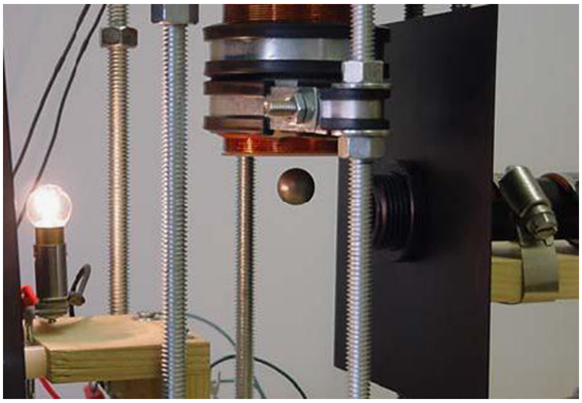

Fig. 9. Ball maglev system.

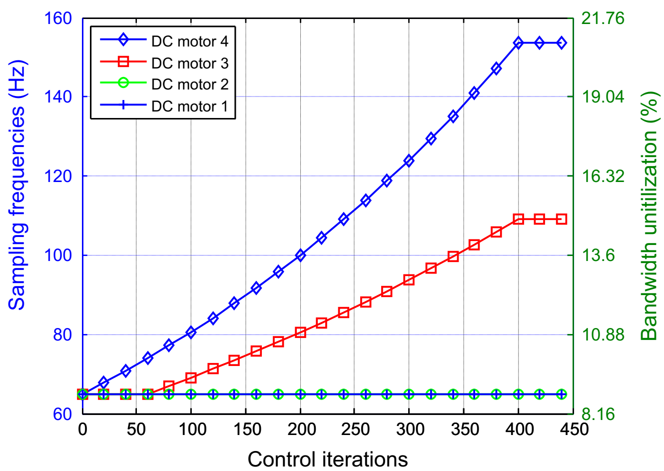

Fig. 10. Profile of the sampling-frequency and BU changes for each DC motor during the experiments with the ball maglev system.

Table 3
Statistic comparison of the exponential and quadratic approximations.

Experiments	Exponential approximation	Quadratic approximation
Mean	395.42	1004.5
Std dev	125.33	526.96

maglev system as a plant that had a fixed sampling frequency. From Table 2 and Eq. (1), the total BUs of the exponential and quadratic approximations are 54.26% and 54.04%, respectively, with the ball maglev system taking approximately 45% of the total BU. Similarly, the total BU of the experiments may not be exactly the same as the ones calculated from Table 3. However, it will be close to the optimal value eventually. From Fig. 10, the final total BU is 53.45%. Note that DC motors 1 and 2 have the same optimal sampling frequencies as in Table 2, so they are overlapped in Fig. 10.

From the experiments, one can see that the exponential approximation represents the practices more closely compared to the quadratic approximation in the sense of the system performance. However, the quadratic approximation takes less computational efforts to solve the optimization objective function and has a closed-form optimal solution. Both the exponential and quadratic approximations could find the optimal sampling frequencies that exhaust about 98% of the total network bandwidth available to the NCS with or without the fixed sampling-frequency plant, the ball maglev system in our experiments.

Statistic comparisons of the exponential and quadratic approximations are given in Table 3. These statistical values are based on the experimental results in Fig. 7. From Table 3, the accuracy of the exponential approximation is about 60% better than that of the quadratic approximation for experiments. Although the exponential approximation is more accurate when capturing the system performance of an NCS, it does not have an analytic closed-form optimal solution that can be implemented on-line. In contrast to the exponential approximation, the quadratic one has a closed-form optimal solution to the objective PIF that can be implemented on-line. Hence, the algorithm of bandwidth allocation and scheduling with quadratic approximation has a better efficiency in real-time operation. However, the accuracy of the quadratic approximation is less than that of the exponential approximation. Moreover, the exponential approximation gives an explicit measurement of the effects of high sampling frequency on the NCS whereas the quadratic approximation only indicates a coupled performance measurement of the NCS.

5. Conclusions

In this paper, the optimal bandwidth allocation and scheduling of NCSs was investigated. The BU of each plant was defined in terms of its sampling frequency. Two nonlinear approximations, exponential and quadratic, were formulated to describe the system performance governed by the DIAE versus the sampling frequencies. Based on the convexity of the proposed approximations, the optimal solution could be obtained from a nonlinear-programming perspective. The optimal sampling frequencies were obtained by solving the approximations with the KKT conditions. Within various network bandwidth saturation thresholds based on different real-time scheduling algorithms, the proposed approximations could find the optimal BU for each plant in the NCS. Later in the paper, experimental results verified the effectiveness of the proposed approximation models. In the experiments, the total BU of the NCS could approach up to 98% of the total available network bandwidth. Therefore, the proposed approximations and the scheduling algorithms can maximize the BU so that the plants can be scheduled along with the system PIFs being optimized.

Acknowledgments

This work was supported in part by the TAMU Program to Enhance Scholarly and Creative Activities under Grant no. 2010-SAC-8779.

References

Al-Hammouri, A. T., Branicky, M. S., Liberatore, V., & Phillips, S. M. (2006). Decentralized and dynamic bandwidth allocation in networked control systems. In *Proceedings of the 20th International Parallel and Distributed Processing Symposium, (IPDPS'06)*. pp. 25–29.

Arzen, K. E., Cervin, A., Eker, J., & Sha, L. (2000). An introduction to control and scheduling co-design. *IEEE Conference on Decision and Control*, *5*, 4865–4870.

Belzarena, P., Ferragut, A., & Paganini, F. (2009). Network bandwidth allocation via distributed auctions with time reservations. In *Proceedings of the IEEE INFOCOM'09*. pp. 2816–2820.

Branicky, M. S., Phillips, S. M., & Zhang, W. (2002). Scheduling and feedback co-design for networked control systems. In *Proceedings of 41st IEEE conference on decision and control. Vol.* 2, pp. 1211–1217.

Franklin, G. F., Powell, J. D., & Workman, M. L. (2010). *Digital control of dynamic systems* (6th ed.). MA: Addison-Wesley, Inc.

Griva, I., Nash, S. G., & Sofer, A. (2008). *Linear and nonlinear optimization* (2nd ed.). PA: Society for Industrial and Applied Mathematics.

Heemels, W. P. M. H., Teel, A. R., Wouw, N. V. D., & Nešić, D. (2010). Networked control systems with communication constraints: Tradeoffs between transmission intervals, delays and performance. *IEEE Transactions on Automatic Control*, *55*(8), 1781–1796.

Karatsuba, E. A. (1991). Fast evaluation of transcendental functions. *Problemy Peredachi Informatsii*, *27*(4), 76–99.

Kim, B. K. (1998). Task scheduling with feedback latency for real-time control systems. In *Proceedings of the 5th international conference on Real-Time Computing Systems and Applications (RTCAS)*. pp. 37–41.

Lee, M. H. (2009). *Real-time networked control with multiple clients (M.S. thesis)*. Texas A&M University.

Lian, F.-L. (2001). *Analysis, design, modeling, and control of networked control systems (Ph.D. dissertation)*. University of Michigan.

Lin, Q. (2004). *Dynamic scheduling of real-time control systems (M.S. thesis)*. National University of Singapore, Singapore.

Lin, Y. -C., & Lian, F. -L. (2012). Data reduction and bandwidth allocation for video-based network system. In *Proceedings of the IEEE international conference on information and automation*. pp. 116–121.

Liu, J. W. S. (2000). *Real-time systems*. NJ: Prentice Hall.

Mantegazza, P. (2012). DIAPM RTAI—real-time application. [Online] Available at: ⟨http://www.rtai.org⟩.

Park, H. S., Kim, Y. H., Kim, D.-S., & Kwon, W. H. (2002). A scheduling method for network-based control systems. *IEEE Transactions on Control Systems Technology*, *10*(3), 318–330.

Park, M. (2007). Non-preemptive fixed priority scheduling of hard real-time periodic tasks. In *Proceedings of the 7th international conference on computational science*. pp. 882–888.

Peng, C., Yue, D., Gu, Z., & Xia, F. (2009). Sampling period scheduling of networked control systems with multiple-control loops. *Mathematics and Computer in Simulation*, *79*(5), 1502–1511.

Schleef, D. (2012). Linux control and measurement device interface. [Online] Available at: ⟨http://www.comedi.org⟩.

Seto, D., Lehoczky, J. P., Sha, L., & Shin, K. G. (2001). Trade-off analysis of real-time control performance and schedulability. *Real-Time Systems*, *21*(3), 199–217.

Spiser, M. (2012). *Introduction of the theory of computation* (3rd ed.). KY: Cengage Learning.

Wang, J., & Liu, W. -D. (2011). Control and scheduling co-design of networked control system. In *Proceedings of the IEEE Conference on Signal Processing, Communications and Computing (ICSPCC)*. pp. 1–4.

Zhang, L., Gao, H., & Kaynak, O. (2013). Network-induced constraints in networked control systems—A survey. *IEEE Transactions on Industrial Informatics*, *9*(1), 403–416.