

Verificación Experimental de la Diagramabilidad de Tareas Esporádicas de Tiempo Real

Rodrigo Santos y Jorge Santos

Dep. Ingeniería Eléctrica, Instituto de Ciencias e Ingeniería de Computación,

Universidad Nacional del Sur, 8000 Bahía Blanca, Argentina

e-mail: iesantos@criba.edu.ar

Resumen: En el presente trabajo se demuestran las condiciones que debe cumplir un conjunto de tareas esporádicas (v.g. alarmas) de tiempo real para que el sistema sea diagramable. Se determina luego la expresión del máximo tiempo de ranura que conserva la diagramabilidad. Experimentalmente se determina el tiempo de crisis del sistema en función de la relación tiempo de ranura/máximo tiempo de ranura que asegura la diagramabilidad. Los resultados experimentales son hipérbolas que confirman las hipótesis elaboradas teóricamente.

1. INTRODUCCION

La definición más aceptada de Sistema de Tiempo Real es la expuesta en [1]: Es un sistema en el cual los resultados no sólo deben ser correctos desde un punto de vista aritmético-lógico sino que además deben producirse antes de un cierto instante denominado *vencimiento* (deadline). Cuando se pierde un vencimiento se dice que el sistema está en crisis.

Los sistemas de tiempo real más estudiados en la literatura de los últimos años están constituidos por un conjunto de m tareas periódicas y quedan caracterizados cuantitativamente mediante la expresión:

$$S(m) = \{(C_1, T_1, V_1), (C_2, T_2, V_2), \dots, (C_m, T_m, V_m)\}$$

en la que C_i , T_i y V_i representan el tiempo de ejecución, el período y el vencimiento de la tarea i . A menudo se acepta como hipótesis que para todo i , $V_i = T_i$. El tiempo se considera ranurado. La duración de una ranura, denominada tiempo de ranura, se nota t_r . Se dice que el sistema es *diagramable* cuando ninguna tarea pierde su vencimiento. Qué tarea es ejecutada en cada momento es decidido por el diagramador, que en general forma parte del Sistema Operativo.

Además de las tareas sincrónicas están presentes tareas de otras características, las denominadas asincrónicas y las esporádicas [2]. Las tareas asincrónicas son tareas que no son de tiempo real en el sentido de que no tienen vencimiento. Mientras que en el caso de las tareas sincrónicas el buen

funcionamiento del sistema se logra cuando todas y cada una de ellas son ejecutadas antes de su vencimiento, en el caso de las tareas asincrónicas se considera que el funcionamiento del sistema es mejor cuanto menor es el tiempo que transcurre entre el instante en que la tarea está lista para ser ejecutada y el instante en que finaliza su ejecución. Cómo acortar este tiempo en un ambiente de tiempo real ha sido estudiado, por ejemplo, en [3, 4].

Las tareas esporádicas, en cambio, son tareas de tiempo real en el sentido de que tienen vencimiento, pero no son estrictamente periódicas, y puede considerarse conceptualmente que sus períodos son mucho mayores que sus vencimientos. En general están asociadas a *alarmas*, tareas con vencimientos cortos y de características tales que el hecho de no ser atendidas puede producir resultados catastróficos, incluso la pérdida de vidas humanas.

La forma en que el diagramador otorga el uso del procesador a las diferentes tareas está regida por una relación de orden lineal denominada *disciplina de prioridades*. Las tareas, ordenadas por la relación, constituyen la denominada *pila de prioridades*. Antes de cada ranura el despachador recorre la pila de arriba hacia abajo y entre todas las tareas listas para ser ejecutadas otorga el uso del recurso a la que tiene la posición más alta en la pila.

Las tres disciplinas más efectivas son Rueda Cíclica Justa, Prioridades Fijas y Menor Tiempo al Vencimiento. De las tres, la más utilizada es la de Prioridades Fijas en su versión de Períodos Monotónicos Crecientes (PMC). En [8] se demostró que si un sistema es diagramable bajo alguna disciplina de prioridades fijas, también lo es con la pila de tareas ordenadas por períodos monotónicos crecientes. PMC fue adoptada por el Departamento de Defensa de los EE. UU. convirtiéndose así en una norma *de facto* [9]. Cuando los vencimientos son menores que los períodos, la pila se ordena por Vencimientos Monotónicos Crecientes, VMC.

El objetivo del presente trabajo es el estudio experimental de la diagramabilidad de tareas esporádicas. Se probará la existencia de una ranura máxima para la cual el sistema es diagramable. Además se mostrará que a medida que el tamaño de la ranura aumenta, el tiempo que le lleva al sistema entrar en crisis desciende hipérbolicamente. Se brindarán para esto datos relevados experimentalmente y se realizará sobre ellos el ajuste de la curva que mejor los aproxime. En la sección 2 se presenta el Método de las Ranuras Vacías [7]; en la sección 3 se describe brevemente el dispositivo experimental; en la sección 4 se brinda un resumen de las medidas realizadas; en la sección 5 el ajuste de las curvas y finalmente en la sección 6 se realiza la discusión de los resultados.

2. METODO DE LAS RANURAS VACIAS

2.1 Tareas de Tiempo Real Sincrónicas

El tiempo se considera ranurado salvo que se indique lo contrario. Cada ranura equivale a una unidad de tiempo y es indivisible. Las ranuras, notadas t , toman valores $1, 2, \dots$.

Para el sistema $S(n)=\{1, 2, \dots, m\}$, se define la función trabajo como:

$$W_m(t) = \sum_{i=1}^m \left\lceil \frac{t}{T_i} \right\rceil C_i \quad (1)$$

La función trabajo es monotónicamente creciente.

En [7] se demuestra que la primera ranura libre en un sistema $S(i-1)$, notada $e_{1(i-1)}$:

$$e_{1(i-1)} = \text{menor } t \in [1, +\infty] | t = 1 + W_{i-1}(t)$$

El sistema $S(m)$ operando bajo la disciplina PMC es diagramable sss:

$$\forall i \in \{1, 2, \dots, m\} \quad \sum_{i=1}^{m-1} \frac{1}{T_i} \leq 1 \quad T_i \geq e_{1(i-1)}$$

La primera condición asegura que el sistema $S(i-1)$ es no-saturado, es decir que hay al menos una ranura libre para ser ocupada por la tarea i que se va a incorporar. La segunda condición asegura que el período de la tarea i es mayor que la primera ranura libre de $S(i-1)$ y que, en consecuencia el sistema es libre de crisis.

2.2 Tareas de Tiempo Real Esporádicas

En lo que sigue se considerarán tareas esporádicas unitarias generadas en sincronismo con el comienzo de una ranura. Como en las mismas, el pseudointervalo entre ejecuciones es bastante mayor que el vencimiento, puede aceptarse como hipótesis que, dado un conjunto de tareas esporádicas, el menor período es mayor que el mayor vencimiento, notado V_{\max} . En consecuencia, si las tareas se ordenan por períodos monotónicos crecientes, se cumple:

$$T_1 > V_{\max}$$

Debido a esto la función trabajo se reduce a una única ejecución de las $(i-1)$ tareas precedentes.

Teorema 1: El sistema es diagramable sss:

$$V_i \geq (i+1)t_r$$

Demostración: Debido a que el menor período es mayor que el mayor vencimiento, la función trabajo de las $(i-1)$ primeras alarmas debe ser considerada sólo en el intervalo $[1, V_{\max}]$. En el mismo, dicha función se

reduce a una única ejecución de cada una de esas alarmas. En consecuencia, el sistema es diagramable (con la hipótesis de generación de alarmas al comienzo de ranura) si:

$$V_i \geq (1+(i-1))t_r = i \cdot t_r$$

Si, más de acuerdo con la que sucede en el mundo real, se elimina la hipótesis de que la generación de la alarma es sincrónica con el comienzo de la ranura, es necesario agregar una ranura a la expresión anterior. El sistema de tareas esporádicas será, en consecuencia, diagramable sss:

$$\forall i \quad V_i \geq (i+1)t_r \quad \square$$

Es evidente que, para un conjunto $\{V_i\}$ de cardinalidad dada, cuanto mayor sea t_r más difícil será la diagramabilidad del sistema. Es por ello que es de interés determinar cuál es el máximo tiempo de ranura, notado $\max t_r$, para el cual el sistema es diagramable. El máximo tiempo de ranura que asegura la diagramabilidad de la tarea i será notado $\max t_{ri}$.

Teorema 2: El máximo tiempo de ranura que satisface la diagramabilidad del conjunto de tareas esporádicas ordenadas por Vencimientos Monotónicos Crecientes, notado $\max t_r$, es:

$$\max t_r = \min \left\{ \max \{t_{ri}\} \right\} \quad i = 1, 2, \dots, m$$

Demostración: Por el Teorema 1, el máximo tiempo de ranura que asegura la diagramabilidad de la tarea i , notada $\max t_{ri}$ es

$$\max t_{ri} = \frac{V_i}{i+1}$$

En consecuencia

$$\max t_r = \min \left\{ \max \{t_{ri}\} \right\} \quad i = 1, 2, \dots, m \quad \square$$

Corolario: El sistema es diagramable si se cumple:

$$\forall i \in \{1, 2, \dots, m\} \quad V_i \geq (i+1) \max t_r$$

Los vencimientos medidos en ranuras se expresan como:

$$V_i [\text{ran}] = \left\lfloor \frac{V_i [\text{ms}]}{\max t_r [\text{ms}]} \right\rfloor$$

2.3 Planteo de la hipótesis teórica

La forma en que se relaciona el tiempo que un sistema de tareas esporádicas tarda en entrar en crisis, llamado tiempo de crisis (t_c), es una función hiperbólica (de algún grado y ganancia) de la longitud de ranura cuando ésta es mayor que la máxima que aún hace diagramable al sistema.

La idea intuitiva que racionaliza la hipótesis es que cuando el tiempo de ranura es el máximo que conserva la diagramabilidad del sistema, el tiempo de crisis es ∞ . A medida que el tiempo de ranura aumenta, los tiempos de crisis se acortan y tienden asintóticamente a cero, medidos en unidades físicas de tiempo.

Para la verificación experimental de la hipótesis se utilizó el Banco de Ensayos de Tiempo Real descrito a continuación. Se realizaron ensayos sobre distintas configuraciones del sistema y en cada caso se midió el t_c (en ranuras) en función de la relación adimensional entre el tiempo de ranura utilizado en el ensayo y el máximo tiempo de ranura que aún preserva la diagramabilidad del sistema. En tales condiciones, la hipérbola debe ser obviamente asintótica a los ejes $t_c=1$ y $\frac{t_r}{\max t_r} = 1$

3. DISPOSITIVO EXPERIMENTAL

El Banco de Ensayo de Tiempo Real cuenta con seis motores de corriente continua a los cuales se puede modificar la tensión de alimentación y con esto la velocidad. Los motores hacen girar un índice entre dos sensores. Al llegar a un sensor el motor debe invertir la marcha para no alcanzar otro sensor que marca el fin de carrera. El movimiento de los motores y la lectura de los sensores es efectuada por una PC mediante una placa de Entrada/Salida Diot 48.

Se desarrolló un software que trabaja sobre la PC y realiza el control de los motores y la lectura de los sensores. Básicamente el programa consta de nueve módulos. Un primer módulo es el encargado de inicializar las variables y cerrar el sistema una vez que se detiene el ensayo. Seis módulos, completamente independientes, son los encargados de controlar el movimiento de cada uno de los motores. El octavo módulo es el encargado de leer los sensores a intervalos regulares de tiempo dados por el reloj de la máquina, coincidentes con el tiempo de ranura. Sobre la base de esta lectura, y de acuerdo con una pila de prioridades previamente determinada, despacha uno de los seis módulos que realizan el control de los motores si es que alguno de ellos requiere atención. Un último módulo, sin acción sobre los motores, corresponde al caso de que la ranura quede ociosa por falta de requerimientos.

4. MEDICIONES Y RESULTADOS

4.1 Descripción de los ensayos

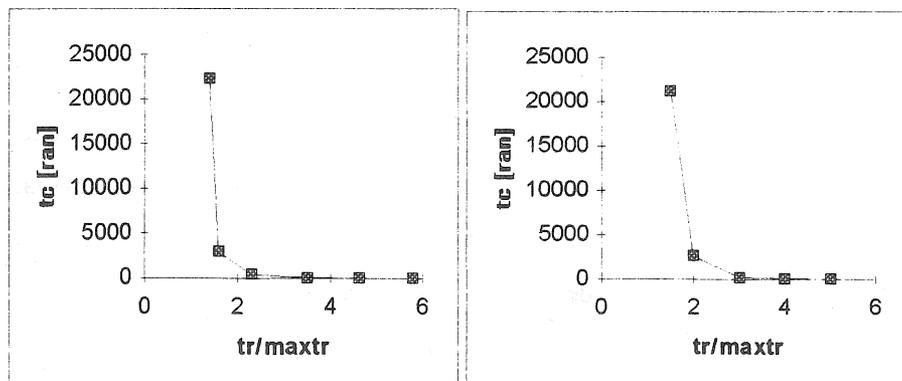
Se establecieron cinco configuraciones. Cada configuración consiste en dar a los motores diferentes tensiones de alimentación a fin de lograr vencimientos y períodos distintos en cada caso.

Para cada una de las configuraciones se realizaron los siguientes pasos:

1. Se caracterizó al sistema compuesto por los seis motores. Es decir se midió el vencimiento y el pseudo período de cada uno de ellos con un Analizador Lógico HP1650. El tiempo de ejecución puede simularse siempre igual a una ranura pues hay al menos tres órdenes de magnitud entre la velocidad de los dispositivos mecánicos y el software que corre en la PC.
2. Sobre la base de los datos obtenidos en el paso anterior se determinó el $\max t_r$.
3. Se incrementó el tamaño de la ranura y se midió el tiempo que el sistema tarda en entrar en crisis para cada tamaño.
4. Se determinó, a partir de los datos obtenidos en el paso anterior, la ley de respuesta.

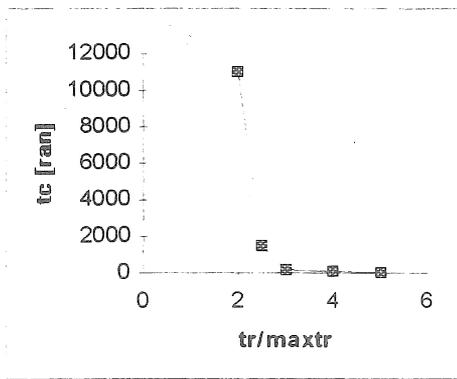
4.2 Presentación de los datos

Se presentan a continuación las curvas obtenidas para las cinco configuraciones. En las abscisas se representa la relación $t_r/\max t_r$ y en las ordenadas el tiempo de crisis t_c [ran]. Los valores graficados son el promedio de los ensayos realizados para cada valor de la relación. El número de ensayos varía entre 10 y 24 para cada punto.

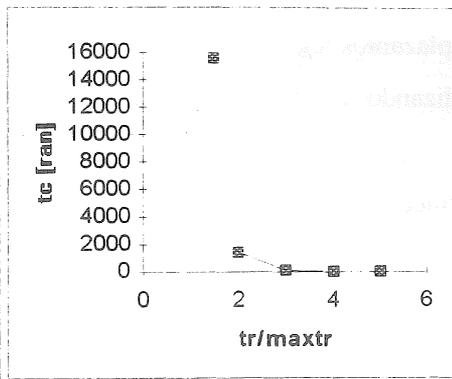


Configuración 1

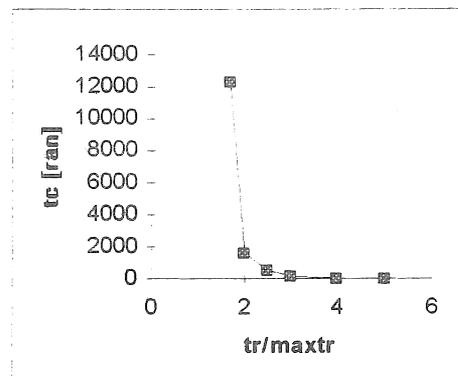
Configuración 2



Configuración 3



Configuración 4



Configuración 5

5. AJUSTE DE CURVAS

En los gráficos de la sección anterior se comprueba que la ley de respuesta del sistema es hiperbólica, tal cual fue anticipado en la hipótesis de trabajo.

Las curvas tienen dos asíntotas. Una se encuentra en las ordenadas con valor 1 ya que el sistema tardará, cualquiera sea el tamaño de la ranura, al menos una ranura en entrar en crisis. La otra asíntota se encuentra en las abscisas, también con valor 1 ya que el sistema tendrá un $t_c = \infty$ cuando se tome $t_r = \max t_r$. Con estas dos consideraciones podemos decir que la curva pertenece a la siguiente familia:

$$t_c = \frac{k}{(x-1)^m} + 1$$

Hay que determinar entonces el valor de las constantes k y m para cada caso. Para ello linealizamos la curva aplicando logaritmo natural a ambos miembros de la ecuación:

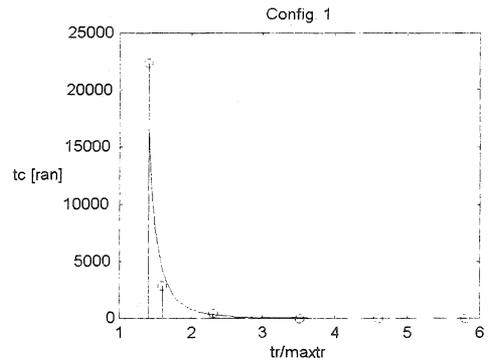
$$\log(t_c - 1) = \log(k) - m \cdot \log\left(\frac{t_r}{\max t_r} - 1\right)$$

Si reemplazamos $\log(t_c - 1)$ por la variable v , y $\log(x - 1)$ por la variable u , obtenemos la ecuación de una recta. Utilizando el método de regresión lineal se determinaron las constantes k y m para cada configuración.

A continuación se presentan los resultados que se obtuvieron para cada configuración. Se dan los valores de m y k y se representa la hipérbola que mejor aproxima los datos experimentales.

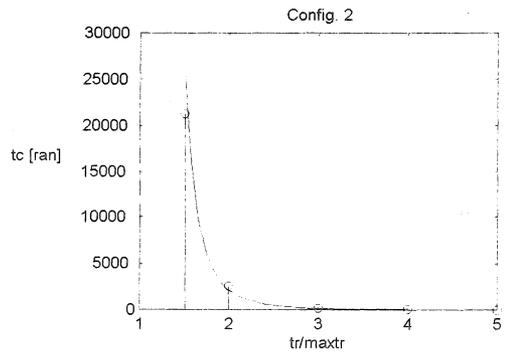
Configuración 1

$$t_c = \frac{820}{\left(\frac{t_r}{\max t_r} - 1\right)^{3.27}} + 1$$



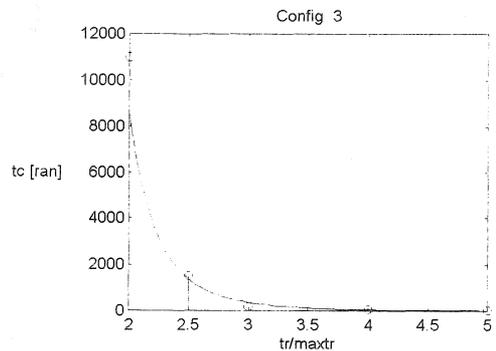
Configuración 2

$$t_c = \frac{2147.2}{\left(\frac{t_r}{\max t_r} - 1\right)^{3.56}} + 1$$



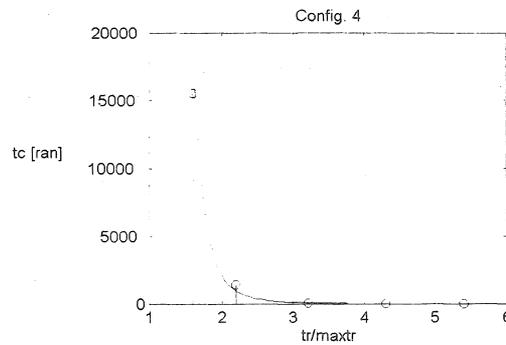
Configuración 3

$$t_c = \frac{8581}{\left(\frac{t_r}{\max t_r} - 1\right)^{4.55}} + 1$$



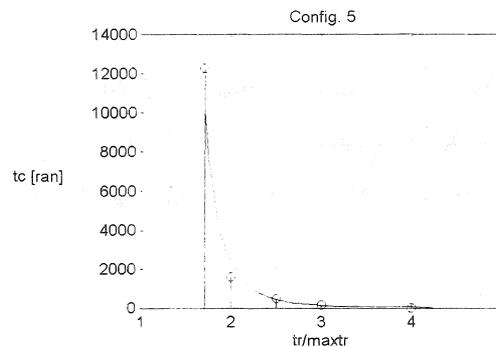
Configuración 4

$$t_c = \frac{2091.4}{\left(\frac{t_r}{\max t_r} - 1\right)^{4.24}} + 1$$



Configuración 5

$$t_c = \frac{2332.5}{\left(\frac{t_r}{\max t_r} - 1\right)^{4.1}} + 1$$



6. DISCUSION DE RESULTADOS

Debe hacerse notar que al medir el tiempo de crisis en ranuras y expresarlo en función de una variable adimensional ($t_r/\max t_r$), el análisis se independiza de la variable física tiempo. Se trata de esta manera de generalizar resultados buscando encontrar similitudes entre sistemas que se parecen si se los analiza en unidades-ranura aunque difieren cuando se los analiza en unidades físicas de tiempo.

Los resultados obtenidos confirman la hipótesis teórica elaborada intuitivamente, en el sentido de que las curvas que representan el tiempo de crisis en función de la relación $t_r/\max t_r$ son hipérbolas. Sin embargo, el hecho de que las mismas tengan distinta ganancia (k) y distinto grado (m) complica la extracción de conclusiones. En principio, m crecientes producen hipérbolas más agudas y k crecientes alejan las curvas de la intersección de las asíntotas. Hipérbolas menos agudas o más alejadas del origen representan configuraciones más robustas. Esto quiere decir que para una misma relación $t_r/\max t_r$, el sistema tarda más tiempo en entrar en crisis. Las configuraciones más robustas corresponden en principio a una distribución de valores de vencimientos más espaciados, esto es con más ranuras libres entre vencimientos. Sin embargo, debe hacerse notar que al tener distintas ganancias y distintos órdenes, las curvas pueden cruzarse en un punto cuyas ordenadas satisfacen simultáneamente ambas ecuaciones. Esto

significa que para un cierto intervalo de valores de la variable independiente, una configuración tiene tiempos de crisis mayores que otra, pero el sentido de la desigualdad se invierte al cruzar la abscisa de intersección. La mayoría de las curvas, tanto las relevadas experimentalmente como las ajustadas matemáticamente, confirman lo expuesto.

Profundizar en las razones que explican la inversión de la desigualdad constituye uno de los temas futuros de trabajo.

Referencias

1. J. A. Stankovic, A serious problem for next generation systems, *IEEE Computer*, 21 (10) (1988), pág. 10-19.
2. J. Orozco, J. Santos, Diagramación de Tareas en Sistemas Operativos de Tiempo Real PMC con Inversión de Prioridades, *Anales 23 JAIHO*, septiembre 1994.
3. B. Sprunt, L. Sha, J. Lehoczky, Aperiodic Task Scheduling for Hard-Real-Time Systems, *Real Time Systems Journal*, 1 (1) (1989), pág. 27-60.
4. J. P. Lechoczky, L. Sha, Y. Ding, The Rate Monotonic Scheduling Algorithm: Exact Characterization and Average Case Behavior, *Proc. Real-Time Systems Symp.*, IEEE CS, Los Alamitos, CA, 1989.
5. J. Santos, J. Orozco, Rate Monotonic Scheduling in Hard-Real Time Systems, *Information Processing Letters*, 48 (1993), pág. 39-45.
6. J. Santos, E. Ferro, J. Orozco, R. Cayssials, A Heuristic Approach to the NP Hard Real-Time Multitask-Multiprocessor Assignment Problem Using the Empty Slots Method and Rate Monotonic Schedule, *Real Time Systems Journal*, (en prensa).
7. J. Santos, M. L. Gastaminza, J. Orozco, D. Picardi, O. Alimenti, Priorities and Protocols in Hard Real-Time LANs: Implementing a Crisis-Free System, *Computer Communications*, 14 (9) (1991), pág. 46-61.
8. C. L. Liu y J. W. Layland, Scheduling Algorithms for Multiprogramming in Hard Real-Time Environment, *J. ACM*, 20 (1) (1973), pág. 46-61.
9. R. Obenza, Rate Monotonic Analysis for Real-Time Systems, *IEEE Computer*, 26 (3) (1993), pág. 73-74.

Agradecimientos:

Los autores agradecen a los ingenieros Edgardo Ferro, Omar Alimenti y Javier Orozco, las sugerencias recibidas a lo largo del proyecto.

CV de Autores

Rodrigo M. Santos. Ingeniero Electrónico, graduado en la Universidad Nacional del Sur en 1997. Ha sido Becario de Introducción a la Investigación para Alumnos Avanzados (1996) y para Graduados (1997).

Jorge Santos. Ingeniero Industrial, graduado en la Universidad de Buenos Aires, en 1951, con Diploma de Honor. Actualmente es Investigador Principal del CONICET y Profesor Titular Consulto del Departamento de Ingeniería Eléctrica y Director del Instituto de Ciencias e Ingeniería de Computación de la Universidad Nacional del Sur. Ha sido Research Fellow de la Universidad de Manchester, UK y Premio Konex 1993 en la especialidad Ingeniería de Computación. En el marco del Programa Multinacional de la OEA dictó cursos de postgrado en la Universidad de Chile y en la Universidad Técnica F. Santa María, en 1971 y 1973 respectivamente.