Priority Tasks Allocation through the Maximum Entropy Principle

Miguel A. Toledo C.
Instituto Politécnico Nacional
CIC
Unidad Profesional Adolfo López Mateos
Col. Lindavista, C.P. 07738
MEXICO

J. Jesús Medel J.
Instituto Politécnico Nacional
CIC
Unidad Profesional Adolfo López Mateos
Col. Lindavista, C.P. 07738
MEXICO

Abstract: Priority scheduling assigns priorities via some task policy. In this work we introduce a priority tasks allocation model for soft-real-time systems. Task entities are analyzed as continuous stochastic processes with time restrictions specified by probability density functions with stationary parameters over long periods of time. We present a technique for task allocation based on the maximum entropy principle as an element of tasks differentiation. The entropy measurement is proposed for hierarchy task allocation.

Key–Words: Soft real time systems, Priority tasks allocation, Maximum entropy, Computational task, Stochastic processes.

1 Introduction

The term “tasks” is referred, in computer systems, to those processes which are executed in servers performing a specific action. Tasks are software entities that aim to respond to the information generated by real-time systems, peripheral devices or internal processes [1]. When timing and correctness are requirements for task completion, then we are dealing with real time tasks. In multimedia or communication processes, when a typical deadline constraint has to be, statistically accomplished, we call these tasks soft-real-time tasks. In fact, we are referring to task timing constraints as tasks processes.

Since its earliest definitions, tasks have been classified as: periodic, semiperiodic, aperiodic or sporadic by various authors, all of these, according to the analysis context; see [1],[10], [11], or [17]. Periodic tasks execute their invocation within regular time intervals; semiperiodic tasks have a variation in their invocation time, meanwhile an aperiodic task is invoked at irregular times. A sporadic task is an aperiodic task with a defined minimum inter-arrival time, [2]. There are stipulated assumptions to handle these tasks in scheduling algorithms. In most of the cases, task hierarchy is established prior to execution.

Scheduling, in real time systems, concerns to the determination of a temporal ordering for some specified timing, precedence and resource requirements. Depending on the nature of the application, scheduling may be classified into different kinds: static scheduling, dynamic scheduling and mixed scheduling [11].

Static scheduling assumes prior knowledge of the relevant characteristics of all tasks, which may be taken in the temporal ordering. When this information is available, priority ordering is assigned to tasks. However, any change in tasks requires complete rescheduling. On the other hand, dynamic scheduling algorithms are designed to work with unpredictable arrival times and possible uncertainties in execution time or deadline. Priority assignation is done in line, that is, priority allocation is done at the time a task arrives.

Priority ordering only has sense when tasks are preemptable. A task is preemptable by other tasks with a higher priority if it can be interrupted and reassumed later meanwhile its overall goal can be achieved. In dynamic scheduling algorithms, tasks assigned priorities vary over time from request to request, but in a statistical way the priority tasks ordering remain the same.

Tasks priority allocation depends mostly on the timing constraints; such as arriving time, computation time or deadline. Our work in tasks scheduling, based on priority ordering, aims to deal with these constraints in an optimal way. In dynamic priority scheduling algorithms, where tasks characteristics are known at the time they arrive, statistical knowledge of the process is required.

In stationary systems where statistical characteristics over long period of time remain stable; it is possible to determine task timing performance based on their statistics. Task process outcomes can be described by some density distribution function with defined parameters. Probability does not say much about individual events, but describes a faceable level of task predictability. Meanwhile the tasks processes repeat over an infinite number of times, some of these relevant variables are observed; incoming channel, inter-arrival time or period, relative deadline and, computations streams. For our analysis, task processes are considered stochastic stationary processes with defined probability density function; likewise task statistical behavior can be described. In soft-real-time systems, it is statistically required that task processes be completed on their deadlines. This means that a statistical time service is acceptable.

While the system evolves in time, task knowledge increases. So that, a measurement of uncertainty can be described in probability with the maximum entropy level. That is, the information is more instructive, if it consists of mean values of task variables. Therefore, the analysis of task processes estimate a measurement of the degree in which a system is schedulable. One of such measurement is the tasks priority allocation which is the main focus of this work.

The principle purpose of this work is to introduce a scheduling theory based on continuous stochastic task processes for soft-real-time. In this work we determine a priority task allocation technique based on the maximum entropy principle. In section 6 we present a simulation result for task hierarchy with this technique to show how the allocation process is done while the system evolves.

2 Probability Model

In this section, we analyze the basic timing requirements based on the Queuing Theory [7], [8]. Task sets are considered independent, continuous time stochastic variables with known probability distribution functions. Queuing Theory was developed to model and predict stochastic system behavior with resource contention. This theory is based on allowing randomness in task arrivals, execution times and deadlines. It focuses on global system performance measurements, which are usually computed under equilibrium assumptions. That is, the probability does not say anything about an individual outcome, but refers to an ensemble of all possible outcomes were the task is experimented in a sufficient number of times.

For soft-real-time systems, main timing constraints are defined to model stochastic task processes. Next, the following assumptions are made in order to simplify the model, nevertheless they can be modified in future works.

- A task becomes active just an instant after it arrives.
- All tasks are preemptive and overhead times are negligible.
- A Deadline is the time between two consecutive arrivals.
- Starting time and precedence are considered nulls.

Let $(\Omega, \mathcal{F}, P)$ be the probability space that describes task processes. The Ω set throw outcomes during the sample time. Here $\mathcal{F}$ is the σ-field of sets in Ω. Then P is the probability measure function in this space. Task constraints are random variables generated by a random variable family indexed in time, [12], [13].

3 The Maximum Entropy Principle

In 1948 Claude Shannon [15] published his Theory of Communication; in which he derives a measurement of uncertainty, denoted “entropy”. Referring to some systems with certain physical or conceptual entities, where the messages they produce have a meaning. Shannon proposed a measurement of how much information is “produced” by these processes, or better, at which rate information is produced. In probability analysis, entropy establishes the uniformity of a random variable over a range.

The entropy is a measure function of the uncertainty of an event outcome. The measurement of a set of all possible events whose probabilities of occurrence are $p_1, p_2, ..., p_n$ is denoted by $H(p_1, p_2, \cdots, p_n)$. Entropy, such as a uniformity measurement, is required to accomplish the following properties:

- H should be continuous in the p_i’s.
- If all the p_i are equal, $p_i = 1/n$, then H should be a monotonic increasing function of n. With equal likely events there is more choice, or uncertainty, when there are more possible events.
- If a choice is broken down into successive choices, the original H should be the weighed sum of the individual entropy value of each choice.

Next, we present the basic Shannon’s theorem for Information Theory.

Theorem 1 (Entropy.) *The only function H that satisfies the three properties above is in the following form: (see Shannon [15])*

$$H(p_1, p_2, \cdots, p_n) = -K \sum_{i=1}^{n} p_i \ln p_i. \qquad (1)$$

The form $H = -\sum p_i \ln p_i$ is recognized as entropy for the probabilities p_i's because of the mathematical similarity with the thermodynamical definition of entropy [4]. Typically, K is taken as a unit and the logarithm is taken as the natural ones. In an analogous manner the entropy of a continuous distribution with density function $p(x)$ is given by:

$$H(x) = -\int_{-\infty}^{\infty} p(x) \ln p(x)\ dx. \tag{2}$$

For task processes with known density distributions; mathematical maximization techniques, like Lagrange Multipliers, are used to determine a measurement of certainness based on the maximum entropy principle [14]. We will use this approach to calculate some measurements in order to meet with task timing constraints.

4 Timing Constraints

Timing requirements for soft-real-time tasks are those where, in task service, something missing of a deadline constraint decreases the performance of the system but does not jeopardize its correct behavior [1]. That is, a statistical distribution of response times is acceptable [10].

Definition 2 (Real-time Task). *A real-time task is a stochastic process described by $\tau = \{A(t,\omega), C(t,\omega), D(t,\omega)\}$, characterized by the family of random variables; Arrival time A_t, Computation time C_t, and Deadline time D_t; defined on the same probability space $(\Omega, \mathcal{F}, P)$, with $t > 0$ and $\omega \in \Omega$.*

For a task process, $\tau(t,\omega)$, we apply a simple $M/M/1$ model. The density functions defined below represent the timing constraints under analysis. For them, a measurement of uncertainty called entropy is developed. Through the method of Lagrange Multipliers a unique probability distribution will be found, for the considered requirements [5]. Stochastic variable constraints obey to statistical task behavior met in Queuing Theory. The maximum entropy will represent the measurement of uniformity for the task constraints set [3].

Definition 3 (Computation Time). *Computation time is a stochastic process $C(t,\omega)$ of τ; defined as the time necessary by the processor to execute the task without interruption. Let $f_C(t)$, for $t > 0$, be the computation time density function subject to the constraints:*

$$\begin{aligned} \int_0^\infty f_C(t)\ dt &= 1, \\ \int_0^\infty t \cdot f_C(t)\ dt &= \lambda, \end{aligned} \tag{3}$$

$$H(C_t) = -\int_0^\infty f_C(t) \ln f_C(t)\ dt. \tag{4}$$

The Computation time is a random variable with a exponential type density function, thus,

$$f_C(t) = \frac{1}{\lambda}\ e^{-t/\lambda}. \tag{5}$$

The mean Computation time, $\lambda > 0$, will represent the time unit for the rest of the paper. Here, we want to emphasize the fact that the processor speed gives the reference for all calculations. So that, timing requirements for a task will be analyzed in terms of this λ.

Definition 4 (Arrival Time.) *Arrival time is a stochastic process $A(t,\omega)$ of τ, defined as the required waiting time to observe an arrival occurrence. This event occurs at a rate $K_a\lambda > 0$. It is also referred to as release time if the task becomes ready for execution. If $f_A(t)$ represents the probability density function for this time A_t, then it must satisfy the constraints:*

$$\begin{aligned} \int_0^\infty f_A(t)\ dt &= 1, \\ \int_0^\infty t \cdot f_A(t)\ dt &= K_a\lambda. \\ H(A_t) &= -\int_0^\infty f_A(t) \ln f_A(t)\ dt, \end{aligned} \tag{6}$$

Arrival times also obey an exponential type distribution function of the form,

$$f_A(t) = \frac{1}{K_a\lambda}\ e^{-t/(K_a\lambda)}. \tag{7}$$

Furthermore, arrival times and deadline times are independent random variables. Also, these processes have K_a and K_d multiple of the calculation rate λ. For them,

$$\begin{aligned} K_a &> 1, \\ K_d &> 1, \\ K_a &\geq K_d. \end{aligned} \tag{8}$$

Definition 5 (Deadline). *Deadline is a stochastic process $D(t,\omega)$ of τ, defined as the time when a task must be completed in order not to decrease system performance level. It is described as the time before the next request occurs plus the previous arrival task time with density function $f_A(t)$. If $f_D(t)$ represents the probability density function for this time D_t, then it must satisfy the constraints:*

$$\begin{aligned} \int_0^\infty f_D(t)\,dt &= 1, \\ f_D(t) &= \int_0^x f_A(t)\,f_D(x-t)\,dt, \\ H(D_t) &= -\int_0^\infty f_D(t)\,\ln f_D(t)\,dt. \end{aligned} \tag{9}$$

When $K_d = K_a$ the deadline density function of D_t is a of the Gamma type

$$f_D(t) = \frac{1}{K_d^2 \lambda^2}\,(t\,e^{-t/K_d\lambda}). \tag{10}$$

The probability function for a task process has a known distribution over the time. A measurement of uniformity for this density function can be expressed by the maximum entropy [5]. A higher entropy value correspond to a lower rate K_a. We establish entropy measurements for arrival time and deadline constraints as the basic forms of priority task allocation.

Theorem 6 (Maximum arrival entropy). *Let $f_A(t)$ be the density of the arrival time of a task, it satisfies the following constraints:*

1. $\int_0^\infty f_A(t)\,dt = 1,$
2. $\int_0^\infty t \cdot f_A(t)\,dt = K_a\lambda,$
3. $H(A_t) = -\int_0^\infty f_A(t) \ln f_A(t)\,dt.$

Then the maximum entropy for the function $f_A(t)$, with pre-specified first moment in $[0,\infty)$; also denoted by $H_{A,i}$, is given by

$$H(A_t) = \ln(K_a\lambda) + 1. \tag{11}$$

Proof: In a probabilistic sense, the arrival time density function $f_A(t)$ is an exponential type function, [8] and [9]. Using the method of Lagrange multipliers, the density distribution that satisfies the constraints 1, 2 and 3, is the one described in theorem 6.

$$\frac{\partial}{\partial f}(-f_A \ln f_A) + \mu \frac{\partial}{\partial f_A}(f_A) + \lambda \frac{\partial}{\partial f}(t \cdot f_A) = 0.$$

The associated maximal entropy is then obtained by,

$$\begin{aligned} H(A_t) &= \\ &= -\int_0^\infty \frac{1}{K_a\lambda} e^{-t/(K_a\lambda)} \ln \frac{1}{K_a\lambda} e^{-t/(K_a\lambda)}\,dt \\ &= \ln(K_a\lambda) \int_0^\infty \frac{1}{K_a\lambda} e^{-t/(K_a\lambda)}\,dt \\ &+ \frac{1}{K_a\lambda} \int_0^\infty \frac{t}{K_a\lambda} e^{-t/(K_a\lambda)}\,dt. \end{aligned}$$

Integrating

$$H(A_t) = \ln(K_a\lambda) + 1.$$

■

Theorem 7 (Maximum task deadline entropy). *Let $f_D(t)$ be the density function of the deadline time of a task satisfying the following constraints:*

1. $\int_0^\infty f_D(t)\,dt = 1,$
2. $f_D(t) = \int_0^x f_A(t)\,f_A(x-t)\,dt,$
3. $H(D_t) = -\int_0^\infty f_D(t) \ln f_D(t)\,dt.$

The maximum entropy for $f_D(t)$ variable with pre-specified first moment in $[0,\infty)$ is $H_{D,i}$ given by

$$H(D_t) = \gamma + \ln(K_a\lambda) + 1. \tag{12}$$

Proof: The density function of the deadline time variable is calculated by the convolution between the previous and next task arrival time density functions; i.e.

$$f_D(t) = \int_0^x f_A(t)\,f_A(x-t)\,dt.$$

Thus, the density distribution that satisfies the constraints 1 and 2, is the one described in theorem 7. The associated maximal entropy is then obtained by

$$H(D_t) = -\int_0^\infty f_D(t) \ln f_D(t)\,dt.$$

This yields

$$H(D_t) = \gamma + \ln K_a\lambda + 1.$$

■

5 Priority Task Allocation

In previous sections we establish a probabilistic criteria for task classification in dynamic algorithms according to based on tasks temporal constraints. As seen in the investigations in this area, periodicity and deadline constraints have been considered bases on task planning algorithms.

A main task property is the period. In Liu and Lyland pioneer analysis, [10], the length of successive tasks is a constant called period p. A periodic task is said to have a regular release time, or it is regular time triggered. If a task does not occur with this criterion is called a nonperiodic or a sporadic task. Task periodicity plays a relevant role in scheduling algorithms because of the assumptions considered in its analysis; in order to have the expected results [6]. Periodicity is considered as a measurement of task order timely speaking. A higher hierarchy task corresponds to a higher periodicity task, which is defined by the task arrival distribution function.

Deadline property is considered in dynamic scheduling as(in) the EDF algorithm [16]. By means of the processor utilization factor, task priority is selected by the absolute deadline. Tasks with earlier deadlines will have higher priorities.

Definition 8 (Hierarchy in Probabilistic Sense). *Let $J_i = J(\tau_i)$ be the task parameter assigned that describes the relative importance among tasks in the system,*

$$\max\{J_i\} = \begin{cases} \min\{H_{A,i}\} \\ \min\{H_{D,i}\} \end{cases} \quad (13)$$

where, $H_{M,i}$ will be the maximum entropy policy of density function f_i, calculated from: the arrival time $f_A(t)$, or deadline $f_D(t)$ functions.

Theorem 9 (Tasks Hierarchy based on Maximum Entropy). *Let τ_i be a task process with arrival time $f_{A,i}(t)$, $f_{C,i}(t)$ and $f_{D,i}(t)$. Task hierarchy of J_i respect to J_j according with the maximum entropy principle is established by*

$$J_i > J_j \quad \text{if and only if} \quad H_{M,i} < H_{M,j}. \quad (14)$$

Proof: Results of theorems 6 and 7 relate the maximum entropy measurement with the distribution density mean for the task parameters under consideration. According to the inequality $H_{M,i} < H_{M,j}$, hierarchy gives the inequality $J_i > J_j$. ■

6 Simulation Results

Assigning hierarchy to task processes is a first step in task planning due to the relevance of classifying tasks in dynamic scheduling algorithms. That is, while tasks arrive and we have enough information about their probabilistic parameters, we are able to distinguish those tasks that must be served first from the others.

In this work a simulation algorithm for task classification, using the maximum entropy principle, is presented. Here we take tasks arriving time samples from which we observed their statistical parameters during the simulation time. With the first probability moment values we applied task classification based on maximum entropy principle in order to give them a relative hierarchy giving an order of tasks to be processed by the server.

In figure 1, a simulation outcome is showed. Here, the simulation algorithm is tested for three tasks with different first probability moments in which arriving task density distributions are stationary. Moreover, the tasks classification is presented for these task samples giving them a priority value in function of maximum entropy certainness measurement.

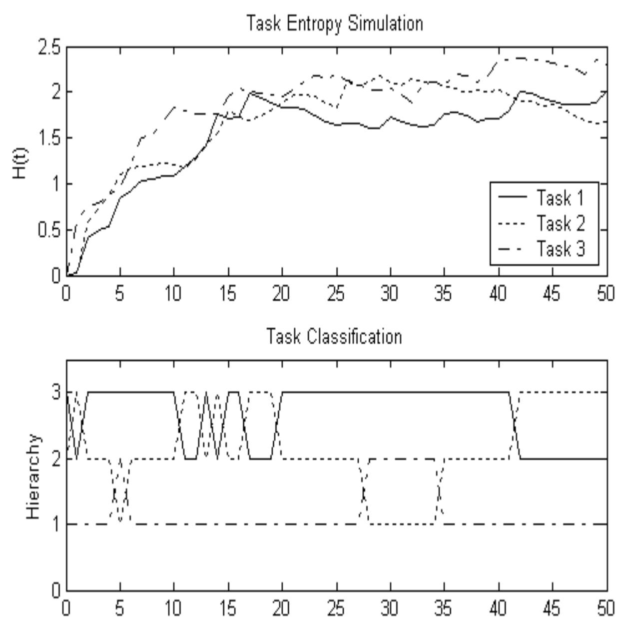

Figure 1. A tasks allocation priority simulation outcome

We observe from the tasks hierarchy values for figure 1, that, while the simulation time evolves, these values remain unchanged because of the information gathered from tasks.

7 Summary and Conclusions

In this paper the priority rule problem for task planning is analyzed. Assigning priority to tasks based upon the maximum entropy principle let us deal with tasks with differences in their statistical parameters. A remarkable result of this work is the priority allocation method, in relation with tasks periodicity and deadline constraints in a similar manner.

Based on the maximum entropy level as an element of differentiation among tasks, this theory gives a continuous analysis in task planning processes. We have modeled soft real-time systems as tasks processes with stationary distribution functions. This analysis establishes the basis for scheduling algorithms in future works.

References:

[1] G. Buttazzo. *Hard Real-time Computing Systems*. Kluwer Academic Publishers, Norwell, MC, USA, 1997.

[2] D. Isovic and G. Fohler. Efficient scheduling of sporadic, aperiodic, and periodic tasks with complex constraints. *in Real-Time Systems Symposium, 2000. Proceedings. The 21st IEEE*, 27(30):207–216, November 1991.

[3] E. T. Jaynes. Information theory and statistical mechanics. *The Physical Review*, 4(106):620–630, May 1957.

[4] E. T. Jaynes. On the rationale of maximum-entropy methods. *Proceedings of the IEEE*, 49(70):939– 952, September 1982.

[5] S. M. Johansson, M. Resource allocation under uncertainty using the maximum entropy principle. *IEEE Transactions on Information Theory*, 51(12):4103– 4117, December 2005.

[6] a. S. L. Lehoczky, J. The rate monotonic scheduling algorithm: Exact characterization and average case behavior. *in Proceedings of the IEEE Real-Time Systems Symposium*, December 1989.

[7] J. P. Lehoczky. Real-time queuing theory. *in Proceedings of the 17th Real-Time System Symposium*, pages 186–195, December 1996.

[8] J. P. Lehoczky. Real-time queuing network theory. *in Proceedings of the 18th Real-Time System Symposium*, pages 58–67, December 1997.

[9] T.-H. Lin and W. Tarng. Scheduling periodic and aperiodic tasks in hard real-time computing systems. *Performance Evaluation Review*, 19(1):31–38, May 1991.

[10] C. L. Liu and J. W. Layland. Scheduling algorithms for multiprogramming in a hard-real-time environment. *Journal of ACM*, 20(1), 1973.

[11] N. Nissanke. *Real-time systems*. Prentice Hall, Great Britain, 1997.

[12] A. Papoulis. *Probability, random variables, and stochastic processes*. McGraw-Hill, 1965.

[13] E. Parzen. *Stochastic processes*. Holden-day, 1962.

[14] F. Reza. *An introduction to information theory*. Dover Publications, NY, USA, 1994.

[15] C. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27, July 1948.

[16] B. G. Spuri, M. Scheduling aperiodic tasks in dynamimc priority systems. *Journal of Real-Time Systems*, 2(10), 1996.

[17] J. Stankovic. Real-time and embedded systems. *Journal of ACM Computing Surveys*, 29(1), March 1996.