Jornada de Seguimiento de Proyectos, 2005
Programa Nacional de Tecnologías Informáticas

Web services, parallelism, evolution and complex systems: NadeWeb (nuevos algoritmos distribuidos y evolutivos en la web, new distributed evolutive algorithm on the web) TIC2003-09481-C04

Juan Julián Merelo Guervós *
Departamento ATC, Universidad de Granada

Ken Sharman †
ITI, Universidad Politécnica de Valencia

Isabel Román ‡
Departamento Economía Financiera y Contabilidad, Universidad de Granada

Abstract

In this paper we present a half-time evaluation of results obtained in the NadeWeb coordinated project. This projects intends to evaluate and enhance the performance of the DRM Distributed Resource Machine, a highly scalable distributed virtual machine, as well as use it for evolutionary computation tasks and apply it to real-world problem such as the evaluation of company risks or the analysis of word streams looking for some temporal structure.
Keywords: Distributed computation, web services, performance evaluation, evolutionary computation, human-computer interface, adaptive computation applications.

1 Project Objectives

The main objectives of the NadeWeb project has been deepening the understanding of the DRM, Distributed Resource Machine, a P2P virtual machine for general computation, and, specifically, to measure its performance in a distributed computation environment. DRM was released as a part of the DREAM project [Paechter. et al., 2000], but the IST-FET project that spawned it did not include extensive testing and evaluation. Further studies proved that its scaling behavior did not correspond to what theoretically should be expected [Arenas, 2005], and

*Email: jmerelo@geneura.ugr.es
†Email: ken@iti.upv.e
‡Email: iroman@ugr.es

specially Maribel Garcia Arenas' PhD thesis; thus, the objective of this project was to extensive evaluate its performance, debug and correct flaws, and finally, use it for evolutionary computation, as the original project intended. DREAM also included the release of the JEO library [Arenas et al., 2004, Arenas et al., 2002b, Arenas et al., 2002a, Arenas et al., 2001], which was adapted for distributed computation, but had several problems. First, it used DRM for statistics collection, but not for migrating objects under evolution; second, it used a classic evolutionary computation paradigm: a generational island model, which did not really take advantage of the ad-hoc and asynchronous nature of the DRM virtual machine. That is why a second objective of the project was to design a new evolutionary/adaptive algorithm that really leveraged DRM. Since this project is a coordinated one, it included several subgroups more interested in the application of these algorithms, as well as advancing the state of the art in those particular applications. These applications were human-computer interfaces, natural language processing and bankruptcy prediction. The objective was to improve results in those areas, as well as use them as testbeds for the algorithms developed.

2 Success level reached in the project

Since we are barely halfway through the project, there have not been many results; mainly, researchers that have been hired for the project are getting up-to-date on the project topics. One of them was hired by March 2004, and the other one by September 2004. There have been also some changes in the project staff; a new researcher has entered the project, another one has changed status and location, and finally, one has left the project. However, a few objectives have been reached.

2.1 DRM performance evaluation

Two problems have been solved using evolutionary computation on the DRM: the classical OneMax and knapsack problem. The main objective was to check what were its scaling properties, when new nodes were added. Results for up to 32 nodes are shown in figure 1.

However, at this point in time, it is not clear what are the causes of this lack of speedup. It could be due to an implementation defect, or it could be essential to the nature of the newscast algorithm: items on each node cache are sent time and again while they have not been substituted by another; this could be fixed by changing the cache size, but it could indicate a deeper problem.

At any rate, the newscast algorithm has been extensively studied, and as an alternative for further work, its reimplementation using web services or libraries such as JXTA [JXTA,] will be considered.

2.2 Genetic programming on the DREAM platform

With support from the Granada-ATC subproject, the ITI-UPV subproject has focused on genetic programming [Riolo and Worzel, 2003], updating the GP modules in JEO and applying them to their main research axis: brain-computer interfaces (BCI).

In fact, many of the structures obtained would have been competitive in the BCI competition that is held every year.

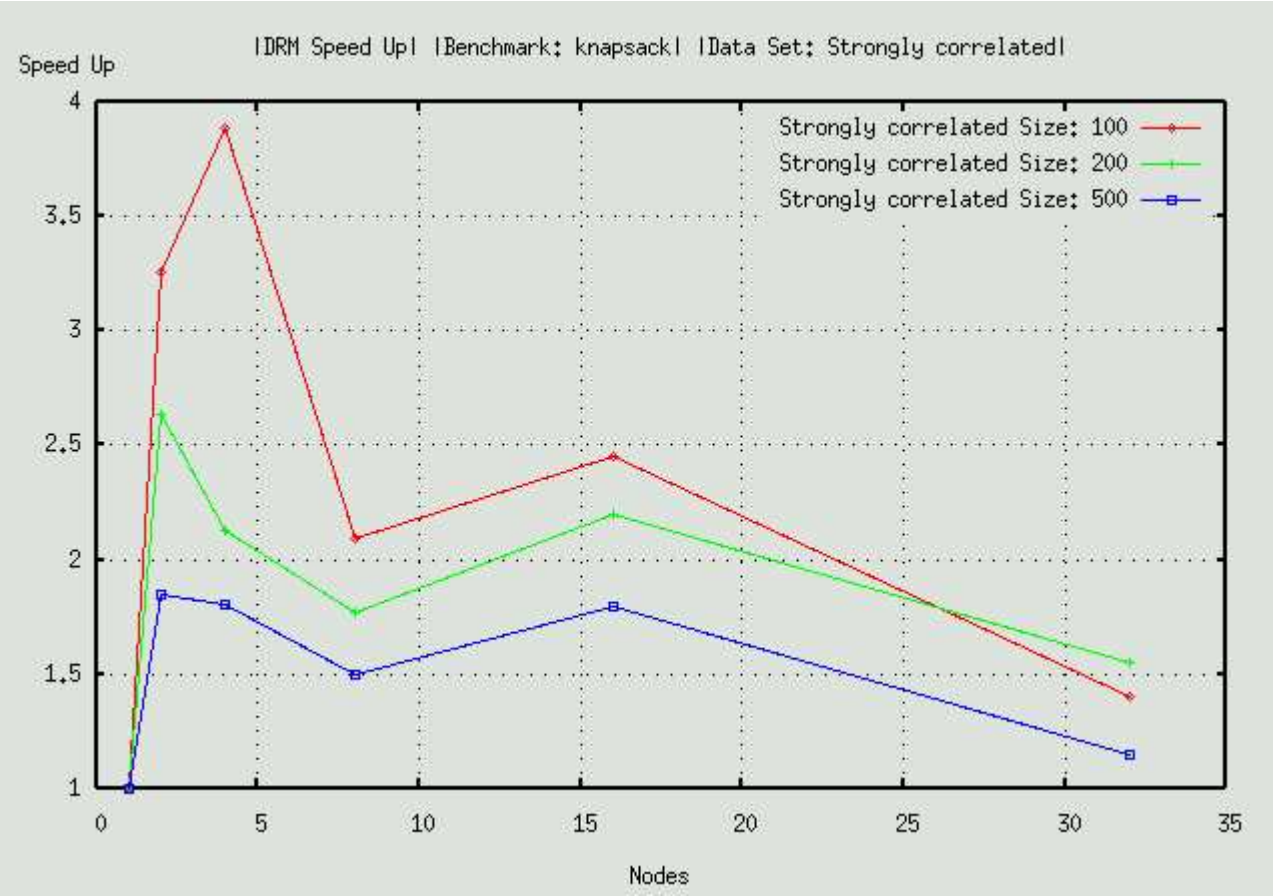

Figure 1: DRM speed-up for the knapsack problem solved using evolutionary computation; nodes are increased along the x axis, while the y axis shows speed up. Speedup for more than 16 machines is clearly lower than expected.

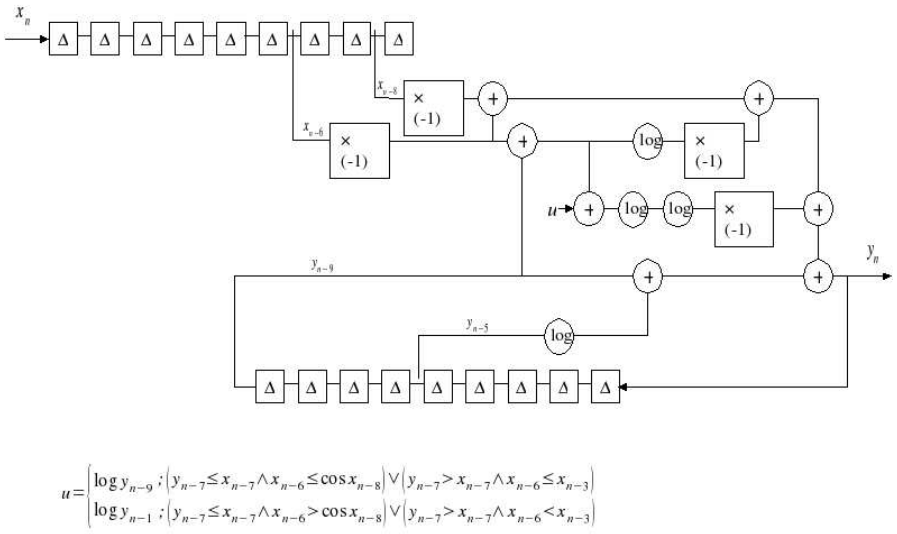

Figure 2: One of the best structures computed using genetic programming

2.3 Memetic algorithms in genetic programming

Genetic programming includes several constants, whose change via mutation takes a very long time; combining this global algorithm with a local algorithm such as simulated annealing can yield better results in less time. The ITI-UPV group is testing such combination on generated signals; it still takes some time, making it a good candidate for parallelization.

2.4 Novelty detection in text streams

In constantly changing text streams, such as weblogs, news sites or Internet forums, it is almost impossible to know what topics are *hot* in a particular moment. Some algorithms have been developed to follow text streams and *fire* when some topic is being mentioned more than usual [Kleinberg, 2002]; however, these algorithms, approached from the point of view of optimization, have not been measured extensively, and, besides, its efficiency, as well as predictive power, could probably be improved using global optimization algorithms such as evolutionary algorithms. The number of times a term appears in a text stream (in this case, a set of weblogs) is shown in figure 3

A model for these streams, which consists of different *states*; (the higher, the *hotter* is this issue at a particular point in time), has been created by comparing different measures of proximity, and finally fitting using an evolutionary algorithm; the combination of both techniques yields a fit that is better than the one found by Kleinberg previously. An example of fits with different distance measures is found in figure .

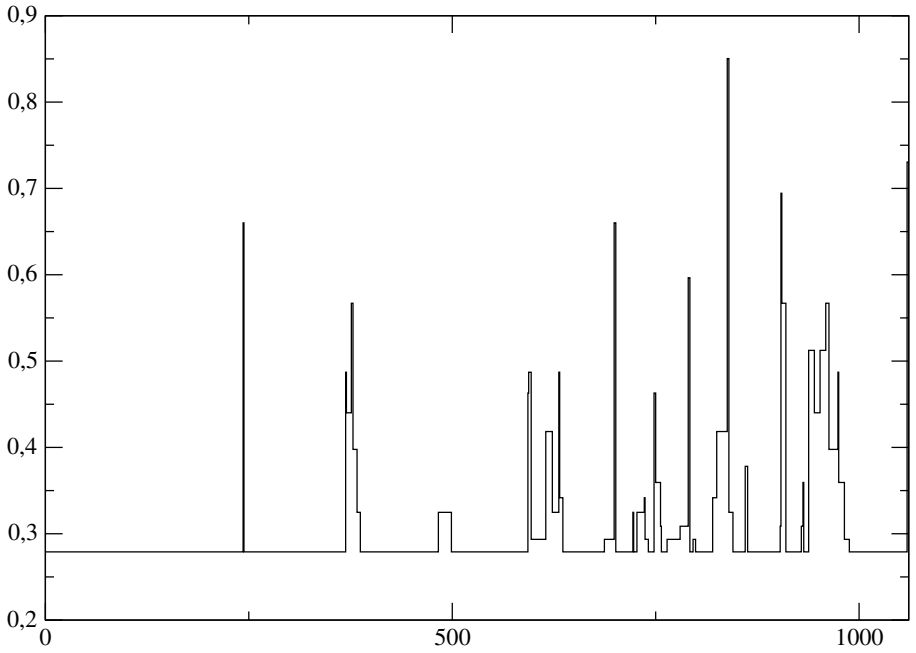

Figure 3: Number of times the word RSS appears in a text stream vs. time. This figure shows the bursty structure of this kind of streams.

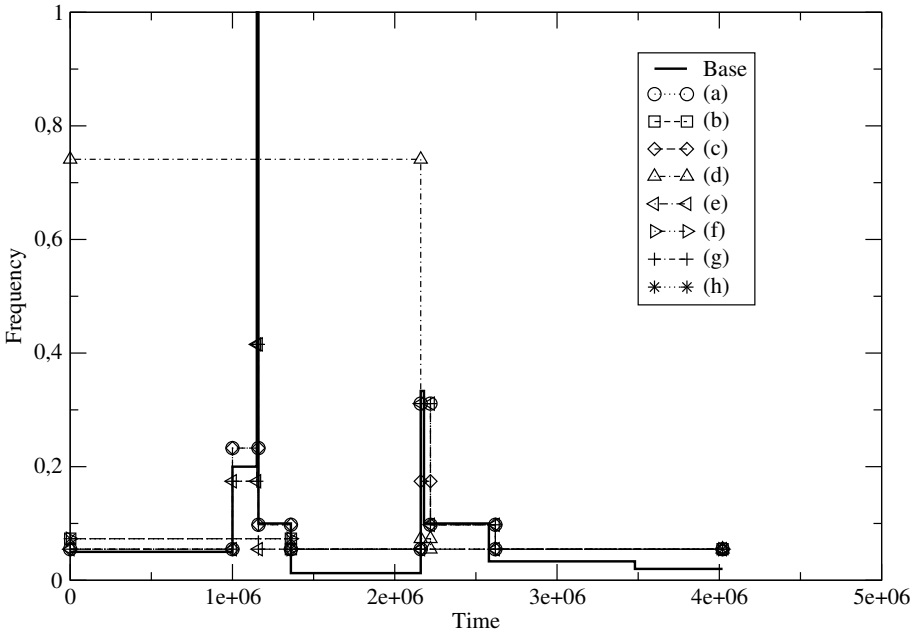

Figure 4: Fitting different measures of proximity to a text stream that uses the same representation as in figure 3. In this case, distance labelled as (g) yields the best result; this will be the one that is finally used.

2.5 Bankruptcy prediction

Some advances have been made in the selection of variables involved in the prediction of bankruptcy of industrial firms. The Granada-EFC group has made some headway in that area, and found out that indebtedness, funding costs and economic return on investment are the best predictors of firm health. For the time being, no adaptive algorithms are being used; in a future stage, these problems will be approached using the DREAM/JEO platform.

3 Results indicators

There are two persons that have been hired for the project, Juan Luis Jiménez Laredo and Eva Alfato Cid; Juan Luis is currently engaged in the PhD program at the University of Granada. Additionally, two persons, Juan José Samper, whose PhD is in an advanced state, and Juan Torrecillas, who has also signed up for the PhD program of the ATC dept. at the University of Granada, have profited from their interaction with members of the project.

A website has been created for diffusion of the project results, , and is maintained by Juan Luis at the Granada-ATC subproject.

Several papers and technical reports have been published by each subproject:

- ITI-UPV has published the following technical reports [Cid, 2005d, Cid, 2005c, Cid, 2005b, Cid, 2005a, Cid, 2004a, Cid, 2004b], mainly on the implementation of a signal classification system using DREAM and JEO; all of them are in Spanish. Besides, a paper has been accepted in a national congress: [Alfaro-Cid et al., 2005].

- The Granada (which includes a professor from Madrid) subproject has published the following papers in a national multidisciplinary congress: [Romero, 2005b, Arenas, 2005, Tricas, 2005, Romero, 2005a]. Two have been published in an international congress: [Gustavo Romero and Merelo, 2005, Castillo et al., 2005]. Several papers have been submitted to international journals, and are in the review phase.

- The EFC group has published a paper in an international congress [Gómez et al., 2005], and another one submitted to a national accounting congress [De la Torre et al., 2005]

Finally, some groups, specially Granada-ATC, maintain links with several other universities: Zaragoza, Málaga, Rey Juan Carlos I in Madrid, but no serious attempt to initiate an international project has been made. We intend to release software as free software as soon as it is well-packaged and we can write a good roadmap for the future development of the platform.

References

[Alfaro-Cid et al., 2005] Alfaro-Cid, E., Esparcia-Alcázar, A., and Sharman, K. (2005). Clasificación de señales de electroencefalograma usando programación genética. In *Proceedings of the IV Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados MAEB'05)*, Granada, Spain.

[Arenas et al., 2002a] Arenas, Foucart, Schoenauer, and Merelo (2002a). Computación evolutiva en Java: JEO. pages 46–53. Universidad de Extremadura.

[Arenas et al., 2001] Arenas, M., Foucart, L., Merelo-Guervós, J.-J., and Castillo, P. A. (2001). JEO: a framework for Evolving Objects in Java. In *Actas Jornadas de Paralelismo*. Universidad Politécnica de Valencia.

[Arenas et al., 2004] Arenas, M. G., Castillo, P. A., Merelo, J. J., Rateb, F., and V. Rivas, G. R. (2004). Un nuevo enfoque evolutivo: Sistema Multi-ACU. pages 396–403.

[Arenas et al., 2002b] Arenas, M. G., Dolin, B., Merelo-Guervós, J.-J., Castillo, P. A., de Viana, I. F., and Schoenauer, M. (2002b). JEO: Java Evolving Objects. In *Poster Accepted at GECCO 2002*. Available from http://geneura.ugr.es/pub/papers/MPP104.ps.gz, CiteSeer context.

[Arenas, 2005] Arenas, V. M. M. J. J. M. G. J. L. J. L. M. G. (2005). Algoritmos evolutivos en java: resolución del TSP usando DREAM. In *Actas XVI Jornadas de Paralelismo, incluido en CEDI'2005. Granada, septiembre 2005. En prensa, actas editadas por Thompson*.

[Castillo et al., 2005] Castillo, P., Romero, G., Arenas, M., Merelo, J., and Prieto, A. (2005). Designing a control system for an autonomous robot using an evolutionary algorithm. In [IWANN 2005, 2005].

[Cid, 2004a] Cid, E. A. (2004a). Aumento de la velocidad de convergencia usando un sistema distribuido (dream). informe técnico iti-sac-004. ITI-SAC-REPORT ITI-SAC-004, Instituto Tecnológico de Informática.

[Cid, 2004b] Cid, E. A. (2004b). Implementación de un clasificador de señales usando programación genética. informe técnico iti-sac-003. ITI-SAC-REPORT ITI-SAC-003, Instituto Tecnológico de Informática.

[Cid, 2005a] Cid, E. A. (2005a). Clasificación de señales de electroencefalograma usando programación genética. informe técnico iti-sac-006. ITI-SAC-REPORT ITI-SAC-006, Instituto Tecnológico de Informática.

[Cid, 2005b] Cid, E. A. (2005b). Implementación de fdas en jeo y su aplicación al problema de clasificación de señales de electroencefalograma. informe técnico iti-sac-007. ITI-SAC-REPORT ITI-SAC-007, Instituto Tecnológico de Informática.

[Cid, 2005c] Cid, E. A. (2005c). Implementación de una técnica de muestreo aleatorio para reducir la sobreadaptación. informe técnico iti-sac-008. ITI-SAC-REPORT ITI-SAC-008, Instituto Tecnológico de Informática.

[Cid, 2005d] Cid, E. A. (2005d). Programación genética con nodos de aprendizaje. informe técnico iti-sac-009. ITI-SAC-REPORT ITI-SAC-009, Instituto Tecnológico de Informática.

[De la Torre et al., 2005] De la Torre, J., Gómez, M., and Román, I. (2005). Análisis de sensibilidad temporal en los modelos de predicción de insolvencia: una aplicación a las pymes industriales. enviado al XIII Congreso de la Asociación Española de Contabilidad y Administración de Empresas, a celebrar en Oviedo en septiembre de 2005 y a la Revista Española de Financiación y Contabilidad.

[Gustavo Romero and Merelo, 2005] Gustavo Romero, M.G. Arenas, P. C. and Merelo, J. (2005). Evolutionary design of a brain-computer interface. In [IWANN 2005, 2005], page 653.

[Gómez et al., 2005] Gómez, M. E., Román, I., and De la Torre, J. M. (2005). The analysis of the economic risk in small and medium-sized companies of the trade sector. In *28th Annual European Accounting Association Congress, Goteborg*.

[IWANN 2005, 2005] IWANN 2005 (2005). *8th International Work-Conference on Artificial and Natural Neural Networks, IWANN 2003*, Lecture Notes in Computer Science,LNCS. Springer-Verlag.

[JXTA,] JXTA. Project JXTA web site www.jxta.org.

[Kleinberg, 2002] Kleinberg, J. (2002). Bursty and hierarchical structure in streams. In *Proc. 8th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*, pages 91–101. ACM.

[Paechter. et al., 2000] Paechter., B., Bäck, T., Schoenauer, M., Sebag, M., Eiben, A. E., Merelo, J. J., and Fogarty, T. C. (2000). A distributed resource evolutionary algorithm machine (DREAM). In *Evolutionary Computation, 2000. Proceedings of the 2000 Congress on*, volume 2, pages 951–958. IEEE, IEEE Press. Available from http://ieeexplore.ieee.org/iel5/6997/18853/00870746.pdf.

[Riolo and Worzel, 2003] Riolo, R. L. and Worzel, B. (2003). *Genetic Programming Theory and Practice*. Genetic Programming Series. Kluwer, Boston, MA, USA. Series Editor - John Koza.

[Romero, 2005a] Romero, M. G. A. P. C. J. J. M. G. (2005a). Coevolución de perceptrones multicapa y conjuntos de entrenamiento. In *Actas Simposio Metaheurísticas, Algoritmos Evolutivos y Bioinspirados, MAEB2005, incluido en el CEDI2005. En prensa, actas editadas por Thompson.*

[Romero, 2005b] Romero, M. G. A. P. C. J. J. M. G. (2005b). Implementación de un sistema complejo clásico utilizando una librería evolutiva distribuida. In *Actas Simposio sobre Sistemas Complejos, SISCOM, incluido en CEDI'2005. Granada, septiembre 2005. En prensa, actas editadas por Thompson.*

[Tricas, 2005] Tricas, J. J. M. F. (2005). Integración de wikis en una clase de informática: un paso más en la creación de comunidad. In *Actas VI Congreso Nacional de Informática Educativa, Simposio Nacional de Tecnologías de la Información y las Comunicaciones en la Educación, SINTICE'2005 (ADIE). En prensa, actas editadas por Thompson.*