

QoSの公平性を考慮した リアルタイムシステムのリソース配分制御 Resource Allocation Control in Real-Time Systems Based on Fairness of QoS

大阪大学大学院基礎工学研究科 原田 史子, 潮 俊光, NEC システムプラットフォーム研究所 中本 幸一
Fumiko HARADA, Toshimitsu USHIO (Osaka University),
Yukikazu NAKAMOTO (NEC System Platforms Research Laboratories)

Abstract Recently, QoS adaptation control has been studied for avoiding overload situations in real-time systems. In QoS adaptation control, unfairness of QoS can occur though resource constraints are satisfied. We have introduced a resource allocation control method to achieve fair QoS levels in the task set under the guarantee of desired CPU utilization factor. In this paper, we extend this control method in order to achieve fairness of QoS levels in multiple resource real-time systems. The proposed adaptation controller allocates resources to each task with on-line search for the fair QoS level based on errors between the current QoS levels and their average. So, its implementation is easy and an overhead is small.

1 はじめに

リアルタイムシステム上において過負荷状態, リソース不足は予測性の低下やシステム性能の劣化を引き起こす. 過負荷状態時のスケジューリングとして, さまざまな QoS 適応制御手法が提案されている [1, 2]. 一方, リアルタイムシステムの性能を保証するために制御理論を適用する研究がなされている [3, 4]. 筆者らは QoS 適応制御手法のアプローチの一つとして, CPU 利用率を指定された値に保ちながら, 全てのタスクの QoS レベルを公平化する QoS 適応制御を提案した [5]. 本報告では, この QoS 公平化制御を拡張し, システムリソースが複数ある環境下で, QoS レベルを公平化するリソース配分制御を提案する.

2 QoS の公平性

n 種類のタスク集合 $\{\tau_1, \tau_2, \dots, \tau_n\}$ と m 種類のリソース集合 $\{\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_m\}$ からなるリアルタイムシステムを考える. リソース \mathcal{R}_j の総量は R_j とする. タスク τ_i は最大 QoS 要求 QoS_i^{\max} と最小 QoS 要求 QoS_i^{\min} を持ち, 実行時の QoS は $QoS_i \in [QoS_i^{\min}, QoS_i^{\max}]$ を満足しなければならない. τ_i が QoS_i の QoS で実行されるために必要なリソース \mathcal{R}_j の量 r_{ij} が一意に決まると仮定する. 全てのタスクの QoS 要求が満たされていても, τ_1 が最大 QoS レベル QoS_1^{\max} で実行される一方, τ_2 が最小 QoS レベル QoS_2^{\min} でしか実行されないとき, タスクの実行結果が偏り, 不公平性が生じる. このとき

$$Q_i := \frac{QoS_i - QoS_i^{\min}}{QoS_i^{\max} - QoS_i^{\min}}, \quad i = 1, \dots, n \quad (1)$$

とおくと, Q_i は QoS_i が最大 QoS レベルのとき 1, 最小 QoS レベルのとき 0 になるような正規化された QoS レベルで, 最大 QoS 要求に対する実際の QoS の満足度を示す. もし

$$Q_1 = Q_2 = \dots = Q_n \quad (2)$$

を満たすようにリソース配分 $\{r_{11}, \dots, r_{nm}\}$ を決定できれば, 全てのタスクの QoS の満足度が等しくなり, 公平なリソース配分が達成できる. 本報告では, 上式を保証するようなリソース配分を動的に決定する制御を提案する. 以降, 正規化された QoS レベル Q_i を単に QoS レベルと表記する.

3 マルチリソース環境下の QoS 適応制御

総量が $\{R_1, \dots, R_m\}$ のリソース集合 $\{\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_m\}$ に対して QoS レベルを公平化する手法を提案する. 図 1 に示すようなリソース配分制御系を考える. τ_i が QoS レベル Q_i を得るのに必要なリソース \mathcal{R}_j の量 r_{ij} は

$$r_{ij} = \phi_{ij}(Q_{ij}) \quad (3)$$

で一意に決まるとする. 本報告では ϕ_{ij} に以下の仮定をおく.

- ϕ_{ij} は微分同相な単調増加関数である.
- $\phi_{ij}(0) = 0, \phi_{ij}(1) \leq R_j$
- $0 < h_{ij} \leq \frac{d\phi_{ij}}{dQ_i} \leq H_{ij} < \infty$

$m > 1$ の場合, 配分されたリソースが全て消費されるとは限らない. 例えば \mathcal{R}_1 の配分量 r_{i1} が $\phi_{i1}(1) \leq r_{i1}$ である一方, \mathcal{R}_2 の配分量 r_{i2} が $\phi_{i2}(0) = r_{i2}$ を満たす量であれば, 実際に実行されるときは QoS レベル 0 になるためである. リソース配分 $\{r_{i1}, \dots, r_{im}\}$ に対して, 実際に達成できる QoS レベル Q_i とタスクの実行に実際に使われるリソース量 r_{ij}^{act} は次の関係を持つ.

$$Q_i = \min_j \phi_{ij}^{-1}(r_i), \quad r_{ij}^{\text{act}} = \phi_{ij}(Q_i) \leq r_{ij}(k) \quad (4)$$

$$\forall i, \exists j, \text{ s.t. } r_{ij}(k) = r_{ij}^{\text{act}}(k) \quad (5)$$

リソース制約を満たすような公平なリソース配分は無数に存在する. タスクがなるべく高い QoS で実行されることが望ましいため, QoS レベルが公平であるという制約下で, QoS レベルを最大化する必要がある. またリソース使用の効率性から, $r_{ij} = r_{ij}^{\text{act}}$ となるのが望ましい. よって, 次式を満たすリソース配分 r_{ij}^e を決定すれば, 公平性とリソース制約を保証する最大の QoS レベル Q^e を達成できる.

$$\phi_{ij}(r_{ij}^e) = Q^e, \quad i = 1, \dots, n, j = 1, \dots, m \quad (6)$$

$$r_{ij}^{\text{act}} = r_{ij}^e, \quad i = 1, \dots, n, j = 1, \dots, m \quad (7)$$

$$\exists j_u \text{ s.t. } \sum_{i=1}^n r_{ij_u}^e = R_{j_u}, \quad \sum_{i=1}^n r_{ij}^e \leq R_j, \quad j \neq j_u \quad (8)$$

本報告では上式を達成するために次式による制御側を提案する .

$$\lambda(k) = \max_q \left(\frac{\sum_{p=1}^n r_{pq}^{\text{act}}(k)}{R_q} \right) \quad (9)$$

$$r_{ij}(k+1) = \left(1 + \alpha \lambda(k) (1 - Q_i(k)) \right) r_{ij}^{\text{act}}(k) + \beta h_j (\bar{Q}(k) - Q_i(k)) \quad (10)$$

但し

$$\bar{Q}(k) := \frac{1}{n} \sum_{i=1}^n Q_i(k) \quad (11)$$

$$h_j := \min_i h_{ij} \quad (12)$$

とおく . この式は次のように解釈できる . $Q_i(k)$ が平均の QoS より小さければリソース配分量を増やし , 大きければ少なくすることで $Q_1(k), \dots, Q_n(k)$ のばらつきを小さくする . また使われないリソース量に応じて , QoS レベルの小さいタスクほど大きく配分量を増やす .

制御目的を達成するためには , $Q_i(k)$ が Q^e に収束するだけではなく , $r_{ij}(k)$ がリソース制約を満たさなければならない . これに関して次の定理が成り立つ .

定理 1

$$\sum_{i=1}^n r_{ij}(0) \leq R_j, 0 \leq r_{ij}(0) \leq R_j \quad (13)$$

$$\alpha \leq 1 \quad (14)$$

$$0 < \alpha \cdot \max \frac{H_{ij}}{h_j} \cdot \max \frac{H_{ij}}{R_j} \cdot \min \frac{nR_j}{\sum_{i=1}^n h_{ij}} \leq \beta \leq 1 \quad (15)$$

ならば次が成り立つ .

$$\sum_{i=1}^n r_{ij}(k) \leq R_j, 0 \leq r_{ij}(k) \leq R_j, k = 1, 2, \dots \quad (16)$$

$$\lim_{k \rightarrow \infty} \lambda(k) = 1 \quad (17)$$

$$\lim_{k \rightarrow \infty} (\bar{Q}(k) - Q_i(k)) = 0, i = 1, \dots, n \quad (18)$$

これは $r_{ij}(k) \in [0, R_j]$ と $\sum_{i=1}^n r_{ij}(k) \leq R_j$ がスケジュール可能性とリソース制約を保証し , そのもとで最大の公平な QoS レベルが達成できることを意味する .

提案手法を適用して , 三種類のリソースと 6 タスクのシステムの QoS レベル公平化のシミュレーション実験を行った . 繰り返し回数に対する $Q_i(k)$ の変動は図 2 のようになり , QoS レベルが公平化されていることがわかる . また , $\sum_{i=1}^n r_{i2}^{\text{act}}(k)$ は図 3 のように変化し , QoS レベルが最大になっていることが分かる .

4 おわりに

本報告では , リアルタイムタスクの QoS レベルを公平化する制御手法を提案し , 公平化を達成する十分条件を示した . しかしながら示した十分条件は保守的であり , 特にタスク間の h_{ij} の差が大きい場合 , α のとり得る値が著しく小さくなるために収束性が悪くなり得る . 今後の課題はこの十分条件をより緩いものにするここと , 提案手法の実装が挙げられる .

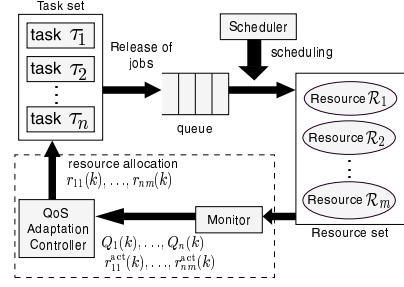


図 1: リソース配分制御システム

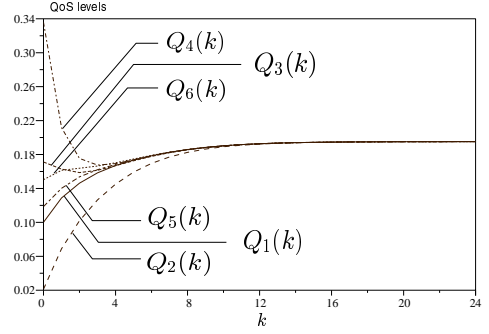


図 2: QoS レベルの変動

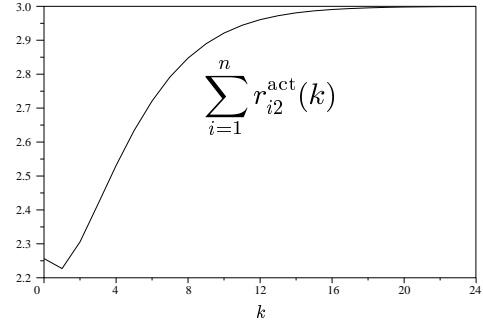


図 3: リソース R_2 の総消費量の変動

参考文献

- [1] G. Buttazzo and L. Abeni, "Adaptive Workload Management Through Elastic Scheduling," Real-Time Systems, vol.23, No. 1, pp.7-24, 2002.
- [2] R. Rajkumar, C. Lee, J. Lehoczky, and D. Siewiorek, "A Resource Allocation Model for QoS Management," Proc. 18th IEEE Real-Time Systems Symposium, pp.298-307, December 1997.
- [3] C. Lu, J. Stankovic, S. Son, and G. Tao, "Feedback Control Real-Time Scheduling: Framework, Modeling, and Algorithms," Real-Time Systems, vol.23, No.1, pp.85-126, 2002.
- [4] L. Palopoli, L. Abeni, and G. Lipari, "On the Application of Hybrid Control to CPU Reservations," HSCC 2003, LNCS 2623 pp.389-404, 2003.
- [5] 原田, 潮, 中本, "QoS レベル公平化に基づくリアルタイムシステムの QoS 適応制御," 信学論, vol.J87-D-I, no.3 pp.364-371, March 2004.