Diciembre 2012
Volumen 10
Número 1

ISSN 1879 - 9532

RIEE&C

REVISTA DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA Y COMPUTACIÓN

Revista publicada por el Depto. de Ing. Eléctrica y Electrónica del Instituto Tecnológico de Sonora

ITSON – Instituto Tecnológico de Sonora
5 de Febrero 818 Sur. C.P. 85000
Ciudad Obregón, Sonora, México.

RIEE&C: Revista de Ingeniería Eléctrica, Electrónica y Computación

Año 8, Volumen 10, Número 1, Diciembre de 2012.

ISSN: 1870-9532

Publicación semestral editada por ITSON a través del
Departamento de Eléctrica y Electrónica.
Esta publicación fue completada en Diciembre de 2012.

Contacto: rieeandc@itson.edu.mx
Internet: http://www.itson.mx/rieeyc

Impreso en el taller del Instituto Tecnológico de Sonora. Fecha de Impresión Diciembre 2012. Tiraje de 50 ejemplares con 19 hojas.

Regulación no lineal de la salida basada en modelos Takagi-Sugeno

Estrada M. Víctor, Castillo T. Bernardino y Bernal R. Miguel A.

Nonlinear output regulation based on Takagi-Sugeno models

Abstract— **Tracking a signal is one of the most important issues in control theory and the output regulation theory is a useful tool. In fact, it is difficult to solve because a set of partial differential equations appears in the nonlinear case. Then, Takagi-Sugeno models are introduced in order to simplify the procedure. This paper shows a systematic approach based on Takagi-Sugeno models and linear matrix inequalities that solve the nonlinear output regulation problem by taking advantage of convex representations for the nonlinear mappings and the time-derivatives of the membership functions, one of the principal contributions that makes it different from previous results. In addition, it is fully based on element-wise linear matrix inequalities which are numerically solvable by convex-optimization methods.**

***Keywords*— Linear Matrix Inequality, Output Regulation Theory, Takagi-Sugeno Models, Trajectory Tracking.**

Resumen— **El seguimiento de trayectoria es uno de los principales problemas en la teoría de control y la teoría de regulación de la salida es una herramienta muy útil. Sin embargo, resulta difícil resolverlo porque aparecen un conjunto de ecuaciones diferenciales parciales en el caso no lineal. Los modelos Takagi-Sugeno han sido introducidos para simplificar el procedimiento. Este artículo muestra un enfoque sistemático basado en modelos Takagi-Sugeno y desigualdades matriciales lineales que solucionan el problema de regulación no lineal de la salida tomando ventaja de las representaciones convexas de los mapeos no lineales y las derivadas de las funciones de membresía, una de las principales contribuciones que lo hacen distinto de trabajos anteriores. Además, está basado en desigualdades matriciales lineales elemento por elemento que pueden ser resueltas numéricamente por métodos de punto-interior.**

***Palabras clave*—Desigualdad Matricial Lineal, Modelo Takagi-Sugeno, Seguimiento de Trayectoria, Teoría de Regulación de la Salida.**

Manuscrito recibido el 13 de marzo de 2012. Este trabajo fue respaldado por el departamento de Ing. Eléctrica y Electrónica del Instituto Tecnológico de Sonora a través del proyecto PROFAPI 1A4005208007 y el Consejo Nacional de Ciencia y Tecnología (CONACYT) por medio de la becas 24176, SIN-37449 y proyecto SEP-CONACYT CB-2011, 168406.

Estrada M. Víctor. es estudiante de maestría en ciencias en ingeniería eléctrica del CINVESTAV del IPN; Av. del Bosque 1145, Col. El Bajío; Zapopan, Jalisco, México; C.P. 45019; Tel: (33) 3777-3600, ext. 1075; e-mail victor.estrada.m@gmail.com.

Castillo T. Bernardino es integrante del Departamento de Ingeniería Eléctrica y Ciencias de la Computación del CINVESTAV del IPN; Av. del Bosque 1145, Col. El Bajío; Zapopan, Jalisco, México; C.P. 45019; Tel: (33) 3777-3600, e-mail toledo@gdl.cinvestav.mx.

Bernal R. Miguel A. del 2011 hasta la fecha se ha de desempeñado como profesor de tiempo completo del Instituto Tecnológico de Sonora en el Departamento de Ingeniería Eléctrica y Electrónica del Instituto Tecnológico de Sonora; Av. Antonio Caso S/N Col. Villa ITSON; Ciudad Obregón, Sonora, México; C.P. 85130; Tel: (644) 4109000, ext. 1200; Fax: (644) 4109001.e-mail miguel.bernal@itson.edu.mx.

I. INTRODUCCIÓN

El seguimiento de una señal es importante en sistemas de control. Existen varias técnicas para lograr tal objetivo, una de ellas es la teoría de regulación, que se basa en técnicas de control en el marco de la geometría diferencial. La teoría de regulación ha sido tratada tanto para sistemas lineales [1] como para no lineales [2, 3, 4].

Esencialmente, el problema de regulación consiste en someter una planta a señales de referencia y/o perturbaciones externas, generadas por un sistema llamado exosistema. Existe solución al problema si es posible encontrar una ley de control tal que en ausencia de perturbaciones el punto de equilibrio de la planta en lazo cerrado sea asintóticamente estable y, además, el error de seguimiento entre la salida de la planta y la señal de referencia tienda a cero asintóticamente [5]. Existen dos casos: a) información completa de los estados, llamada regulación por retroalimentación del estado; b) cuando sólo se tiene información del error de seguimiento, entonces se utiliza regulación por retroalimentación del error [2, 6].

Para el caso lineal, es suficiente y necesario resolver un par de ecuaciones matriciales lineales, llamadas ecuaciones de Francis [1]. La extensión para sistemas no lineales, está basada en la teoría de la variedad central [7] y su solución se consigue al resolver un conjunto de ecuaciones diferenciales parciales conocidas como ecuaciones de Francis-Isidori-Byrnes (FIB) [2].

Trabajar con las ecuaciones FIB, resulta en la mayoría de los casos muy complicado. Para tratar de simplificar el procedimiento se ha trabajado en el marco de sistemas lineales [3, 4, 8, 9]. En años recientes, el modelado en la forma Takagi-Sugeno (TS) y el uso de desigualdades matriciales lineales (LMIs, por sus siglas en inglés) ha sido utilizado para enfrentar las dificultades mencionadas anteriormente [10, 11, 12].

En este trabajo se busca la solución sistemática y completa para el problema de regulación no lineal de la salida (PRNS) por retroalimentación, tanto del estado como del error, a través modelos TS [13] y LMIs. Para esto se asume una estructura TS de las ecuaciones FIB, mismas que originan derivadas de las funciones de membresía que generan inconvenientes [11, 12]. Dificultades similares se presentan en el área de estabilización

por medio de una función no cuadrática de Lyapunov para sistemas en forma TS, las primeras investigaciones [14, 15, 16] solo dan una aproximación. En [17] se presentó por primera vez un análisis completo que toma ventaja de esas derivadas de funciones de membresía (FMs). Este análisis es incluido en las ecuaciones FIB y permite resolver el PRNS por medio de modelos TS y LMIs.

El presente artículo está ordenado como sigue: la sección II muestra el PRNS y notación para modelos TS; la sección III contiene los resultados principales en donde se desarrolla el análisis de las derivadas de FMs para solucionar el problema del regulador difuso; un par de ejemplos son mostrados en la sección IV para verificar la efectividad del enfoque; finalmente, en la sección V se muestran las conclusiones.

II. TEORÍA BÁSICA Y NOTACIÓN

A. Regulación de la salida por retroalimentación del estado

Considere el sistema no lineal

$$\begin{aligned}\dot{x}(t) &= f(x)+g(x)u+p(x)w\\ \dot{w}(t) &= s(w)\\ e(t) &= h(x)+q(w)\end{aligned} \tag{1}$$

donde $x(t)\in X\subset\mathbb{R}^{n\times 1}$ es el vector de estados, $u(t)\in\mathbb{R}^{m\times 1}$ es la entrada del sistema, $w(t)\in W\subset\mathbb{R}^{q\times 1}$ el exosistema y $e(t)\in\mathbb{R}^{o\times 1}$ es el error de seguimiento.

El PRNS por retroalimentación del estado consiste en encontrar una ley de control $u(t)=\alpha(x,w)$ tal que:

A) $\dot{x}(t)=f(x)+g(x)\alpha(x,0)$ tenga el punto de equilibrio $x=0$ exponencialmente estable con $\alpha(x,0)=K^*x$, $K^*\in\mathbb{R}^{n\times m}$.

B) $\exists U\subset X\times W\supset(0,0):\forall\left(x(0),w(0)\right)\in U\Rightarrow\lim_{t\to\infty}e(t)=0$.

En [5] se proponen las siguientes hipótesis:

H1) $w=0$ es un punto de equilibrio estable del exosistema y $\exists\tilde{W}\subset W\supset 0:\ \forall w(0)\in\tilde{W}$ es Poisson-estable.

H2) $\left(f(x),g(x)\right)$ tiene una aproximación lineal estabilizable en $x=0$.

Si se cumplen H1 y H2, el PRNS por retroalimentación del estado tiene solución si y sólo si $\exists\, x=\pi(w)$, $u=\gamma(w)$: $\pi(0)=0$, $\gamma(0)=0$ como mapeos en $W^0\subset W\supset 0$ tales que [5]:

$$\begin{aligned}\frac{\partial\pi}{\partial w}s(w) &= f\left(\pi(w)\right)+g\left(\pi(w)\right)\gamma(w)+p\left(\pi(w)\right)w\\ 0 &= h\left(\pi(w)\right)+q(w).\end{aligned} \tag{2}$$

La ley de control está dada por

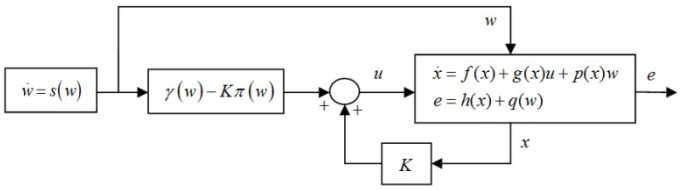


Fig. 1 Diagrama del sistema en lazo cerrado con retroalimentación de estado.

$$u(t)=\alpha(x,w)=\gamma(w)+K^*\left(x-\pi(w)\right) \tag{3}$$

La figura 1 muestra un diagrama a bloques del esquema de control por retroalimentación de estado.

B. Regulación de la salida por retroalimentación del error

En caso de que no se tenga acceso a todos los estados, es necesario hacer un estimado de las variables utilizando la información disponible, la del error. En consecuencia el problema de regulación debe resolverse a través de retroalimentación del error. Para esto considere el sistema no lineal (1), el PRNS por retroalimentación del error consiste en encontrar una ley de control

$$\begin{aligned}\dot{\xi}(t) &= \eta(\xi,e)\\ u(t) &= \theta(\xi)\end{aligned} \tag{4}$$

donde $\xi(t)\in\Xi\subset\mathbb{R}^{(n+q)\times 1}$, además $\xi(t)=\left[\xi_0\quad\xi_1\right]^T$ tales que:

C) $\dot{x}(t)=f(x)+g(x)\theta(\xi)$ y $\dot{\xi}(t)=\eta\left(\xi,h(x)\right)$ tengan un punto de equilibrio $(x,\xi)=(0,0)$ asintóticamente estable.

D) $\exists U\subset X\times\Xi\times W\supset(0,0,0)$ y $\forall\left(x(0),\xi(0),w(0)\right)\in U$ para que el sistema en lazo cerrado

$$\begin{aligned}\dot{x}(t) &= f(x)+g(x)\theta(\xi)+p(x)w\\ \dot{\xi}(t) &= \eta\left(\xi,h(x)+q(w)\right)\\ \dot{w}(t) &= s(w)\end{aligned} \tag{5}$$

cumpla que el $\lim_{t\to\infty}e(t)=0$.

Además de las hipótesis H1 y H2, es necesario considerar una tercera:

H3) $\left(\begin{bmatrix}f(x)+p(x)w\\ s(x)\end{bmatrix},h(x)+q(w)\right)$ es un par detectable en la aproximación lineal $(x,w)=(0,0)$.

En caso de cumplirse H1, H2 y H3 el PRNS por retroalimentación del error tiene solución si y sólo si $\exists\, x=\pi(w)$, $u=\gamma(w)$: $\pi(0)=0$, $\gamma(0)=0$ como mapeos en $W^0\subset W\supset 0$ tales que las ecuaciones en (2) se cumplan.

La ley de control está dada por

$$u(t)=\theta(\xi)=\gamma(\xi_1)+K^*\left(\xi_0-\pi(\xi_1)\right) \tag{6}$$

C. Modelado Takagi-Sugeno (TS) y LMIs

A continuación se definirá el modelo de la forma TS [18] por medio del enfoque sector no lineal [13], que es una representación exacta del modelo no lineal, es decir, una reescritura del mismo. Con esto se busca poder resolver ambos problemas de regulación no lineal de la salida por medio de LMIs. Suponga que (1) tiene p no-linealidades acotadas $nl_j(\cdot)\in\left[\underline{nl_j},\overline{nl_j}\right]$, $j\in\{1,\dots,p\}$ en una región compacta $\Delta\subset X\times W$ incluyendo el origen. Sea $z(\cdot)\in\mathbb{R}^p$ el vector de premisas en el cual las no-linealidades son expresadas (acotadas en una región compacta $\Delta\supset 0$, posiblemente dependiendo de x y w). Las no-linealidades descritas anteriormente son capturadas en los siguientes pesos

$$\omega_0^j(\cdot)=\frac{nl_j(\cdot)-\underline{nl_j}}{\overline{nl_j}-\underline{nl_j}},\quad \omega_1^j(\cdot)=1-\omega_0^j(\cdot), \tag{7}$$

$j\in\{1,\dots,p\}$, con los cuales se construyen las funciones de membresía (FMs) para $i\in\{1,\dots,2^p\}$, $i_j\in\{0,1\}$:

$$h_i=h_{1+i_1+i_2\times 2+\cdots+i_p\times 2^{p-1}}=\prod_{j=1}^{p}\omega_{i_j}^j(z_j) \tag{8}$$

Las FMs tienen la propiedad de suma convexa $\sum_{i=1}^{r}h_i(\cdot)=1$, $h_i(\cdot)\geq 0$ en Δ. Estas sumas, por simplicidad, se escribirán como $\Upsilon_z=\sum_{i=1}^{r}h_i\Upsilon_i$. Otras expresiones como la derivada de la suma convexa $\dot{\Upsilon}_z=\sum_{i=1}^{r}\dot{h}_i\Upsilon_i$ y la doble suma $\Upsilon_{zz}=\sum_{i=1}^{r}\sum_{j=1}^{r}h_i(z(t))h_j(z(t))\Upsilon_{ij}$ aparecerán a lo largo del trabajo.

Además, el símbolo $\prec(\succ)$ representa menor que (mayor que) elemento a elemento en expresiones relacionadas con matrices. También, $<(>)$ se usarán como definido negativo (positivo) en expresiones matriciales.

Lema de Relajación [19]: Sea Υ_{ij}, $i,j\in\{1,\dots,r\}$ una colección de matrices de tamaño adecuado. La doble suma convexa $\Upsilon_{zz}<0$ es garantizada se cumple si lo siguiente:

$$\begin{aligned}&\Upsilon_{ii}<0,\quad \forall i\in\{1,\dots,r\}\\ &\frac{2}{r-1}\Upsilon_{ii}+\Upsilon_{ij}+\Upsilon_{ji}<0,\quad \forall(i,j)\in\{1,\dots,r\}^2,\ i\neq j.\end{aligned} \tag{9}$$

D. Regulador Difuso

La representación exacta de (1) en Δ es:

$$\begin{aligned}\dot{x}(t)&=\sum_{i=1}^{r}h_i(z(t))\left(A_ix(t)+B_iu(t)+E_iw(t)\right)\\&=A_zx(t)+B_zu(t)+E_zw(t),\\ \dot{w}(t)&=\sum_{i=1}^{r}h_i(z(t))S_iw(t)\\&=S_zw(t),\\ e(t)&=\sum_{i=1}^{r}h_i(z(t))\left(C_ix(t)-Q_iw(t)\right)\\&=C_zx(t)-Q_zw(t)\end{aligned} \tag{10}$$

con $A_i\in\mathbb{R}^{n\times n}$, $B_i\in\mathbb{R}^{n\times m}$, $C_i\in\mathbb{R}^{o\times n}$, $E_i\in\mathbb{R}^{n\times q}$, $Q_i\in\mathbb{R}^{o\times q}$, $S_i\in\mathbb{R}^{q\times q}$, $i=1,\dots,r$, matrices de tamaño apropiado derivadas del enfoque de sector no lineal con $r=2^p\in\mathbb{N}$ el número de reglas.

En las siguientes expresiones, el argumento del tiempo será omitido donde sea conveniente.

Ahora, el PRNS tiene solución si y sólo si

$$\begin{aligned}\begin{bmatrix}\dot{\pi}(w)\\0\end{bmatrix}&=\sum_{i=1}^{r}h_i(z(t))\left(\begin{bmatrix}A_i&E_i\\C_i&-Q_i\end{bmatrix}\begin{bmatrix}\pi(w)\\w\end{bmatrix}+\begin{bmatrix}B_i\\0\end{bmatrix}\gamma(w)\right)\\&=\begin{bmatrix}A_z&E_z\\C_z&-Q_z\end{bmatrix}\begin{bmatrix}\pi(w)\\w\end{bmatrix}+\begin{bmatrix}B_z\\0\end{bmatrix}\gamma(w)\end{aligned}, \tag{11}$$

satisface los mapeos no lineales $x=\pi(w)$ y $u=\gamma(w)$. La expresión (11) es equivalente a las ecuaciones FIB (2).

En [11, 12] se asume una estructura de suma convexa para los mapeos no lineales

$$\begin{aligned}\pi(w(t))&=\sum_{j=1}^{r}h_j(z(t))\Pi_jw(t)=\Pi_zw(t)\\ \gamma(w(t))&=\sum_{j=1}^{r}h_j(z(t))\Gamma_jw(t)=\Gamma_zw(t)\end{aligned}, \tag{12}$$

con $\Pi_j\in\mathbb{R}^{n\times q}$, $\Gamma_j\in\mathbb{R}^{m\times q}$, $j\in\{1,\dots,r\}$. Y la reescritura de las leyes de control (3) y (6):

$$u(t)=\alpha(x,w)=\Gamma_zw+K_z(x-\Pi_zw) \tag{13}$$

En Teoría de Regulación la ganancia de estabilización K^* normalmente se diseña primero. En la literatura existen diversos trabajos que solventan este paso. Aquí se adoptará el enfoque Compensador Paralelo Distribuido, en donde $K^*=K_z$.

En caso de no tener toda la información completa, ya sea que alguno de los estados x o del exosistema w no estén disponibles, es necesario resolver el PRNS por retroalimentación del error. Reescribiendo (4) en la región Δ queda

$$\begin{aligned}\dot{\xi}(t) &= F_z\xi + G_z e \\ u(t) &= H_z\xi\end{aligned} \tag{14}$$

con $F_i \in \mathbb{R}^{(n+q)\times(n+q)}$, $H_i \in \mathbb{R}^{1\times(n+q)}$, $i \in \{1,\ldots,r\}$, $G_z \in \mathbb{R}^{(n+q)\times 1}$ definidos como sigue $F_z = \begin{bmatrix} A_z - G_z^0 C_z + B_z K_z & E_z + G_z^o Q_z + B_z\left(\Gamma_z - K_z\Pi_z\right) \\ -G_z^1 C_z & S_z + G_z^1 Q_z \end{bmatrix}$, $G_z = \begin{bmatrix} G_z^0 \\ G_z^1 \end{bmatrix}$ y $H_z = \begin{bmatrix} K_z & \Gamma_z - K_z\Pi_z \end{bmatrix}$. La ganancia G_z es diseñada para que los pares $\left(\begin{bmatrix} A_z & E_z \\ 0 & S_z \end{bmatrix}^T, \begin{bmatrix} C_z & -Q_z \end{bmatrix}^T\right)$ sean asintóticamente estables por medio del enfoque Compensador Paralelo Distribuido. Finalmente la ley de control tiene la forma

$$u(t) = \theta(\xi) = \Gamma_z\xi_1 + K_z\left(\xi_0 - \Pi_z\xi_1\right) \tag{15}$$

En la siguiente sección, la representación exacta TS del sistema no lineal permitirá trabajar con las ecuaciones FIB para resolverlas con técnicas de optimización convexa, es decir, en términos LMI. Resultados presentados anteriormente [11, 12, 16] quedarán como casos particulares de este nuevo enfoque.

III. RESULTADO PRINCIPAL

El siguiente teorema proporciona elementos para resolver el PRNS por medio de LMIs.

Teorema 1 [20]*:* Sean L, $R:\mathbb{R}^a \to \mathbb{R}^{b\times c}$ matrices lineales continuamente diferenciables funciones de un vector de decisión $x \in \mathbb{R}^a$. Si existe solución única $\bar{x} \in \mathbb{R}^a$ para el problema $L(x) = R(x)$, ésta solución es aproximada con una precisión arbitraria por el problema de minimización LMI elemento por elemento:

$$\min \varepsilon > 0 : -\varepsilon \prec L(x) - R(x) \prec \varepsilon .$$

Demostración: De $L(x) = R(x)$ se sigue que $L(\bar{x}) - R(\bar{x}) = \mathbf{0}$ con $\mathbf{0} \in \mathbb{R}^{b\times c}$. Por continuidad y unicidad de la solución $\bar{x}$, esto implica que $\forall\varepsilon > 0$ arbitrariamente pequeño, $\exists\delta > 0 : |x - \bar{x}| < \delta \Rightarrow |L(x) - R(x)| \prec \varepsilon$ con ε arbitrariamente pequeño. Esto lleva directamente a la deseada expresión LMI. □

Sustituyendo (12) en (11) permite reescribir las ecuaciones FIB como sigue:

$$\begin{bmatrix} \Pi_z\dot{w} + \dot{\Pi}_z w \\ 0 \end{bmatrix} = \begin{bmatrix} A_z & E_z \\ C_z & -Q_z \end{bmatrix}\begin{bmatrix} \Pi_z w \\ w \end{bmatrix} + \begin{bmatrix} B_z \\ 0 \end{bmatrix}\Gamma_z w \tag{16}$$

En resultados anteriores [11, 12] el problema de la derivada de las funciones de membresía que aparecen en $\dot{\Pi}_z$ no se considera por se asume $\dot{\Pi}_z = 0$, es decir, $\Pi_j = \Pi$, $\forall j \in \{1,\ldots,r\}$. También se consideran cotas que no son conocidas a priori. En [17, 21] se presenta una manera de tomar ventaja de la información que proporcionan las derivadas de las FMs en otro contexto; pero que mantiene propiedades LMI para ser solucionadas por medio de técnicas de optimización convexa [22, 23]. Los resultados presentados en el siguiente teorema utilizan estas ideas, con lo cual se abre la posibilidad de no despreciar a priori el término $\dot{\Pi}_z$.

Teorema 2: Suponga que H1 y H2 se cumplen, entonces el PRNS por retroalimentación del estado tiene solución si $\exists\, X = X^T \in \mathbb{R}^{n\times n}$, $M_j \in \mathbb{R}^{m\times n}$, $\Pi_j \in \mathbb{R}^{n\times q}$, $\Gamma_j \in \mathbb{R}^{m\times q}$, $j \in \{1,\ldots,r\}$, $\beta_k > 0$, $k \in \{1,\ldots,p\}$, y un $\varepsilon > 0$ arbitrariamente pequeño tales que:

$$\begin{gathered} X > 0, \quad A_z X + B_z M_z + X A_z^T + M_z^T B_z^T < 0, \\ -\varepsilon \prec \begin{bmatrix} A_z\Pi_z + E_z + B_z\Gamma_z - \Pi_z S_z \\ -\sum_{k=1}^{p}(-1)^{d_k^\alpha}\beta_k\left(\Pi_{g_1(z,k)} - \Pi_{g_2(z,k)}\right) \\ C_z\Pi_z - Q_z \end{bmatrix} \prec \varepsilon, \\ \left|\dot{\omega}_0^k\right| \le \beta_k \end{gathered} \tag{17}$$

con $g_1(j,k) = \left\lfloor (j-1)/2^{p+1-k} \right\rfloor \times 2^{p+1-k} + 1 + (j-1) \bmod 2^{p-k}$, $g_2(j,k) = g_1(j,k) + 2^{p-k}$, $\lfloor\cdot\rfloor$ la función *floor* y d_k^α obtenido de $\alpha - 1 = d_p^\alpha + d_{p-1}^\alpha \times 2 + \ldots + d_1^\alpha \times 2^{p-1}$, $\alpha \in \{1,\ldots,2^p\}$. La ley de control viene dada por (13) en donde $K_z = M_z X^{-1}$.

Demostración: Con el primer conjunto de expresiones se encuentra la ganancia de estabilización K_z a través de una función candidata de Lyapunov $V = x^T P x \ge 0$, $P = P^T > 0$ y una la ley de control $u = K_z x$ (sin perturbaciones w) se tiene el siguiente desarrollo:

$$\begin{aligned} \dot{V} &= x^T P\dot{x} + \dot{x}Px^T \\ &= x^T\left(P\left(A_z + B_z K_z\right) + \left(A_z^T + K_z^T B_z^T\right)P\right)x < 0 \\ &\Leftrightarrow P\left(A_z + B_z K_z\right) + \left(A_z^T + K_z^T B_z^T\right)P < 0 \\ &\Leftrightarrow A_z X + B_z M_z + X A_z^T + M_z^T B_z^T < 0 \end{aligned}$$

con $X = P^{-1}$ y $M_z = K_z P^{-1}$. Para la segunda parte de (17), se puede ver que las ecuaciones FIB en (16) pueden reescribirse, después de sustituir $\dot{w}$ y simplificar los términos comunes:

$$\begin{bmatrix} \Pi_z S_z + \dot{\Pi}_z \\ 0 \end{bmatrix} = \begin{bmatrix} A_z\Pi_z + E_z + B_z\Gamma_z \\ C_z\Pi_z - Q_z \end{bmatrix} \tag{18}$$

En [17] se probó que

$$\dot{\Pi}_z = \sum_{k=1}^{p} \dot{\omega}_0^k \left(\Pi_{g_1(z,k)} - \Pi_{g_2(z,k)} \right) \tag{19}$$

con $g_1(j,k)$ y $g_2(j,k)$ como ya se definieron anteriormente. Reemplazando (19) en (18) resulta:

$$\begin{bmatrix} A_z\Pi_z + E_z + B_z\Gamma_z - \Pi_z S_z \\ -\sum_{k=1}^{p} \dot{\omega}_0^k \left(\Pi_{g_1(z,k)} - \Pi_{g_2(z,k)} \right) \\ C_z\Pi_z - Q_z \end{bmatrix} = 0 \tag{20}$$

A partir del Teorema 1 se tienen condiciones suficientes para la ecuación matricial (20) se puede aproximar mediante el problema de minimización sobre un $\varepsilon > 0$ arbitrariamente pequeño, resulta la siguiente expresión:

$$-\varepsilon \prec \begin{bmatrix} A_z\Pi_z + E_z + B_z\Gamma_z - \Pi_z S_z \\ -\sum_{k=1}^{p} \dot{\omega}_0^k \left(\Pi_{g_1(z,k)} - \Pi_{g_2(z,k)} \right) \\ C_z\Pi_z - Q_z \end{bmatrix} \prec \varepsilon \tag{21}$$

Si $\left|\dot{\omega}_0^k\right| \le \beta_k$ para $\beta_k > 0$, entonces (21) resulta en (17) utilizando la propiedad $Y + \dot{\omega}_0^k Z < 0$ si $Y \pm \beta_k \times Z < 0$. □

Teorema 3: Suponga que H1, H2 y H3 se satisfacen, entonces el PRNS por retroalimentación del error tiene solución si $\exists\, X_1 = X_1^T \in \mathbb{R}^{n\times n}$, $X_2 = X_2^T \in \mathbb{R}^{(n+q)\times(n+q)}$ $M_{1j} \in \mathbb{R}^{m\times n}$, $M_{2j} \in \mathbb{R}^{m\times(n+q)}$, $\Pi_j \in \mathbb{R}^{n\times q}$, $\Gamma_j \in \mathbb{R}^{m\times q}$, $j \in \{1,\ldots,r\}$, $\beta_k > 0$, $k \in \{1,\ldots,p\}$, y un $\varepsilon > 0$ arbitrariamente pequeño tales que (17) y

$$\begin{aligned} &X_1 > 0, \quad A_z X_1 + B_z M_{1z} + X_1 A_z^T + M_{1z}^T B_z^T < 0, \\ &X_2 > 0, \quad \bar{A}_z X_2 - \bar{B}_z M_{2z} + X_2 \bar{A}_z^T - M_{2z}^T \bar{B}_z^T < 0, \\ &-\varepsilon \prec \begin{bmatrix} A_z\Pi_z + E_z + B_z\Gamma_z - \Pi_z S_z \\ -\sum_{k=1}^{p} (-1)^{d_k^\alpha} \beta_k \left(\Pi_{g_1(z,k)} - \Pi_{g_2(z,k)} \right) \\ C_z\Pi_z - Q_z \end{bmatrix} \prec \varepsilon, \\ &\qquad \left|\dot{\omega}_0^k\right| \le \beta_k \end{aligned} \tag{22}$$

con $g_1(j,k) = \left\lfloor (j-1)/2^{p+1-k} \right\rfloor \times 2^{p+1-k} + 1 + (j-1) \bmod 2^{p-k}$, $g_2(j,k) = g_1(j,k) + 2^{p-k}$, $\lfloor\cdot\rfloor$ la función *floor* y d_k^α obtenido de $\alpha - 1 = d_p^\alpha + d_{p-1}^\alpha \times 2 + \ldots + d_1^\alpha \times 2^{p-1}$, $\alpha \in \{1,\ldots,2^p\}$. La ley de control obtenida de (14) con $K_z = M_{1z} X_1^{-1}$ y $G_z = \left(M_{2z} X_z^{-1}\right)^T$.

Demostración: Al igual que en el teorema previo, se utiliza el una Función Candidata de Lyapunov para estabilizar los pares (A_z, B_z) y $(\bar{A}_z, \bar{B}_z)$ con $\bar{A}_z = \begin{bmatrix} A_z & E_z \\ 0 & S_z \end{bmatrix}^T$ y $\bar{B}_z = \begin{bmatrix} C_z & -Q_z \end{bmatrix}^T$. El resto de la demostración es idéntico al del teorema 2. □

Observación 1: Para resolver el PRNS las condiciones planteadas en cada teorema se resuelven simultáneamente, proporcionando un enfoque sistemático y gracias a las técnicas de optimización convexa.

Observación 2: Para obtener expresiones LMI de (17) y (22) hace falta aplicar el Lema de Relajación presentado anteriormente para quitar las sumas dobles implicadas haciendo una correcta selección de Υ_{zz}. En [24, 25] se muestran más relajaciones que igualmente pueden aplicarse.

Observación 3: Existen diferentes opciones para el diseño de K^*: a) una ganancia común $K^* = K$ que estabiliza el par $\left(f(x), g(x)\right)$ en $x = 0$; b) el enfoque del Compensador Paralelo Distribuido [11, 12, 13] en donde $K^* = K_z$ y una función cuadrática de Lyapunov $V = x^T P x$; c) Control Difuso No Cuadrático [12, 13, 17] con $K^* = K_z P_z^{-1}$ y una función difusa de Lyapunov $V = x^T P_z^{-1} x$. Las mismas opciones son válidas para la ganancia G^* del observador.

Observación 4: Las LMIs obtenidas de los teoremas 2 y 3, puede ser resueltas como un problema de minimización convexa sobre ε; además, esto representa una cota de cada elemento del error en estado estable $e_{ss} = C_z\Pi_z - Q_z$, como se puede ver de las expresiones (17) y (22), si el mapeo $x = \Pi_z w$ es alcanzado con suficiente precisión.

Observación 5: La cota sobre $\dot{\omega}_0^k$ es verificable a posteriori por medio de la simulación del sistema en lazo cerrado.

Observación 6: Para PRNS por retroalimentación del error se asume que el vector de premisas no puede depender de estados estimados (Caso 1, en [13]).

IV. EJEMPLOS

El primer ejemplo resuelve el PRNS cuando se tiene toda la información, es decir, a través de una retroalimentación del estado, resultando Π_i iguales. En el segundo, no se cuenta con toda la información, entonces se soluciona el PRNS por retroalimentación del error, donde $\dot{\Pi}_z \neq 0$ y es importante una elección correcta $\left|\dot{\omega}_0^k\right| \le \beta_k$.

Ejemplo 1: Sea el siguiente sistema no lineal:

$$\dot{x}(t) = \begin{bmatrix} -1 & x_2^2 \\ 0.76 & 0.5 \end{bmatrix} x(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t) + \begin{bmatrix} 0 & 0.5x_1^2 \\ 0 & 0 \end{bmatrix} w(t), \tag{23}$$

con $x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$ como vector de estado y $w = \begin{bmatrix} w_1 & w_2 \end{bmatrix}^T$ el vector del exosistema que proporcionará las referencias a seguir y perturbaciones a rechazar, cuya dinámica es

$$\dot{w}(t)=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}w(t) \tag{24}$$

Las premisas son $z_1=x_1$ y $z_2=x_2$, entonces el siguiente modelo TS representa exactamente a (23) y (24) en la región $\Delta=\{|x_i|\le 1,|w_i|\le 1\}\supset \mathbf{0}$:

$$\begin{aligned}\dot{x}(t)&=\sum_{i=1}^{4}h_i\left(z(t)\right)\left(A_i x(t)+B_i u(t)+E_i w(t)\right)\\ \dot{w}(t)&=\sum_{i=1}^{4}h_i\left(z(t)\right)S_i w(t),\end{aligned} \tag{25}$$

con $A_1=A_2=\begin{bmatrix}-1 & 1\\ 0.76 & 0.5\end{bmatrix}$, $A_3=A_4=\begin{bmatrix}-1 & 0\\ 0.76 & 0.5\end{bmatrix}$, $B_i=\begin{bmatrix}1\\ 0\end{bmatrix}$, $S_i=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}$, $i=1,\ldots,4$, $E_1=E_3=\begin{bmatrix}0 & 0.5\\ 0 & 0\end{bmatrix}$, $E_2=E_4=\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}$, $\omega_0^1=x_2^2$, $\omega_0^2=x_1^2$, $\omega_1^1=1-\omega_0^1$, $\omega_1^2=1-\omega_0^2$, $h_1=\omega_0^1\omega_0^2$, $h_2=\omega_0^1\omega_1^2$, $h_3=\omega_1^1\omega_0^2$ y $h_4=\omega_1^1\omega_1^2$.

La salida del sistema es $y=x_1$ y se requiere seguir a w_1, entonces $C_i=[1\ \ 0]$, $Q_i=[1\ \ 0]$, $i=1,\ldots 4$. Al aplicar el Teorema 2 y el Lema de Relajación sobre las sumas dobles, las ganancias estabilizadoras resultantes son $K_1=K_2=[-1.4259\ \ -5.6353]$, $K_3=K_4=[-1.4259\ \ -4.6353]$. Las cotas $\beta_k=1\times 10^8$, $k=1,2$ son verificadas a posteriori para cumplir $|\dot{\omega}_0^k|\le\beta_k$ y elegidas tan grandes como se pueda. Los mapeos difusos obtenidos son $\Pi_i=\begin{bmatrix}1 & 0\\ -0.304 & -0.608\end{bmatrix}$, $i=1,\ldots 4$, $\Gamma_1=[1.304\ \ 1.108]$, $\Gamma_2=[1.304\ \ 1.608]$, $\Gamma_3=[1\ \ 0.5]$ y $\Gamma_4=[1\ \ 1]$.

En este caso, tenemos Π_i iguales, por lo tanto la elección de β_k no juega un papel fundamental, ya que el término de (17) $-\sum_{k=1}^{p}(-1)^{d_k^\alpha}\beta_k\left(\Pi_{g_1(z,k)}-\Pi_{g_2(z,k)}\right)$ se ve claramente que es igual a cero al tener Π_i idénticas.

Los resultados en simulación muestran seguimiento exacto de la salida hacia su referencia (ver figura 2) para las condiciones iníciales $x(0)=[0.4\ \ 0.1]^T$ y $w(0)=[-0.8\ \ 0]^T$. Además puede verse que $\dot{\Pi}_z=0$ debido a que $\Pi_1=\Pi_2=\Pi_3=\Pi_4$.

Ejemplo 2: Sea el modelo no lineal

$$\begin{aligned}\dot{x}(t)=&\begin{bmatrix}0.3+0.027x_1^2 & 0.9\\ 0.3 & 0.4x_2^2\end{bmatrix}x(t)+\begin{bmatrix}0\\ 1\end{bmatrix}u(t)\\ &+\begin{bmatrix}0 & 0\\ 0 & 0.3x_2^2\end{bmatrix}w(t)\end{aligned} \tag{26}$$

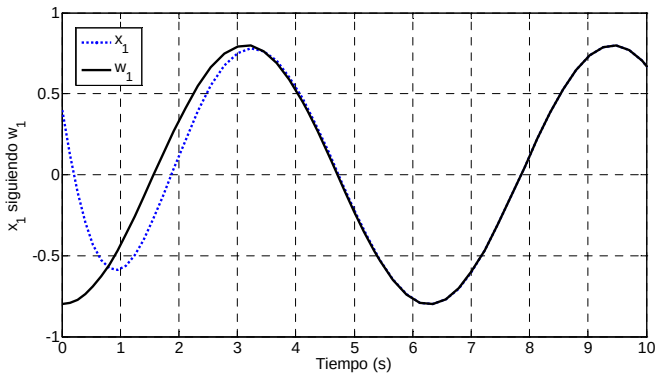


Fig. 2 x_1 siguiendo referencia de w_1 en ejemplo 1.

con $x=[x_1\ \ x_2]^T$ como vector de estado y $w=[w_1\ \ w_2]^T$ el vector del exosistema referencias y/o perturbaciones generadas por

$$\dot{w}(t)=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}w(t) \tag{27}$$

El vector de premisas está formado por $z_1=x_1$ y $z_2=x_2$, entonces el modelo TS (representación exacta) de (26) y (27) en $\Delta=\{|x_i|\le 1,|w_i|\le 1\}\supset \mathbf{0}$ es:

$$\begin{aligned}\dot{x}(t)&=\sum_{i=1}^{4}h_i\left(z(t)\right)\left(A_i x(t)+B_i u(t)+E_i w(t)\right)\\ \dot{w}(t)&=\sum_{i=1}^{4}h_i\left(z(t)\right)S_i w(t),\end{aligned} \tag{28}$$

con $A_1=\begin{bmatrix}0.327 & 0.9\\ 0.3 & 0.4\end{bmatrix}$, $A_2=\begin{bmatrix}0.327 & 0.9\\ 0.3 & 0\end{bmatrix}$, $A_3=\begin{bmatrix}0.3 & 0.9\\ 0.3 & 0.4\end{bmatrix}$, $A_4=\begin{bmatrix}0.3 & 0.9\\ 0.3 & 0\end{bmatrix}$, $B_i=\begin{bmatrix}1\\ 0\end{bmatrix}$, $S_i=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}$, $i=1,\ldots,4$, $E_1=E_3=\begin{bmatrix}0 & 0\\ 0 & 0.3\end{bmatrix}$, $E_2=E_4=\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}$, $\omega_0^1=x_1^2$, $\omega_0^2=x_2^2$, $\omega_1^1=1-\omega_0^1$, $\omega_1^2=1-\omega_0^2$, $h_1=\omega_0^1\omega_0^2$, $h_2=\omega_0^1\omega_1^2$, $h_3=\omega_1^1\omega_0^2$ y $h_4=\omega_1^1\omega_1^2$.

Teniendo $y=x_1$ y a $y_{ref}=w_1$, entonces $C_i=[1\ \ 0]$, $Q_i=[1\ \ 0]$, $i=1,\ldots 4$. Aplicando a (22) el Lema de Relajación las ganancias resultantes son: $K_1=\begin{bmatrix}-2.3578\\ -1.6417\end{bmatrix}^T$, $K_2=\begin{bmatrix}-2.3578\\ -1.2417\end{bmatrix}^T$, $K_3=\begin{bmatrix}-2.3893\\ -1.6570\end{bmatrix}^T$, $K_4=\begin{bmatrix}-2.3893\\ -1.2570\end{bmatrix}^T$,

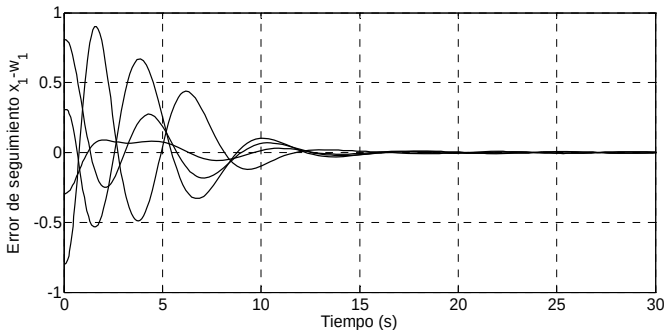


Fig.3 Error de seguimiento x_1-w_1 en ejemplo 2.

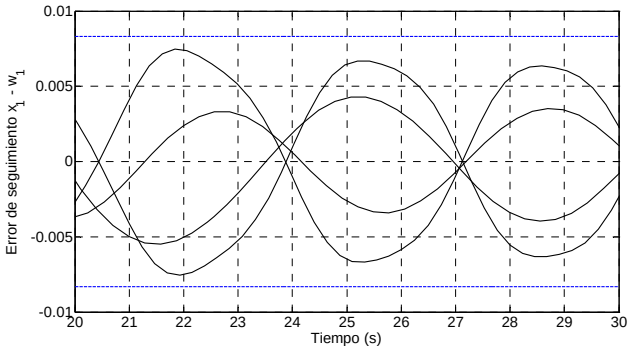


Fig.4 Vista aumentada del error de seguimiento en estado estable en ejemplo 2.

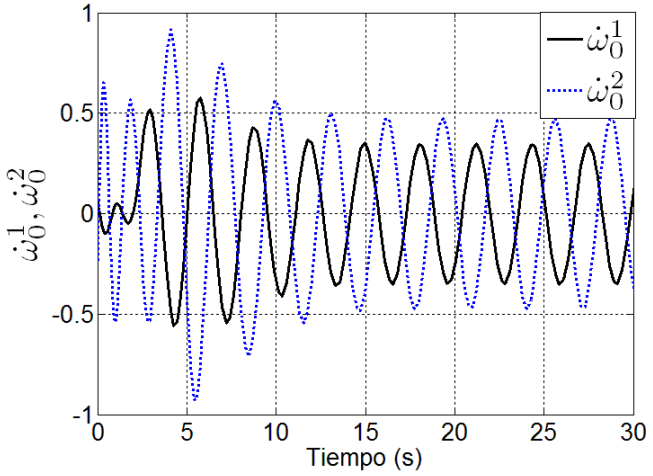


Fig.5 Verificación de las cotas β_k. en el ejemplo2.

$$G_1=\begin{bmatrix}10.9143\\13.3953\\5.0924\\-1.3855\end{bmatrix},\ G_2=\begin{bmatrix}10.8429\\13.2989\\5.0963\\-1.3832\end{bmatrix},\ G_3=\begin{bmatrix}10.7302\\13.2045\\5.0082\\-1.3699\end{bmatrix}\ \text{y}\ G_4=\begin{bmatrix}10.6588\\13.1081\\4.9720\\-1.3676\end{bmatrix}.$$

Las cotas propuestas sobre $\dot{\omega}_0^k$ son $\beta_1=1.5$ y $\beta_2=5$ son verificadas en simulación. Tras el cálculo se comprueba que no hay seguimiento exacto y que $\varepsilon=0.0083$ es el valor mínimo de ε y será la cota del error en estado estable.

Los mapeos calculados son: $\Pi_1=\Pi_2\begin{bmatrix}0.9937 & -0.0026\\-0.3489 & 1.1050\end{bmatrix}$, $\Pi_3=\Pi_4\begin{bmatrix}0.9937 & 0.0029\\-0.3436 & 1.1031\end{bmatrix}$, $\Gamma_1=\begin{bmatrix}-1.2640\\-1.0900\end{bmatrix}^T$, $\Gamma_2=\begin{bmatrix}-1.4027\\-0.3483\end{bmatrix}^T$, $\Gamma_3=\begin{bmatrix}-1.2634\\-1.0859\end{bmatrix}^T$ y $\Gamma_4=\begin{bmatrix}-1.4016\\-0.3444\end{bmatrix}^T$.

Observe que, para este caso aparece $\dot{\Pi}_z\neq 0$ además de que el $\varepsilon_{ss}\le\varepsilon=0.0083$ corresponde a la cota del error de seguimiento en estado estable. La figura 3 muestra la simulación de los resultados obtenidos para varias condiciones iníciales, una vista más cercana del error en estado estable se tiene en la figura 4, en donde las líneas punteadas representan la cota ε y en líneas continuas el valor del error en estado estable para distintas condiciones iníciales. ´Finalmente, la figura 5 ilustra que $\dot{\omega}_0^k$ cumple con las cotas previamente definidas para las condiciones iníciales $x(0)=[0.1\ \ 0.1]^T$, $w(0)=[-0.6\ \ 0]^T$ y $\xi(0)=[0.4\ \ 0.1\ \ -0.3\ \ 0]^T$.

V. CONCLUSIONES

En este trabajo se han presentado condiciones para resolver el problema de regulación difusa. En él se muestra un nuevo método que está completamente basado en LMIs y modelos TS. La principal novedad de este trabajo es la inclusión de las derivadas de las funciones de membresía que aparecen debido a las ecuaciones FIB y a las representaciones convexas de los mapeos, además el PRNS es resuelto de manera sistemática a través de métodos de optimización convexa implementados en software disponible comercialmente. Al tomar en cuenta las derivadas de las FMs, este enfoque engloba resultados anteriores, dejándolos como casos particulares. Con los ejemplos mostrados, se prueba la efectividad del enfoque propuesto. Como trabajo a futuro se plantea investigar la manera de garantizar las cotas sobre las derivadas de las funciones de membresía a priori para el Teorema 2 así como reducir las limitaciones mostradas en el modelado de los mapeos no lineales.

REFERENCIAS

[1] B. A. Francis, "The linear multivariable regulator problem", SIAM Journal of Control and Optimization, Vol. 15, pp. 486-505, 1977.
[2] A. Isidori, "Nonlinear control systems", Springer, Berlin, 1995.
[3] J. Huang, W.J. Rugh, "On the nonlinear multivariable servomechanism problem", Automatica, Vol. 26, pp. 963-972, 1990.
[4] C. I. Byrnes, F. Delli Priscoli and A. Isidori, "Output regulation of uncertain nonlinear systems", Birkhäuser, Boston, 1997.
[5] A. Isidori, C.I. Byrnes, "Output regulation of nonlinear systems", IEEE Trans. on Automatic Control, Vol. 35, pp. 131-140, 1990.
[6] L.E. Ramos-Velasco, S. Čelikovský, V. Kučera, V. López-Morales, "Generalized output regulation problema for a class of nonlinear systems using error feedback", American Control Conference, pp. 1325-1330, Portland, USA, Junio, 2005

[7] J. Carr, "Aplications of centre manifold theory", Springer-Verlag, New York, 1981.
[8] L.E. Ramos-Velasco, "Control de un sistema electromecánico subactuado (Pendubot)" (in Spanish), M.Sc Thesis, CINVESTAV GDL, 1999.
[9] M. Bernal, R. Marquez, V. Estrada-Manzo, B. Castillo-Toledo, "An element-wise linear matrix inequality approach for output regulation problems", artículo aceptado en World Automation Congress, Puerto Vallarta 2012.
[10] X.-J. Ma, Z.-Q- Sun, "Output tracking and regulation of nonlinear system based on Takagi-Sugeno fuzzy model" IEEE Trans. on Systems, Man and Cybertics Part B, Vol. 30(1) pp. 47-59. Feb 2000.
[11] B. Castillo-Toledo, J.A. Meda-Campaña, A. Titli, "A fuzzy output regulator for Takagi-Sugeno fuzzy models", Proc. 2003 IEEE Internat. Symp. On Intelligent Control, Vol. 2, pp. 310-315, Houston, TX, Dec. 2003.
[12] J.A. Meda-Campaña, B. Castillo-Toledo, G. Chen, "Synchronization of chaotic systems from a fuzzy regulation approach", Fuzzy Sets and Systems, Vol. 160, pp. 2860-2875, 2009.
[13] K. Tanaka and H.O. Wang, Fuzzy control systems design and analysis. A linear matrix inequality approach. John Wiley and Sons, New York, USA. 2001.
[14] K. Tanaka, T. Hori, H.O. Wang, "A multiple Lyapunov function approach to stabilization of fuzzy control systems", IEEE Trans. on Fuzzy Systems, Vol. 11 (4), pp 582-589, 2003.
[15] M. Bernal, P. Hušek, V. Kučera, "Non quadratic stabilization of continuous-time systems in the Takagi-Sugeno form", Kybernetika, vol. 42 (6), pp. 665-672, 2006.
[16] J.A. Meda-Campaña, J.C. Gómez-Mancilla, B. Castillo-Toledo, "On the exact output regulation for Takagi-Sugeno fuzzy systems", In Proc. of the 8th IEEE Conf. on Control and Automation, pp. 417-422, Xiamen, China, 2010.
[17] T.M. Guerra and M. Bernal, "A way to escape from the quadratic framework", in Proc. FUZZ-IEEE Conference, pp. 784-789, Jeju, Korea, 2009.
[18] T. Taniguchi, K. Tanaka, and H.O. Wang, "Model construction, rule reduction and robust compensation for generalized form of Takagi-Sugeno fuzzy systems", IEEE Trans. on Fuzzy Systems, vol.9 (4), pp. 525-537, 2001.
[19] H.D. Tuan, P. Apkarian, T. Narikiyo, and Y. Yamamoto, "Parameterized linear matrix inequality techniques in fuzzy control system design", IEEE Trans. on Fuzzy Systems, vol. 9 (2), 2001, pp. 324–332.
[20] M. Bernal, R. Marquez, V. Estrada-Manzo, B. Castillo-Toledo, "Nonlinear outpur regulation via Takagi-Sugeno fuzzy mappings: a full-information LMI approach", artículo aceptado en IEEE World Congress on Computational Intellligence, Brisbane, Australia 2012.
[21] M. Bernal and T.M. Guerra, "Generalized non-quadratic stability of continuous-time Takagi-Sugeno models", IEEE Trans. on Fuzzy Systems, vol. 18 (4), 2010, pp 815-822.
[22] S. Boyd, L.E. Ghaoui, E. Feron, V. Balakrishnam, "Linear matrix inequalities in systems and control theory", SIAM, Philadelphia, PA, 1994.
[23] C. Scherer, "Linear matrix inequalities in control theory", Publicly available from Delft University, 2004.
[24] E. Kim and H. Lee, "New approaches to relaxed quadratic stability condition of fuzzy control systems", IEEE Trans. On Fuzzy Systems, vol. 8 (5), pp. 523-533, 2000.
[25] A. Sala and C. Ariño, "Asymptotically necessary and sufficient conditions for stability and performance in fuzzy control: Applications of Polya's theorem" Fuzzy Sets and Systems, Vol. 158, (24), pp 2671-2686, 2007.

Víctor Estrada M. nació en Zamora, Michoacán en 1987. Es ingeniero mecatrónico por la Universidad de Guadalajara, Lagos de Moreno, Jalisco, en 2009.

Él estudia el segundo año de la maestría en ciencias en ingeniería eléctrica área control automático en el Centro de Investigación y de Estudios Avanazodos del IPN, unidad Guadalajara en Jalisco. Sus áreas de interés son el control de sistemas no lineales, modelado Takagi-Sugeno, desigualdades matriciales lineales.

Bernardino Castillo T. nació en Cd. Ixtepec, Oaxaca. Obtuvo el grado de Doctor en Ciencias Especialidad Control Automático por la Universidad de Roma "La Sapienza" en 1992.

Él es profesor investigador en el Departamento de Ingeniería Eléctrica y Ciencias Computacionales del Centro de Investigación y de Estudios Avanazodos del IPN, en Guadalajara, Jalisco. Sus líneas de investigación: análisis y síntesis de esquemas de control para sistemas, control de procesos por computadora, control de robots.

El Dr. Castillo pertenece al Sistema Nacional de Investigadores con nivel 2 y *Senior Member* en la IEEE.

Miguel A. Bernal R. nació en Guadalajara, Jalisco en 1976. Obtuvo el grado de maestro en ciencias en ingeniería eléctrica por el CINVESTAV unidad Guadalajara en 1999 y el grado de Doctor en Ciencias en control automático por la Czech Technical University, Praga, República Checa en 2005.

Actualmente es profesor investigador categoría B en el Instituto Tecnológico de Sonora, en Ciudad Obregón, Sonora. Del 2006 al 2009 realizó una estancia postdoctoral con el grupo de investigación de sistemas, modelado y control del Laboratorie d' Automatique, de Mécanique et d'Informatique Industrielles et Humaines, Université de Valenciennes et du Hainaut-Cambrésis, Francia. Ha sido profesor en diversas universidades en México. Sus líneas de investigación abarcan análisis y modelado de sistemas no lineales a través de programación convexa para modelos difusos y neuronales.

El Dr. Bernal pertenece al Sistema Nacional de Investigadores con nivel 1.

A design-of-experiments approach to maximize the reliability of adjustable speed drives

Calleja G. Hugo, Chan-Puc Freddy, Toral Homero, Torres M. Emmanuel y Sánchez H. Víctor

Un diseño de experimentos enfocado a maximizar la confiabilidad de *drives* de velocidad ajustable

***Abstract*— There are several proposals to improve the reliability of adjustable speed drives. One is to use an active rectifier and a small-valued capacitor. Another is to use switching patters aimed at reducing the current ripple flowing through the capacitor. A third one is to closely monitor the capacitors, in order to detect its degradation. Manufacturers claim, however, that modern capacitors can perform satisfactorily in this application. This paper presents a methodology to select the design specifications for a standard ASD, in such a way that the capacitor operational life can be maximized. It is based on the design of experiments technique, a statistical tool whose advantage is that the parameters with the highest impact on the operational life can easily be identified at the design stage, and dealt with in a systematic manner.**

The selecting procedure of an electrolytic capacitor is already well known, however there is no explicit link between the electrical variables and the variables of reliability. As a solution to this difficulty is presented a method for relating the electrical behavior with the capacitor operational life.

***Keywords*— Reliability, Drives, Design methodology.**

I. INTRODUCTION

For a long time, a major source of concern in adjustable speed drives (Fig.1) has been the large electrolytic capacitor connected across the DC-link. The current flowing through the capacitor includes harmonics produced both by the rectifier at the line side, and the inverter at the load side. The temperature within the capacitor rises as the current flows through the equivalent series resistance (ESR), degrading the capacitor characteristics and shortening the lifespan [1][2].

Several approaches have been followed in order to improve the reliability. One approach is to drastically reduce the capacitor to values so small that either film or ceramic capacitors can be used, instead of electrolytic ones. Although its feasibility has already been demonstrated, this approach requires an active rectifier as the first stage [3]. Furthermore, the complete elimination of capacitors across the DC link can be achieved, but at the expense of a voltage with a high ripple [4]. Besides the obvious complexity of the added hardware, such converters are highly susceptible to transient over or under voltages on the DC link [5]. Even minor grid voltage imbalance can cause a large second harmonic ripple to appear in the link voltage, possibly interfering with the load. These conditions impose higher restrictions on the control system [6].

Since the temperature rise is produced by the current flowing through the ESR, a second approach is to use optimized switching patterns aimed at reducing the current stresses on the capacitors. The drawback, however, is that the requirements of a current with a lower harmonic content sometimes are in conflict with other operational requirements [7][8].

A third approach followed is to closely monitor the capacitors, in order to timely detect its degradation [9]-[11]. This approach does not provide a higher reliability. Instead, it is focused in avoiding catastrophic failures, thus shortening the down time, and providing a better availability.

An ASD for applications not requiring regenerative operation might benefit from the simplest configuration, with an uncontrolled rectifier as front-end. The advantages are that the requirements imposed on the controller block are greatly simplified, and that the reliability attained with a diode-based rectifier is much higher than that exhibited by an active rectifier built with transistors. In fact, several semiconductor manufacturers offer integrated modules containing the uncontrolled rectifier and a three-phase inverter in a single package (shown within the dotted box in Fig. 1). The input and output stages are unconnected, and an LC filter can be connected in between.

The capacitor C_F is the one blamed as the major source of failures. Manufacturers claim, however, that capacitors for this application have longer lives than previously thought [12]. A common reliability prediction procedure is outlined in the MIL HDBK217-F handbook [13]. The manufacturers' argument is that the handbook relies on statistical data collected throughout the years, and such data might not apply to current-technology capacitors, hence yielding misleading reliability predictions. A study recently published seems to confirm the statement [14].

The design procedure for the LC filter is already well known,

Manuscrito recibido el 21 de abril de 2012. Este trabajo presenta parte de los resultados obtenidos durante la estancia posdoctoral realizada por el Dr. Chan en el periodo julio 2011- agosto 2012y financiada por la Universidad de Quintana Roo.

Calleja G. Hugo hasta la fecha se ha desempeñado como Investigador en el Centro Nacional de Investigación y Desarrollo Tecnológico, e-mail hcalleja@cenidet.edu.mx.

Chan-Puc Freddy hasta la fecha se ha de desempeñado como Profesor investigador en la Universidad de Quintana Roo en la División de Ciencias e Ingenierías; e-mail freddy@uqroo.mx .

Toral Homero hasta la fecha se ha de desempeñado como Profesor investigador en la Universidad de Quintana Roo/DCI e-mail htoral@uqroo.mx.

Torres M. Emmanuel hasta la fecha se ha de desempeñado como Profesor investigador en la Universidad de Quintana Roo/DCI e-mail etorres@uqroo.mx.

Sánchez H. Victor hasta la fecha se ha de desempeñado como Profesor investigador en la Universidad de Quintana Roo/DCI e-mail vsanchez@uqroo.mx.

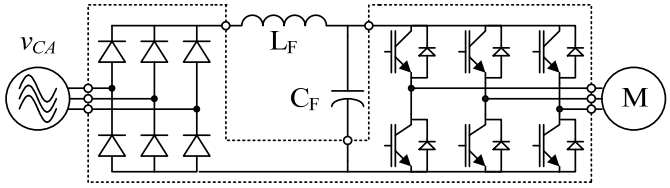


Fig. 1 Standard adjustable speed drive with an electrolytic capacitor in the DC-link.

TABLE I. Expected lifetime for PEH200 electrolytic capacitors, at T_M = 85°C [18]

Diameter (mm)	35	50	65	75	90
L_B (10^3 hours)	20	24	30	40	60

and several guidelines can be followed: specifying the amount of voltage ripple, the cut-off frequency, or the damping factor, among others [15]. The drawback is that the procedure usually employs a desired operating point as input data, without including the capacitor reliability as a design parameter. As a result, the designer does not know what effect a change in the quality factor, for instance, will have on the operational life. There have been a few attempts to provide design criteria that take into account the size and the capacity of the DC-link capacitor [16]. The proposals, however, do not ensure that reliability can be maximized because it is not explicitly included. The most advantageous, least expensive approach is to introduce reliability as early as possible in a product life cycle [17].

This paper presents a methodology to select an operating point that maximizes the capacitor operational life. It is based on the design-of–experiments technique, a statistical tool whose advantage is that the parameters with the highest impact on the operational life, under normal operating conditions, can easily be identified at the design stage. Once the effect of these parameters is identified, the filter can be improved to comply with a reliability target. The analysis is performed for a 5 kW ASD, and four optimization parameters are taken into account: the cut-off frequency f_C, and the quality factor Q of the filter, the capacitor voltage rating V_R, and the carrier frequency f_{PWM} used to modulate the inverter.

II. CAPACITOR OPERATIONAL LIFE

The capacitor operating life can be mathematically described as [12]:

$$L_{OP} = L_B \, f_1(V) \, f_2(\Delta T) \tag{1}$$

where L_B is the expected lifetime at the rated hot-spot temperature T_M, and depends on the capacitor diameter, as can be seen in table I.

The term f_1 takes into account the voltage stress applied to the capacitor, and is given by:

$$f_1(V) = 4.3 - 3.3\frac{V_A}{V_R} \tag{2}$$

where V_A is the voltage applied to the capacitor, and V_R is the rated voltage. In turn, f_2 is a function of the internal temperature rise, and is expressed as:

$$f_2(\Delta T) = 2^{(TM - Th)/B} \tag{3}$$

where B is a constant that depends on the capacitor construction, and T_h is the actual hot-spot temperature. The term T_h depends on the power dissipated by the capacitor which, in turn, depends on the ESR, and on the current flowing through the capacitor. Further, the ESR exhibits a highly nonlinear behavior, as can be noticed in Fig. 2. The plot corresponds to a 4700 μF electrolytic capacitor rated at 450 V [19].

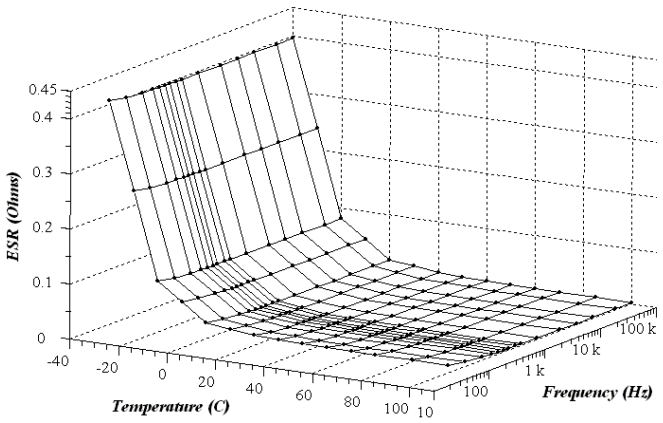


Fig. 2 ESR as a function of the hot-spot temperature, and the frequency.

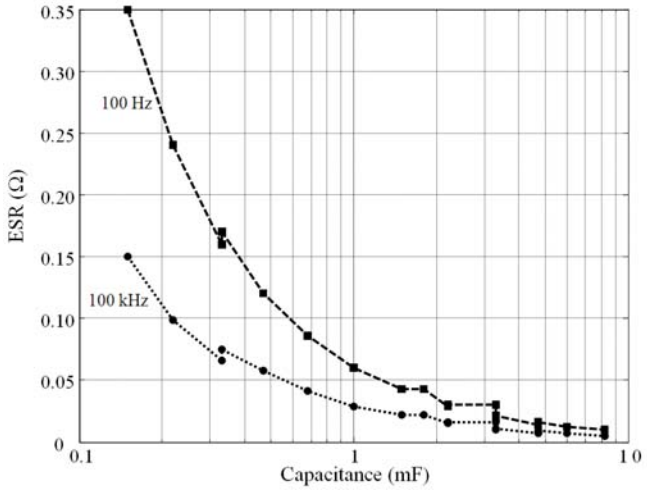


Fig. 3 ESR as a function of the capacitance value.

The ESR also depends on the capacitance value, as can be seen in Fig. 3. The graph corresponds to capacitors rated at 450 V, from the PEH200 family, and shows that the ESR is inversely proportional to the capacitance value. Fig. 4 illustrates the ESR as a function of the rated voltage, for 4700 μF capacitors from the same family. Both graphs correspond to an ambient temperature equal to 20°C.

III. DESIGN-OF-EXPERIMENTS BASED ANALYSIS

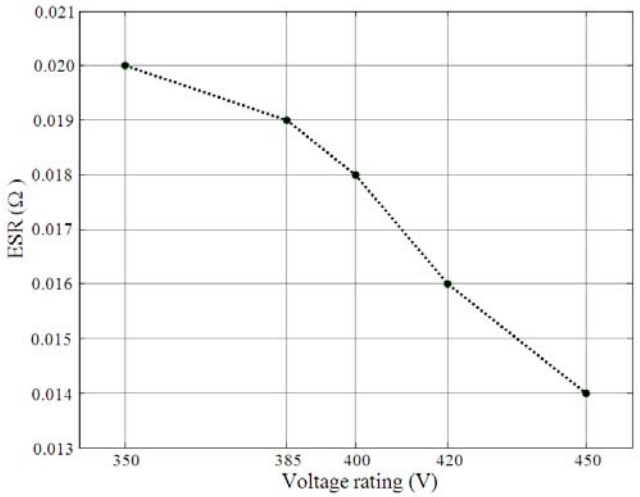


Fig. 4 ESR as a function of the rated voltage.

TABLE II. MINIMUM AND MAXIMUM VALUES OF THE DESIGN PARAMETERS

Parameter	Minimum	Maximum
f_C [Hz]	50	100
Q	2.4	5
V_R [V]	350	450
f_{PWM} [Hz]	900	4500

The Design-of-Experiments (DOE) technique analyzes the simultaneous effect of several factors on an output variable. When properly planned and executed, it helps obtain the maximum amount of information with the minimum amount of work. Using a factorial design all the factors are varied at once, and experiments (or evaluations) are performed for all combinations of levels for all of the factors. This procedure reveals what the effect of one variable is when the other factors are changing.

Interaction occurs when the effect on the response of a change in the level of one factor from low to high depends on the level of another factor. In other words, when an interaction is present between two factors, the combined effect of those two factors on the response variable cannot be predicted from the separate effects. The effect of two factors acting in combination can either be greater (synergy) or less (interference) than would be expected from each factor separately. In a 2-level factorial design, each factor is assigned a minimum and a maximum value. Let n be the number of factors. The 2-level factorial design requires that the experiment be performed $2n$ times. A fractional factorial design requires fewer experiments, and there is a simple relationship for the minimum number of runs required: round up the number of factors to a power of two and then multiply by two [20].

In the current case the output variable is the capacitor operating life, and the factors are the following four design parameters ($n = 4$):

- The cut-off frequency f_C of the filter. It should be low enough to provide a ripple-free DC voltage, and to prevent the backward propagation of the current harmonics generated by the inverter.

- The quality factor Q of the filter. Its value should provide a suitable transient response.

TABLE III. DESIGN MATRIX WITH MINIMUM AND MAXIMUM VALUES

m	f_C [Hz]	Q	V_R [V]	f_{PWM} [kHz]	Lop [10^3 hrs]
1	50	2.4	350	0.9	670
2	50	2.4	450	4.5	1492
3	50	5	350	4.5	326
4	50	5	450	0.9	550
5	100	2.4	350	4.5	657
6	100	2.4	450	0.9	567
7	100	5	350	0.9	70
8	100	5	450	4.5	157

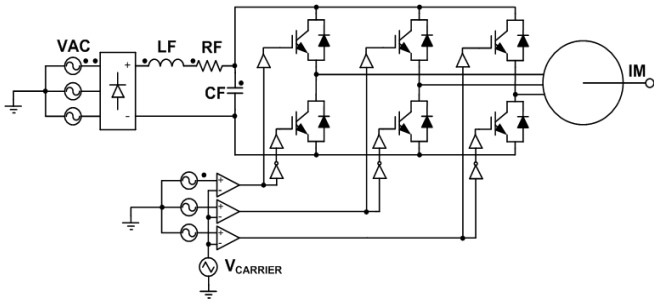


Fig. 5 Circuit schematic.

- The capacitor voltage rating V_R.

- The carrier frequency f_{PWM} used to modulate the inverter. Besides the harmonics at $6\ h\ f_{LINE}$, the current flowing through the capacitor will include components related to this frequency.

According to the DOE technique, it is necessary to select a minimum and a maximum value for each parameter. The selected values are listed in table II. It will be assumed that the voltage at the DC-link is $V_{DC} = 280$ V, the output power is $P_O = 5$ kW, $R_F = R_L/100$, and the ambient temperature is $T_A = 40$C.

A. Design matrix

The first step is to setup an appropriate design matrix, each row containing a different combination of the design parameters. Each combination constitutes an entry, and the operating life will be calculated as many times as entries are in the matrix. Since $n = 4$, if the fractional factorial design strategy is followed, then there are 8 combinations ($m = 8$), as shown in table III**.**

B. Operating life calculation

The inductor and the capacitor are calculated following the procedure described in the previous section, with the input parameters listed in the second and third columns in the design matrix. The resulting circuits are simulated in PSIM using the schematic shown in Fig. 5. A PWM modulation with a triangular carrier is used, and its value is given by the fifth column in the design matrix. The model of a symmetrical 3-phase squirrel-cage induction machine was used in the simulations. The outputs are a set of vectors containing the current flowing through capacitor

TABLE IV. PARTIAL AVERAGES AND SUM OF SQUARES FOR DESIGN PARAMETERS

	f_C [Hz]	Q	V_R [V]	f_{PWM} [kHz]
Avg(max)	363	276	692	658
Avg(min)	760	847	431	464
ΔAvg	-397	-571	260	194
SS	314454	651223	135632	75055

C_F. The current vectors are exported to Matlab, to calculate the hot-spot temperature. following the procedure described in [21].

The inputs for the calculation are the current vector, the ESR matrix for the capacitor, and the corresponding thermal resistance R_θ. The operating life *Lop* is calculated using equations (1), (2) and (3). *Lb,* T_M and *B* are obtained from manufacturer data, and $V_A = V_{DC}$. The results are listed in the rightmost column of the design matrix.

C. Variance analysis

The following procedure follows the guidelines provided by Hicks [22]. The overall arithmetic average $\overline{Lop}$ should be calculated using:

$$\overline{Lop} = \frac{1}{n}\sum_{j=1}^{n} Lop_j \tag{4}$$

For the operational lives listed in the rightmost column, table II, $\overline{Lop}$ = 561.3 x 10^3 hours.

Let *par* represent any of the design parameters. Eight partial averages should be calculated, two for each variable *par*: one with the *Lop* values obtained when the variable is at its maximum, and another with the values obtained when the variable is at its minimum. That is:

$$Avg(max)_{par} = \frac{1}{4}\sum_{j=1}^{4} Lop_j \,|_{par=max} \tag{5}$$

$$Avg(min)_{par} = \frac{1}{4}\sum_{j=1}^{4} Lop_j \,|_{par=min} \tag{6}$$

The difference is:

$$\Delta Avg_{par} = Avg(max)_{par} - Avg(min)_{par} \tag{7}$$

and the corresponding Sum-of-Squares SS_{par} is:

$$SS_{par} = 2\,\Delta Avg_{par}{}^2 \tag{8}$$

The results are listed in table IV. As an example, the *Avg(max)* for f_C is obtained averaging the *Lop* values for rows *m* = 4 through 8 in the design matrix.

As previously stated, the term "interaction" refers to the combined effect of two variables on the output. At this point, the number of interactions n_I is selected considering the non-repetitive, binary combinations of two variables. There are three interactions in an analysis with four variables. The first interaction is "$f_C\,Q$ OR $V_R f_{PWM}$". The second one is "$f_C\,V_R$ OR $Q\,f_{PWM}$". The third is "Q V_R OR $f_C\,f_{PWM}$". Further, there will be a high value H, and a low value L associated with each interaction. Let *int* represent any of the interactions. At this point it is necessary to calculate the same terms that were calculated for the variables. That is:

TABLE V. PARTIAL AVERAGES AND SUM OF SQUARES FOR INTERACTIONS

			Interactions		
m		*Lop* [10^3 hrs]	$f_C\,Q$ OR $V_R f_{PWM}$	$f_C\,V_R$ OR $Q f_{PWM}$	$Q\,V_R$ OR $f_C f_{PWM}$
1	▪▪▪	670	H	H	H
2	▪▪▪	1492	H	L	L
3	▪▪▪	326	L	H	L
4	▪▪▪	550	L	L	H
5	▪▪▪	657	L	L	H
6	▪▪▪	567	L	H	L
7	▪▪▪	70	H	L	L
8	▪▪▪	157	H	H	H
		Avg(H) (10^3)	597	430	508
		Avg(L) (10^3)	525	692	614
		ΔAvg (10^3)	72	-262	-105
		SS (10^6)	10358	137708	22176

$$Avg(max)_{int} = \frac{1}{4}\sum_{j=1}^{4} Lop_j \,|_{int=H} \tag{9}$$

$$Avg(min)_{int} = \frac{1}{4}\sum_{j=1}^{4} Lop_j \,|_{int=L} \tag{10}$$

$$\Delta Avg_{int} = Avg(max)_{int} - Avg(min)_{int} \tag{11}$$

$$SS_{\text{int}} = 2\,\Delta Avg_{\text{int}}{}^2 \tag{12}$$

The numerical results are listed in table V. As an example, the term *Avg(H)* for the interaction $f_C\,Q$ OR $V_R f_{PWM}$ is calculated averaging rows 1, 2, 7 and 8.

The error ε is calculated using:

$$\varepsilon = \sum_{j=1}^{3} (SS_{\text{int}})_j \tag{13}$$

The values listed in table V yield ε = 170242 x 10^6.

There are several degrees of freedom: DF_{par} for parameters, DF_{int} for interactions, and DF_{ε}. for the error. Since each parameter takes only two different values, the following simplifications can be applied: DF_{par} = 1, DF_{int} = 1 and $DF_{\varepsilon} = n_I$ = 3. The next step is to obtain the mean square MS values for the parameters, the interactions, and the error. With the aforementioned simplifications, the mean square value for variables is $MS_{par} = SS_{par}$. The mean square value for interactions is $MS_{int} = SS_{int}$. The mean square value for the error is:

TABLE VI. RATIO F_{PAR} FOR THE DESIGN VARIABLES

	f_C [Hz]	Q	V_R [V]	f_{PWM} [kHz]
F_{par}	5.54	11.48	2.39	1.32

$$MS_{\varepsilon} = \frac{\varepsilon}{DF_{\varepsilon}} = 56750 \times 10^{6} \tag{14}$$

For each parameter, the ratio F_{par} is calculated as:

$$F_{par} = \frac{MS_{par}}{MS_{\varepsilon}} \tag{15}$$

The numerical results are listed in table VI.

Let α represent the level at which the designer is willing to risk in concluding that a significant effect is not present when in actuality it is. The term $F(\alpha,DF_{par},DF_{\varepsilon})$ corresponds to the critical value of the statistical *F*-distribution, and is tabulated in most statistical books [23]. If the calculated F_{par} ratio is greater than the tabulated value of $F(\alpha,DF_{par},DF_{\varepsilon})$, then the parameter does have a significant effect on the operational life, and should be included in the optimization procedure.

Let α = 0.2 (that is, a 20% risk level). The tabulated value is $F(\alpha,DF_{par},DF_{\varepsilon}) = 2.05$. Comparing this value with those listed in table VI, it turns out that $F_{PWM} < 2.05$. Therefore, the modulation frequency has a minor effect on *Lop*, and can be excluded from further consideration. Also, the largest F_{par} corresponds to the variable with the highest impact on the operational life which, in this case, is the quality factor *Q*. Once the variables with the highest impact have been identified, the next step is to find out how should the variables be changed in order to improve *Lop*. The following prediction equation can be used for this purpose:

$$Lop = \overline{Lop} + \sum_{par=1}^{3} \frac{\Delta Avg_{par}}{\mathrm{F}_{(\alpha,\mathrm{DF}_{\mathrm{par}},\mathrm{DF}_{\varepsilon})}} \tag{16}$$

where $\overline{Lop}$ is calculated with equation (4). The summation includes the three terms in table VI larger than $F(\alpha,DF_{par},DF_{\varepsilon})$. Using the previously calculated values:

$$Lop = 561.3 - \frac{397}{2.05}\bigg|_{fC} - \frac{571}{2.05}\bigg|_{Q} + \frac{260}{2.05}\bigg|_{VR}$$

$$Lop = 561.3 - 193_{fC} - 278_{Q} + 127\big|_{VR} \tag{17}$$

To improve the operational life, the variables with positive coefficients should be increased, and the variables with negative coefficients should be reduced. According with these results, in order to improve the capacitor operating life, both the quality factor and the cut-off frequency should be reduced, and the voltage rating should be increased.

IV. DISCUSSION

It is instructive to compare the operational lives obtained with the different design values listed in table VII, representing a gradual optimization procedure. The first row corresponds to the shortest operating life, and will be used as a reference. The rightmost column lists the improvement obtained, when compared with the previous step.

TABLE VII. OPERATIONAL LIVES COMPARISON

f_C [Hz]	Q	V_R [V]	f_{PWM} [kHz]	Lop [10^3 hrs]	ΔLop [10^3 hrs]
100	5	350	900	85	---
100	5	350	4.5	116	33
100	5	450	4.5	157	41
50	5	450	4.5	442	285
50	2.4	450	4.5	1492	1050

The first row corresponds to a worst-case design, and the operational life attained is rather short. In the second row f_{PWM} is increased, up to 4.5 kHz, yielding a 33 x 10^3 hrs improvement in *Lop*. In the third row the voltage rating is increased to 450 V, yielding ΔLop = 41 x 10^3 hrs. In the fourth row the cut-off frequency is halved, and the resulting improvement in *Lop* is much larger. Finally, in the fifth row the quality factor is reduced down to 2.4, and the *Lop* improvement in the largest. Although f_{PWM} was not included in the prediction equation, nevertheless for comparison purposes its effect was included in table VII.

It is apparent that the larger coefficients in the prediction equation yield the larger ΔLop. Therefore, if in a particular application the calculated lifetime of the capacitor does not fulfill the requirements, the best thing to do is to redesign the filter with a smaller quality factor. This will yield a larger capacitor and a smaller inductor (larger capacitors imply lower ESR values, as illustrated by Fig. 3). Designing a filter with a lower *fc* will also produce a higher capacitance value, although in this case a larger inductor is also obtained

Capacitors rated at higher voltages are built in larger cans, thus producing both a reduction in the hot-spot temperature (due to a lower ESR, as shown in Fig. 4), and an increase in the term $f_1(V)$, equation (2).

It must be pointed out that the numerical values presented are for steady state operation, not taking into account the transients that occur during motor start-up. Also, the simulations were performed with a flicker-free well-balanced AC supply. In adjustable speed drives connected to unbalanced supplies the capacitors exhibit a much shorter life, because the unbalance produces low-order current harmonics.

It is worth pointing out that the maximum carrier frequency f_{PWM} used for the calculations was not very high. A higher value can be used, but the numerical results obtained are practically the same. This occurs because, as can be noticed from Fig. 2, the ESR does not change significantly at frequencies above 2 kHz.

In this case in was assumed that f_C, Q, V_R, f_{PWM} were "free" parameters and could be modified at will within reasonable limits. There is no doubt that the output frequency generated by the inverter also affects the operational life. This frequency, however, usually depends on the operating point desired, and cannot be modified at will because it depends on the load requirements. Therefore, variables such as the output frequency must be excluded from the analysis.

It should be noted that the prediction equation does not yield an actual value of the operational life expected. Rather, it represents the magnitude of the impact that each variable has on

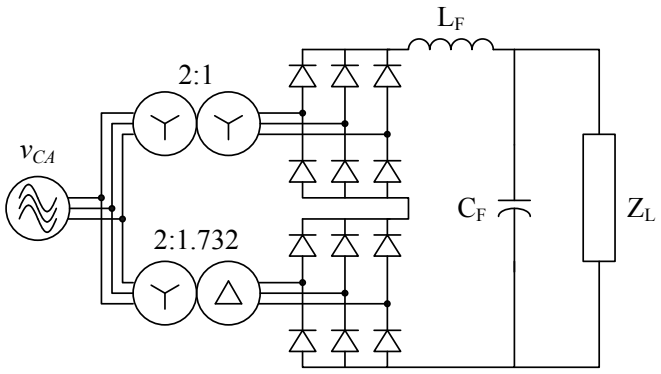


Fig. 6 Adjustable speed drive with a 12-pulse front-end.

Lop. Also, the operational lives calculated are just predictions obtained using statistical data gathered by the manufacturer. The coefficients in equation (17) resemble the sensitivity factors used in other applications, such as the design of analog filters. In order to perform the sensitivity analysis it is necessary to have a mathematical expression relating cause and effect. In many cases the relationship is not straightforward, and it is much easier to apply the method described herein.

The methodology can be applied in a straightforward manner to other configurations, such as the ASD with a 12-pulse front-end illustrated in Fig. 6. To maintain the same voltage level at the DC-link, two step-down voltage transformers were used at the inputs. A filter with a higher cut-off frequency can be used in this circuit, which exhibits the same trends obtained in the prediction equation, but with longer operational lives. As an example, using the values listed in the first row, table VII, yields L_{OP} = 2955 x 10^3 hours. A further advantage of the circuit is the power-factor improvement.

V. CONCLUSIONS

This paper presents a methodology to improve the operational life that can be achieved by an electrolytic capacitor in an adjustable speed drive. The methodology is based on the design-of-experiments technique, and was applied to an ASD rated at 5 kW. The results show that the operational life can be extended by decreasing the cut-off frequency, and the quality factor, and by increasing the voltage rating. The carrier frequency does not have a major effect on the operational life.

Clearly, the same conclusions could be achieved following a heuristic approach. The advantage of the methodology followed herein is that several variables can be dealt with in a systematic manner at the design stage, providing a sturdier design in a much shorter time. The methodology is computationally intensive but, nowadays, that does not represent a major obstacle. It can be applied to other configurations, such as the ASD with a 12-pulse front-end, or an ASD with AC input reactors instead of the inductor at the DC-link.

The MIL HDBK 217F procedure was selected for the reliability calculations because it can be applied without actually building the converter, and is still most widely accepted in the aerospace and military industry, although it is generally viewed as pessimistic.

REFERENCES

[1] W. Sarjeant, I. Clelland, and R. Price, Capacitive Components for Power Electronics, Proceedings of the IEEE, Vol. 89, No. 6, June 2001, pp. 846-855.

[2] P.D. O'Connor. Practical Reliability Engineering. John Wiley and Sons, Chichester, West Sussex, England, 2002, Chapter 9, Electronic systems reliability, pp. 247-289.

[3] Xiong Liu, Peng Wang, Poh Chiang Loh, Blaabjerg, F. and Mingyu Xue, Six switches solution for single-phase AC/DC/AC converter with capability of second-order power mitigation in DC-link capacitor, Energy Conversion Congress and Exposition (ECCE), 2011 IEEE, pp. 1368-1375.

[4] Grabner, C., Electrical drives without DC link converter component and recuperation capability, International Conference on Power Engineering, Energy and Electrical Drives, 2009, POWERENG '09, pp. 128-133.

[5] Maheshwari, R., Munk-Nielsen, S., Henriksen, B., Obel, P.M. and Kragh, H. Active damping technique for small DC-link capacitor based drive system, 2010 IEEE International Symposium on Industrial Electronics (ISIE), pp. 1205- 1209.

[6] G. Hwang, P. Lehn and M. Winkelnkemper, Control of Grid Connected AC-DC Converters with Minimized DC Link Capacitance Under Unbalanced Grid Voltage Condition, 2007 European Conference on Power Electronics and Applications, Aalborg, Denmark 2007, pp. 1-10.

[7] O. Pyrhen, H. Saren, K. Rauma and O. Laakkonen, Verification of Frequency Converter with Small DC-Link Capacitor, EPE2005 Conference proceedings.

[8] M. Huber, W. Amrhein, S. Silber, M. Reisinger, G. Knecht and G. Kastinger, Ripple Current Reduction of DC Link Electrolytic Capacitors by Switching Pattern Optimisation, 36th Annual IEEE Power Electronics Specialists Conference Record, Recife, Brazil, 2005, pp. 1875-1880.

[9] Anderson, J.M., Cox, R.W. and Noppakunkajorn, J., An on-line fault diagnosis method for power electronic drives, 2011 IEEE Electric Ship Technologies Symposium (ESTS), pp. 492- 497.

[10] D.-C. Lee, K.-J. Lee, J.-K. Seok and J.-W. Choi, On-Line Capacitance Estimation of DC-Link Capacitor by Input Current Injection for PWM Converters, EPE 2003, Conference Proceedings.

[11] A. Imam, D. Divan, R. Harley and T. Habetler. Real-Time Condition Monitoring of the Electrolytic Capacitors for Power Electronics Applications, 22nd. IEEE Applied Power Electronics Conference, Anaheim, California, 2007. Pp. 1057-1061.

[12] S. Parler, Reliability of CDE Aluminum Electrolytic Capacitors. Application Note, Cornell-Dubilier Electronics (Available at http://www.cde.com/tech/reliability.pdf)

[13] Department of Defense, *Reliability Prediction of Electronic Equipment*, Military Handbook 217-F, USA, 1991.

[14] D. Hirschmann, D. Tissen, S. Schröder and R. De Doncker. Reliability Prediction for Inverters in Hybrid Electrical Vehicles. IEEE Transactions on Power Electronics, Vol. 22, No. 6, November 2007, pp. 2511-2517.

[15] K. Rajashekara, V. Rajagopalan, A. Sevigny, J. Vithayathil. DC Link Filter Design Considerations in Three-Phase Voltage Source Inverter-Fed Inductor Motor Drive System. IEEE Transactions on Industry Applications. Vol. 23, No. 4, July-August 1987, pp. 673-680.

[16] M. Winkelnkemper and S. Bernet. Design and optimization of the DC-link capacitor of PWM voltage source inverter with active front-end for low-voltage drives. EPE 2003 Conference Proceedings.

[17] Mark Levin and Ted Kalal. "*Improving product Reliability. Strategies and Implementation*". John Wiley and sons. Chichester, England 2003.

[18] PEH200 capacitors data sheet, Evox Rifa.

[19] Capacitor PEH200YV447BM ESR Matrix, Evox Rifa.

[20] J. K. Telford, A Brief Introduction to Design of Experiments. Johns Hopkins APL Technical Digest, Volume 27, Number 3 (2007), pp. 224-232.
[21] M. L. Gasperi, Life Prediction Modeling of Bus Capacitors in AC Variable-Frequency Drives. IEEE Transactions on Industry Applications, Vol. 41, No. 6, November/December 2005, pp. 1430-1435.
[22] C.R. Hicks, *"Fundamental Concepts in the Design of Experiments"* Holt, Rinehart and Winston, Inc, New York, 1982.
[23] J. Wesley Barnes*, "Statistical Analysis for Engineers and Scientists: A computer-based approach",* Mc Graw Hill, 1994, pp. 366.

Calleja G. Hugo recibió el título de Doctor en Ciencias en Ingeniería Electrónica en el cenidet en 2000. Actualmente es profesor del centro nacional de investigación y desarrollo tecnológico, cenidet, adscrito al departamento de electrónica. Sus áreas de interés son instrumentación electrónica para convertidores de potencia, confiablidad en sistemas fotovoltaicos.

Chan-Puc Freddy nació el 5 de marzo de 1973 en Mérida Yucatán. Obtuvo el grado de Ingeniero en Electrónica por el Instituto Tecnológico de Mérida en 1991. Obtuvo el grado de Maestro en Ciencias en Ingeniería Electrónica y el Doctorado en ciencias en Ingeniería Electrónica en el cenidet de Cuernavaca en 1999 y 2008 respectivamente. Desde el 2000 funge como profesor investigador de la Universidad de Quintana Roo. Sus áreas de interés son la electrónica de potencia y la conversión de energía. Es miembro del Sistema Nacional de Investigadores.

Torres M. Emmanuel, nació el 2 de Septiembre de 1979 en Chetumal Quintana Roo, obtuvo el título de Ingeniero eléctrico por el Instituto Tecnológico de Chetumal, en la ciudad de Chetumal Quintana Roo, México en el 2002. Obtuvo el grado de Maestro en Ciencias en Ingeniería Eléctrica por el CINVESTAV Unidad GDL, en la ciudad de Guadalajara Jalisco en el 2006. Del 2002 al 2004 laboró en la Industria de la Construcción en el diseño y supervisión de instalaciones eléctricas en baja y media tensión. A partir del 2007 es profesor Investigador de la Universidad de Quintana Roo en la ciudad de Chetumal Quintana Roo, México. Sus líneas de investigación de interese son: análisis y control del generador de inducción auto-excitado, calidad de energía y ahorro de energía.

Sánchez H. Víctor ingeniero en Electrónica por el Instituto Tecnológico de Orizaba (1996), Maestro en Ciencias en Ingeniería Electrónica por el Centro Nacional de Investigación y Desarrollo Tecnológico (2000) y Dr. en Ingeniería Eléctrica por el CINVESTAV Unidad GDL (2011). De 2002 a la fecha es profesor investigador de la Universidad de Quintana Roo. Sus áreas de estudio son sistemas de generación de energía eléctrica a partir de fuentes renovables, convertidores de potencia de alta eficiencia y sistemas de generación distribuidos.

Toral Homero recibió el grado de Doctor y Maestro en Ciencias en Ingeniería Eléctrica con opción en Telecomunicaciones por el CINVESTAV Unidad Guadalajara en 2010 y 2006 respectivamente. En el 2002 recibió el grado de Ingeniero Electrónico por el Instituto Tecnológico de la Laguna. Sus aéreas de interés incluyen evaluación de desempeño y modelado de sistemas de comunicación. Desde 2010 es Profesor Investigador de la Universidad de Quintana Roo. Actualmente es miembro del Sistema Nacional de Investigadores.

Una aproximación epidémica para el problema de direccionamiento de consultas semánticas en redes p2p estructuradas

Colmenares G. Luis E. y Solís L. Eder

An epidemic approach for the semantic query routing problem in the structured p2p networks

***Abstract*— In this paper, we propose an epidemic model, which is used to search for data in a structured *p2p* network. Structured *p2p* networks, carry out deterministic searches without associating the interests of each node. Their efficiency is demonstrated by performing a comparison between the proposed epidemic algorithm, and the algorithm that uses *DHT* Bamboo. The requests are based on *Zipf's* law, like they were *World Wide Web*. By utilizing epidemic algorithms, it is possible to replicate data, and therefore, add up semantics to the information on the network, and as a result, reduce the number of rounds or hops performed to accomplish such a search, in a distributed system such as a structured *p2p* networks.**

***Keywords*— DHT, fanout, epidemic model, structured p2p.**

***Resumen*— En el presente trabajo se propone un modelo epidémico, que se utiliza para realizar búsquedas de datos en una red p2p estructurada. Las redes p2p estructuradas realizan búsquedas deterministas sin relacionar los intereses de cada nodo. La eficiencia se demuestra realizando una comparación de la propuesta del algoritmo epidémico con el algoritmo DHT que utiliza *Bamboo*. Las peticiones realizadas se basan en la ley de Zipf como si fuera el World Wide Web. Con la utilización de un algoritmo epidémico es posible replicar datos, y así agregar semántica a la información en la red y como resultado se reduce el número de rondas o saltos para las búsquedas en un sistema distribuido como son las redes p2p estructuradas.**

***Palabras clave*— DHT, fanout, epidemic model, structured p2p.**

I. INTRODUCCIÓN

Los Algoritmos Epidémicos tienen gran popularidad en la diseminación de información en sistemas distribuidos de gran escala, particularmente en sistemas punto a punto (*peer-to-peer, p2p*) que funcionan a través de Internet o en redes *adhoc*. Una cuestión importante en el uso de los algoritmos epidémicos y los sistemas *p2p* ha sido la importancia de los intereses de cada nodo a la hora de recibir o compartir una determinada información.

La técnica de diseminación de información se basa en imitar la forma de expansión de las epidemias. Estos algoritmos reproducen la manera en que se difunden las enfermedades contagiosas o los rumores en un entorno social, de manera que un individuo infectado de una enfermedad pasa los gérmenes de la misma a otro individuo con el cual mantiene contacto directo o indirecto. Estableciendo la analogía con un sistema distribuido, una nueva información recibida es distribuida de manera aleatoria a nuevos nodos o procesos, y éstos reproducen nuevamente el mismo ciclo, evitando que haya un único servidor o *cluster* a cargo del proceso de transmitir dicha información [1].

En ciencias de la computación, la utilización de estos algoritmos epidémicos ha sido analizada en aplicaciones como la detección de fallos, la agregación de datos, el monitoreo y descubrimiento de recursos y la replicación de bases de datos [2].

El problema que se aborda en este trabajo es el direccionamiento de consultas semánticas en redes *p2p* [3] mediante la realización de copias de documentos en nodos de una red *p2p* estructurada, es decir, la replicación de datos para simular una *CDN* (*Content Delivery Network*) con la implementación de un algoritmo epidémico.

El problema de direccionamiento de consultas semánticas (*Semantic Query Routing Problem*, *SQRP*) consiste en tener una palabra clave en la consulta y la decisión de un peer y sus peer vecinos de reenviar la consulta. Lo que distingue a éste tipo de búsquedas de otras existentes (tales como: búsquedas de imágenes, texto libre, entre otras.) es que en éstas los datos de búsqueda necesitan tener una descripción semántica asociada con ellos [4].

Se realiza una comparación de la eficiencia del algoritmo con el modelo epidémico infecta por siempre, con el algoritmo de la herramienta *Bamboo*[5], que utiliza una *DHT (Distributed Hash Table).* El escenario es una red *p2p* y se utilizan un número de rondas para llegar a un nodo de la red. Cada nodo de la red representa un servidor de la *CDN* [6].

II. TRABAJO RELACIONADO

Manuscrito recibido el 21 de Mayo de 2012. Este trabajo fue respaldado por la Facultad de Ciencias de la Computación y Vicerrectoría de Investigación de Posgrado de la Benemérita Universidad Autónoma de Puebla.

Colmenares G. Luis E. hasta la fecha se ha de desempeñado como Profesor de Tiempo Completo, se encuentra actualmente realizando investigación y docencia en el área de Sistemas de Tiempo Real, Sistemas Distribuidos y Cómputo Ubicuo en la Facultad de Ciencias de la Computación de la Benemérita Universidad Autónoma de Puebla BUAP, Apdo. Postal J-32, Ciudad Universitaria, Puebla, México. (Teléfono: (52) 222-2 29-55-00 Ext. 7214, e-mail: lecolme@cs.buap.mx).

Solís L. Eder es estudiante de la Facultad de Ciencias de la Computación y actualmente está realizando su trabajo de Tesis.

A. Redes p2p

Las redes *p2p* tienen dos funciones principales: la búsqueda y la compartición de ficheros. Estas redes, se han asociado a la búsqueda de ficheros. En las primeras arquitecturas, se utilizaba el mecanismo del índice central, donde todos los usuarios se registraban en un servidor central que servía para encontrar los contenidos. Las búsquedas se hacían en el servidor central, y las transferencias de datos entre los clientes afectados. Éste tenía el problema de escalabilidad: el servidor central se convertía en un cuello de botella al verse saturado de peticiones de múltiples usuarios.

Para solucionar este problema, la siguiente generación de sistemas *p2p* apareció con la red *Gnutella* [7]. El esquema utilizado es inundar con mensajes hasta que encuentre el contenido solicitado, a estas redes se les llama no estructuradas y consiste en una red de nodos conectados anárquicamente entre sí, los cuales no dependen de ningún servidor centralizado. No obstante, el problema de la localización de los recursos es un problema no determinista. Para encontrar un determinado recurso, un nodo envía un mensaje de búsqueda a cada uno de los nodos a los que está conectado. Estos, a su vez, realizan la misma operación, de forma que la búsqueda se expande por una cantidad de nodos exponencial desde el nodo origen, "inundando" la red de mensajes.

Los retos clave de estos sistemas son:

1) Evitar los cuellos de botella que se pueden producir en determinados nodos y por tanto, se distribuyen las responsabilidades de igual forma entre los nodos existentes.
2) Adaptarse a las continuas entradas y salidas (y también caídas) de nodos. Para ello, hay que dar responsabilidades a los nodos que entran y redistribuyen las responsabilidades de los nodos que salen de la red.

Por último, está la generación de redes *p2p* estructurada, que pueden utilizar las clásicas tablas de *hash*, pero en este caso, los *buckets* de hash son los nodos físicos de la red. Este tipo de servicio *p2p* se denomina comúnmente como *DHT*. Las *DHT* proporcionan mecanismos para añadir, borrar o localizar claves de hash. Se construyen por encima de las redes sobrepuestas (*overlay*) *p2p* y suelen ser eficientes, resistentes a fallos, y autoorganizativas. Encima de estas redes se pueden construir, además, otros servicios interesantes, como enrutamiento y localización de objetos descentralizados, servicios de *multicast* y *anycast* escalables. En particular, la abstracción *DHT* almacena pares clave-valor. El valor siempre se guarda en el nodo de la red *overlay*, al cual le pertenece el hash de la clave. Su estructura interna es un grafo (anillo, árbol, o lugares geográficos), lo cual satisface la condición que el número de saltos necesarios para encontrar un determinado valor en la *DHT* es típicamente O(*log n*), donde *n* es el número total de nodos en el sistema. En las redes *p2p* estructuradas como *Bamboo*, un nodo tiene dos tipos de vecinos lejanos y cercanos. Los vecinos lejanos (*table routing*) y los vecinos cercanos (*leaf set*) ayudan a que el límite máximo sea el O(log *n*) [8].

El uso de estándares abiertos para la creación de redes *p2p* estructuradas basadas en *DHT* como *Bamboo* es uno de los más usados debido a que considera el *churn-rate*[8] . El *churn-rate* es el proceso de conexión y desconexión de los *peers*. Dentro de la *DHT*, el mecanismo de búsqueda, proporciona un modelo con

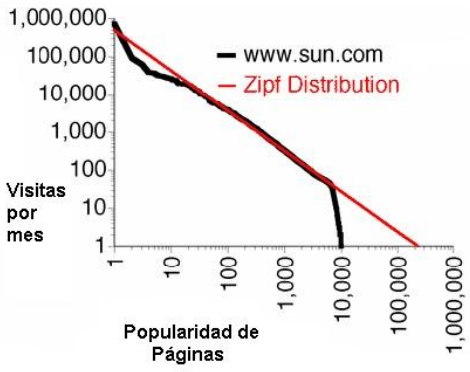


Fig. 1 Ley de Zipf.

búsquedas deterministas, que ocultan al usuario: petición de *routing*, costos de *churn-rate*, balanceo de carga y disponibilidad.

En la actualidad existe software como Emule [9] y Bittorrent [10] que tienen como objetivo el intercambio de ficheros. En el trabajo de Lavastida [11] se hace especial énfasis en la categorización de cada nodo asignándole un peso y se realiza la distribución de la información en nodos específicos.

B. Problema del Enrutamiento de Consultas Semánticas

El problema del enrutamiento de consultas semánticas, consiste en una red representada por el grafo de conexiones (G), un conjunto de palabras-clave (R) distribuida en sus nodos y un conjunto de consultas semánticas (*C*) emitidas dinámicamente por los nodos como sigue:

Una consulta semántica puede ser originada desde cualquier nodo en el tiempo T_0, asumiendo un tiempo de reloj con unidades fijas. Un nodo que origina la consulta o recibe la consulta de otro nodo en el tiempo (T_0+i) puede procesar la consulta localmente y/o reenviar una réplica de la consulta a un conjunto de vecinos inmediatos en el tiempo T_0+i+1. El procesamiento de la consulta termina cuando todas las palabras-clave esperadas son encontradas.

Se busca maximizar la cantidad de recursos encontrados y minimizar la cantidad de saltos realizados por una entidad de búsqueda para encontrar dichos recursos [4, 11].

C. Ley de Zipf

El lingüista norteamericano *George Zipf*, formuló la ley de *Zipf* para las frecuencias de las palabras en los textos en inglés. La palabra más común, *the*, aparece el doble de veces que la segunda, *of*, el triple que la tercera, el cuádruple que la cuarta y así sucesivamente. Pero es aplicable a muchos otros fenómenos. El último que han descubierto es el ajedrez: la jugada más usada ocurre el doble de veces que la segunda más usada, el triple que la tercera; en resumen, la ley de *Zipf* se representa mediante la ecuación (1*)*.

$$P_n \approx 1/n^a \qquad (1)$$

donde P_n representa la frecuencia de una palabra ordenada *n-ésima* y el exponente *a* es próximo a *1*. Esto significa que el segundo elemento se repetirá aproximadamente con una

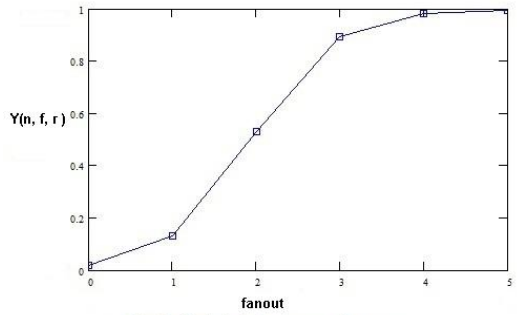


Fig. 2 Modelo Infecta por siempre.

frecuencia de *1/2* de la del primero, y el tercer elemento con una frecuencia de *1/3* y así sucesivamente: En la Figura 1, se puede visualizar la ley de *Zipf*, en el eje *x*, representa la popularidad de páginas con respecto al eje *y*, las visitas por mes que se realizan en el *WWW* a páginas con alta popularidad.

III. Modelos Matemáticos Epidémicos

A. Modelos de Población finita

Incluye la incorporación de una población de tamaño *N*, donde X_r es el número de individuos infectados en la *r-ésima* ronda de propagación epidémica [1, 2].

En cada ronda cada individuo infeccioso con probabilidad P_k intentará contaminar a *k* miembros de la población total. Estos *k* miembros se elegirán aleatoriamente desde la población total.

Los casos son:

1) Infectar y morir: modelo en el que un individuo intenta contaminar a otros por una sola ronda, y luego se detiene.
2) Infectar siempre: modelo en el que los individuos infectados siguen siendo infecciosos en todo momento.

B. Descripción del modelo infecta por siempre

En el modelo infectar por siempre, los individuos infecciosos intentan contaminar a *f* miembros en cada ronda, la fórmula (2), representa el número esperado de miembros infectados en *r* rondas.

$$Y_r \approx \frac{1}{1+ne^{-fr}} \qquad (2)$$

donde:

Y_r : Probabilidad de infección después de un número de rondas *r*.
n : Número de población.
f : Infectados por ronda (*fanout*).
r : Número de rondas.

En el modelo infecta por siempre, cada nodo infectado intenta contaminar a *f* nodos por cada ronda hasta que toda la población esté infectada, al proceso de nodos infectados por ronda se le conoce como *fanout.* Además un nodo ya contagiado nunca muere y sigue infectando a sus vecinos a pesar que ya están contagiados, es decir, nunca termina de contagiar.

Así, como se puede apreciar en la Figura 2, la proporción del número de individuos infectados con el número de los no infectados aumenta exponencialmente rápido en promedio, por un factor de e^f en cada ronda. En este caso el *fanout =2* y realiza 5 rondas para infectar un total de 48 nodos.

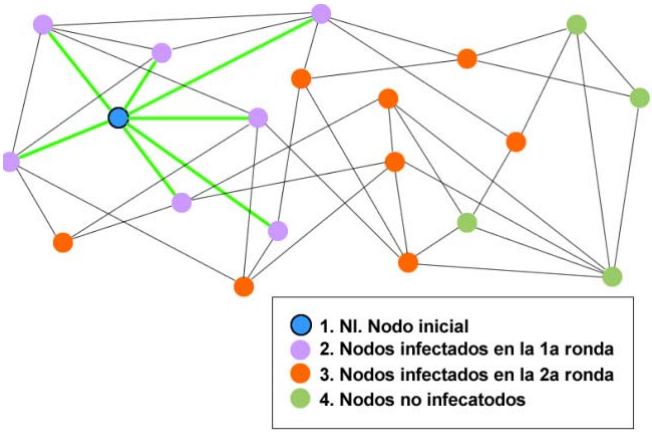


Fig. 3 Comportamiento del modelo infecta y muere.

C. Descripción del modelo infecta y muere

En la figura 3, se muestra el comportamiento del modelo epidémico infecta y muere en una red. Se toma un nodo inicio, y en la 1ª ronda de la infección se contaminan a todos sus vecinos del nodo inicial. En la *2ª* ronda se contaminan a los vecinos de cada uno de los nodos infectados en la *1ª* ronda. Continua la infección hasta *n* rondas necesarios para contaminar a todos los nodos de la red. Después de infectar a los nodos vecinos, cada nodo debe morir, esto es, una vez que un nodo infecta a sus nodos vecinos, este no puede ser infectado nuevamente en una ronda posterior.

En la fórmula (3) que es una ecuación de punto fijo del modelo infecta y muere, la variable π simboliza la proporción acumulativa de nodos eventualmente infectados en cada ronda y *f* representa el *fanout*, que son los nodos infectados en cada ronda. La utilización de este algoritmo no se considera porque los nodos infectados resultantes fueron menos significativos para rondas y número de nodos pequeños.

$$\prod \approx 1 - e^{\pi f} \qquad (3)$$

IV. Descripción del Algoritmo

El algoritmo epidémico propuesto realiza la propagación de un mensaje en la red de nodos. El algoritmo inicia a partir de cualquier nodo aleatorio y las peticiones siguen la ley de *Zipf* que infectan a un número fijo de nodos, el cual está definido por el *fanout*. Este proceso se repite tantas veces el número de rondas. En la realización de este algoritmo se ocuparon diversos lenguajes de programación como lo son java, *bash*, *awk* y *perl*. Este algoritmo consta de tres fases para llevar a cabo la infección en la población: Historial, Propagación y el Gestor.

A. Historial

Esta etapa del proceso ayuda a conocer los nodos que se han infectado en una ronda. Este proceso es conocido como el

TABLA I. RESULTADOS CON *FANOUT* = 4

Estado	*% epidémico*
1	8.33
2	35.41
3	91.66
4	97.916
5	100

historial de un nodo, los nodos infectados se vuelven propagadores de la infección. El historial de un nodo, es la información que se almacena en cada nodo y así se puede conocer los antecedentes de las peticiones. Para la elaboración de esta parte se realizó el código script en *perl*.

Con el historial se toman los nodos infectados y se vuelven propagadores de la infección.

Los ficheros generados con la herramienta *Bamboo* son trazas que se realizan para conocer el funcionamiento de cada nodo durante la búsqueda de información. Para poder usar el archivo *check-obj-ptrs* se necesitan los ficheros *experiments* generados por *Bamboo*. Por esta razón se diseñó un programa en *bash*, que tiene como función localizar y asignar como parámetros de entrada los *experiments* al script *check-obj-ptrs* que se han modificado.

Una vez obtenido las trazas que corresponden al número de saltos que un nodo recorre de un nodo a otro nodo en la consulta semántica, se puede realizar la obtención del recorrido total, solamente sumando todos los recorridos de los nodos involucrados y así obtener el número de saltos totales.

B. Propagación

En esta segunda etapa, se obtiene la información de los nodos contagiados en una ronda, mediante el uso del historial para que así se pueda expandir la infección a toda la red distribuida. El programa que obtiene al conjunto de nodos está elaborado en el lenguaje *AWK*, por ser un lenguaje especializado en el procesamiento de textos para tomar el archivo que nos genera el historial y procesarlo de una manera sencilla.

En esta etapa se optó por utilizar este lenguaje de programación porque utiliza expresiones regulares que se asociaron con los ficheros de salida que se obtienen de la herramienta de *Bamboo*. Una vez obtenido los ficheros salida se utiliza *Gnuplot* para la graficación de los resultados.

C. Gestor

Es un script que se encarga de manipular las dos etapas anteriores, es decir, de la coordinación para efectuar la propagación de la infección. Esta coordinación se realiza a través de la petición de un número que describe las rondas que se deben de realizar para expandir la infección en una red y un *fanout* que está representado en *Bamboo* como las réplicas de un documento. Las replicas en los documentos se pueden hacer de dos a tres replicas en toda la red debido a que exponencialmente representa una media de replicas promedio para una red de 48 nodos.

D. Propuesta solución

El algoritmo se inicializa en un nodo aleatorio que pertenece a la red, y se asigna un número de rondas, este nodo se convierte en el origen de la infección y lo propaga a todos los demás nodos. Después de la primera ronda se genera el historial, que proporcionan los nodos infectados, para que estos nodos se conviertan en el nuevo origen de la infección en la red, hasta que toda la red quede totalmente propagada se termina, o si no sucede esto, se termina hasta que se cumpla el número de rondas que se establecieron desde un principio, ver el pseudocódigo 1.

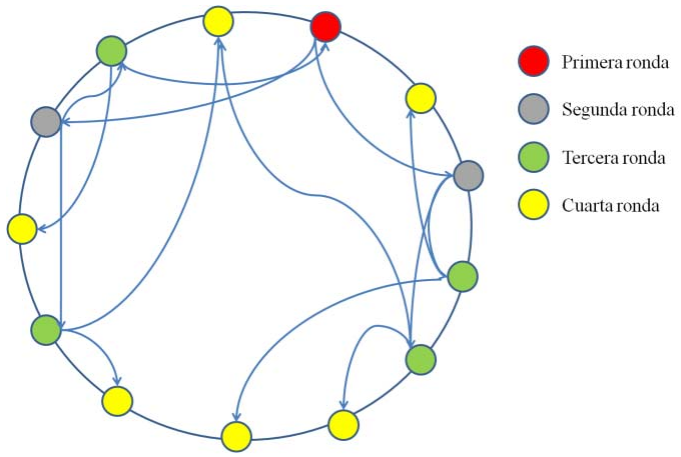


Fig. 4 Algoritmo epidémico en cuatro rondas.

Pseudocódigo 1: **Algoritmo epidémico**

1. Inicio
2. Inicializa la infección en un nodo perteneciente a la red.
3. Crea historial de infectados
4. Se verifica que existan nodos que no estén infectados.
   - 4.1 Para 1 hasta número de rondas-1
     - 4.1.1 Lee archivo historial.
     - 4.1.2 Inicia la infección con el nuevo conjunto de infectados.
     - 4.1.3 Actualizar el historial de infectados.
     - 4.1.4 Se verifica todos los nodos estén infectados.
       - 4.1.4.1 Si es así, detiene la infección.
   - 4.2 Fin Para
5. Fin

La tabla I, permite visualizar el funcionamiento del algoritmo epidémico cuando el *fanout* es igual a 4.

En la Figura 4, se muestra una imagen representativa del algoritmo epidémico, en donde se pueden ver los nodos infectados por ronda en una herramienta *DHT* de una topología circular como *Bamboo*.

V. EVALUACIÓN DEL ALGORITMO EPIDÉMICO EN UNA RED *P2P* ESTRUCTURADA

El escenario que se utiliza es el siguiente, 48 (2^6) nodos en una red *p2p* estructurada, el número de saltos para llegar a un nodo en una *DHT* es de 6 saltos máximo por que el número de

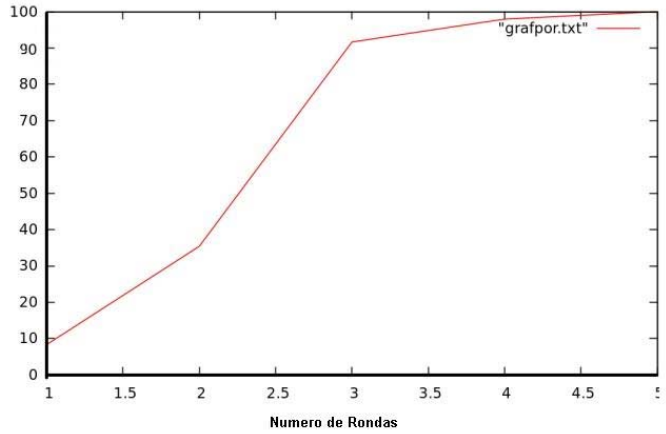


Fig. 5 Red de 48 nodos.

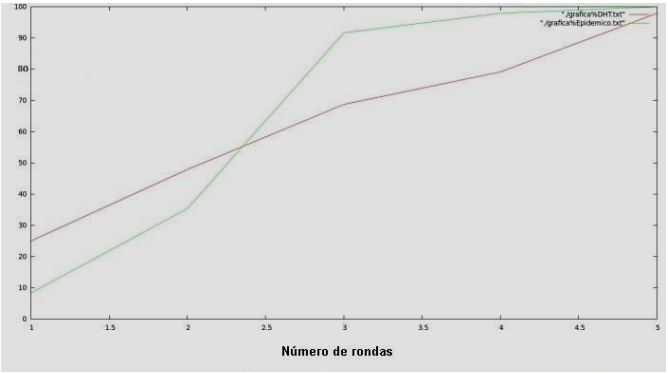


Fig. 6 Comparación de los algoritmos con fanout = 2.

TABLA II. RESULTADOS CON *FANOUT* = 2

Estado	%DHT	*% epidémico*
1	25	8.33
2	47.91	35.41
3	68.75	91.66
4	79.16	97.916
5	97.91	100

nodos está entre $2^5 < 48 < 2^6$ por el $O(\log n)$. Además se utilizan de 10 a 100 peticiones basadas en la ley de *Zipf*, y se realizan las peticiones de manera aleatoria desde cualquier nodo de la topología.

Se realizaron múltiples pruebas del algoritmo epidémico, desde 10 hasta 48 nodos, en las cuales se nota que el crecimiento de la replicación de datos se comporta de manera exponencial. Es decir:

$$N^k \tag{4}$$

Donde, en la fórmula (4):

N = *fanout*

k = número de rondas

En la Figura 5, se muestra una gráfica de una red con 48 nodos y un *fanout* de 4, el eje *Y* representa el porcentaje de nodos

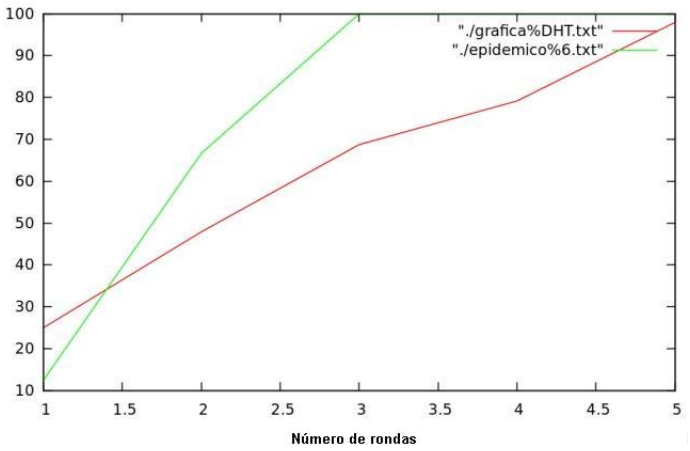


Fig. 7 Comparación de los algoritmos con fanout = 6.

TABLA III. RESULTADOS CON *FANOUT* = 6

Estado	%DHT	*% epidémico*
1	25	12.5
2	47.91	66.67
3	68.75	100
4	79.16	100
5	97.91	100

visitados en una ronda, y en el eje de las *X* es el número de rondas.

A. Comparación con el algoritmo DHT de Bamboo en una red p2p estructurada

Como parte de esta investigación se busca demostrar que el algoritmo epidémico actúa de manera más eficiente en la búsqueda de datos que el método *DHT* de *Bamboo*. Por esto, se utilizaron los dos algoritmos con 48 nodos para obtener la gráfica de la Figura 6 y los datos obtenidos se muestran en la Tabla II. Donde el eje de las *Y* representa el porcentaje de número de nodos recorridos o conocidos en un salto o ronda, y el eje *X* representa el número de rondas.

En la gráfica de la Figura 7, se puede apreciar que el algoritmo epidémico crece más rápido en las últimas rondas, aunque en las primeras rondas su crecimiento fue más lento que el algoritmo de *DHT* de *Bamboo*. Sin embargo, si el *fanout* es un número mayor, el algoritmo epidémico será mejor que la *DHT*.

Se muestra un ejemplo en la Figura 7, en una red de 48 nodos y un *fanout* de 6:

En la tabla III, el algoritmo de *DHT* no tiene variación como lo tiene el algoritmo epidémico, es debido a que el *fanout* afecta solo al algoritmo epidémico. Aunque el *fanout* para la *DHT* sería algo similar a los vecinos cercanos y vecinos lejanos que contiene la red *p2p* estructurada de *Bamboo.*

VI. CONCLUSIONES Y TRABAJO A FUTURO

En el desarrollo de esta investigación, se demostró que los algoritmos epidémicos son directamente proporcional al *fanout*, es decir, que tienden a ser exponencial entre más grande sea este

número. Aunque se realizaron pruebas con *fanout*= 2, 4 y 6, hacerlo con un número más grande resultaría inmediato la infección a todos los nodos. Además el *fanout* representa el interés que los nodos tienen entre sus vecinos es decir la semántica de la información. El algoritmo *DHT* de *Bamboo* realizó las búsquedas en menor número en las primeras *k* rondas comparado con el algoritmo epidémico, debido a que las DHT mantienen un $O(log\ n)$. El número de nodos de una red *p2p* estructurada es del orden 2^n el número de saltos máximo es *n*, el utilizar un número de nodos más grande representa que el algoritmo epidémico sea aún más rápido en la búsqueda de información que el algoritmo DHT, debido a que la DHT no contiene información semántica en sus nodos.

En este primer trabajo se utiliza un algoritmo epidémico para demostrar que es posible replicar información y así agregar semántica a la información de la red y como resultado el tiempo de las búsquedas sea menor en un sistema distribuido como son las redes *p2p* estructuradas.

El uso de los algoritmos epidémicos es importante porque se recorren los caminos de la redes con amplia difusión, lo que genera un tráfico en la red, sin embargo si se utilizan algoritmos epidémicos con búsquedas semánticas, se reduce el trafico de mensajes y se pueden emplear para eliminar el correo *spam* o reducir el número de peticiones en el *WWW*.

Como trabajo a futuro se propondría realizar escenarios, con otras topologías de redes, con tráfico en la red, con protocolos de transporte TCP o UDP, y así generar nuevas mediciones de los algoritmos epidémicos que han sido presentados en este trabajo.

REFERENCIAS

[1] Tim Daniel Hollerung, Peter Bleckmann,(2004), *Epidemic Algorithms Topic 9*, University Padeborn, August, 4th, 2004.

[2] P.T. Eugster, R. Guerraoui, A.-M. Kermarrec, L. Massoulie. (2004), *From Epidemics to Distributed Computing*. *IEEE Computer*, Vol. 37, No. 5, May 2004, pp. 60-67.

[3] Aguirre M. Tesis de Maestría, Algoritmos de Reconocimiento de Patrones que Guíen la Búsqueda de Información en Redes Complejas, noviembre del 2008 en Tamaulipas, México.

[4] Michlmayr E., *Ant Algorithms for Self-Organization in Social Networks*. Tesis Doctoral, Women's Postgraduate College for Internet Technologies (WIT), Institute of Software Technology and Interactive Systems, Universidad de Tecnología de Vienna, 2007.

[5] Sean Rhea, Brighten Godfrey, Brad Karp, John Kubiatowicz, Sylvia Ratnasamy, Scott Shenker, Ion Stoica, Harlan Yu. *OpenDHT: A Public DHT Service and Its Uses*. Proceedings of ACM SIGCOMM 2005, August 2005, ["Online"]. Disponible, http://www.bamboo-dht.org y http://opendht.org.

[6] Balachander Krishnamurthy, Craig Wills, Yin Zhang. (2001), *On the Use and Performance of Content Distribution Networks*, ACM Sigcomm Internet Measurement workshop.

[7] M. Ripeanu. (2001) *Peer-to-Peer Architecture Case Study: Gnutella Network*. Technical Report– 2001-26, Department of Computer Science, University of Chicago, 24 de Julio del 2001.

[8] Sean Rhea, Dennis Geels, Timothy Roscoe, and John Kubiatowicz, (2004). *Handling Churn in a DHT*. *Proceedings of the USENIX Annual Technical Conference*, June 2004

[9] Y. Kulbak, D Bickson.(2005), *The eMule Protocol Specification,* Technical Report – 2005-03, Leibniz Center, School of Computer Science and Engineering, The Hebrew University, 01/2005.

[10] Johan Pouwelse, Pawel Garbacki, Dick Epema, Henk Sips. (2005) *Peer-to-Peer Systems IV*, Volume 3640/2005 de Lecture Notes in Computer Science, The Bittorrent P2P File-Sharing System: Measurements and Analysis, págs. 205-216. Springer Berlin/Heildeberg.2005.

[11] Zilia Lavastida-López, Yudivián Almeida-Cruz, (2009) *Propuesta de un modelo para el intercambio automático de información en redes P2P*, VI Jornadas para el Desarrollo de Grandes Aplicaciones de Red (JDARE'09). Computación como Servicio, Desarrollo de Grandes Aplicaciones de Red. Alicante, España, octubre 15-16, 2009. ISBN 978-84-613-4894-7, GrupoM. Alicante. 2009. Páginas: 333-350.

Colmenares G. Luis E., nació en Tuxtla Gutiérrez Chiapas, México el 9 de Abril de 1969. Realizó sus estudios de la Licenciatura en Computación en la Benemérita Universidad Autónoma de Puebla en la Facultad de Ciencias Físico-matemáticas. Los estudios de Maestría en la Universidad de las Américas Puebla obteniendo el título de Maestro en Ciencias. El doctorado fue realizado en la Universidad Politécnica de Cataluña en Barcelona España, en la especialidad de Sistemas distribuidos en el Departamento de arquitectura de computadores.

Actualmente está como profesor investigador de tiempo completo en la Facultad de Ciencias de la Computación de la Benemérita Universidad autónoma de Puebla. Imparte asignaturas de las currículas de Licenciatura, Ingeniería y Postgrado. Pertenece al cuerpo académico de Sistemas de Información Promep, y colabora con la Dirección General de Innovación educativa. Esta en diferentes proyectos de Sistemas Distribuidos, Sistemas de Tiempo Real, Computo Pervasivo y Cómputo Ubicuo.

Luis Enrique Colmenares es miembro de SOMECYTA (Sociedad Mexicana en Ciencia y Tecnología Aeroespacial. Además es miembro de la red temática Conacyt "Tecnologías de la información y comunicaciones" de 2011-2015. Pertenece además al padrón de Investigadores de la Benemérita Universidad autónoma de puebla. Ganador del tercer lugar del equipo que fue mentor en la final nacional de diseño de software de Imagine Cup 2010 y En el 2011 recibió dos reconocimientos por la Benemérita Universidad Autónoma de Puebla Pue. México. En el 2012 participo como mentor en la final nacional de diseño de software en el Imagine Cup 2012.

Eder Solís López, nació en Puebla, Pue. México el 29 de junio de 1990. Realizo estudios de preparatoria en el Centro Escolar Niños Héroes de Chapultepec, preparatoria incorporada a la Benemérita Universidad Autónoma de Puebla. Ingreso a la Licenciatura en Ciencias de la Computación en la BUAP (2008). Ha asistido a diferentes congresos como: 50 años de la computación en México (UNAM 2008); Congreso de Computación, Informática, Biomédica y Electrónica (Universidad de Guadalajara 2009); 7° Congreso Nacional de Ciencias de la Computación (BUAP 2010); Congreso Nacional de Enlace Tecnológico (BUAP 2010); Eleventh Mexican International Conference on Computer Science (UAEM 2011); Twelfth Mexican International Conference on Computer Science (Universidad de Guanajuato 2012). Tiene conocimientos del área de programación en los lenguajes C, C#, Java, Visual Basic, PHP, mySQL y UML. A lo largo de su carrera ha tomado cursos extracurriculares de programación, redes e inglés.

A finales del año 2011 obtuvo la certificación en el lenguaje de programación Java "Oracle Certified Professional Java SE 6 Programmer". Para el año 2012 participo en la competencia tecnológica de Microsoft: "Imagine Cup 2012**"** con el equipo "DIGIMFILT" en la categoría "Diseño de Software"; el proyecto es un filtro de contenido pornográfico basado en el procesamiento digital de imágenes para el "Internet Explorer" donde se aplicaron tecnologías de Microsoft y el algoritmo RSOR de Pedro Ivan Tello Flores para la detección de desnudos en imágenes digitales; llegó a la final nacional en el mes de Abril en las instalaciones de Microsoft México en la ciudad de México.

En la actualidad se encuentra realizando su trabajo de Tesis para obtener el título de Licenciado en Ciencias de la Computación.

Evaluación de implementaciones en software de algoritmos para la multiplicación escalar en criptografía de curvas elípticas

Vega C. Karina, Cortina R. Antonio y Morales S. Miguel

Evaluation of software implementations of algorithms for scalar multiplication in elliptic curve cryptography

***Abstract*— This work presents an evaluation of different algorithms to implement in software the most time demanding operation in ECC, the scalar multiplication. The results presented in this work could help a designer to select the most appropriate method when implementing ECC-based cryptographic schemes such as encryption or digital signatures. Different scalar multiplication algorithms for ECC defined over prime and binary fields are considered, using affine coordinates to represent the elliptic curve points. The evaluation was performed over two computing platforms. The first one is a computer powered with an Intel Core i5 processor at 2.50 GHz. The second one is a mobile device LG P500h with an ARM processor at 600 MHz. The performance evaluation consisted of comparing the timing for computing a scalar multiplication in an ECC key generation scheme. For ECC defined over binary fields, it was found that the *NAF* and *wNAF* algorithms run 1.6 times faster than the *Double&Add* method both on the PC and the mobile device. This same speed-up is observed when these two algorithms are implemented for elliptic curves defined over the prime field and executed on the PC. However, when the *wNAF* algorithm is executed on the mobile device, it is 1.2 times slower that the *Double&Add* method and 1.5 times slower than the *NAF* algorithm.**

***Keywords*— Finite fields, Cryptographic schemes based on elliptic curve, Scalar multiplication.**

***Resumen*— Este trabajo presenta una evaluación de diversas implementaciones en software de algoritmos para calcular la operación más demandante en esquemas criptográficos basados en Criptografía de Curvas Elípticas (ECC), la multiplicación escalar. Los resultados presentados en este trabajo podrán servir como un punto de referencia para quienes implementan esquemas criptográficos basados en ECC tales como esquemas de cifrado o de firma digital. En esta investigación se usaron diferentes algoritmos definidos tanto en campo primo como en campo binario empleando coordenadas afines. Las pruebas se efectuaron en una computadora de escritorio con un procesador Intel Core i5 con 2.50 GHz de velocidad, así como en el dispositivo móvil LG P500h con un procesador ARM a 600 MHz. Tras evaluar los tiempos de ejecución para la generación de una llave pública dada una llave privada previamente establecida, se encontró que los algoritmos *NAF* y *wNAF* se ejecutan 1.6 veces más rápido que el método de Suma y Doblado, tanto en la PC como en el dispositivo móvil. Sin embargo, en el caso particular del método *wNAF* sobre campo primo implementado en el dispositivo móvil, se observó que éste se ejecuta 1.2 veces más lento que el método de Suma y Doblado y 1.5 veces más lento que *NAF*.**

***Palabras clave*— Campos finitos, Esquemas criptográficos basados en curva elíptica, Multiplicación escalar**

Manuscrito recibido el 20 de Agosto de 2012. Este trabajo fue respaldado por la Universidad Politécnica de Cd. Victoria, Tamaulipas.

Vega C. Ana K. hasta la fecha se ha de desempeñado como estudiante en el programa de Maestría en Ingeniería con especialidad en Tecnologías de la Información de la Universidad Politécnica de Cd. Victoria; Av. Nuevas Tecnologías 5902 Parque TECNOTAM, Carretera Victoria - Soto la Marina Km. 5.5; Ciudad Victoria, Tamaulipas, México; C.P. 87138; (e-mail anakarynavega@hotmail.com).

Cortina R. Antonio hasta la fecha se ha de desempeñado como estudiante en el programa de Maestría en Ingeniería con especialidad en Tecnologías de la Información de la Universidad Politécnica de Cd. Victoria; Av. Nuevas Tecnologías 5902 Parque TECNOTAM, Carretera Victoria - Soto la Marina Km. 5.5; Ciudad Victoria, Tamaulipas, México; C.P. 87138;(e-mail antoniocr06@hotmail.com).

Morales S. Miguel hasta la fecha se ha desempeñado como Profesor de Tiempo Completo de la Universidad Politécnica de Cd. Victoria; Av. Nuevas Tecnologías 5902 Parque TECNOTAM, Carretera Victoria - Soto la Marina Km. 5.5 Ciudad Victoria, Tamaulipas; C.P. 87138; Tel: (834) 1720383, Fax: (834) 1720388; (e-mail mmoraless@upv.edu.mx).

I. INTRODUCCIÓN

Gracias al avance de la tecnología, hoy en día es posible contar con plataformas móviles y de escritorio con capacidades de procesamiento cada vez superiores, lo que ha dado lugar a aplicaciones más potentes, las cuales en muchas ocasiones demandan la garantía de servicios de seguridad informática, tales como confidencialidad, integridad, autenticación, etc., debido a que la información que administran es de carácter sensible, por lo que es indispensable el contar con un esquema criptográfico confiable, robusto y eficiente, que permita brindar dichos servicios. Recientemente las técnicas de criptografía basadas en curvas elípticas (ECC) y emparejamientos bilineales han demostrado la viabilidad de proveer servicios de seguridad informática de manera eficiente tanto en plataformas de escritorio como en el dominio de las aplicaciones móviles [1, 2], debido a que permiten la utilización de longitudes de llaves más cortas que esquemas tradicionales de criptografía, como el algoritmo RSA. Al usar llaves más cortas, el espacio de memoria utilizado para el almacenamiento de llaves se reduce considerablemente así como también los requerimientos de ancho de banda para la transmisión de las llaves, además de ser capaces de proveer niveles de seguridad confiables [3].

En la literatura se han realizado estudios comparativos de algoritmos que calculan la operación más demandante en esquemas criptográficos basados en ECC, la multiplicación escalar, en base a la cantidad de operaciones que dichos algoritmos efectúan [4, 5]. En este artículo se describe el trabajo experimental para evaluar el desempeño de algoritmos para la multiplicación escalar a partir de los tiempos de ejecución, haciendo una evaluación bajo las mismas condiciones de prueba, es decir, los algoritmos fueron evaluados sobre la misma plataforma de cómputo, ya sea PC o dispositivo móvil, y usando las mismas rutinas de software.

La aportación principal de este trabajo consiste en proveer un análisis de los tiempos de ejecución de diversos algoritmos existentes para el cálculo de la multiplicación escalar en ECC, que sirva como un punto de referencia para quienes implementan esquemas criptográficos basado en ECC tanto en computadoras de escritorio como en dispositivos móviles.

Se realizó una revisión de la literatura y se seleccionaron diversos algoritmos tanto para campos binarios *GF(2^m)* como para campos primos *GF(p)*. La implementación y pruebas se efectuaron tanto en un ordenador de escritorio, como en un dispositivo móvil.

Las secciones restantes del artículo se ordenan como sigue: el marco teórico de esta investigación se describe en la sección II; La multiplicación escalar y los algoritmos para calcular esta operación se describen en la sección III. Los resultados obtenidos se detallan en la sección IV; finalmente en la sección V se presentan las conclusiones de este trabajo.

II. CRIPTOGRAFÍA DE CURVAS ELÍPTICAS (ECC)

La ECC pertenece a la criptografía asimétrica [6], esto debido a que se utilizan dos claves distintas: una pública y una privada, donde el conocimiento de la clave pública no permite determinar la clave privada.

ECC fue propuesta de manera independiente en 1985 por Neal Koblitz [7] y Victor Miller [8]. Desde entonces una gran cantidad de investigaciones se han realizado para tener implementaciones eficientes y seguras de estos esquemas criptográficos. La Criptografía de Curvas Elípticas ha permitido explorar nuevos criptosistemas, tal como la técnica de emparejamientos bilineales [1].

A. Campos finitos

Una curva elíptica sobre un campo *K* es un conjunto formado por el punto al infinito ∞ y los puntos P=(x,y)∈K×K que satisfacen la ecuación de Weistrass (1) [4]:

$$E(K): y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 \qquad (1)$$

La curva elíptica $E(K)$ junto con el punto al infinito (∞) forman el grupo aditivo $\{E(K) \cup \infty, +\}$, donde el punto ∞ es el elemento identidad $(P + \infty = P)$. Para aplicaciones criptográficas, el campo K es un campo finito. Si Kes un campo primo *GF(p)* la ecuación de la curva elíptica es (2):

$$E(GF(p)):\ y^2 = x^3 + ax + b \qquad (2)$$

Cuando la curva elíptica está definida en el campo binario *GF(2^m)*, la ecuación es (3):

$$E(GF(2^m)): y^2 + xy = x^3 + ax^2 + b \qquad (3)$$

B. Problema del logaritmo discreto elíptico

Dada una curva E en un campo finito, podemos representar la operación principal llamada multiplicación escalar en ECC de la siguiente manera:

$$Q = dP$$

donde:

$\boldsymbol{P}$ *y* $\boldsymbol{Q}$**:** Son puntos de una curva elíptica.

$\boldsymbol{d}$: Es un escalar secreto.

Dicho esto podemos definir al Problema de Logaritmo discreto Elíptico (PLDE) como determinar el escalar d, dado los puntos P y Q.

La seguridad basada en ECC es basada en la dificultad de resolver este problema. En general la dificultad del PLDE resulta ser más difícil que otros problemas como el de factorización de enteros y logaritmo discreto [9].

C. Estándares

Un criptosistema de curva elíptica se define a partir de una tupla *T*, la cual es un conjunto de parámetros que define un campo finito de trabajo, la ecuación de una curva elíptica (ver ecuación 1) definida sobre el campo finito sobre la que se va a trabajar, un generador de la curva elíptica, entre otros valores.

Para el campo primo *GF(p)* los parámetros son una séxtupla de valores denotados como [3]:

$$T = (p, a, b, G, n, h)$$

donde:

$\boldsymbol{p}$: Es un número primo grande.

$\boldsymbol{a}, \boldsymbol{b}$: Son coeficiente que definen la curva E en *GF(p)*.

$\boldsymbol{G}$: Es el generador de un subgrupo cíclico en la curva elíptica *E(GF(p))*.

$\boldsymbol{n}$: Es el orden de G (n es el entero más pequeño tal que $nG = \infty$).

$\boldsymbol{h}$: Es el cofactor de la curva y se define como el $número\ de\ puntos\ /n$. Este valor es opcional.

Para el campo Binario *GF(2^m)* la tupla T consiste de siete parámetros necesarios definidos como [3]:

$$T = (m, f(x)\ a, b, G, n, h)$$

donde :

$\boldsymbol{m}$: Es el entero que especifica el orden del campo finito que se está usando.

$\boldsymbol{f(x)}$: Es el polinomio irreducible de grado m.

Todos los demás valores tienen una definición similar al caso de $GF(p)$.

Existen diversos estándares que especifican un conjunto de valores para cada elemento de la tupla *T*. Dichos parámetros han sido derivados a través de métodos ampliamente estudiados y por ello se consideran seguros. Dentro de los estándares se

encuentran IEEE [10], NIST [11], ANSI [12, 13], ISO [14], SECG [15], entre otros. Los estándares de ECC en su mayoría son compatibles además de contar con vectores de prueba con los cuales se puede verificar los resultados.

D. Esquemas Criptográficos

Para lograr contar con los servicios de seguridad que las aplicaciones móviles y de escritorio necesitan, es indispensable utilizar esquemas criptográficos. Entre los principales basados en ECC se pueden mencionar [16]:

1) Esquema de Diffie-Hellman en Curvas Elípticas (ECDH). Esquema que permite la generación y el intercambio de llaves entre dos entidades.

2) Esquema de Cifrado Integrado en Curva Elíptica (ECIES). Utilizado para el cifrado de datos. Dentro de este esquema se utiliza ECDH para generar dos llaves simétricas, una utilizada para cifrar el texto claro y otra para autentificar el texto cifrado.

3) Algoritmo de Firma Digital de Curva Elíptica (ECDSA). Este esquema puede ser divido en dos etapas, la primera sería la generación de firma digital y la segunda la verificación de firma digital, donde en cada una de estas etapas se vería implicado el uso de una llave privada y una función hash.

Algoritmo 1: Generación de firma con ECDSA

Requiere: *(GF(q),a,b,G,n,h)* llave privada d_a, mensaje *m*
Salida: Firma digital *(r, s)*
1: Seleccionar $k \in [1, n-1]$
2: $(X_1, Y_1) \leftarrow kP$
3: $r \leftarrow X_1 \ mod \ n$. Si $r = 0$ regresar al paso 1
4: $e \leftarrow H(m)$
5: $s \leftarrow k^{-1}(e + d_a r) mod \ n$. Si $s = 0$ regresar al paso 1
6: Regresar (r, s)

Algoritmo 2: Verificación de firma con ECDSA

Requiere: *(GF(q),a,b,G,n,h)*, llave pública D_a, mensaje *m*, firma digital *(r, s)*
Salida: Aceptación o rechazo de la firma digital *(r, s)*
1: Verificar que *(r, s)* $\in [1, n-1]$
2: $e \leftarrow H(m)$
3: $w \leftarrow s^{-1} \ mod \ n$
4: $u_1 \leftarrow ew \ mod \ n$; $u_2 \leftarrow rw \ mod \ n$
5: $(X_1, Y_1) \leftarrow u_1 G + u_2 D_a$
6: $v \leftarrow X_1 \ mod \ n$
7: **Si** $v = r$ Regresar "Firma aceptada"
9: **Si no** Regresar "Firma rechazada"

III. MULTIPLICACIÓN ESCALAR

Para una comprensión más sencilla de la multiplicación escalar, su realización puede ser dividida en tres capas independientes [3].

A. Capa superior

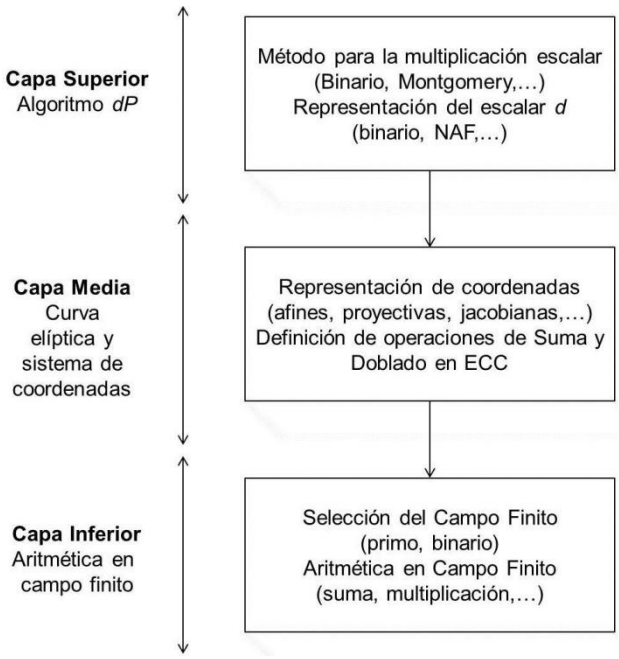


Fig. 1 Modelo de capas para la implementación de la multiplicación escalar.

En la capa superior se encuentran los diferentes métodos para efectuar la multiplicación escalar.

La multiplicación escalar dP donde d es un entero en el intervalo $[1, n-1]$ y P es un punto en la curva elíptica E definida sobre el campo finito, es el resultado de sumar $P + P + P + \cdots + P$, $d - 1$ veces [17]. Esta operación es realizada aplicando una secuencia de Sumas-ECC y Doblados-ECC.

El método básico para efectuar la multiplicación escalar para cuando P es un punto desconocido es el método de Suma y Doblado, donde la representación binaria del escalar d es usada. Existen dos algoritmos para calcular *dP* en función de cómo se recorren los bits del escalar *d*, de izquierda a derecha o de derecha a izquierda.

Multiplicación Doblado y Suma Izquierda-Derecha. El algoritmo 3 realiza en promedio $(m-1)/2$ Sumas-ECC y $m-1$ Doblados-ECC, debido a que el número esperado de unos en la representación binaria del escalar d es $t/2 \approx m/2$, donde t es el número de bits del escalar d.

Algoritmo 3: Multiplicación Doblado y Suma Izquierda-Derecha

Requiere: $d = (d_{t-1}, \ldots, d_1, d_0)_2$, $P \in E(GF_q)$
Salida: dP
1: $Q \leftarrow P$
2: **Para** $i = t-2$ hasta 0 **hacer**
3: $Q \leftarrow 2Q$
4: **Si** $d_i = 1$ **entonces**
5: $Q \leftarrow Q + P$
6: **Fin Si**
7: **Fin Para**
8: Regresar (Q)

Ya que este método itera de izquierda a derecha y el bit más significativo del escalar d es uno, el ciclo se realiza desde t-2 hasta 0, ahorrando así una Suma-ECC y un Doblado-ECC.

Multiplicación Doblado y Suma Derecha-Izquierda. El algoritmo 4 realiza en promedio $m/2$ Sumas-ECC y m Doblados-ECC. Debido a que este método se realiza de derecha a izquierda, se debe comenzar a iterar desde 0 hasta $t-1$.

Algoritmo 4: Multiplicación Doblado y Suma Derecha-Izquierda
Requiere: $d = (d_{t-1}, \dots, d_1, d_0)_2$, $P \in E(GF_q)$
Salida: dP
1: $Q \leftarrow \infty$
2: **Para** $i = 0$ hasta $t-1$ **hacer**
3: **Si** $d_i = 1$ **entonces**
4: $Q \leftarrow Q + P$
5: **Fin Si**
6: $P \leftarrow 2P$
7: **Fin Para**
8: Regresar (Q)

En los algoritmos 3 y 4, se efectúa un Doblado-ECC de manera incondicional y la operación de Suma-ECC se lleva a cabo solo si el bit d_i es uno.

Para mejorar el tiempo de ejecución de la multiplicación escalar se han diseñado nuevos métodos, algunos de los cuales logran disminuir la cantidad de Sumas-ECC, al reducir el número de dígitos diferentes de cero del escalar d, como por ejemplo los métodos *NAF* [18] y *wNAF* [19, 20], que recodifican d de forma que se incrementan los símbolos con los que se puede representar d.

Multiplicación Binaria NAF. El método *NAF* (Non-Adjacent Form) de un entero positivo d es una expresión de la forma $d = \sum_{i=0}^{l-1} d_i 2^i$ donde $d_i \in \{0, \pm 1\}, d_{l-1} \neq 0$ y al menos dos bits consecutivos d_i son diferentes de cero. La longitud de la representación *NAF* es representada por l.

Algunas de las propiedades de la representación *NAF* de un entero positivo *d* son [4]:

1. d tiene una única representación NAF denotada por $NAF(d)$
2. $NAF(d)$ tiene menos dígitos diferentes de cero que la representación binaria de d
3. La longitud de $NAF(d)$ es a lo más un bit mayor que la longitud de la representación binaria de d
4. Si la longitud de $NAF(d)$ es l, entonces $2^l/3 < d < 2^{l+1}/3$
5. La densidad promedio de dígitos diferentes de cero de $NAF(d)$ es aproximadamente 1/3

El valor $NAF(d)$ puede ser calculado eficientemente usando el algoritmo 5.

Algoritmo 5: Representación NAF
Requiere: Entero positivo d
Salida: $NAF(d)$
1: $i \leftarrow 0$
2: **Mientras** $d \geq 1$ **hacer**
3: **Si** d es impar **entonces**
4: $d_i \leftarrow 2 - (d \bmod 4), d \leftarrow d - d_i$
5: **Sino**
6: $d_i \leftarrow 0$
7: $d \leftarrow d/2, i \leftarrow i + 1$
8: **Fin Si no**
9: **Fin Mientras**
10: Regresar $(d_{i-1}, d_{i-2}, \dots, d_1, d_0)$

Posteriormente el algoritmo 6 se puede utilizar para obtener la multiplicación escalar. Este algoritmo realiza en promedio $(m-1)/3$ Sumas-ECC o restas de puntos y $m-1$ Doblados-ECC, gracias a que la densidad de dígitos diferentes de cero es de dos ceros por cada dígito diferente cero.

Algoritmo 6: Multiplicación binaria NAF de Izquierda-Derecha
Requiere: d, $P \in E(GF_q)$
Salida: dP
1: Obtener $NAF(d) = \sum_{i=0}^{l-1} d_i 2^i$
2: $Q \leftarrow P$
3: **Para** $i = l-2$ hasta 0 **hacer**
4: $Q \leftarrow 2Q$
5: **Si** $d_i = 1$ **entonces** $Q \leftarrow Q + P$
6: **Si** $d_i = -1$ **entonces** $Q \leftarrow Q - P$
7: **Fin Para**
8: Regresar (Q)

Como el método va de izquierda a derecha, nuevamente puede ahorrarse una Suma-ECC o una resta de puntos y un Doblado-ECC al igual que el algoritmo 1.

Multiplicación Binaria wNAF. El método *wNAF* (window Non-Adjacent Form) permite reducir aún más la densidad de dígitos diferentes de cero del escalar d, por lo que disminuye la cantidad de Sumas-ECC requeridas para multiplicación escalar.

Algunas de las propiedades de la representación *wNAF* de un entero d son [4]:

1. d tiene una única representación *wNAF* denotada por $NAF_w(d)$
2. $NAF_2(d) = NAF(d)$
3. La longitud de $NAF_w(d)$ es a lo más un dígito mayor que la longitud de la representación binaria de d
4. La densidad promedio de dígitos diferentes de cero de $NAF_w(d)$ es aproximadamente $1/(w+1)$

El valor $NAF_w(d)$ puede ser calculado mediante el algoritmo 7.

Algoritmo 7: Representación wNAF
Requiere: Entero positivo d
Salida: $NAF_w(d)$
1: $i \leftarrow 0$
2: **Mientras** $d \geq 1$ **hacer**
3: **Si** d es impar **entonces**
4: $d_i \leftarrow d \bmod 2^w, d \leftarrow d - d_i$
5: **Si no**
6: $d_i \leftarrow 0$
7: $d \leftarrow d/2, i \leftarrow i + 1$
8: **Fin Si no**
9: **Fin Mientras**
10: Regresar $(d_{i-1}, d_{i-2}, \dots, d_1, d_0)$

La función *mods* es definida de manera distinta en el campo primo y binario. Los algoritmos 8 y 9 calculan esta función.

Algoritmo 8: Función mods campo primo
Requiere: d, w
Salida: $d \bmod 2^w$
1: **Si** $d \bmod 2^w \geq 2^w/2$

2: Regresar (d mod 2^w) $- 2^w$
3: **Si no**
4: Regresa d mod 2^w

Algoritmo 9: Función mods campo binario
Requiere: d, w
Salida: $d\ mods\ 2^w$
1: **Si** d $\geq 2^{w-1}$
2: Regresar (d mod 2^w) $- 2^w$
3: **Si no**
4: Regresa d mod 2^w

El algoritmo 10 muestra la multiplicación escalar de izquierda a derecha utilizando la representación *wNAF*. Este algoritmo realiza en promedio $m/(w+1)$ Sumas-ECC o restas de puntos y m Doblados-ECC.

Algoritmo 10: Multiplicación wNAF de Izquierda-Derecha
Requiere: ancho w, entro positivo $d, P \in E(GF_q)$
Salida: dP
1: Obtener $wNAF(d) = \sum_{i=0}^{l-1} d_i 2^i$
2: Precalcular $P_i = iP$ para $i \in \{1,3,\dots,2^{w-1}-1\}$
3: $Q \leftarrow \infty$
4: **Para** $i = l-1$ hasta 0 **hacer**
5: $Q \leftarrow 2Q$
6: **Si** $d_i \neq 0$ **entonces**
7: **Si** $d_i > 0$ **entonces**
8: $Q \leftarrow Q + P_{ki}$
9: **Si no**
10: $Q \leftarrow Q - P_{-ki}$
11: **Fin Si**
12: **Fin Si**
13: **Fin Para**
14: Regresar (Q)

B. Capa media

La capa media corresponde al sistema de coordenadas en las cuales los puntos de la curva elíptica son representados. Existen diferentes tipos de coordenadas con las que es posible representar los puntos en la curva elíptica. Las más populares son las coordenadas afines, en donde cada punto en la curva es representado por el par *(x, y).* En este trabajo se utiliza este tipo de representación.

Esta capa además define como son realizadas las operaciones de suma y doblado en la curva elíptica.

Cuando la curva está definida sobre el campo primo $E(GF(p))$, las operaciones Suma-ECC y Doblado-ECC son definidas como sigue [21]:

Sea $a, b \in GF(p)$ que satisfacen la ecuación $4a^3 + 27b^2 \neq 0$. Sea $P, Q, R \in E(GF(p))$. $P = (x_1, x_1)$

$Q = (x_2, x_2)$ $R = (x_3, x_3)$

1. $P + Q = \infty$, si $P = \infty$ o $Q = \infty$
2. $Suma - ECC(P \neq \pm Q)$

$R = P + Q$, donde

$$x_3 = \lambda^2 - x_1 - x_2$$
$$y_3 = \lambda(x_1 - x_3) - y_1$$
$$\lambda = \frac{(y_2 - y_1)}{(x_2 - x_1)}$$

3. $Doblado - ECC(P = Q)$

$R = P + Q = 2P$, donde

$$x_3 = \lambda^2 - 2_{x_1}$$
$$y_3 = \lambda(x_1 - x_3) - y_1$$
$$\lambda = \frac{3x_1^2 + a}{2y_1}$$

Cuando la curva elíptica está definida sobre el campo finito $E(GF(2^m))$, las operaciones de Suma-ECC y Doblado-ECC se realizan como sigue [21]:

Dado los puntos:

$P = (x_1, x_1), Q = (x_2, x_2),$ $R = (x_3, x_3) \in E(GF(2^m))$

1. $P + Q = \infty$, si $P = \infty$ o $Q = \infty$
2. $Suma - ECC(P \neq \pm Q)$

$R = P + Q$, donde

$$x_3 = \lambda^2 + \lambda + a$$
$$y_3 = \lambda(x_1 + x_3) + x_3 + y_1$$
$$\lambda = \frac{(y_2 - y_1)}{(x_2 - x_1)}$$

3. $Doblado - ECC(P = Q)$

$R = P + Q = 2P$, donde

$$x_3 = \lambda^2 + \lambda + a$$
$$y_3 = x_1^2 + \lambda x_3 + x_3$$
$$\lambda = x_1 + \frac{y_1}{x_1}$$

C. Capa inferior

Finalmente la capa inferior involucra aritmética en campo finito. La implementación eficiente de estas operaciones aritméticas impactan el desempeño de la multiplicación escalar.

Para realizar la Suma-ECC y el Doblado-ECC de la capa media se requieren operaciones de la capa inferior como suma, multiplicación, inversión (puede ser sustituida por la división) y elevación al cuadrado.

Para $GF(2^m)$ la implementación de las operaciones en campo finito dependen de una base, la cual puede ser polinomial, normal o dual [21]. En bases polinomiales, los elementos de $GF(2^m)$ son vistos como polinomios $A(x)$ de grado $m-1$, con coeficientes en $GF(2) = \{1,0\}$. Una base en $GF(2^m)$ es representada por $\{1, t, t_1, t_2, \dots, t_{m-1}\}$, donde t es un cuadrado de un polinomio irreducible $F(x)$ de grado m [9] y fue la base utilizada en la realización de pruebas de este artículo. La aritmética en $GF(2^m)$ con base polinomial es aritmética de polinomios modulo $f(x)$. Para $GF(p)$, la aritmética es implementada como aritmética de enteros modulo p.

IV. IMPLEMENTACIÓN Y RESULTADOS

La implementación de la multiplicación escalar y la realización de pruebas se llevo a cabo en una computadora de escritorio y en un dispositivo móvil, cuyas características se describen en la tabla I.

TABLA I. CARACTERÍSTICAS DE CÓMPUTO DE LA PC Y EL DISPOSITIVO MÓVIL

Parámetros	**Especificaciones**	
	PC	Móvil
Procesador	Intel Core i5 2.5GHz	ARM 600 MHz
SO	Windows 7	Android 2.2
Memoria RAM	4 GB	170 MB

TABLA II. ESPECIFICACIÓN DE PARÁMETROS DE LA CURVA SECP256R1

Parámetro	Valor
$\boldsymbol{p}$	FFFFFFFF 00000001 00000000 00000000 00000000 FFFFFFFF FFFFFFFF FFFFFFFF
$\boldsymbol{a}$	FFFFFFFF 00000001 00000000 00000000 00000000 FFFFFFFF FFFFFFFF FFFFFFFC
$\boldsymbol{b}$	5AC635D8 AA3A93E7 B3EBBD55 769886BC 651D06B0 CC53B0F6 3BCE3C3E 27D2604B
$\boldsymbol{G}$	(6B17D1F2 E12C4247 F8BCE6E5 63A440F2 77037D81 2DEB33A0 F4A13945 D898C296, 4FE342E2 FE1A7F9B 8EE7EB4A 7C0F9E16 2BCE3357 6B315ECE CBB64068 37BF51F5)
$\boldsymbol{n}$	FFFFFFFF 00000000 FFFFFFFF FFFFFFFF BCE6FAAD A7179E84 F3B9CAC2 FC632551
$\boldsymbol{h}$	01

La plataforma de programación utilizada fue Java Edición Estándar (o Java SE 7) y el entorno de desarrollo utilizado fue Eclipse.

Un aspecto fundamental al momento de seleccionar las curvas elípticas con las que se realizó la etapa de implementación y pruebas es el nivel de seguridad que éstas ofrecen, es por ello que siguiendo la recomendación del SECG [15], este trabajo se enfocó en desarrollar un análisis de un nivel de seguridad de 128 bits tanto para campo primo como para campo binario. Los resultados alcanzados pudieron ser corroborados en los vectores de prueba que el estándar SECG [15] proporciona.

Para la obtención de resultados, y debido a la variación del tiempo respecto a la generación de la llave pública, se optó por seguir el teorema del límite central el cual menciona que para obtener una distribución normal es suficiente evaluar un mínimo de 30 muestras [22]. Para la realización de pruebas, en este artículo se tomaron 32 muestras de cada uno de los cuatro algoritmos analizados para el cálculo de la multiplicación escalar (ver tablas III y V), y posteriormente se obtuvo el promedio de cada uno de ellos.

A. Campo primo

Para evaluar los métodos de multiplicación escalar definido sobre el campo primo, para la tupla $T = (p, a, b, G, n, h)$ se utilizaron los valores recomendados para la curva secp256r1, la cual está recomendada en el estándar SECG [15], y es compatible con IEEE [10] y esta también recomendada en ANSI [12, 13] y NIST [11]. Los valores para cada elemento de la tupla *T* se muestran en la tabla II.

Para la implementación de la multiplicación escalar sobre campo primo se utilizaron clases incluidas en la versión 1.2 de la Arquitectura Criptográfica de Java o JCA, específicamente las clases incluidas en el paquete java.security.spec el cual brinda soporte para esquemas criptográficos y que fue utilizado para implementar las operaciones de la capa superior y la capa media de la multiplicación escalar. También, se utilizó la clase `BigInteger` con la cual fue posible implementar las operaciones requeridas en la capa inferior de la multiplicación escalar, ya que permite realizar operaciones sobre números grandes necesarios para la representación de parámetros criptográficos.

TABLA III. TIEMPO DE EJECUCIÓN DE LA MULTIPLICACIÓN ESCALAR SOBRE CAMPO PRIMO

	Tiempo de ejecución (ms)	
	PC	Móvil
S&D Izq-Der	26.50	678.16
S&D Der-Izq	27.03	742.16
NAF	21.09	556.61
wNAF	10.59	885.97

TABLA IV. ESPECIFICACIÓN DE PARÁMETROS DE LA CURVA SECT283R1

Parámetro	Valor
$\boldsymbol{f(x)}$	$x^{283} + x^{12} + x^7 + x^5 + 1$
$\boldsymbol{a}$	00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000001
$\boldsymbol{b}$	027B680A C8B8596D A5A4AF8A 19A0303F CA97FD76 45309FA2 A581485A F6263E31 3B79A2F5
$\boldsymbol{G}$	(05F93925 8DB7DD90 E1934F8C 70B0DFEC 2EED25B8 557EAC9C 80E2E198 F8CDBECD 86B12053 03676854, FE24141C B98FE6D4 B20D02B4 516FF702 350EDDB0 826779C8 13F0DF45 BE8112F4)
$\boldsymbol{n}$	03FFFFFF FFFFFFFF FFFFFFFF FFFFFFFF FFFFEF90 399660FC 938A9016 5B042A7C EFADB307
$\boldsymbol{h}$	02

En la Tabla III se muestran los tiempos de ejecución para la multiplicación escalar definida sobre campo primo, tanto en la computadora de escritorio como en el dispositivo móvil.

Como puede apreciarse en la Tabla III el método que menor tiempo de ejecución demanda en el ordenador de escritorio es el *wNAF*, mientras que en el dispositivo móvil es el *NAF*. En la mayoría de los casos la multiplicación escalar se calcula en menos de un segundo sobre el procesador ARM, el cual es un tiempo que puede ser tolerado en la mayoría de las aplicaciones móviles.

B. Campo binario

Para evaluar los métodos de multiplicación escalar definido sobre el campo binario, para la tupla $T = (m, f(x)\ a, b, G, n, h)$ se utilizaron los valores recomendados para la curva sect283r1, la cual está recomendada en el estándar SECG [15], y es compatible con ANSI [12, 13], IEEE [10] y esta también recomendada en NIST [11]. Los valores para cada elemento de la tupla T se muestran en la tabla IV.

En cuanto a la implementación de la multiplicación escalar sobre campo binario se utilizó el paquete java.security.spec que proporciona clases e interfaces para la especificación de parámetros de algoritmos criptográficos y la librería de código

TABLA V. TIEMPO DE EJECUCIÓN DE LA MULTIPLICACIÓN ESCALAR SOBRE CAMPO BINARIO

	Tiempo de ejecución (ms)	
	PC	Móvil
S&D Izq-Der	86.18	7408.5
S&D Der-Izq	91.93	8285.3
NAF	60.25	4977.9
wNAF	59.09	4947.7

abierto FlexiProvider [23], ya que brinda soporte para operaciones de polinomios, indispensables en este tipo de campo por lo cual fue posible realizar las operaciones respectivas a la capa media y la capa inferior del modelo de capas de la multiplicación escalar.

En la Tabla V se muestran los tiempos necesarios para la ejecución de la multiplicación escalar, tanto en la computadora de escritorio como en el dispositivo móvil.

Como se puede apreciar el método con un menor tiempo de ejecución sobre campo binario es *wNAF* al realizarse la implementación en el ordenador de escritorio al igual que para el dispositivo móvil. Este valor incluye además el tiempo de generación de la representación *wNAF* del escalar *d*, lo que permite confirmar que al reducir la densidad de dígitos diferentes de cero del escalar *d* mediante el método *wNAF* es posible lograr incrementar la eficiencia en el desempeño de la multiplicación escalar.

De acuerdo a los resultados obtenidos la Criptografía de Curvas Elípticas sobre campo primo resulta ser más eficiente en comparación con el campo binario, tanto para PC como para móvil. En los resultados obtenidos en la PC se puede apreciar que la mayoría de los métodos implementados en campo binario son en promedio 2 o 3 veces más lentos que en campo primo exceptuando el método *wNAF* el cual es 5 veces más lento que su versión en campo primo. Mientras que en los resultados obtenidos del dispositivo móvil se puede observar que los métodos en campo binario son entre 8 y 11 veces más lentos que los métodos en campo primo con excepción del método *wNAF* el cual es 5 veces más lento que su versión en campo primo. Estas diferencias en el tiempo de ejecución de los algoritmos para el cálculo de la multiplicación escalar nos muestran que dentro de las implementaciones en software es mejor opción el uso del campo primo, sin embargo también es importante la elección correcta del método para la multiplicación escalar dado a que el uso de pre-caculos como en el caso de *wNAF* indica mayor uso de recursos.

V. CONCLUSIONES

Debido al gran interés del uso de esquemas criptográficos basados en curva elíptica, resulta crucial el análisis de las operaciones subyacentes para la implementación de dicha técnica, con la finalidad de lograr el mejor desempeño posible.

En este artículo presentamos los resultados obtenidos por la implementación de cuatro diferentes técnicas algorítmicas para el cálculo de la multiplicación escalar de curva elíptica sobre campos finitos primos y binarios

Los resultados obtenidos reflejan que los métodos *NAF* y *wNAF* son más rápidos, debido a que en ambos métodos el escalar *d* es transformado y la densidad de dígitos diferentes de cero resulta menor que en su representación binaria, que es la que usan los métodos de Suma y Doblado. Esto implica que se necesiten menos operaciones aritméticas para el cálculo de la multiplicación escalar, ya que al disminuir la densidad de dígitos diferentes de cero en el escalar *d*, el número de Sumas-ECC se reduce.

Se realizó un análisis de los tiempos de ejecución del ordenador de escritorio contra los obtenidos en el dispositivo móvil. Podemos mencionar que al ejecutar los algoritmos para la multiplicación escalar utilizando el campo primo, el tiempo de ejecución en el dispositivo móvil fue en promedio 26 veces más lento que el tiempo consumido por la PC con excepción del método *wNAF* que fue 83 veces más lento en el móvil en comparación con la PC. Al ejecutar los algoritmos para la multiplicación escalar utilizando el campo binario, el tiempo requerido por el móvil fue en promedio 85 veces más lento que la PC. Gracias a estas comparaciones es que podemos concluir que el uso del campo primo en implementaciones de software sobre plataformas móviles es más adecuado que el uso del campo binario.

Con los resultados obtenidos podemos puntualizar que múltiples aplicaciones móviles pueden soportar el cálculo de las operaciones sobre curvas elípticas, lo cual permite el uso de ECC en estos dispositivos con características de procesamiento restringidas.

Como trabajo futuro se explorarán algoritmos para cálculo de la multiplicación escalar definidos sobre otros campos finitos, tales como GF(3^m). También se sugiere probar técnicas algorítmicas más complejas que realicen el pre-procesamiento del escalar *d*, tales como el método *Frac-wNAF* (fractional window Non-Adjacent Form) [24], el método *Fixed-base NAF windowing* [4], o el método *Sliding window* [5]. Además de un análisis detallado de los métodos y las capas de la multiplicación escalar para su implementación sobre arquitecturas paralelas tales como los GPUs (Graphics Processing Unit).

REFERENCIAS

[1] Y. Kawahara, T. Takagi, and E. Okamoto, Efficient Implementation of Tate Pairing on a Mobile Phone Using Java. In Computational Intelligence and Security, vol. 2, pp. 1247 - 1252, Berlin, 2006.

[2] A. Weimerskirch, C. Paar, and S. Chang Shantz. Elliptic Curve Cryptography on a Palm OS Device. In V. Varadharajan and Y. Mu, editors, The 6th Australasian Conference on Information Security and Privacy, vol. LNCS 2119, pp. 502-513, Berlin, 2001.

[3] M. Morales-Sandoval, A reconfigurable and interoperable hardware architecture for elliptic curve cryptography, Tesis de Doctorado, Instituto Nacional de Astrofísica, Óptica y Electrónica, México, 2008.

[4] D. Hankerson, A. J. Menezes, and S. Vanstone. Guide to Elliptic Curve Cryptography, USA, 2003, Springer-Verlag New York, Inc.

[5] P. Longa, Accelerating the Scalar Multiplication on Elliptic Curve Cryptosystems over Prime Fields, Tesis de Maestría, School of Information Technology and Engineering University of Ottawa, Canada, 2007

[6] N. Ferguson and B. Schneier, Practical Cryptography. Wiley, 2003.

[7] N. Koblitz. Elliptic curve in cryptography. American Mathematical Society J. Comput. Math., pp. 207–209, 1987.

[8] V. S Miller. Use of elliptic curves in cryptography. In Lecture notes in computer sciences; 218 on Advances in cryptology CRYPTO 85, pp. 417–426, USA, 1986. Springer-Verlag New York, Inc.

[9] A. J. Menezes, T. Okamoto, and S. A. Vanstone, Reducing elliptic curve logarithms to logarithms in a finite field, IEEE Transactions on Information Theory , 1639 (1993).

[10] IEEE P1363. Standard Specifications for Public-Key Cryptography. Institute of Electrical and Electronics Engineers, 2000.

[11] NIST, Recommended Elliptic Curves for Federal Government Use, 1999. [En línea] Disponible: http://csrc.nist.gov/groups/ST/toolkit/documents/dss/NISTReCur.pdf
[12] ANSI X9.62-1998: Public Key Cryptography for the Financial Services Industry: the Elliptic Curve Digital Signature Algorithm (ECDSA). American Bankers Association, 1999.
[13] ANSI X9.63-199x: Public Key Cryptography for the Financial Services Industry: Key Agreement and Key Transport Using Elliptic Curve Cryptography. American Bankers Association, October, 1999. Working Draft.
[14] ISO/IEC 18033-3:2005: International Standard. [En línea] Disponible: http://webstore.iec.ch/p-preview/info_isoiec18033-3%7Bed1.0%7Den.pdf
[15] D. R. Brown, SEC 2: Recommended Elliptic Curve Domain Parameters, Certicom Research, 2010. [En línea] Disponible: http://www.secg.org/collateral/sec2_final.pdf
[16] A. Trujillo-Vázquez, Criptografía basada en ECC y AES para dispositivos con recursos restringidos, Tesis de Maestría, Universidad Politécnica de Cd. Victoria, México, 2011.
[17] G. L. Hernández, Paralelización de la multiplicación escalar en curvas elípticas en una arquitectura multinúcleo de Intel, Tesis de Maestría, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, México, 2010.
[18] Harsandeep Brar and Rajpreet Kaur. Design and Implementation of Block Method for Computing NAF. International Journal of Computer Applications pp. 37-41, 2011.
[19] Fan R. On The Efficiency Analysis of wNAF and wMOF, Tesis de Maestría. Technische Universität Darmstadt,Alemania, 2005.
[20] K. Okeya, and T. Takagi, "The Width-w NAF Method Provides Small Memory and Fast Elliptic Scalar Multiplications Secure against Side Channel Attacks," CT-RSA 2003: The Cryptographers' Track at the RSA Conference 2003, USA, 2003.
[21] Z. Cheng, Simple Tutorial on Elliptic Curve Cryptography, School of Computing Science, 2004.
[22] J. Gil-Flores, Aplicación Del Método Bootstrap al Contraste De Hipótesis En La Investigación Educativa, Universidad de Sevilla *Revista de Educación* núm. 336, pp. 251-265, 2005.
[23] FlexiProvider 2012. [En línea] Disponible: http://www.flexiprovider.de/overview.html
[24] Bodo Moller: Improved Techniques for Fast Exponentiation. In: ICISC 2002, LNCS, vol.2587, pp. 298-312. Springer Heidelberg, 2003.

Karina Vega Castillo, nació en Cd Victoria, Tamaulipas el 15 de agosto de 1986. Obtuvo el grado de Ingeniero en Sistemas Computaciones por el Instituto Tecnológico de Ciudad Victoria en el estado de Tamaulipas en el año 2009. Ella actualmente, se encuentra desarrollando su proyecto de investigación para obtener el grado de Maestra en Ingeniería en la Universidad Politécnica de Ciudad Victoria, Tamaulipas. Su área de especialidad es la seguridad informática y el desarrollo de esquemas criptográficos.

Antonio Cortina Reyes nació en Cd. Victoria, Tamaulipas el 06 de abril de 1988. Obtuvo el grado de Ingeniero en Sistemas computacionales en el Instituto Tecnológico de Cd. Victoria en el estado de Tamaulipas en el año 2011.El actualmente se encuentra desarrollando su proyecto de investigación para obtener el grado de Maestro en Ingeniería en la Universidad Politécnica de Ciudad Victoria, Tamaulipas. Su área de especialidad es la seguridad informática y el desarrollo de esquemas criptográficos.

Miguel Morales Sandoval nació en Calpan, Puebla el 07 de octubre de1978. Recibió el grado de licenciado en Ciencias de la Computación en 2002 por la Universidad Autónoma de Puebla. Recibió el grado de Maestro y Doctor en Ciencias en 2004 y 2008 respectivamente, ambos por el Instituto Nacional de Astrofísica, Óptica y Electrónica, ubicado en Tonantzintla, Puebla. Desde diciembre de 2008 es profesor investigador en la Universidad Politécnica de Victoria, en Ciudad Victoria, Tamaulipas donde dirige tesis de maestría e imparte diversas materias de posgrado y licenciatura en el área de las Tecnologías de Información. Desde 2003 realiza proyectos de investigación en las áreas de la criptografía, con especialidad en Criptografía de Curvas Elípticas y el diseño digital de arquitecturas hardware usando dispositivos reconfigurables y lenguajes de descripción de hardware. Ha publicado alrededor de 20 artículos en revistas y congresos especializados, y participa en el desarrollo de proyectos de investigación aplicada. El Dr. Morales-Sandoval es miembro del Sistema Nacional de Investigadores y cuenta con la distinción de perfil deseable por PROMEP.

Sistema de control de emersión e inmersión del vehículo Sub-Chaac

Sotelo O. Arturo, García O. Manuel D.J., Coria D.R. Luis N., Vázquez L. Carlos E. y Ortega C. Jorge A.

Emersion and immersion control system for the Sub-Chaac vehicle

***Abstract*— In this work we propose a solution in order to control the immersion and emersion of an unmanned underwater vehicle. This vehicle has been named "Sub-Chaac" honoring to the Mayan god associated to water and rain. So as to measure environment variables Sub-Chaac has instrumentation on-board allowing it to find out its operative conditions. Sub-Chaac has been designed over several stages, just to mention some of them: a) Mechanical design, b) Ballast control, c) Speed control, d) Power control, e) Electronic Instrumentation and f) Communications and telemetry. This work is about the development of the leveling and deepness control subsystem, which is assisted by the instrumentation subsystem.**

***Keywords*— Emersion, Immersion Underwater vehicle, Unmanned.**

***Resumen*— En este trabajo se propone una solución para el control de emersión e inmersión de un vehículo subacuático no tripulado. El vehículo ha sido nombrado "Sub-Chaac" en honor a la deidad Maya asociada al agua y la lluvia, y se referirá a él de esta forma en el resto del documento. El Sub-Chaac cuenta con instrumentación para medir algunas variables de su entorno y poder determinar sus condiciones de operación. Existen diversas etapas en el diseño del Sub-Chaac, por mencionar algunas: a) Diseño mecánico, b) Control de lastre, c) Control de velocidad, d) Control de potencia, e) Instrumentación electrónica y f) Comunicaciones y telemetría. Este trabajo trata de parte del desarrollo del subsistema de control de lastre, consistente en el control de nivelación y profundidad, el cual es asistido por el subsistema de instrumentación.**

***Palabras clave*— Emersión, Inmersión Vehículo subacuático, No tripulado.**

Manuscrito recibido el 12 de Julio de 2012. Este trabajo fue respaldado por el departamento de Ing. Eléctrica y Electrónica del Instituto Tecnológico de Tijuana, por el proyecto DGEST 3416.10-P y el cuerpo académico ITT-CA-6.

Sotelo O. Arturo, hasta la fecha se ha de desempeñado como Profesor de Tiempo Completo del Instituto Tecnológico de Tijuana, en el Departamento de Ingeniería Eléctrica y Electrónica; Blvd. Industrial S/N, Mesa de Otay; Tijuana, B.C., México; C.P. 22500; Tel/Fax: (664) 6244743; (e-mail asotelo@tectijuana.mx).

García O. Manuel D.J. , hasta la fecha se ha de desempeñado como jefe del departamento y Profesor de Tiempo Completo del Instituto Tecnológico de Tijuana, en el Departamento de Ingeniería Eléctrica y Electrónica; Blvd. Industrial S/N, Mesa de Otay; Tijuana, B.C., México; C.P. 22500; Tel/Fax: (664) 6244743; (e-mail mdejgaro@yahoo.com).

Coria D.R. Luis N. hasta la fecha se ha de desempeñado como Profesor/Investigador de Tiempo Completo del Instituto Tecnológico de Tijuana, en el Departamento de Ingeniería Eléctrica y Electrónica; Blvd. Industrial S/N, Mesa de Otay; Tijuana, B.C., México; C.P. 22500; Tel/Fax: (664) 6244743; (e-mail mdejgaro@yahoo.com)..

I. INTRODUCCIÓN

El diseño y construcción de vehículos subacuáticos es un tema de interés debido sus aplicaciones, principalmente: en el estudio de perfiles de columna de agua en emersiones verticales mediante sonar [1,5], en la generación de mapas para exploración arqueológica submarina [2], en tareas de monitorización colectiva para determinar la trayectoria óptima para llegar a un objeto sumergido determinado [3], en la búsqueda y rescate de objetos sumergidos [4]. La operación habitual del vehículo requiere de un sistema electromecánico que permita realizar la inmersión y emersión del vehículo, así como el control de movimiento del mismo. Dentro de este marco diversos trabajos abordan técnicas aplicadas para este fin, por ejemplo, en los trabajos [6,7] se diseñan controladores con métodos clásicos; se aplican controladores por modos deslizantes en [8]; en el trabajo de Meldrum [9] se explora la aplicación de un GPS para determinar la posición de un vehículo en el diseño del controlador.

El Instituto Tecnológico de Tijuana a través del Departamento de Ingeniería Eléctrica y Electrónica ha puesto en marcha un proyecto consistente en el diseño y construcción de un vehículo subacuático no tripulado para exploración. El vehículo fue denominado Sub-Chaac (Chaac: Deidad Maya relacionada con el agua y la lluvia), y la construcción del prototipo se encuentra dentro de las metas que se persiguen en el proyecto DGEST "Vehículo Subacuático No Tripulado".

Para lograr la implementación del Sub-Chaac es necesario desarrollar diversos subsistemas: a) Casco, b) Control inmersión y emersión, c) Control de movimiento, d) Etapa de potencia, d) Instrumentación electrónica, e) Comunicaciones y telemetría y d) Administración de energía. En este documento se describen los detalles técnicos del diseño, así como la descripción de la implementación y resultados obtenidos en el desarrollo del sistema para inmersión y emersión del Sub-Chaac.

El documento está organizado como sigue: En la Sección II se presentan los preliminares en los que se basa el desarrollo del sistema de control del Sub-Chaac. En la Sección III se muestra el desarrollo del sistema de control y una descripción de cada uno de los elementos utilizados. La Sección IV muestra los resultados obtenidos. Finalmente la Sección V incluye las conclusiones del trabajo realizado.

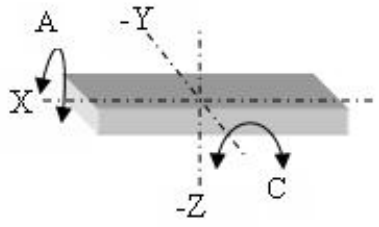


Fig. 1 Ilustración de los ejes: longitudinal X, Transversal Y, y Normal Z. Así como de los movimientos de Cabeceo C, y de Alabeo A.

A. Aspectos generales

Para la implementación del hardware del sistema se hace uso de una red maestro-esclavo de microcomputadoras interconectadas mediante una red I^2C. La parte principal del sistema de control es la computadora de navegación (*CN*), basada en un microcontrolador ARM7 de Atmel, cuya función es controlar la profundidad, nivelación, dirección y velocidad del Sub-Chaac. Adicionalmente se utilizan microcontroladores microchip para implementar la red de sensores que permiten la adquisición y preprocesamiento de las señales presión, temperatura e inclinación, así como para controlar el sentido de giro de los propulsores y de las bombas de llenado/vaciado de las cámaras de lastre.

Para determinar la desviación del Sub-Chaac respecto de la horizontal, se hace uso de un acelerómetro de tres ejes en conjunto de un giroscopio para determinar la velocidad angular, los cuales entregan las lecturas a través de I^2C. La desviación respecto de esta referencia comúnmente se les denomina de alabeo y cabeceo. El alabeo es la desviación al girar sobre el eje longitudinal del vehículo, *X*; el cabeceo es la desviación al girar sobre el eje transversal, *Y*, como se muestra en la Fig. 1. Estos movimientos presentan características similares entre ellos, así que resolviendo el alabeo se puede tomar como base para resolver el cabeceo. El ángulo de inclinación se puede medir en base a la componente de aceleración producida por la acción de la gravedad sobre el eje *X* y *Y* del vehículo.

Adicionalmente, al encontrarse el vehículo en movimiento las lecturas de los acelerómetros resultaran afectadas por los cambios de la velocidad en los tres ejes, resultando en una estimación incorrecta de la desviación respecto a la horizontal. Conociendo la magnitud y dirección de la velocidad con que cambia la desviación es posible compensar obteniendo una estimación más acertada. Para fusionar estas variables y obtener la desviación respecto de la horizontal, el filtro Kalman permite obtener una estimación inmune a los errores de la instrumentación y a los movimientos propios del vehículo [10]. Para estimar la inclinación con la menor cantidad de ruido, deberá ser posible describir el proceso mediante un sistema lineal. De acuerdo con [11] un sistema lineal esta descrito por la Ecuación de estado (1) y Ecuación de salida (2).

$$x_{k+1}=Ax_k + Bu_k + w_k \quad (1)$$

$$y_k=Cx_k + z_k \quad (2)$$

TABLA I.
DIRECCIONES DE LOS MÓDULOS ESCLAVOS EN LA RED I^2C

Dirección	*Descripción del dispositivo*
0x02	Sensor de inclinación
0x04	Amplificadores de potencia
0x06	Sensor de Profundidad y temperatura

Donde *k* es el índice de tiempo, *x* es el denominado estado del sistema, *u* es la entrada conocida al sistema, *y* es la medida de salida; finalmente según [12, 13], *w es* el denominado ruido del proceso y *z* el ruido de la medición, respectivamente.

II. DESARROLLO DEL SISTEMA DE CONTROL

El protocolo de comunicación I^2C permite que la *CN* se comunique con el resto del hardware usando las mismas líneas de comunicación que los dispositivos esclavos. Estos dispositivos tienen una dirección específica de acceso, así es como la *CN* puede determinar los valores de las variables y le permite enviar las consignas para los actuadores. Las direcciones propuestas para los dispositivos en el esquema de I^2C se muestran en la Tabla I.

En la Fig. 2 se muestra el diagrama esquemático de la implementación del sistema de control del Sub-Chaac. El microcontrolador mostrado en la parte central es encargado de enviar las señales de control a la etapa de potencia (*EP*), es importante señalar que aunque solo se muestra un solo amplificador (*AP*) en el circuito, el sistema cuenta con un total de cuatro, uno para cada motor/bomba. Estos circuitos se encuentran configurados como esclavos en la red de comunicación I^2C.

A grandes rasgos, el algoritmo para el control de inmersión/emersión se describe a continuación:

I. La *CN* lee los datos procesados de profundidad y temperatura que provienen de sensor remoto de la dirección 0x06, así como los datos procedentes del sensor de inclinación con la dirección 0x02.
II. Para controlar la posición horizontal y vertical del vehículo, los datos se procesan en la *CN* por medio de un algoritmo sencillo que evalúa once condiciones para determinar qué motores se deben activar y en qué orden tienen que operar.
III. El resultado del procesamiento de la información se envía al microcontrolador encargado de la etapa de potencia de los motores de las bombas, con la dirección 0x04. En base de los valores recibidos e envía la señal a cada *AP* para activar o desactivar las bombas en el sentido requerido para llevar al Sub-Chaac a la profundidad y nivelación deseada.

La computadora de navegación es la encargada de tomar la decisión de la profundidad a la que se debe situar el vehículo, así como de nivelarlo. La *CN* cuenta con un microprocesador ARM 7 de 32 bits y un núcleo que le permite ejecutar aplicaciones en multitareas. Adicionalmente, posee un sistema de archivos que habilita el almacenar en un archivo la información recopilada para su futuro análisis.

Uno de los acelerómetros mide, respecto del vehículo, en dirección paralela al eje normal, *Z*; los otros dos acelerómetros miden en dirección paralela a los eje *X* y *Y* con sus respectivos giroscopios que miden la velocidad angular. El ángulo de

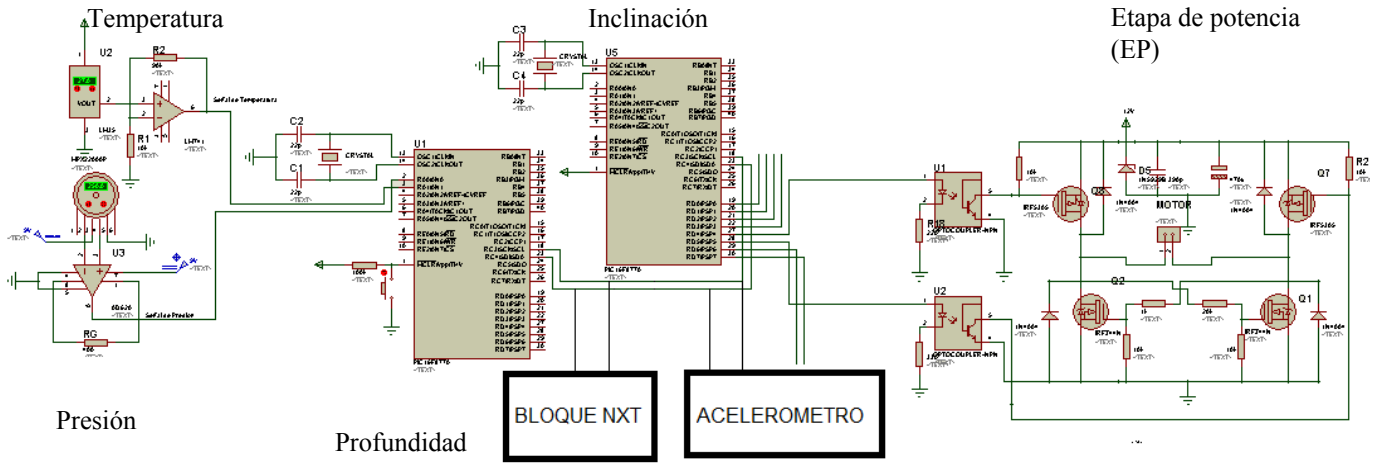


Fig. 2 Diagrama esquemático de la electrónica del sistema de control de inmersión/emersión del Sb-Chaacs.

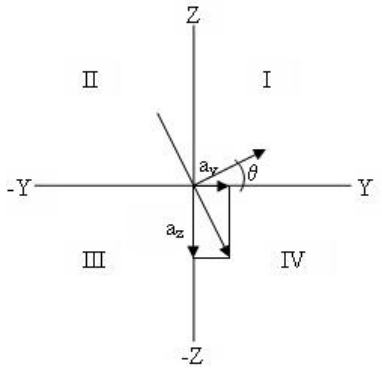


Fig. 3 Determinación del ángulo θ mediante las componentes de aceleración, visto desde la parte frontal.

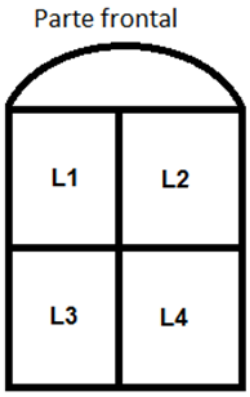


Fig. 4 Localización de las cámaras de lastre del Sub-Chaac, para lograr profundidad y nivelación.

inclinación de determina mediante la lectura de dos acelerómetros, como muestra la Fig. 3.

De acuerdo con [14] el ángulo está determinado por la Ecuación (3) y el signo del ángulo será positivo cuando la resultante de las componentes de aceleración este apuntando en los cuadrantes III a IV, y será negativo cuando apunte en los cuadrantes I a II.

$$\theta = a\tan\left(\frac{a_y}{a_z}\right) \tag{3}$$

Para lograr la inmersión/emersión y nivelación, el Sub-Chaac cuenta con cuatro cámaras de lastre, como se muestran en la Fig. 4. Los actuadores utilizados son cuatro bombas peristálticas, cuyo sentido se determina al cambiar la polaridad mediante los *EP*, esto con el objetivo de introducir y extraer el agua de los tanques de lastre del Sub-Chaac.

Una propuesta de solución para resolver el problema de inmersión/emersión y nivelación, consistió en analizar cada una de las combinaciones posibles entre los valores de la inclinación de los ejes principales del Sub-Chaac. La respuesta de los motores dependerá de la aceleración registrada en los ejes *X*, *Y* y *Z* de los acelerómetros y compensadas mediante los giroscopios y el filtro de Kalman: La activación de los motores se realizara cuando el valor de los ejes supere la tolerancia de inclinación para ambos ejes. Las condiciones posibles se muestran en la Tabla II.

Una vez activados los motores estos se mantendrán en este estado hasta que la inclinación en los ejes *X* y *Y* sean menores que el valor de tolerancia, esto para lograr una ventana de histéresis para los motores, y evitar al activación y desactivación constante de los motores y reducir el consumo de energía de las baterías.

Para controlar la profundidad a la que se debe ubicar el Sub-Chaac, se plantean tres condiciones, las cuales se analizarán cuando el vehículo se encuentre nivelado en su posición horizontal. Estas tres condiciones se establecen en la Tabla III.

Para determinar la profundidad se usa un batímetro, el cual mide la presión absoluta que produce el agua sobre el casco del Sub-Chaac y realiza compensación de la densidad tomando en cuenta la temperatura del agua, este trabajo fue desarrollado en otra etapa del proyecto realizado en paralelo.

TABLA II. CONDICIONES PARA ACTIVAR/DESACTIVAR EL BOMBEO PARA CONTROLAR LA POSICIÓN HORIZONTAL DEL SUB-CHAAC

Condición 1	*Estado de bombeo*	*Condición 5*	*Estado de bombeo*
L1	-1	L1	-1
L2	0	L2	-1
L3	0	L3	1
L4	1	L4	1

Condición 2	*Estado de bombeo*	*Condición 6*	*Estado de bombeo*
L1	1	L1	1
L2	0	L2	1
L3	0	L3	-1
L4	-1	L4	-1

Condición 3	*Estado de bombeo*	*Condición 7*	*Estado de bombeo*
L1	0	L1	-1
L2	-1	L2	1
L3	1	L3	-1
L4	0	L4	1

Condición 4	*Estado de bombeo*	*Condición 8*	*Estado de bombeo*
L1	0	L1	1
L2	1	L2	-1
L3	-1	L3	1
L4	0	L4	-1

-1: Sacar Agua; 0: Parar; 1: Meter Agua;

TABLA III. CONDICIONES PARA ACTIVAR/DESACTIVAR EL BOMBEO PARA CONTROLAR LA PROFUNDIDAD DEL SUB-CHAAC

Condición 10	*Estado de bombeo*	*Condición 9*	*Estado de bombeo*
L1	-1	L1	1
L2	-1	L2	1
L3	-1	L3	1
L4	-1	L4	1

Condición 11	*Estado de bombeo*
L1	0
L2	0
L3	0
L4	0

-1: Sacar Agua; 0: Parar; 1: Meter Agua;

Para el llenado y vaciado de las cámaras de lastre se usaron 4 bombas peristálticas, en la Fig. 5 se puede apreciar la colocación de las bombas de las cámaras de lastre frontales y los ductos para el llenado y vaciado. Sobre la cámara seca es posible ver la implementación provisional de la electrónica para la medición de profundidad y de la *EP* que activan las bombas.

En la Fig. 6 se puede apreciar la colocación de las bombas peristálticas de las cámaras posteriores, y sobre la cámara seca se puede apreciar la computadora de navegación.

Mientras el Sub-Chaac se encuentre operando, la computadora de navegación genera una bitácora de las variables que se involucran en el proceso: profundidad, cabeceo, alabeo y

Fig. 5 Localización del sistema de control y bombeo instalado dentro del casco del Sub-Chaac.

Fig. 6 Computadora de navegación y equipo de bombeo instalado dentro del casco del Sub-Chaac.

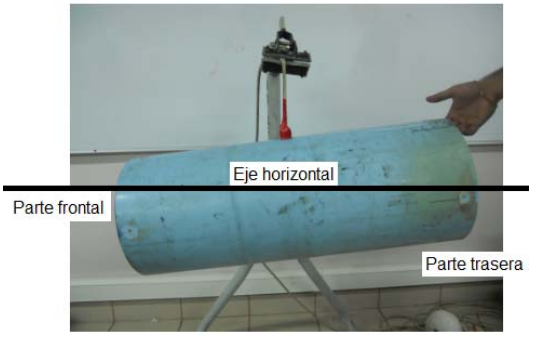


Fig. 7 Representación del movimiento de cabeceo positivo.

estado de activación de las motobombas. El intervalo de tiempo entre muestras se propuso en 0.1s. De este registro se analizarán los datos para posteriormente evaluar el tiempo de respuesta y determinar las características del sistema de control, lo que se estudiara en trabajos futuros.

III. RESULTADOS

Una vez instalado el sistema de control dentro del vehículo se estableció una profundidad de 0 cm para probar primero las condiciones de nivelación horizontal, se estableció una tolerancia de >10° y <-10° para la inclinación en el eje *X* de >12° y <-12° para la inclinación en el eje *Y*, y un nivel de histéresis para cuando la inclinación sea menor a 5° en ambos ejes. Se introduce un movimiento de cabeceo hacia abajo, como se puede observar en la Fig. 7, lo que se registra como un ángulo positivo de

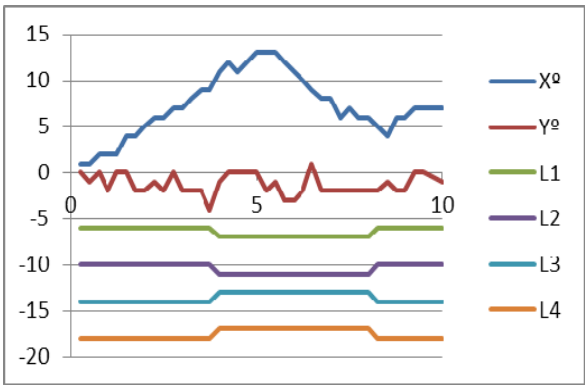


Fig. 8 Gráfica del ángulo de inclinación durante cabeceo (Xº) positivo y ciclos de activación/desactivación de bombas peristálticas, para lograr la nivelación del Sub-Chaac.

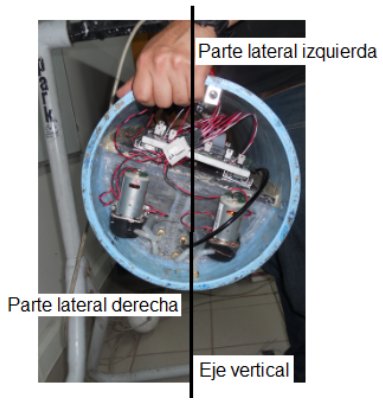


Fig. 9 Representación del movimiento de alabeo a la izquierda.

inclinación. El sistema responde activando las bombas para compensar la inclinación y nivelar el vehículo, en la gráfica de la Fig. 8 se puede observar cuando cambia la inclinación por encima de 10º en el eje *X*, se activan las bombas (*L1, L2*) para vaciar y (*L3, L4*) para llenar las cámaras de lastre y la desactivación cuando el ángulo es menor a 5º, logrando nivelar el vehículo dentro de la tolerancia establecida.

A continuación se introduce un movimiento de alabeo a la izquierda, como se muestra en la Fig. 9, lo que se registra como un ángulo negativo de inclinación. A lo que el sistema responde activando las bombas para compensar la inclinación y nivelar el vehículo, en la gráfica de la Figura 10 se puede observar cuando cambia la inclinación por encima de 13º en el eje *Y*, se activan las bombas (*L1, L3*) para vaciar y (*L2, L4*) para llenar las cámaras de lastre y la desactivación cuando el ángulo es menor a 5º, logrando nivelar el vehículo dentro de la tolerancia establecida.

Una vez comprobado el correcto funcionamiento del algoritmo de nivelación, se procede a verificar el control de profundidad, para lo cual se fijara una profundidad de 25 cm, al encontrarse el vehículo en la superficie, las bombas (*L1* a *L4*) se activaran iniciando el llenado las cuatro cámaras de lastre. En la gráfica de la Fig. 11 se puede apreciar la desactivación de las bombas cuando la profundidad alcanza 21 cm, se mantienen desactivadas mientras le Sub-Chaac sigue descendiendo hasta

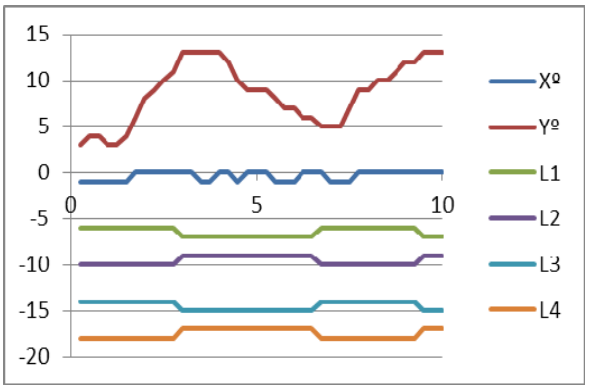


Fig. 10 Gráfica del ángulo de inclinación durante alabeo (Yº) positivo y ciclos de activación/desactivación de bombas peristálticas, para lograr la nivelación del Sub-Chaac.

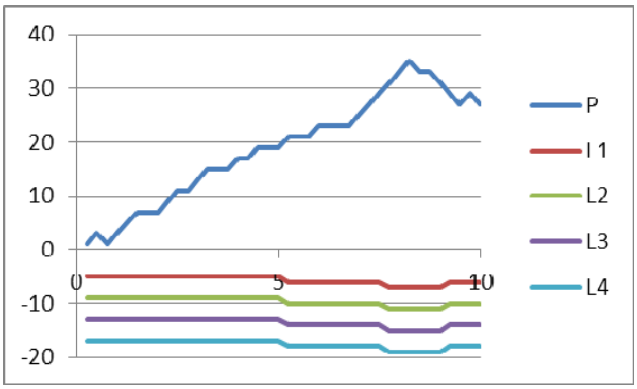


Fig. 11 Gráfica profundidad y ciclos de activación/desactivación de bombas peristálticas, para lograr la profundidad consignada del Sub-Chaac.

Fig. 12 Gráfica del ángulo de inclinación durante cabeceo (Xº) positivo y ciclos de activación/desactivación de bombas peristálticas, para lograr la nivelación del Sub-Chaac.

alcanzar 29 cm, entonces las bombas se activan nuevamente para extraer agua de las cámaras hasta lograr la profundidad de 25 cm.

Mientras el Sub-Chaac se encuentra bajo prueba, la computadora de navegación genera una bitácora de las variables que se involucran en el proceso: profundidad, cabeceo, alabeo y estado de activación de las motobombas,

En la Fig. 12 se muestra al Sub-Chaac durante de la prueba de nivelación en flotación positiva.

IV. Conclusión

El sistema de control funciono dentro de los parámetros esperados, nivelando satisfactoriamente el vehículo a

compensando las perturbaciones de cabeceo y alabeo producidos por el medio en el que se encuentra inmerso, así también se estabilizó la profundidad consignada. Demostrando con esto que la solución planteada para controlar la nivelación y profundidad es adecuada para un vehículo subacuático no tripulado, donde no es requerido posicionamiento riguroso del vehículo. Así mismo la arquitectura de en red I^2C, permite la escalabilidad del sistema para incorporar sensores adicionales, según sean requeridos por la aplicación.

RECONOCIMIENTOS

Agradecemos los profesores/investigadores del departamento de Ingeniería Eléctrica y Electrónica que participan en el desarrollo del proyecto del vehículo subacuático no tripulado, así como a los estudiantes, becarios y residentes que se encuentran involucrados en el desarrollo de los diferentes módulos que componen el Sub-Chaac.

Este proyecto fue financiado parcialmente por la DGEST 3416.10-P.

REFERENCIAS

[1] Gomáriz, S., Prat, J., Ruiz, A.G., Sole, J., Gayá, P., Del Rio, J. Development of a low-cost autonomous oceanographic observation vehicle (2009) OCEANS '09 IEEE Bremen: Balancing Technology with Future Needs.

[2] Conte, G., Gambella, L., Scaradozzi, D., Zanoli, S., Caiti, A., Calabrò, V., Alcocer, A., Alves, J., Cardeira, B., Cunha, R., Curado, F., Oliveira, P., Oliveira, A., Pascoal, A., Rufino, M., Sebastïo, L., Silvestre, C. Underwater vehicle technology in the European research project VENUS 1 (2009) Underwater Technology, 28 (4), pp. 175-185.

[3] Kulkarni, I.S., Pompili, D. Coordination of autonomous underwater vehicles for acoustic image acquisition (2008) Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM, pp. 27-34.

[4] Murphy, A.J., Landamore, M.J., Birmingham, R.W. The role of autonomous underwater vehicles for marine search and rescue operations (2008) Underwater Technology, 27 (4), pp. 195-205.

[5] Masmitja, I., Masmitja, G., González, J., Shariat-Panahi, S., Gomariz, S. Development of a control system for an autonomous underwater vehicle (2010) 2010 IEEE/OES Autonomous Underwater Vehicles, AUV 2010.

[6] Nègre, A., Zhang, H., Marceau, O., De Saporta, B., Laneuville, D., Dufour, F. Stochastic control for underwater optimal trajectories (2012) IEEE Aerospace Conference Proceedings.

[7] Khan, I., Bhatti, A.I., Khan, Q., Ahmad, Q. Sliding mode control of lateral dynamics of an AUV (2012) Proceedings of 2012 9th International Bhurban Conference on Applied Sciences and Technology, IBCAST 2012, pp. 27-31.

[8] Meldrum, D.T., Haddrell, T. GPS in autonomous underwater vehicles (1994) IEE Conference Publication, (394), pp. 11-17.

[9] Sun, Y., Li, W., Qin, Z., Chen, H., Li, J. Application of modified self-adaptive Kalman filter in integrated navigation system of autonomous underwater vehicle (2011) Applied Mechanics and Materials, 79, pp. 298-303.

[10] K. Ogata, Sistemas de Control en Tiempo Discreto, 2nd ed. México, México: Prentice Hall, 2009.

[11] D. Simon, "Kalman Filtering," Embedded Systems Programming, vol. 14, no. 11, pp. 72-79, June 2001.

[12] M. Darby and M. Nikolau, "A Parametric Programming Approach to Moving-Horizon State Estimation," Automatica, vol. 43, no. 5, pp. 885-891, May 2007.

[13] A. Sotelo Orozco, L.N. Coria de los Ríos, M.J. Garcia Ortega, C.E. Vazquez Lopez, and S. A. Puga Guzman, "Estimación de Posición Angular con el Filtro de Kalman Utilizando Sensores de Bajo Costo.," in VII Congreso Internacional en Innovación y Desarrollo Tecnológico, Cuenavaca, Morelos, 2009.

Arturo Sotelo Orozco nació en el Distrito Federal, México en 1966. Ingeniero en Comunicaciones y Electrónica egresado de la Escuela Superior de Ingeniería Mecánica y Eléctrica del Instituto Politécnico Nacional en 1990. Obtuvo el grado de Maestría en Sistemas Digitales por el Centro de Investigación en Desarrollo de Tecnología Digital del Instituto Politécnico Nacional en 1997 y Candidato a Doctor en la Universidad Politécnica de Valencia, España.

Él ha trabajado en el desarrollo de sistemas de telemetría y automatización de procesos de manufactura en la industria privada, en paralelo ha sido profesor de tiempo parcial: Desde 2010 es profesor de tiempo completo en el Instituto Tecnológico de Tijuana, en el departamento de ingeniería eléctrica y electrónica, en la carrera de electrónica. Sus principales áreas de interés son en el área biomédica con aplicaciones en el área de epilepsia experimental, y en los vehículos de guía autónoma.

El M.C. cuenta con la distinción de perfil deseable PROMEP.

Manuel de Jesús García Ortega nació en Rio grande, Zacatecas en 1964, Ingeniero en Comunicaciones y Electrónica egresado de la Escuela Superior de Ingeniería Mecánica y Eléctrica del Instituto Politécnico Nacional en 1990. Obtuvo el grado de Maestría en Sistemas Digitales por el Centro de Investigación en Desarrollo de Tecnología Digital del Instituto Politécnico Nacional en 2003.

Él ha trabajado en la industria de telecomunicaciones en el área de redes mundiales y metropolitanas. Desde 2008 es profesor-investigador en el Departamento de Ingeniería Eléctrica y Electrónica del Instituto Tecnológico de Tijuana. Sus áreas de interés se enfocan en los sistemas de comunicaciones con aplicaciones a los vehículos de guía autónoma.

El M.C. cuenta con la distinción de perfil deseable PROMEP.

Luis Néstor Coria de los Ríos, nació en el Salto, Durango en 1978, obtiene el grado de Ingeniero en Electrónica en el Instituto Tecnológico de Durango en el 2000, obtiene el grado de M.C. con Especialidad en Sistemas Digitales en el CITEDI-IPN en 2005. Obtuvo el grado de Doctor en Comunicaciones y Electrónica en Instituto Politécnico Nacional en el 2010.

Él ha tenido experiencia profesional en la industria de servicios relacionados con los sistemas computacionales. Actualmente es Profesor Investigador en el Departamento de Ingeniería Eléctrica y Electrónica del Instituto Tecnológico de Tijuana dese 2010. Su principal área de interés es el control y los sistemas caóticos, en estas áreas cuenta con diversas publicaciones en revistas indexadas.

El Dr. obtuvo el grado de ingeniero con medalla al mérito académico y el grado de doctor con mención honorifica, también cuenta con la distinción de perfil deseable PROMEP.

Carlos Edgar Vázquez López nació en el D.F, México, recibió el título de Ingeniero en Comunicaciones y Electrónica de la ESIME-IPN en 1991. Obtiene el grado de Maestro en Ciencias con especialidad en sistemas Digitales del CITEDI-IPN en 1997.

Desde hace 20 años se desempeña como docente y ha ocupado diferentes cargos administrativos en el Instituto Tecnológico de Tijuana. Actualmente es titular de las materias: Sistemas Digitales, Optoelectrónica y Física de Semiconductores.

Jorge Alberto Ortega Camacho nació en la ciudad de Tijuana en 1989 actualmente es alumno de 9^0 semestre de la carrera de Ingeniería Electrónica del Instituto Tecnológico de Tijuana, en fase de residencia profesional.

RIEE&C

Revista de Ingeniería Eléctrica, Electrónica y Computación

AGRADECE AL GRUPO DE REVISORES QUE COLABORARON EN LA PRESENTE EDICIÓN

Adolfo Espinoza Ruiz
Adolfo Soto Cota
Adrián Macías Estrada
Armando García Berumen
Erica Cecilia Ruiz Ibarra
Guillermo Morales Luna
Javier Pérez Ramirez
Jorge Ernesto Cota Félix
José Antonio Beristáin Jiménez
José Manuel Campoy Salguero
Juan Carlos Murrieta Lee
Juan José Padilla Ybarra
Manuel Domitsu Kono
Ramón García Hernández
Raymundo Márquez Borbón

ISSN 1870 - 9532

RIEE&C

Revista de Ingeniería Eléctrica, Electrónica y Computación

INVITACIÓN A FORMAR PARTE DEL COMITÉ REVISOR

El Instituto Tecnológico de Sonora a través del Departamento de Ingeniería Eléctrica y Electrónica pone en marcha la Revista RIEE&C (ISSN: 1870-9532) con el fin de coadyuvar a la difusión científica y cultural en el país en las áreas de Electrónica, Eléctrica y Sistemas Computacionales. RIEE&C es un espacio donde se pueden publicar resultados científicos y de desarrollo tecnológico tanto para investigadores como para estudiantes de nivel superior.

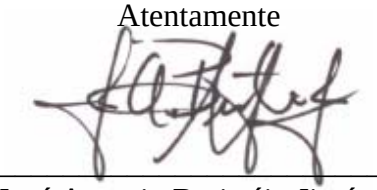
En esta revista se publican artículos de investigación con resultados originales y deseamos que dichos artículos estén sujetos a un estricto arbitraje realizado por investigadores líderes en su especialidad, adscritos a instituciones nacionales y extranjeras. Lo anterior con el objetivo de ser incluida en el índice de revistas reconocidas por el Consejo Nacional de Ciencia y Tecnología, CONACYT.

Por lo anterior le hacemos una cordial invitación para que forme parte del grupo de revisores de la revista RIEE&C. En busca de indexarse, RIEE&C crea un expediente de cada uno de sus revisores, por lo que en caso de aceptar la invitación, se le solicita envíe su currículum vitae así como una copia de la cédula profesional o copia del título del máximo grado académico obtenido. Es importante que su currículum contenga datos como:

- Número de cédula profesional.
- Grado académico.
- Áreas de interés.
- Centro de trabajo.

Estos datos deberán ser enviados a la dirección de correo electrónico rieeandc@itson.edu.mx. También le invitamos a conocer nuestra página electrónica en la dirección http://www.itson.mx/rieeyc.

Atentamente



José Antonio Beristáin Jiménez
Editor en Jefe de la RIEE&C
Instituto Tecnológico de Sonora
Departamento de Ing. Eléctrica y Electrónica

ISSN 1870 - 9532

RIEE&C

Revista de Ingeniería Eléctrica, Electrónica y Computación

INFORMACIÓN PARA LOS AUTORES

RIEE&C, Revista de Ingeniería Eléctrica, Electrónica y Computación (ISSN: 1870-9532) se publica semestralmente. Se aceptan artículos originales en aplicaciones de la ciencia, desarrollo de nueva tecnología o soluciones eficientes de ingeniería, siempre y cuando no hayan sido publicados o estén bajo consideración para publicarse en alguna otra revista.

Todos los documentos deberán enviarse al editor en jefe de RIEE&C, José Antonio Beristáin Jiménez, por e-mail a: rieeandc@itson.edu.mx.

Los artículos se revisarán por especialistas en el área y dictaminarán si el artículo es apropiado para su publicación. No se deberán enviar manuscritos directamente a los editores asociados.

Los manuscritos enviados por los autores deberán seguir el formato que muestra la guía para el autor, la cual se encuentra en la página de internet http://www.itson.mx/rieeyc.


ITSON
Educar para
Trascender