

Resilience to Churn of a Peer-to-Peer Evolutionary Algorithm

-PostPrint-

J.L.J. Laredo*

Department of Architecture and Computer Technology,
University of Granada, Spain
E-mail: juanlu@geneura.ugr.es

*Corresponding author

P.A. Castillo

Department of Architecture and Computer Technology,
University of Granada, Spain
E-mail: pedro@atc.ugr.es

A.M. Mora

Department of Architecture and Computer Technology,
University of Granada, Spain
E-mail: amorag@geneura.ugr.es

C. Fernandes

LASEEB-ISR/IST,
University of Lisbon, Portugal
E-mail: cfernandes@laseeb.org

J.J. Merelo

Department of Architecture and Computer Technology,
University of Granada, Spain
E-mail: jmerelo@geneura.ugr.es

Abstract:

In this paper we analyse the resilience of a Peer-to-Peer (P2P) Evolutionary Algorithm (EA) subject to the following dynamics: computing nodes acting as peers leave the system independently from each other causing a collective effect known as *churn*. Since the P2P EA has been designed to tackle large instances of computationally expensive problems, we will assess its behaviour under these conditions, by performing a scalability analysis in five different scenarios using the Massively Multimodal Deceptive Problem as a benchmark.

In all cases, the P2P EA reaches the success criterion without a penalty on the runtime. We show that the key to the algorithm resilience is to ensure enough peers at the beginning of the experiment; even if some of them leave, those that remain contain enough information to guarantee a reliable convergence.

Keywords: Peer-to-Peer Computing; Evolutionary Computation; Fault tolerance analysis; Scalability.

Reference to this paper should be made as follows: Laredo, J.L.J., Castillo, P.A., Mora, A.M., Fernandes, C., Merelo, J.J. (2009) 'Resilience to Churn of a Peer-to-Peer Evolutionary Algorithm', Int. J. High Performance Systems Architecture, Vol. 1, Nos. 4, pp. 260–268. DOI:10.1504/IJHPSA.2008.024210

Biographical notes: Juan Lu3s Jim3nez Laredo is a PhD. student at the Department of Architecture and Computer Technology at the University of Granada (ATC-UGR). His research interests include Distributed System and Evolutionary Computation.

Pedro A. Castillo Valdivieso is currently Associate Professor at ATC-UGR. His main research interests include Neural Networks and Evolutionary Computation.

Antonio M. Mora Garc3a is a PhD student at ATC-UGR. His researching topics are focused on Ant Colony Optimization Algorithms, Multi-Objective problems and Self-Organizing maps.

Carlos M. Fernandes is a PhD. student at LASEEB-ISR/IST, University of Lisbon. His research interests include Evolutionary Algorithms, Swarm Intelligence and Dynamic Optimization.

Juan Juli3n Merelo Guerv3s is assistant professor at ATC-UGR. His main research interests include Volunteer and Parasitic Computing, specially applied to Evolutionary Computation, and

1 INTRODUCTION

By the term Peer-to-Peer (P2P) Evolutionary Algorithms (EA) we are going to refer to a spatially structured EA in which the population structure is defined by a P2P overlay network. As a spatially structured EA (see Tomassini (2005) for a survey), it can be modelled as a graph in which the vertices are individuals and edges represent relationships between them. A graph can be easily mapped to the topology of a network, consequently, an EA can be easily distributed.

The Evolvable Agent (EvAg) model, which is an instance of a P2P EA, has been shown in Laredo et al. (2008b) to be a suitable approach for tackling large instances of hard optimization problems via massive scalability of P2P systems. As a general property of optimization problems, the evaluation cost scales with respect to the size. Hence, large instances imply a bigger computational cost on the evaluation function. Additionally, the problem complexity increases with size, making the problem more difficult to solve. Resolution methods based on population, as EAs, require larger population sizes in order to tackle such instances with enough reliability (Harik et al. (1997)), and P2P systems are a large and mostly free source of computing resources.

However, Evolutionary Computing has recently entered this area and there are still many challenging issues. Merelo et al. (2008) shows that a P2P system is subject to the peers dynamics: Peers join the system, contribute some resources and leave it afterwards. The independent arrival and departure of thousands of peers causes a collective effect called *churn* which has been modelled by Stutzbach and Rejaie (2006). *Churn* is an inherent property of P2P systems and has to be taken into account in the design of any P2P application. In this paper we will assess a P2P EA under different *churn* scenarios.

The key to our study is the EvAg model presented in Laredo et al. (2008a). It consists of a fine grained approach for parallelizing EAs in which there is a population of concurrent and self-scheduled agents performing the evolutionary steps of selection, variation and evaluation of individuals. The population structure is based on the gossiping protocol newscast, a P2P protocol which builds an overlay network with small-world properties (e.g. Jelasty and van Steen (2002); Spyros et al. (2004)). Such a kind of small-world graphs have been shown, by Giacobini et al. (2006), to be suitable as population structure for an EA.

When tackling an optimization problem, the population size of an EA (i.e. number of nodes-individuals¹ in the graph) shows a dependence on the population structure and scales according to the given optimization problem. Therefore, we analyse the adequacy of our algorithm for large problem instances by scaling the Massively Multimodal Deceptive Problem (MMDP), proposed by Goldberg et al. (1992), from small to large sizes. In order to assess the conditions in which our algorithm is fault

¹In this work, we will refer equally to the terms individual and node, since each individual has its own schedule and could potentially be placed in a different node; moreover, every individual is a node in the overlay network that is mapped to the physical network; when referring to a node in the physical network we will use the term *computing* node.

tolerant, we have performed a scalability analysis using five scenarios. In the first one all nodes stay during the whole experiment while we set several intensities of *churn* in the rest. Note that we do not add any special feature to make *EvAg* fault tolerant, rather this study focuses on fault tolerance as an emergent property of the collective dynamics.

In all the scenarios, we have established that the algorithm has to reach success (i.e. find the problem optimum) with a 98% of reliability. To that end, we obtain the adequate population size using a method based on bisection, proposed by Sastry (2001), which establishes the minimum population size able to solve a problem with a given reliability (a success rate of 0.98 in our case). In spite of the departure of nodes, possibly containing valid solutions, our approach is able to reach success in all the cases (in the worst case the experiment ends with a 0.4% of the initial population). In all the test cases, the population size scales with curves of the same order. Nevertheless, the curves are shifted by a constant which depends on the *churn*; the more extreme the *churn* conditions, the bigger the constant is. The fact of yielding the problem optimum with enough reliability points to the redundancy of nodes as a natural mechanism for fault tolerance. Assuming no restrictions in the amount of available peers, the runtime of the algorithm scales independently of the *churn* scenario and the population size which confirms that the *EvAg* model is resilient and fault tolerant.

Investigating scalability is of extreme importance when changing from a "toy problem" test environment to real-world problems which may require very large chromosomes to codify the solutions. However, real environments are restrictive for the estimation of an adequate population size (e.g. using bisection method). Additionally to the previous study, we have tried to gain some insight on the algorithm performance when the optimum population size is unknown in order to address practitioners on this issue. We have analysed how the success rate varies depending on the problem instance and the *churn* intensity. As a very general rule, the increase of the population size always benefits the search without an additional cost on the runtime.

The rest of the paper is structured as follows. Section 2 introduces some related work in distributed Evolutionary Computing. The overall description of the EvAg model is presented in Section 3. We expose, in Section 4, the methodology and the experimental setup. In Section 5, we analyse the results of the experiments. Finally, conclusions and some future work lines are proposed in Section 6.

2 RELATED WORK

The idea of distributed Evolutionary Algorithms was proposed very early (e.g. Grefenstette (1981)); nowadays, parallel EAs are approached mainly under three methodologies: master-slave, islands and fine grained spatially structured EAs. Tomassini (2005) makes a good review of the different state of the art models with a special focus in fine grained approaches as the one presented in this paper.

Nevertheless, P2P EAs are more recent and not all the models fit with the issues involving P2P systems as shown below:

- In the master-slave mode the algorithm runs on the master node and the individuals are sent for evaluation to the slaves, in an approach usually called *farming*. Such an architecture does not match decentralized structures and the master represents a single point of failure.
- Cantú-Paz (2000) describes the island model in which several panmictic EAs (islands) process their own population and exchange individuals between islands with a certain rate and frequency. This coarse grained approach has been shown by Laredo et al. (2008c) to be very sensitive to parameter calibration and P2P systems do not provide a priori knowledge of the global environment that an island model would need in order to set parameters such as the number of islands, the population size per island and the migration rate.
- Finally, fine grained and spatially structured approaches are more suitable for decentralization as stated by Wickramasinghe et al. (2007) and Laredo et al. (2007b), where the key underlying idea is that individuals evolve within a defined set of neighbours. Following this line, we presented (Laredo et al. (2008a)) a formal model for P2P EAs, it is the *Evolvable Agent model* that we analyse in this paper under *churn* conditions. The model uses the gossiping protocol newscast (Jelasity and van Steen (2002)) as population structure. Newscast was proposed within the DREAM project (Arenas et al. (2002)), one of the pioneers in distributed P2P EAs. The island-based parallelization of DREAM was shown in Laredo et al. (2007a) to be insufficient for tackling large-scale decentralized scenarios. Nevertheless, the gossiping protocol newscast forms a communication graph with small-world features and such a kind of graphs have been shown to be suitable as population structure for Genetic Algorithms in several works (Giacobini et al. (2005); Preuss and Lasarczyk (2004); Giacobini et al. (2006)).

Thus, from all the schemes mentioned above, fine grained approaches seem to be the most suitable for P2P computing, and newscast, as a neighbourhood structure, is scalable and robust (Spyros et al. (2004)). However, there is no evidence on whether an EA based on it might be resilient to churn or scalable by extension, hence the study that we are proposing.

Within the same context, there are some works addressing the resilience of distributed EAs to unreliable environments, Fernandez de Vega (2006) advanced that Evolutionary Algorithms are fault tolerant because its nature and design. However, the employed methodologies, so far, do not have a direct correspondence with a real non-forced scenario. E.g. Hidalgo et al. (2007) analyse the fault tolerance of the island model dying out up to the 50% of the computational resources. Lombrana and Fernandez de Vega (2007) use two rates for processors failures, respectively 2% and 10% every generation. Scriven et al. (2008a) propose a distributed Multi-Objective Particle Swarm Optimisation and a P2P approach in Scriven et al. (2008b). Nevertheless, they use failure rates of 0%, 5%, 10% and 20% which, at the end, are ad-hoc values.

Our study is based on the analysis by Stutzbach and Rejaie (2006) of the peers dynamics. That is, we model faults following the behaviour of real P2P systems. Therefore, the key contribution of this paper is proving that our proposal allows massive scalability of a distributed EA being resilient to *churn*.

3 OVERALL MODEL DESCRIPTION

The overall procedure of our approach consists of a population of Evolvable Agents (EvAg) whose main design objective is to carry out the main steps of evolutionary computation: selection, variation and evaluation of individuals (A.E. Eiben and J.E. Smith (2003)). Each EvAg is a node within a newscast topology in which the edges define its neighbourhood. For the sake of simplicity, we assume a newscast node as a peer. However, a peer could hold several nodes in practice.

3.1 EVOLVABLE AGENT

An *Evolvable Agent (EvAg)* itself is an EA composed of a single individual (Laredo et al. (2007b, 2008a)). In spite of the model not having a *population* in the canonical sense, adjacent EvAgs provide each other with the genetic material that they require to evolve. Therefore, we talk about a population of EvAgs instead of a population of individuals.

Algorithm 1 shows the pseudo-code of an EvAg where the agent owns an evolving solution (S_t).

Algorithm 1 Evolvable Agent

```

 $S_t \leftarrow$  Initialize Agent
loop
  Sols  $\leftarrow$  Local Selection(newscast) See algorithm 2
   $S_{t+1} \leftarrow$  Recombination(Sols)
   $S_{t+1} \leftarrow$  Mutation( $S_{t+1}$ )
  Evaluate( $S_{t+1}$ )
  if  $S_{t+1}$  better than  $S_t$  then
     $S_t \leftarrow S_{t+1}$ 
  end if
end loop

```

The selection takes place locally into a given graph neighbourhood where each agent select other agents' current solutions (S_t). Selected solutions are stored in *Sols* ready to be recombined. Within this process a new solution S_{t+1} is generated. If the newly generated solution S_{t+1} is better than the old one S_t , it replaces the current solution.

3.2 POPULATION STRUCTURE

In principle, our method places no restrictions in the choice of population structure, although this choice will have an impact on the dynamics of the algorithm since it establishes the environmental selection pressure. As it has been previously said, we apply the newscast protocol as graph structure. Within this section we do not enter on the dynamics but on its functioning elements (see Jelasity and van Steen (2002); Spyros et al. (2004) for further details). Algorithm 2 shows the pseudo-code

of the main tasks in the self-organized process which builds the newscast graph. Each node maintains a cache with one entry per node in the network at most. Each entry provides information about a neighbour node: A time-stamp of the entry creation (it allows the replacement of old items), and an agent identifier.

Algorithm 2 Newscast protocol in node $EvAg_i$

Active Thread

```

while  $EvAg_i$  not finished do
  sleep  $\Delta T$ 
   $EvAg_j \leftarrow$  Random selected node from  $Cache_i$ 
  send  $Cache_i$  to  $EvAg_j$ 
  receive  $Cache_j$  from  $EvAg_j$ 
   $Cache_i \leftarrow$  Aggregate ( $Cache_i, Cache_j$ )
end while

```

Passive Thread

```

while  $EvAg_i$  not finished do
  wait  $Cache_j$  from  $EvAg_j$ 
  send  $Cache_i$  to  $EvAg_j$ 
   $Cache_i \leftarrow$  Aggregate ( $Cache_i, Cache_j$ )
end while

```

Local Selection(newscast)

```

 $[EvAg_h, EvAg_k] \leftarrow$  Random selected nodes from  $Cache_i$ 

```

There are two different tasks that the algorithm carries out within each node. The active thread which initiates communications and the passive thread that waits for the answer. In addition, the local selection procedure provides the EvAg with other agents' current solutions ($EvAg_h(S_t)$ and $EvAg_k(S_t)$). After ΔT time each $EvAg_i$ initiates a communication process (active thread). It selects randomly an $EvAg_j$ from $Cache_i$ with uniform probability. Both $EvAg_i$ and $EvAg_j$ exchange their caches and merge them following an aggregation function. In our case, the aggregation consists of picking up the newest items (newscast) for each cache entry in $Cache_i$, $Cache_j$ and merging them into a single cache that $EvAg_i$ and $EvAg_j$ will share. We have fixed ΔT to once per evaluation.

The cache size plays an important role in the newscast algorithm. It represents the maximum number of connections (edges) that a node could have. For example, a topology with n nodes and a cache size of n , will lead to a complete graph topology. Therefore, the cache size is smaller than the number of nodes (typically around $\log(n)$) in order to get small-world features such as a small characteristic path length and a high clustering coefficient. We have fixed the cache size to 20 based on the study of performance for different cache sizes in Laredo et al. (2008a). Additionally, the work by Spyros et al. (2004) shows that newscast is robust for a cache size of 20 in spite of node failures; the graph partition happens under high percentages of node removals, at that point, most of the nodes still remain in a large cluster (e.g. with a 90% of removed nodes, $\sim 92\%$ of the nodes remain in the largest cluster).

4 METHODOLOGY AND EXPERIMENTAL SETUP

In this section we try to assess the impact of the *churn* in our P2P EA approach when tackling the Massively Multimodal Deceptive Problem (MMDP) proposed by Goldberg et al. (1992).

We have focused on the Success Rate (SR) as a metric of the performance of an EA (i.e. times that the algorithm finds the optimum solution out of all trials (A.E. Eiben and J.E. Smith (2003))). In this study, there are three variables that affect the SR: The size of the problem (k), which will conduct to a scalability analysis, the intensity of *churn* (λ) described in Section 4.1 and the population size (P).

Any of the variables have the following influence on the SR if we assume a fixed value for the rest of them. In the case of the problem instance, the bigger the size, the lower the SR. With respect to *churn*, the more departures of nodes, the lower the SR. Finally, the bigger the population size, the higher the SR.

If we consider λ and k as two independent variables under the condition of obtaining a SR of 0.98, the population size can be expressed as a function $f(\lambda, k) = P$. In order to find a fixed condition and obtain P , the bisection method has been used (explained in Section 4.3). We will analyse the curves produced by $f(\lambda, k)$ when $\lambda = 1, 5, 10, 50, \infty$ (where ∞ stands for no churn) and $k = 2, 4, 8, 16, 32, 64$.

Furthermore, the runtime of the algorithm is also analysed as a function of the three variables $g(\lambda, k, P)$ since one of the goals in any distributed EA is to reduce the time for obtaining a solution.

Finally, this empirical study has been carried out using a benchmark problem in which estimating the optimal population size is possible. Nevertheless, such an estimation could turn into unapproachable when switching to hard time consuming real problems. Therefore, we have performed a last case of study using three different policies for the population size and analysing how the SR varies.

4.1 MODELING CHURN

There are two main group-level properties of *churn* which characterize the behaviour of all participating peers: The *inter-arrival time* and the *session length*, respectively, the time between two sessions and the time from the beginning to the end of a session.

In this study we have assumed that all nodes start at the same time with a certain *session length*. Following the work by Stutzbach and Rejaie (2006), the *session length* can be modelled randomly from a Weibull distribution using the following formula:

$$X = \lambda(-\ln(U))^{\frac{1}{s}} \quad (1)$$

where U is drawn from the uniform distribution, s stands for the shape and λ for the scale of the Weibull distribution. The analysis in Stutzbach and Rejaie (2006) exposes that the *session length* of different P2P systems fit with a shape of $s \approx 0.40$ but the values of λ differ. Additionally, the simulator driven experiments define the time unit as a simulator cycle which could apply for different time metrics in real time. Thus, Figure 1 depicts the following values for $\lambda = 1, 5, 10, 50, \infty$. It shows the

complementary cumulative distribution functions (CCDF), representing the percentage of remaining nodes at each moment of the experiment for the different values of λ (e.g. in the cycle 10, $\sim 8\%$ of the peers remain for $\lambda = 1$ and $\sim 100\%$ for $\lambda = 50$).

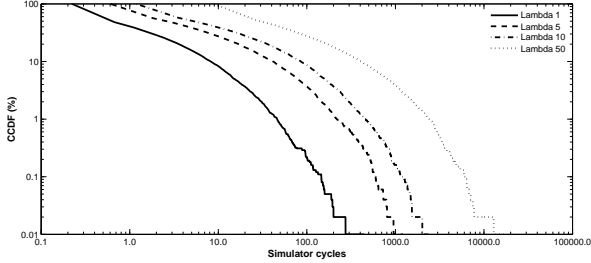


Figure 1: Complementary cumulative distribution functions

In spite of the initial assumption, the set of *churn* scenarios allows a worst case analysis in which the system loses peers until none is left. Once that a node leaves the experiment, it does not rejoin again (i.e. there is no *inter-arrival*).

4.2 MASSIVELY MULTIMODAL DECEPTIVE PROBLEM (MMDP)

The Massively Multimodal Deceptive Problem (MMDP) (Goldberg et al. (1992)) is designed to be difficult for an EA which has to find the optimum in spite of deceptive attractors. It is composed of k subproblems of 6 bits each one (s_i). Depending of the number of ones (unitation) s_i takes the values shown in Table 1.

Unitation	Subfunction value
0	1.000000
1	0.000000
2	0.360384
3	0.640576
4	0.360384
5	0.000000
6	1.000000

Table 1: Basic deceptive bipolar function (s_i) for MMDP

The fitness value is defined as the sum of the s_i subproblems with an optimum of k (equation 2). The search space is composed of 2^{6k} combinations from which there are only 2^k global solutions with 22^k deceptive attractors. Hence, a search method will have to find a global solution out of 2^{5k} . Additionally, such a complexity is specially remarkable under deceptive conditions.

We consider several instances from low to high difficulty using $k = 2, 4, 6, 8, 10, 16, 32, 64$.

$$f_{MMDP}(\vec{s}) = \sum_{i=1}^k fitness_{s_i} \quad (2)$$

4.3 A METHOD FOR ESTIMATING THE POPULATION SIZE

The bisection method (Sastry (2001)) was used to determine the optimal *EvAg* population size P , that is, the lowest P for which MMDP is solved in 98% of the runs. To find it, mutation rate is set to 0, so as to search a minimum population size such that using random initialization it is able to provide enough building blocks to converge to the optimum without other mechanism than recombination.

Algorithm 3 depicts the method based on bisection. The method begins with a small population size which is doubled until the algorithm ensures a reliable convergence. We define the reliability criterion as the convergence of the algorithm to the optimum 49 out of 50 times (0.98 of Success Rate). After that, the interval (min, max) is halved several times and the population size adjusted within such a range. min and max stand respectively for the minimum and maximum population size estimated.

Algorithm 3 Method based on Bisection

```

P = Initial Population Size
while Algorithm reliability < 98% do
  min = P ; max, P = Double (P)
end while
while  $\frac{max-min}{min} > \frac{1}{16}$  do
   $P = \frac{max+min}{2}$ 
  (Algorithm reliability < 98%) ? min = P : max = P
end while

```

4.4 EXPERIMENTAL SETUP

The table 2 shows the setting used for the whole set of experiments. The *EvAg* has been set using standard operators in EAs (see e.g. A.E. Eiben and J.E. Smith (2003)). All results are obtained on the average of 50 independent runs which were conducted in PeerSim Simulator².

Churn scenarios

s	0.40
λ	1, 5, 10, 50
	No churn

MMDP instances

k	2, 4, 6, 8, 10, 16, 32, 64
-----	----------------------------

EvAg Setup

Population size	Estimated by Bisection
Selection of Parents	Binary Tournament + individual
Recombination	DPX, $p_c = 1.0$
Individual Length	$6 \times k$

Newscast setup

Cache Size	20
ΔT	once per evaluation

Test bed

PeerSim simulator

Table 2: Parameters of the experiments

²<http://peersim.sourceforge.net/>. Accessed on December 2008.

5 RESULTS

Figure 2 shows the respective runtimes of the algorithm for $k = 2, 4, 8, 16, 32, 64$. They are the result of averaging 50 runs in every possible *churn* scenario (i.e. $\lambda = 1, 5, 10, 50$). That way, every intersection between a vertical line and a CCDF represents the percentage of individuals in P for which the algorithm is expected to end on a given instance k and a given scenario λ . The effects of *churn* are more pernicious as the instances scale since the runtime increases (e.g. for a $\lambda = 50$, $k = 2$ ends with a 100% of the initial population while $k = 64$ ends with a $\sim 35\%$ of P). Additionally, the lower the λ , the more departures of nodes (e.g. when $k = 64$, $\lambda = 50$ ends with a $\sim 35\%$ of the initial population and $\lambda = 1$ with a $\sim 0.4\%$).

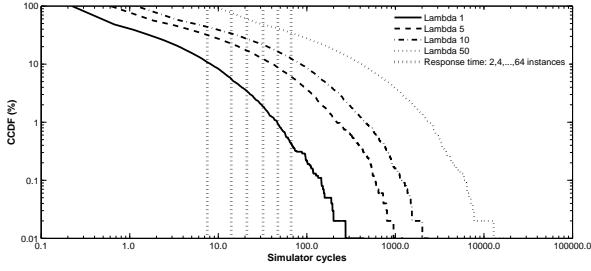


Figure 2: Complementary cumulative distribution functions and average runtime

Under these conditions, Figure 3 shows the scalability of the population size (P) as different curves of k , that is, k scales and λ remains fixed in $f(\lambda, k) = P$. The scalability of $f(\lambda, k)$ fits with a complexity order of $O(k^2)$ in any of the scenarios, that is, *churn* does not damage the scalability order (i.e. the curves are just shifted by a constant which is *churn* dependent) and a reliable convergence can be guaranteed by ensuring enough resources. This fact points to the robustness of the Evolvable Agent model.

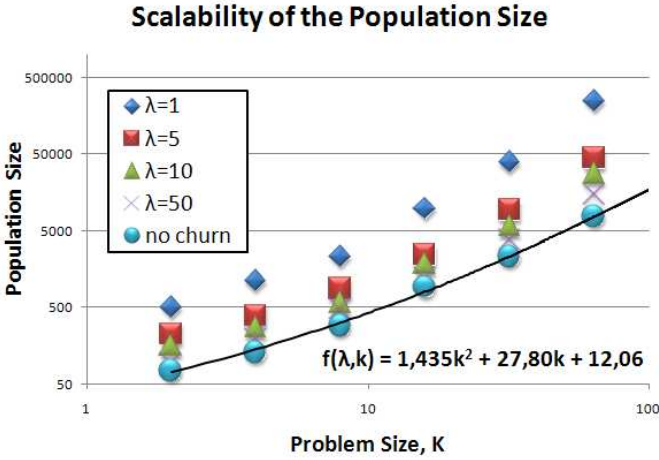


Figure 3: Scalability of the population size $f(\lambda, k) = P$ for the MMDP in 5 scenarios of churn (λ)

Additionally, Figure 4 shows the runtime of the algorithm

$g(\lambda, k, P)$. g is independent from λ and P , with an order $O(k^{0.6})$. In this case, *churn* and the population size do not affect the scalability (neither shifting the curves with a constant) and the runtime is completely dependent on the problem instance k . That means that we can expect the same runtime under any *churn* scenario if we ensure enough resources to satisfy the 0.98 of SR condition. Therefore, the algorithm is resilient to *churn*.

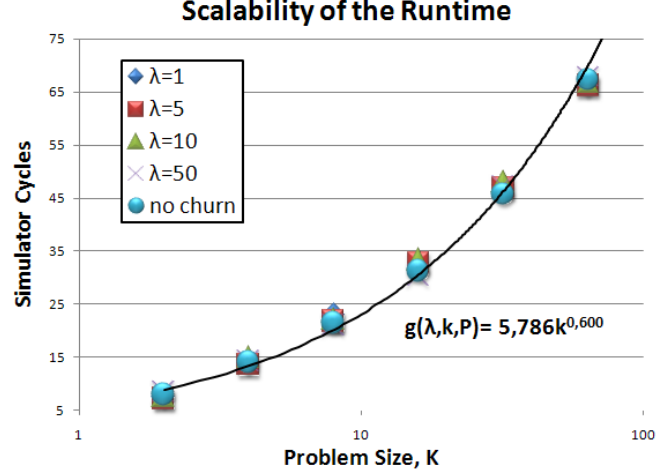


Figure 4: Scalability of the runtime $g(\lambda, k, P)$ for the MMDP in 5 scenarios of churn (λ) using P estimated in $f(\lambda, k)$

The Figure 5 shows the scalability of the computational effort with *no churn*: population size remains constant along of a run. Such a scenario is the one that spends less evaluations since *no churn* needs smaller population sizes. However, the algorithm in a *churn* scenario loses nodes very quickly as k scales (e.g. Figure 2), then, it might scale better than using a fixed population size. If that would be the case, *churn* would turn into a good alternative to save computing resources in large-scale problem instances. We find this hypothesis of the maximum interest and will try to address it in future works.

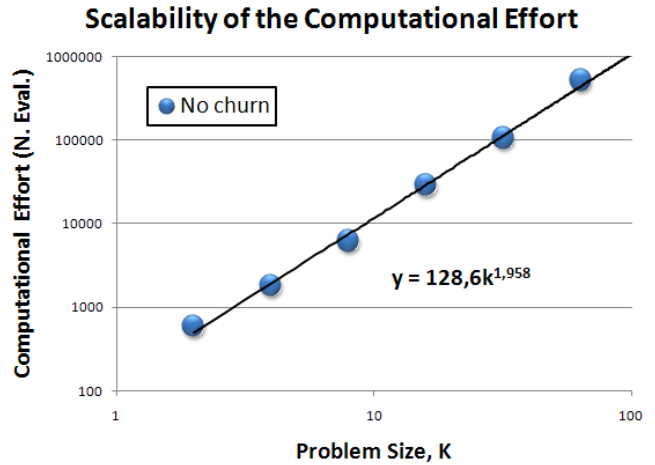


Figure 5: Scalability of the computational effort (or number of evaluations) for the no churn case

Finally, a last performed case of study assumes that the optimal population size can not be estimated under *churn*. This way, we intend to address how the SR varies depending on the *churn* (λ) and the problem instances (k).

We have considered three different policies to establish the population size P .

1. Assuming that the population size with *no churn* can be estimated. P is the optimal population size in a *no churn* scenario.
2. P is four times the optimal size in a *no churn* scenario.
3. Assuming that P can not be estimated but there are many resources in the P2P system. P is an ad-hoc big population size i.e. $P = 200000$.

Figure 6 shows three graphs of the respective policies. They show that the SR decreases with respect to both λ and k . But the change is specially noticeable for small values of λ (e.g. for $\lambda = 5, 10, 50$ policies 2 and 3 show a reliable convergence to the optimum which decreases with $\lambda = 1$). Therefore, it would be very convenient to analyse the *churn* of a P2P system before deploying a P2P EA. If we pay attention to the policy 3 (i.e. $P = 200000$) we can see how the increase of the population size always benefits the search, the worst value here is a SR of 0.9 with $k = 64$ and $\lambda = 1$.

Besides, Figure 7 shows the runtimes for the three policies. It confirms the results on Figure 4, the runtime of the algorithm is independent from λ , P and the SR. Therefore, it can be expressed as a function of k and do not depend on the availability of resources (P). However, the increase of P guarantees a better SR in any scenario.

6 CONCLUSIONS

In this paper we have analysed a P2P Evolutionary Algorithm and proved its adequacy for tackling large problem instances under five *churn* scenarios. We have used the Massively Multimodal Deceptive Problem as a benchmark, but, given that it has been designed to be difficult for EAs, these results should be easily extended to more general discrete or combinatorial optimization problems.

Our approach deals with P2P issues such as decentralization, large-scalability and *churn*. To this end, the population structure is managed by the newscast gossiping protocol. Through the experimental results we conclude that large instances of hard optimization problems can be tackled in P2P systems using the Evolvable Agent (EvAg) model in spite of aggressive *churn* conditions.

The population size scales with polynomial order with respect to the problem size which demands for a big amount of resources. Besides, the expected runtime of the algorithm scales with fractional power with respect to the problem size which makes the algorithm efficient. The approach shows to be resilient to *churn*, once that the estimated population size guarantees a reliable convergence to the problem optimum, the departure of nodes does not inflict a penalization on the runtime.

As future lines of work, we intend to assess the impact of both, the latency between peers and the use of the gossiping algorithm, on the runtime performance. We expect that the idle processing time decreases as the problem instances scale (i.e. bigger instances require a bigger computational time while the communication time can be assumed as fixed). Additionally, we will analyse such an approach taking into account bootstrapping and heterogeneous nodes. Finally, through the course of this investigation, we have formulated an hypothesis about the loss of nodes as a good alternative to save computational effort in large-scale problems. Such an hypothesis will be tested using different probability distributions.

Acknowledgements

This work has been supported by the Spanish MICYT project TIN2007-68083-C02-01, the Junta de Andalucia CICE project P06-TIC-02025 and the Granada University PIUGR 9/11/06 project.

REFERENCES

- M.G. Arenas, Pierre Collet, A.E. Eiben, M. Jelasity, J.J. Merelo, Ben Paechter, Mike Preuss, and Marc Schoenauer. A framework for distributed evolutionary algorithms. In *Parallel Problem Solving from Nature - PPSN VII, Granada, Spain*, number 2439 in Lecture Notes in Computer Science, LNCS, pages 665–675. Springer-Verlag, 2002.
- George Harik, David E. Goldberg, Erick Cantú-paz, Brad L. Miller. The gambler’s ruin problem, genetic algorithms, and the sizing of populations. In *IEEE International Conference on Evolutionary Computation*, pages 7–12. IEEE Press, 1997.
- Erick Cantú-Paz. *Efficient and Accurate Parallel Genetic Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA, 2000.
- Agoston E. Eiben and J. E. Smith. *Introduction to Evolutionary Computing*. SpringerVerlag, 2003.
- Mario Giacobini, Mike Preuss, and Marco Tomassini. Effects of scale-free and small-world topologies on binary coded self-adaptive CEA. In *Evolutionary Computation in Combinatorial Optimization – EvoCOP 2006*, volume 3906 of LNCS, pages 85–96, Budapest, 10-12 April 2006. Springer Verlag.
- Mario Giacobini, Marco Tomassini, and Andrea Tettamanzi. Takeover time curves in random and small-world structured populations. In *GECCO ’05: Proceedings of the 2005 conference on Genetic and evolutionary computation*, pages 1333–1340, New York, NY, USA, 2005. ACM.
- David E. Goldberg, Kalyanmoy Deb, and Jeffrey Horn. Massive multimodality, deception, and genetic algorithms. In *Parallel Problem Solving from Nature, 2*, Amsterdam, 1992. Elsevier Science Publishers, B. V.

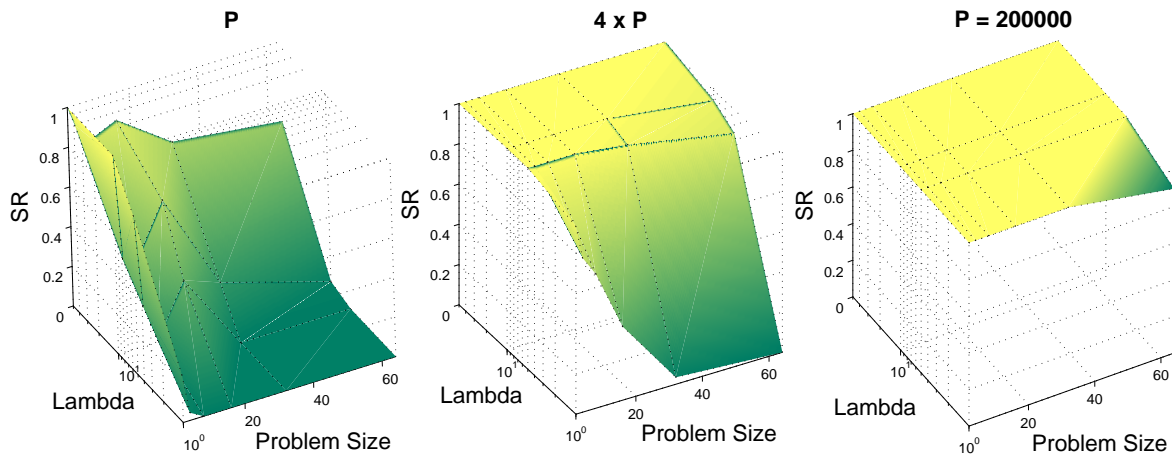


Figure 6: Success Rate (SR) as a function of λ , k and P

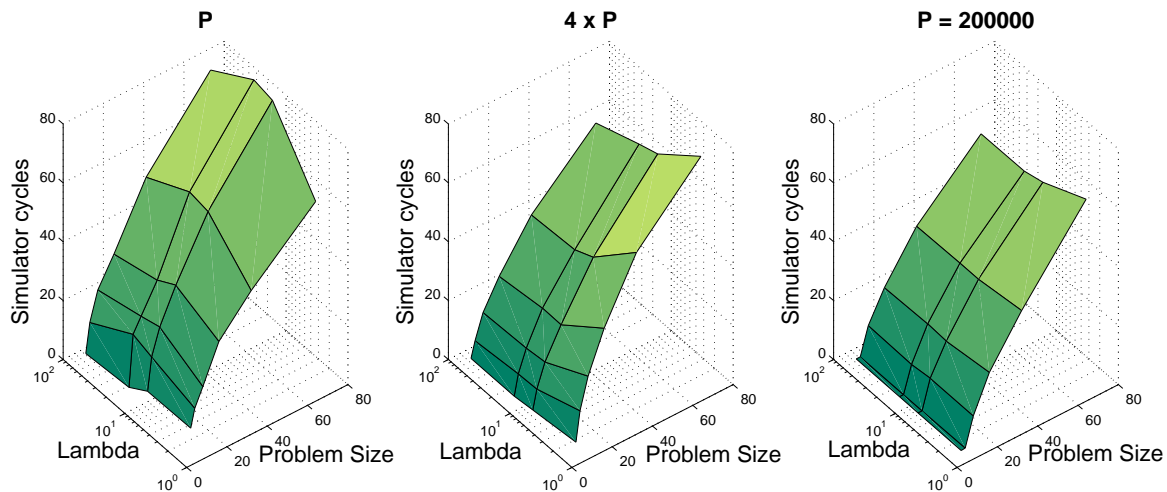


Figure 7: Runtime

Daniel Lombrana Gonzalez and Francisco Fernandez de Vega. On the intrinsic fault-tolerance nature of parallel genetic programming. In *PDP '07: Proceedings of the 15th Euro-micro International Conference on Parallel, Distributed and Network-Based Processing*, pages 450–458, Washington, DC, USA, 2007. IEEE Computer Society.

J.J. Grefenstette. Parallel adaptive algorithms for function optimization. Technical Report CS-81-19, Vanderbilt University, Computer Science Department, Nashville, 1981.

J. Ignacio Hidalgo, Juan Lanchares, Francisco Fernández de Vega, and Daniel Lombrana. Is the island model fault tolerant? In *GECCO '07: Proceedings of the 2007 GECCO conference companion on Genetic and evolutionary computation*, pages 2737–2744, New York, NY, USA, 2007. ACM.

Márk Jelasity and Maarten van Steen. Large-scale newscast computing on the Internet. Technical Report IR-503, Vrije

Universiteit Amsterdam, Department of Computer Science, Amsterdam, The Netherlands, October 2002.

J. L. J. Laredo, E. A. Eiben, M. Schoenauer, P. A. Castillo, A. M. Mora, and J. J. Merelo. Exploring selection mechanisms for an agent-based distributed evolutionary algorithm. In *GECCO '07*, pages 2801–2808, New York, NY, USA, 2007b. ACM Press.

J.L.J. Laredo, P.A. Castillo, A.M. Mora, and J.J. Merelo. Exploring population structures for locally concurrent and massively parallel evolutionary algorithms. In *IEEE Congress on Evolutionary Computation (CEC2008), WCCI2008 Proceedings*, pages 2610–2617. IEEE Press, Hong Kong, June 2008a.

J.L.J. Laredo, A.E. Eiben, M. van Steen, P.A. Castillo, A.M. Mora, and J.J. Merelo. P2P Evolutionary Algorithms: A suitable approach for tackling large instances in hard optimization problems. In Emilio Luque, Tomas Margalef, and

- Domingo Benitez, editors, *Euro-Par*, volume 5168 of *Lecture Notes in Computer Science*, pages 622–631. Springer, 2008b.
- Juan Luís Jiménez Laredo, Pedro A. Castillo, Antonio Miguel Mora, and Juan Julián Merelo Guervós. Evolvable agents, a fine grained approach for distributed evolutionary computing: walking towards the peer-to-peer computing frontiers. *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, 12(12):1145–1156, 2008c.
- Juan Luís Jiménez Laredo, Pedro A. Castillo Valdivieso, Ben Paechter, Antonio Miguel Mora, Eva Alfaro-Cid, Anna Esparcia-Alcázar, and Juan Julián Merelo Guervós. Empirical validation of a gossiping communication mechanism for parallel EAs. In *EvoWorkshops*, volume 4448 of *Lecture Notes in Computer Science*, pages 129–136. Springer, 2007a.
- J.J. Merelo, P.A. Castillo, J.L.J. Laredo, A.M. Mora, and A. Prieto. Asynchronous distributed genetic algorithms with Javascript and JSON. In *IEEE Congress on Evolutionary Computation (CEC2008)*, *WCCI2008 Proceedings*, pages 1372–1379. IEEE Press, Hong Kong, June 2008.
- Mike Preuss and Christian Lasarczyk. On the importance of information speed in structured populations. In *PPSN*, volume 3242, pages 91–100. Springer, 2004.
- K. Sastry. Evaluation-relaxation schemes for genetic and evolutionary algorithms. Technical Report 2002004, University of Illinois at Urbana-Champaign, Urbana, IL., 2001.
- I. Scriven, D. Ireland, A. Lewis, S. Mostaghim, and J. Branke. Asynchronous multiple objective particle swarm optimisation in unreliable distributed environments. In *IEEE Congress on Evolutionary Computation (CEC2008)*, *WCCI2008 Proceedings*. IEEE Press, Hong Kong, June 2008a.
- I. Scriven, A. Lewis, D. Ireland, and J. Lu. Decentralised distributed multiple objective particle swarm optimisation using peer to peer networks. In *IEEE Congress on Evolutionary Computation (CEC2008)*, *WCCI2008 Proceedings*. IEEE Press, Hong Kong, June 2008b.
- Daniel Stutzbach and Reza Rejaie. Understanding churn in peer-to-peer networks. In *Proceedings of the 6th ACM SIGCOMM on Internet measurement (IMC 2006)*, pages 189–202, New York, NY, USA, 2006. ACM Press.
- Marco Tomassini. *Spatially Structured Evolutionary Algorithms: Artificial Evolution in Space and Time (Natural Computing Series)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.
- Francisco Fernandez De Vega. A fault tolerant optimization algorithm based on evolutionary computation. In *DEPCOS-RELCOMEX '06: Proceedings of the International Conference on Dependability of Computer Systems*, pages 335–342, Washington, DC, USA, 2006. IEEE Computer Society.
- Spyros Voulgaris, Márk Jelasity, and Maarten van Steen. *Agents and Peer-to-Peer Computing*, volume 2872 of *Lecture Notes in Computer Science (LNCS)*, chapter A Robust and Scalable Peer-to-Peer Gossiping Protocol, pages 47–58. Springer Berlin / Heidelberg, 2004.
- W. R. M. U. K. Wickramasinghe, M. van Steen, and A. E. Eiben. Peer-to-Peer evolutionary algorithms with adaptive autonomous selection. In *GECCO '07*, pages 1460–1467, New York, NY, USA, 2007. ACM Press.