Proceedings of the 2000 Winter Simulation Conference
J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

ON THE USE OF CONTROL VARIATES IN THE SIMULATION OF MEDIUM ACCESS CONTROL PROTOCOLS

Andrés Suárez-González
Cándido López-García
José C. López-Ardao
Manuel Fernández-Veiga

Escola Técnica Superior de Enxeñería de Telecomunicación
Universidade de Vigo
SPAIN

ABSTRACT

Simulation is an essential tool for performance evaluation of communication networks. We are interested in the waiting time W of packets. The Control Variates method takes profit of the knowledge about another stochastic process strongly correlated with W to reduce the uncertainty in the estimation of its mean. We analyze the usefulness of the cycle time as a control stochastic process for Medium Access Control (MAC) protocols with polling service discipline, showing its potential and drawbacks. We propose a control variate that overcomes the disadvantages of cycle time and show its behavior in a case study. This new control variate will also be useful in the case of other MAC protocols.

1 INTRODUCTION

With the widespread deployment of local and metropolitan area networks that can support very high data transmission rates and a wide range of services, the task of performance evaluation is becoming increasingly complex.

Local and metropolitan area networks consist of a number of devices that are connected by a common transmission medium. They are characterized by its MAC protocol, which is the algorithm used by every node to determine when it may transmit. Performance evaluation is based on studying a variable that represents the system behaviour. We are interested in the packet waiting time (in steady state) in the case of infinite buffers. This quantity is random in nature and can be represented by a stochastic process $W = \{W_i ; i = 1, \ldots, \infty\}$ for each traffic source, and we assume it is a covariance-stationary process. We are interested in studying its mean, $\mathrm{E}(W) = \mu$, a basic parameter of the performance of a MAC protocol.

Unfortunately, the complexity of MAC protocols makes impossible the analytical resolution of an accurate enough mathematical model, except for simplified cases. Thus, simulation is an essential tool for performance evaluation of MAC protocols. The simulation technique commonly used for analyzing a communication network is the *discrete-event simulation*.

We estimate the mean waiting time μ from the observations of a single simulation run. So, simulation is a computer-based statistical sampling experiment. Through simulation we may measure a sequence of size n of W and estimate its mean by the average

$$\overline{W}[n] \equiv \frac{1}{n} \cdot \sum_{i=1}^{n} W_i \qquad (1)$$

Since W is a correlated stochastic process, there are some techniques to compute confidence intervals for (1)—see for example Bratley et al. (1987)—. One of them is the batch means method. This method estimates the variance of (1) through m batches of size l $(n = m \cdot l)$

$$\overline{W_i}[l] \equiv \frac{1}{l} \cdot \sum_{j=(i-1)\cdot l+1}^{i \cdot l} W_j$$

under the asumption that $\left\{\overline{W_i}[l] ; i = 1, \ldots, \infty\right\}$ behaves asymptotically as $l \to \infty$ like a gaussian renewal process. We use a simple algorithm due to Law and Carson (1979) as described in Law and Kelton (1982) for the experiments commented in this paper. This way, after gathering a group of $m = 400$ batches of size l of W, $\left\{\overline{W_i}[l] ; i = 1, \ldots, 400\right\}$, we calculate the one lag coefficient of autocorrelation; when it is below certain level we may compute the confidence interval over the 40 batches of size $10 \cdot l$, $\left\{\overline{W_i}[10 \cdot l] ; i = 1, \ldots, 40\right\}$.

An important disadvantage of a simulation approach (as opposed to an analytic one) toward network performance studies is the amount of time needed to simulate a model. The simulation run length must ensure that the estimate of the variable under study has a given quality. For that reason, the computational cost increases as the complexity of the model grows. Although important research efforts have been spent to develop efficient simulation techniques, they have not been commonly used in simulation of computer networks. In this paper, we consider the use of a variance reduction technique: Control Variates (CV). Variance reduction techniques exploit the statistical properties of the simulation model to reduce the uncertainty in the output.

A control variate is a random magnitude that is correlated with the performance measure under study and whose expected value is known. The control variate can be an internal control if it is generated as a part of the simulation, or an external control if it is generated with an auxiliary simulation.

The rest of the paper is organized as follows.

In subsection 1.1 we sum up the CV method and related statistical theory.

In section 2 we explore one internal stochastic process proposed as control in López-García et al. (1995) for polling systems. We study its potential and its drawbacks.

In section 3 we design a new internal stochastic process starting from the analysis of the behavior of the cycle time, not limited to polling systems. In subsection 3.1 we propose a synchronization method with W batches to use both the control variates and the batch means method simultaneously.

In section 4 we present the results obtained by the new control stochastic process when we simulate a system with known mean waiting time.

In section 5 we present the conclusions and expose future work to be done.

1.1 Control Variates

The purpose of variance reduction techniques is to obtain an estimator of μ with less variance than (1)—see for example Bratley et al. (1987) or Frost, LaRue and Shanmugan (1988)—.

The Control Variates method takes profit of the knowledge about a stochastic process C (whith known mean $\mathrm{E}(C) = \nu$) strongly correlated with W to estimate its mean, μ, defining the controled stochastic process $Y = \{Y_i; i = 1, \dots, \infty\}$ as

$$Y_i \equiv W_i - \beta \cdot (C_i - \nu)$$

so we hope that its average $\overline{Y}[n]$ will have less variance than (1). The controled stochastic process Y with smallest variance is obtained from

$$\beta^* = \frac{\mathrm{Cov}(W, C)}{\mathrm{Var}(C)}$$

As this value is usually unknown, it is estimated through $\widehat{\beta}[n]$ from the same samples of W and C used to compute $\overline{W}[n]$ and $\overline{C}[n]$ in

$$\overline{Y}[n] = \overline{W}[n] - \widehat{\beta}[n] \cdot \left(\overline{C}[n] - \nu\right) \tag{2}$$

As a consequence, although W and C would be renewal processes, $\overline{Y}[n]$ will be generally a biased estimator of μ.

Nevertheless, if $(W, C) \equiv \{(W_i, C_i); i = 1, \dots, \infty\}$ are i.i.d. and distributed as a multivariate normal, Lavenberg, Moeller and Welch (1981) show that (2) is an unbiased estimator of μ. They also develop an unbiased estimator $\widehat{\sigma^2}_{\overline{Y}[n]}$ of $\mathrm{Var}\left(\overline{Y}[n]\right)$, and show that

$$\frac{\overline{Y}[n] - \mu}{\widehat{\sigma^2}_{\overline{Y}[n]}}$$

has a Student's t distribution with $n-2$ degrees of freedom. Lavenberg et al. (1982) also show that the loss in potential variance reduction when the optimum coefficient β^* is estimated by $\widehat{\beta}[n]$ is $\frac{n-2}{n-3}$.

If W and C are both correlated stochastic processes under a joint functional central limit theorem asumption, Loh (1997) shows, applying the previous result to m batches of size l ($n = l \cdot m$) of W and C, that

$$\frac{\overline{Y}[m, l] - \mu}{\widehat{\sigma^2}_{\overline{Y}[m,l]}}$$

behaves asymptotically as $l \to \infty$ like a Student's t random variable with $m-2$ degrees of freedom. Hence, it is possible to use both control variates and batch means methods simultaneously.

2 CYCLE TIME CONTROL PROCESS

In a network with shared transmission medium and polling service discipline, the cycle time stochastic process C is made up of the successive times needed for polling once all nodes of the network. The mean value of this stochastic process is

$$\mathrm{E}(C) = \nu = \frac{u_{\mathrm{a}}}{1-\rho} \tag{3}$$

where ρ is the network load factor, and u_{a} is the time needed for polling all nodes when none of them transmits. In the sequel u_{a} is considered constant.

Intuitively, C seems a good choice as a control variable for W, as first proposed in López-García et al. (1995). Both C and W take large values during high-load periods, and small values during low-load periods.

The only problem is how to synchronize the batches of W with those of C, since the ratio of their respective amount of values (sample size) can be other than one. There could be several cycles per packet (of one node) in average, or conversely, several packets (of one node) per cycle.

We have tried to synchronize the batches of the stochastic process C with the batches of W allowing the batches of C to be of different sizes. Batch j of C has $N_j[l]$ samples, that is, the number of polls (of the node where we are measuring W) during the measure of $\overline{W_j}[l]$.

If the system has Poisson arrivals, service times are a renewal stochastic process, and the service discipline is exhaustive, then the mean waiting time—in Takagi (1986)—is

$$\mathrm{E}(W) = \mu = \frac{u_\mathrm{a} \cdot (1 - \frac{\rho}{\mathsf{N}}) + \mathsf{N} \cdot \lambda \cdot \mathrm{E}\left(S^2\right)}{2 \cdot (1-\rho)} \quad (4)$$

where λ is the arrival rate to a node, N is the number of nodes, S is the service time random variable and $\rho = \mathsf{N} \cdot \lambda \cdot \mathrm{E}(S)$ is the load factor. We are able to estimate the coverage obtained by the averages computed from the simulations, using only the batch means method or using also control variates.

In table 1 we present the results of simulating 1000 times a polling network of $\mathsf{N} = 40$ nodes with Poisson arrivals, constant service times and load factor $\rho = 0.95$ for three values of 95% confidence intervals width, relative to the computed average $\gamma_\mathrm{r} \equiv \Delta\overline{W}[n]_{95\%} / \overline{W}[n]$. We have selected a constant value of 1 for the coefficient β in this experiment, so there is no bias due to estimation of β.

Table 1: Cycle Time Behavior for Exhaustive Service

γ_r	5%	2.5%	1.25%
% Coverage non-CV	92.4	93.1	95.7
Sample size	26893.28	112753.06	466463.12
% Coverage with CV	70.6	82.8	91.1
% Sample size reduction	53.88	59.37	59.14
$\widehat{\theta}$	0.19	0.1	0.049

We may observe how the different sizes of C batches induce a high bias on the controled average estimator, which is higher (less coverage of the CV method) the higher the variation coefficient

$$\theta \equiv \frac{\sigma_{\overline{N}[m,l]}}{\mathrm{E}\left(\overline{N}[m,l]\right)}$$

is. Also, from the high reduction in sample size whichever the coverage, we may deduce that a high correlation exists between W and C.

We find that the bias of the CV method in this case is directly due to

$$\mathrm{E}\left(\frac{\overline{C_j}[l]}{\overline{N_j}[l]}\right) \neq \frac{\mathrm{E}\left(\overline{C_j}[l]\right)}{\mathrm{E}\left(\overline{N_j}[l]\right)} = \nu$$

and we assume that the high correlation between W and C is due to the factor

$$\frac{1}{1-\rho} \quad (5)$$

driving the behavior of both (3) and (4) near overload, that is

$$\lim_{\rho\to 1} \mu \cdot (1-\rho) = \lim_{\lambda \to \frac{1}{\mathsf{N}\cdot\mathrm{E}(S)}} \frac{\mathsf{N} \cdot \lambda \cdot \mathrm{E}\left(S^2\right)}{2} = \frac{\mathrm{E}\left(S^2\right)}{2 \cdot \mathrm{E}(S)} \quad (6)$$

and

$$\lim_{\rho\to 1} \nu \cdot (1-\rho) = u_\mathrm{a}$$

are both finite positive constants.

From these results we conclude that a better control stochastic process is needed, a process easier to synchronize with W but that retains the high correlation observed with W.

3 PROPOSED CONTROL PROCESS

We propose to extend the factor (5) to constitute a new control stochastic process, replacing ρ by its estimation

$$\widehat{\rho_i} \equiv \frac{B_i}{T_i}$$

where T_i is the duration of a given interval and B_i is the time along T_i where the system is busy (transmiting packets) of T. We then obtain that

$$\frac{1}{1-\widehat{\rho_i}} = \frac{T_i}{T_i - B_i} = \frac{T_i}{L_i}$$

where L_i is the time along T_i where the system is idle. As packets arrive to a node at mean rate λ, the expected value between them is $1/\lambda$, each packet share (from one node) in simulation time. If we select L_i as each packet share (from one node) in idle time, that is the constant value

$$\mathsf{L} \equiv \frac{1-\rho}{\lambda}$$

then

$$D_i \equiv \frac{T_i}{\mathsf{L}}$$

defines the control stochastic process D with known mean

$$\mathrm{E}(D) = \frac{1}{1-\rho}$$

Moreover, we propose to use T as a control stochastic process, which is a scaled version of D, and its mean value is

$$\mathrm{E}(T) = \frac{1}{\lambda}$$

An additional advantage of this new control stochastic process is its wider usability, because it only needs a MAC protocol sharing a common medium, which does not need to be a polling service dicipline.

3.1 Synchronization With W

The ratio of the amount of values (sample size) of T and W, tends to one as the simulated time increases. They are not perfectly synchronized in time, though. Due to the stochastic nature of the queueing systems, we will have different number of values of T and W, and hence a different number of batches $\overline{T_i}[l]$ and $\overline{W_i}[l]$. Nevertheless, we take profit of the batch means method to help in the synchronization, and for this purpose we design a strategy of five points explained below.

First, we measure the stochastic process T through the auxiliar stochastic process $K = \{K_i; i = 1, \ldots, \infty\}$ where K_i is the instant (simulation clock) when the system has been idle exactly $i \cdot \mathsf{L}$ units of time. Batches of K of size l are represented by the first value of the batch, that is, $K_i[l] \equiv K_{(i-1)\cdot l+1}$. We easily obtain the batches of T from those of K by differentiation:

$$\overline{T_i}[l] = \frac{K_{i+1}[l] - K_i[l]}{l}$$

Moreover, doubling the size of batches of T is simply done by deleting one of every two stored values of K.

Second, to be able to synchronize in a suitable way both stochastic processes, we mark each batch $\overline{W_j}[l]$ with the arrival time to the queue of its first packet, $A_j[l] \equiv A_{(j-1)\cdot l+1}$, where $A = \{A_i; i = 1, \ldots, \infty\}$ is the arrival process. This way, we may delete some initial values of W to limit the impact of the transient period of the simulation, and we still are able to synchronize both stochastic processes in an easy way. Moreover, we do not have to begin to construct pairs at the begining or at the end of both sequences, as we will see in the fourth point.

Third, and also for the purpose of better synchronization, we store K with a batch size smaller ($l/2$) than that of W (l), that is, we store more values of K ($2m = 800$) than batches of W ($m = 400$). Nevertheless, the maximum number of batches of W drives the growing of both batch sizes. We also allow some extra amount of stored values of K (400) to deal with the non-perfect synchronization of both processes, though we will discard the initial values of K if the maximum amount of storage for K ($2m + 800 = 1200$) is reached. In this way we do not have to do any explicit deletion of values of K due to the transient period; values of K are discarded due to the storage limit. The synchronization task selects the usable values of K.

Fourth, to obtain a better synchronization in average, we propose to match both sets of batches in pairs beginning from the middle batch of W, $\overline{W_{m/2}}[l]$, and pairing it with the closest value to $A_{m/2}[l]$, $K_{i'}[l]$, and continuing toward both sides from there, as we see in figure 1 for $m = 400$ (indexes of K and T batches are those selected for pairing).

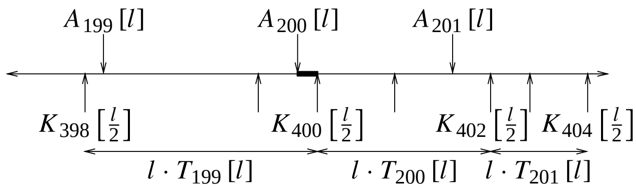

Figure 1: Pairing Process

And fifth, given that, in general, we will not obtain the same number of batches of both stochastic processes T and W, we estimate the coefficient $\widehat{\beta}\left[m', 10 \cdot l\right]$, the variance of the average $\widehat{\sigma^2}_{\overline{Y}[m',10\cdot l]}$ and the confidence interval, with a batch size higher than that used to compute the average $\overline{Y}[m, l]$ (that uses $\widehat{\beta}\left[m', 10 \cdot l\right]$ estimation); for example, we may have $m = 397$ paired batches of size l for computing the average, but then only $m' = 39$ paired batches of size $10 \cdot l$ for computing the confidence interval. We use more values of W to compute its average than to compute its confidence interval width, and so we hope that the method is slightly pessimistic (computed confidence intervals will be wider than should be).

4 CASE STUDY

We have designed and run a group of experiments more exhaustive than those for the cycle time. Here we focus on the results of tables 2 to 5 on the coverage and reduction in sample size (duration of simulation) obtained by using the proposed control stochastic process T.

We run groups of 1000 simulations of different configurations of a polling service network with Poisson arrivals, constant service demands (service rate of 100 Mbps and packet size of 4486 bytes), and latency of 6.25 μs between any adjacent pair of nodes. We select different number of nodes, $\mathsf{N} \in \{2, 10, 20, 40\}$, different values for the load factor of the network, $\rho \in \{0.8, 0.9, 0.95\}$ and specify different 95% confidence interval requirements, $\gamma_{\mathrm{r}} \in \{0.04, 0.02, 0.01\}$. Each node in the network is a traffic source, except in the configurations with $\mathsf{N} = 2$ where only one node is active. We also discard measured values of W

Table 2: T Behavior for Exhaustive Service $\mathsf{N} = 2$

ρ	0.8	0.9	0.95	γ_r
% coverage non-CV	*93.4 ±1.54*	*92.9 ±1.59*	*91.9 ±1.69*	±4%
	92.9 ±1.59	94.1	94.4	±2%
	93.7	94.8	94.0	±1%
% coverage CV	95.2	94.9	94.9	±4%
	94.3	95.3	95.3	±2%
	94.4	96.2	95.6	±1%
sample size non-CV	153441	557760	2153630	±4%
	621354	2287297	8957975	±2%
	2501727	9318613	35719291	±1%
% reduction CV	20.9	22.2	21.9	±4%
	26.7	29.3	31.1	±2%
	32.5	35.3	36.0	±1%

Table 3: T Behavior for Exhaustive Service $\mathsf{N} = 10$

ρ	0.8	0.9	0.95	γ_r
% coverage non-CV	95.4	94.6	*93.64 ±1.34*	±4%
	95.2	94.5	94.2	±2%
	94.8	95.4	95.3	±1%
% coverage CV	96.7	96.7	96.1	±4%
	96.4	96.9	96.3	±2%
	96.0	96.8	96.3	±1%
sample size non-CV	22599	66010	231696	±4%
	90198	264867	944286	±2%
	360651	1068804	3790894	±1%
% reduction CV	15.1	20.3	21.7	±4%
	20.8	28.5	31.5	±2%
	24.5	33.8	37.2	±1%

Table 4: T Behavior for Exhaustive Service $\mathsf{N} = 20$

ρ	0.8	0.9	0.95	γ_r
% coverage non-CV	94.5	94.1	94.2	±4%
	94.9	95.3	94.9	±2%
	94.9	95.4	95.3	±1%
% coverage CV	96.5	96.7	97.3	±4%
	95.9	97.0	96.8	±2%
	95.9	96.7	97.5	±1%
sample size non-CV	13511	34665	115924	±4%
	53245	139266	472054	±2%
	212528	554258	1871606	±1%
% reduction CV	12.1	18.3	21.6	±4%
	16.8	27.5	33.0	±2%
	20.2	33.1	39.3	±1%

Table 5: T Behavior for Exhaustive Service $\mathsf{N} = 40$

ρ	0.8	0.9	0.95	γ_r
% coverage non-CV	94.8	*93.74 ±0.952*	94.5	±4%
	95.0	95.1	94.4	±2%
	94.9	95.2	95.8	±1%
% coverage CV	96.5	96.8	97.3	±4%
	96.2	97.0	97.1	±2%
	95.9	96.5	96.3	±1%
sample size non-CV	8380	18528	57428	±4%
	33254	74082	227431	±2%
	132014	292699	906853	±1%
% reduction CV	7.22	16.5	22.0	±4%
	12.7	26.1	32.9	±2%
	15.7	31.5	40.7	±1%

for the first 20 seconds in each simulation, in order to limit the impact from the transient period.

In every active node we measure the coverage of the computed confidence interval (1 if it contains the mean value, 0 otherwise) and the sample size at the end of the simulation (when all of the nodes fulfil their requirements of confidence interval).

In one simulation the coverage sample may be not only 1 or 0 (except when $\mathsf{N} = 2$), but it may be a value in between, with increments of $1/\mathsf{N}$. From the 1000 samples of the coverage for a given configuration we are able to estimate a 95% confidence interval (being the widest of all of them ±1.69%); we show this 95% confidence interval (in italic font) in the following tables if it is below and does not contain the theoretic value of 95%.

The mean sample size for each configuration is estimated taking into account every active node. We have also computed their 95% confidence interval from the 1000 samples, being the widest interval relative to the average, $\Delta\overline{n}_{95\%}/\overline{n}$, equal to ±1.76%. We have also computed the 95% confidence interval for the reduction in sample size by the CV method, being the widest equal to ±1.21%.

We observe that only the non-CV method gets coverages perceptibly lower than the theoretic one, on tables 2, 3 and 5. Using T as control stochastic process it gets better coverages than those obtained by the batch means method alone in all the configurations studied, so the problem showed by the cycle time is clearly avoided. Moreover, CV method with T seems to be slightly pesimistic in general, as predicted.

For a given number of nodes N, that is, for each one of the tables 2 to 5, the reduction in sample size is higher the higher the sample size is. This is due to a higher load factor ρ or to a narrower confidence interval requirement γ_r. That is, we get higher reductions in longer simulations, when they are more useful.

5 CONCLUSIONS AND FUTURE WORK

As we have shown, the new control stochastic process works quite well for a polling exhaustive service discipline with

Poisson arrivals. We hope that it should work well for those systems where we expect a behavior of $\mathrm{E}(W)$ similar to that of expression (6), that is

$$\lim_{\rho \to 1} \mathrm{E}(W) \cdot (1 - \rho) = \mathtt{c} \in (0, \infty)$$

We are working at present in the evaluation of the control stochastic process T for other MAC protocols.

Usually MAC protocols cannot achieve a network utilization of 1, but only another limit $\rho < \rho_{\max} < 1$. In this case, we are working on extending the proposed control stochastic process using this value or a beforehand estimation of it, through the modification of (5) to $1/(\rho_{\max} - \rho)$.

ACKNOWLEDGMENTS

This work was partially supported by the project XUGA32204A97 of *Xunta de Galicia*.

REFERENCES

Bratley, P., B. L. Fox, and L. E. Schrage. 1987. *A guide to simulation*. 2d ed. New York: Springer-Verlag.

Frost, V. S., W. W. LaRue, and K. S. Shanmugan. 1988. Efficient techniques for the simulation of computer communication networks. *IEEE Journal on Selected Areas in Commun.* 6(1):146–157.

Lavenberg, S. S., and P. D. Welch. 1981. A Perspective on the use of control variables to increase the efficiency of mont carlo simulations. *Management Science* 27:322–335.

Lavenberg, S. S., T. L. Moeller, and P. D. Welch. 1982. Statistical results on control variables with application to queueing Nntwork simulation. *Operations Research* 30:182–202.

Law, M., and J. S. Carson. 1979. A sequential procedure for determining the length of a steady-state simulation. *Operations Research* 27: 1011–1025.

Law, A. M., and W. D. Kelton. 1982. *Simulation Modeling and Analysis*. New York: McGraw-Hill.

Loh, Wing Wah. 1997. *On the Method of Control Variates*. Ph.D. Thesis. Operations Research Department. Standford University.

López-García, C., J. Vinyes-Sanz, A. Suárez-González, C. López-Ardao. 1995. Speeding up the simulation of an FDDI network. *European Simulation Multiconference*.

Takagi, H. 1986. *Analysis of polling systems*. The MIT Press, Cambridge, MA.

AUTHOR BIOGRAPHIES

ANDRÉS SUÁREZ-GONZÁLEZ is an associate professor in the *Departamento de Tecnoloxías das Comunicacións* at *Universidade de Vigo* (DTC-UV). He received a Ph.D. degree in telecommunication engineering from *Universidade de Vigo* this year. He is a member of ACM. His email addresses is `<asuarez@ait.uvigo.es>` and their research group web address is `<www-gris.ait.uvigo.es>.`

CÁNDIDO LÓPEZ-GARCÍA is an associate professor in the DTC-UV. He received a Ph.D. degree in telecommunication engineering from *Universidad Politécnica de Madrid* in 1995. His email address is `<candido@ait.uvigo.es>`.

JOSÉ C. LÓPEZ-ARDAO is an associate professor in the DTC-UV. He received a Ph.D. degree in telecommunication engineering from *Universidade de Vigo* in 1999. He is a member of ACM and IEEE. His email address is `<jardao@ait.uvigo.es>`.

MANUEL FERNÁNDEZ-VEIGA is an assistant professor in the DTC-UV. He is a telecommunication engineer from *Universidade de Vigo*. He is a student member of IEEE. His email address is `<mveiga@ait.uvigo.es>`.