What is the Exact Speedup Factor for Fixed Priority Pre-emptive versus Fixed Priority Non-pre-emptive Scheduling?

Robert I. Davis\textsuperscript{1}, Oliver Gettings\textsuperscript{1}, Abhilash Thekkilakattil\textsuperscript{2} Radu Dobrin\textsuperscript{2}, Sasikumar Punnekkat\textsuperscript{2}
\textsuperscript{1}Real-Time Systems Research Group, Department of Computer Science, University of York, York, UK.
\textsuperscript{2}Malärdalen Real-Time Research Center, Malärdalen University, Sweden.

robert.davis@york.ac.uk, oliver.gettings@york.ac.uk, abhilash.thekkilakattil, radu.dobrin, sasikumar.punnekkat {at} mdh.se

Background

The performance of real-time scheduling algorithms can be compared in a number of different ways. Empirical techniques typically rely on generating a large number of task sets with parameters chosen from some appropriate distributions. The performance of the scheduling algorithms is then compared by determining task set schedulability according to exact or sometimes sufficient schedulability tests and plotting a graph of the success ratio (i.e. the proportion of task sets that are deemed schedulable) at different utilisation levels. An alternative theoretical method of comparing real-time scheduling algorithms is to determine the resource augmentation bound or speedup factor \([6]\) required. This approach focuses on those task sets that are particularly difficult to schedule using one algorithm but easy to schedule using another.

The Speedup Factor \(S(A,B)\) comparing two real-time scheduling algorithms \(A\) and \(B\) is given by the minimum factor by which the speed of the processor needs to be increased to ensure that any task set that is schedulable according to algorithm \(B\) is guaranteed to be schedulable by algorithm \(A\). When comparison is made against an optimal algorithm (OPT), then \(S(A,OPT)\) is referred to as the sub-optimality of algorithm \(A\).

Combining the utilisation bounds for fixed priority pre-emptive (FP-P) and EDF pre-emptive (EDF-P) scheduling from the seminal paper of Liu and Layland \([8]\) shows that the speedup factor \(S(FP-P, EDF-P) = 1/\ln(2) \approx 1.44270\) for implicit deadline task sets. Since EDF-P is an optimal uniprocessor scheduling algorithm \([4]\) this result also determines the sub-optimality of FP-P for implicit deadline task sets. In 2009, Davis et al. \([1]\) derived the exact sub-optimality of FP-P for constrained-deadline task sets; \(S(FP-P, EDF-P) = 1/\Omega \approx 1.76322\) (where \(\Omega\) is the mathematical constant defined by \(\ln(1/\Omega) = \Omega\), hence, \(\Omega \approx 0.567143\)). In 2009, Davis et al. \([2]\) gave upper and lower bounds of \(S(FP-P, EDF-P) = 2\) and \(1/\Omega \approx 1.76322\) for the case of arbitrary deadline task sets. In 2010 \([3]\) they gave upper and lower bounds for the non-pre-emptive case of \(S(FP-NP, EDF-NP) = 2\) and \(1/\Omega \approx 1.76322\) for implicit, constrained, and arbitrary deadline task sets. In 2013, Thekkilakattil et al. \([9]\) provided an upper bound on the sub-optimality of non-pre-emptive EDF (i.e. \(S(EDF-NP, EDF-P)\) parameterised by the shortest task deadline and the longest task execution time.

Open Problem: Speedup Factor for FP-P v. FP-NP

EDF-P is an optimal uniprocessor scheduling algorithm \([4]\) and so dominates EDF-NP, FP-P, and FP-NP. Further, EDF-NP dominates FP-NP \([5]\); however, there are no such dominance relationships between pre-emptive and non-pre-emptive fixed priority scheduling. Hence there are non-trivial speedup factors in both directions: \(S(FP-P, FP-NP)\) and \(S(FP-NP, FP-P)\). An exact speedup factor for the latter problem has recently been derived by the authors and is under submission. The former problem, determining the exact speedup factor needed so any task set that is schedulable under fixed priority non-pre-emptive scheduling is guaranteed to be schedulable under fixed priority pre-emptive scheduling is to the best of our knowledge open and forms the focus of this abstract.

Upper bounds on \(S(FP-P, FP-NP)\): since EDF-P dominates FP-NP, then the exact values of \(1/\ln(2) \approx 1.44270\), \(1/\Omega \approx 1.76322\), and 2 for \(S(FP-P, EDF-P)\) provide upper bounds on \(S(FP-P, FP-NP)\) for implicit, constrained and arbitrary deadline task sets respectively.

Theorem 1: A lower bound on \(S(FP-P, FP-NP)\) for constrained or arbitrary-deadline task sets is \(S = \sqrt{2}\).

Proof: Consider the following task set scheduled on a processor of speed 1.
\[
\begin{align*}
\tau_A & : C_A = \sqrt{2} - 1, D_A = 1, T_A = 1, \\
\tau_B & : C_B = (2 - \sqrt{2})/2, D_B = \sqrt{2}, T_B = \infty, \\
\tau_C & : C_C = (2 - \sqrt{2})/2, D_C = \sqrt{2}, T_C = \infty
\end{align*}
\]

As the task set has constrained deadlines, then under FP-P scheduling, Deadline Monotonic Priority Order (DMPO) \([7]\) is optimal. It is easy to see that the task set is only just schedulable with this priority ordering, since any increase in execution times would cause task \(\tau_C\) to miss its deadline — see Figure 1. Next consider the same task set scheduled under FP-NP on a processor of lower speed: \(f = \sqrt{2} (2 - \epsilon)\) where \(\epsilon\) takes an infinitesimally small value. Again assume the task priorities are in DMPO. Now the sum of the execution times of task \(\tau_A\) and task \(\tau_B\) (or task \(\tau_A\) and \(\tau_C\)) can be expressed as follows:
\[
2(\sqrt{2} - 1)(2 - \epsilon)/2\sqrt{2} + (2 - \sqrt{2})(2 - \epsilon)/2\sqrt{2} = (2 - \epsilon)/2 < 1
\]
Hence task $\tau_A$ is schedulable when blocked by either of tasks $\tau_B$ or $\tau_C$. Further, once the first jobs of tasks $\tau_A$ and $\tau_B$ have executed, the first job of task $\tau_C$ is able to start executing before the second job of task $\tau_A$ is released. Thus the first job of task $\tau_C$ has a response time of:

$$\frac{\sqrt{2}}{2} + \frac{(2-\varepsilon)}{\sqrt{2}} + \frac{\sqrt{2}}{2} = \frac{2}{\sqrt{2}} - \frac{\sqrt{2}}{2}$$

and so meets its deadline at $\frac{\sqrt{2}}{2}$. Continuing on through the busy period, the second job of task $\tau_A$ completes at: $\varepsilon + 2 < 2$ meeting its deadline. At this point the busy period ends. Thus we have shown that all of the tasks are schedulable. Task $\tau_A$ has a worst-case response time of $\frac{\sqrt{2}}{2}$ and tasks $\tau_B$ and $\tau_C$ have worst-case response times of $2 - \varepsilon < \sqrt{2}$. Hence the task set is schedulable at speed $f$ under FP-NP scheduling. The speedup factor required such that this task set is schedulable under FP-P therefore tends to 2 as $\varepsilon$ tends to zero □

Figure 1: FP-P schedule

Figure 2: FP-NP schedule

Figure 3: Empirical results for $S(FP-P, FP-NP)$

References