Evolutionary Market Agents and Heterogeneous Service Providers: Achieving Desired Resource Allocations

Peter R. Lewis, Paul Marrow and Xin Yao

Abstract—In future massively distributed service-based computational systems, resources will span many locations, organisations and platforms. In such systems, the ability to allocate resources in a desired configuration, in a scalable and robust manner, will be essential. We build upon a previous evolutionary market-based approach to achieving resource allocation in decentralised systems, by considering heterogeneous providers. In such scenarios, providers may be said to value their resources differently.

We demonstrate how, given such valuations, the outcome allocation may be predicted. Furthermore, we describe how the approach may be used to achieve a stable, uneven load-balance of our choosing. We analyse the system’s expected behaviour, and validate our predictions in simulation.

Our approach is fully decentralised; no part of the system is weaker than any other. No cooperation between nodes is assumed; only self-interest is relied upon. A particular desired allocation is achieved transparently to users, as no modification to the buyers is required.

I. INTRODUCTION

Future massively distributed service-based computational systems will allow computational resources to span many locations, organisations and platforms, connected through the Internet [1]. There is a need to find novel ways to understand and autonomically allocate, manage and control resources in these large, decentralised and dynamic systems [2]. Building on the simple idea of bilateral exchange in human society, markets may be used as a mechanism with which to control and allocate resources [3] in such complex computational systems.

From an engineering perspective, any resource allocation problem will have an objective: a desired allocation or outcome. A common example of this is load-balancing, where it is desired that the task of providing a resource be shared evenly between a group of providing nodes. More complex objectives involve stable, uneven load-balances, where the resource allocation takes account of other factors. Such factors might include users’ preferences over quality of service issues, underlying costs to the service provider, or differences in the ability of nodes to provide an equivalent resource. This paper proposes an approach to deal with the latter of these cases.

Classically, resource allocation objectives are achieved in a centralised manner, often relying on a single node responsible for load-balancing [4], [5]. Scalability, however, is a critical factor in load-balancing systems, and though decentralised load-balancing mechanisms have been proposed (e.g. [6]), these rely upon cooperation between nodes within the system; something which we may not be able to assume.

In previous work [7], we proposed an evolutionary, market-based approach to achieve even load-balancing between service providing nodes in a fully decentralised system without the presence of cooperation. Our approach, which makes use of an artificially created market for the resource, relies purely upon self-interest, and no individual node has any desire in favour of a load-balanced outcome. In our approach, service providing nodes quote prices for equivalent resources and service users select from whom to purchase the desired resource. Evolutionary market agents, acting on behalf of service providing nodes, evolve the price quoted by their respective seller over time. The payoff from the live market is used as the fitness function for each seller’s evolutionary algorithm.

However, the approach described in [7] considers that all service providing nodes have an equivalent ability to provide the resource. In many real-world scenarios this will not be the case. For example, one node may be able to perform twice as much work as another.

In this kind of scenario, with heterogeneous service provider abilities, the approach in [7] is not able to account of these factors. Its use will result in an even resource allocation between the service providing nodes, regardless of their abilities. This is therefore likely to be an inefficient outcome. In this paper we extend the approach in order to take advantage of this information in providing an appropriate, desired outcome allocation.

Our approach falls into the broad category of market-based control, a methodology which has been applied to the allocation of resources in various real-world scenarios. Clearwater provides a useful introduction to the use of computational markets in scenarios such as bandwidth allocation and air-conditioning control [3], and a useful review may be found in [8].

Cliff and Bruten [8] note that a large proportion of market-based control systems, however, either rely on a central auctioneer, or require human intervention. Therefore, though much of the computation is done by individual agents and is distributed, these systems are often not decentralised. They argue that this leads to a brittleness of the system.

A number of distributed auction mechanisms have also been proposed [9], [10], [11], [12], which reduce the fragility associated with reliance upon a single node, provide more scalability and allow for dynamic composition of auctions. However, similarly to the vulnerabilities of the Internet’s
domain name system [13], failure at certain points in the network may well cripple wider functionality, at best.

Kuwabara et al. [14] propose what we believe to be the most decentralised market-based approach to the allocation of resources. Here, no auctioneer, specialist or market-maker is used; prices are set solely by the sellers and advertised via a broadcast mechanism. Rational buyers then decide the quantity to purchase from each seller, in order to maximise their payoff. However, unlike our sellers’ strategies, those used by Kuwabara et al. are not driven by self-interest.

Non-market-based approaches to the problem of decentralised resource allocation in the presence of heterogeneous service providing nodes also exist in the specific domain of downloading replicated files. Dynamic parallel access schemes [15], [16] make use of client self-interest in a similarly decentralised way to increase the speed of file downloads. Our approach has some similarities with this mechanism, though is more generic and hence applicable to services other than the downloading of files.

A. Problem Formulation

We consider a scenario consisting of a set of service providing nodes, \( S \), each member of which provides an equivalent, quantitatively divisible service, the resource \( \pi \), which may vary only in price. We assume that the members of \( S \) cannot be relied upon to cooperate. We then imagine a large population of service users or buyers, \( B \), each member of which aims to consume some of the resource \( \pi \), at regular intervals.

In previous work [7], our objective was to balance the load evenly, such that all the service providers in \( S \) provide an equal amount of \( \pi \) across the population of service users. In this paper however, we wish to predict what outcome resource allocations will be reached, given heterogeneous seller valuations. Furthermore, we wish to understand how the approach may be extended in order to achieve a stable, uneven load-balance of our choosing. This is indeed a trivial problem when cooperation may be assumed, however we wish to achieve this using self-interest, in a fully decentralised system. Additionally, the possibility of complex buyer decision functions means that there may not be a straightforward mapping between valuations and outcome allocation.

At a given instant, a service provider, \( s_i \in S \), advertises \( \pi \) at the price \( p_{s_i}^\alpha \) per unit. Each service user, a buyer in this case, then purchases some of the resource \( \pi \), should it be in their interest to do so at the price offered. The system iterates, with service providers able to adapt their prices to the market conditions over time. The actual provision of \( \pi \) may be regarded as instantaneous, such that it does not interfere with this mechanism.

For simplicity at this stage, we assume that the system proceeds in discrete time-steps, that each buyer \( b_j \in B \) desires exactly one unit of \( \pi \) per time-step, and that each and every \( s_i \in S \) has sufficient quantity of \( \pi \) available to satisfy all the buyers in \( B \) should it be so requested. This final assumption is commonplace in the provision of information-based services and is present in other related work such as [14]. We do not believe that it alters the underlying behaviour being demonstrated.

Each time-step, each buyer, if it chooses to buy, may purchase any amount of \( \pi \) from any number of service providers in \( S \), subject to the constraint that the total amount purchased per time-step is equal to exactly one unit. If no offer from any \( s_i \in S \) is in its interest, the buyer will instead opt to purchase nothing. We therefore define \( q_{ij} \) to be the instantaneous quantity bought by buyer \( b_j \) from seller \( s_i \). The constraints mean therefore that \( \sum_{i=1}^{S} q_{ij} \in \{0, 1\} \) for all \( b_j \in B \).

The quantity of \( \pi \) sold by a given seller \( s_i \) at a given time-step, its load, \( l_{si} \), is therefore:

\[
l_{si} = \sum_{j=1}^{B} q_{ij}.
\]

We may therefore describe outcome resource allocations in terms of \( l_{si} \), for \( s_i \in S \).

B. Evolutionary Market Agents

Competition between sellers is driven by the co-evolution of their respective evolutionary market agents; their evolutionary fitness being analogous to their payoff from interactions in the market, \( P_{si} \), for seller \( s_i \in S \). An evolutionary market agent, operating on behalf of a particular service providing node, has the self-interested objective of maximising its fitness and hence payoff. Using evolutionary computation techniques, the agent evolves the market position of its host over time. As in our previous work [7], in this model a market position consists simply of price. Therefore each individual represents a real-valued price. For each interaction in the market, the price encoded by an individual is adopted, and the resulting payoff provides its fitness.

The evolutionary algorithm for seller \( s_i \)’s agent proceeds as follows:

1) Decide upon the design parameters to be used: initial price range \([p_{\min}, p_{\max}]\), population size and mutation factor, \( \alpha \). In the simulations described, \( p_{\min} = 0 \), and \( p_{\max} = 500 \). A population size of 10 was used, with a mutation factor, \( \alpha = 0.1 \).
2) Generate an initial population, \( Pop \), and set \( k = 1 \). Each individual in \( Pop \) is a real value, drawn from the uniform random distribution \([p_{\min}, p_{\max}]\).
3) Initial fitness testing
   a) Set the seller’s offer to the value of the first individual in \( Pop \), and enter the market for one market time-step. Record the seller’s payoff, \( P_s \), as that individual’s fitness.
   b) Repeat for the next individual in \( Pop \), until all initial individuals have been evaluated in the market.

\[\text{2009 IEEE Congress on Evolutionary Computation (CEC 2009)}\]
II. HETEROGENEOUS SERVICE PROVIDING NODES

In the original model, the assumptions were that all service providing nodes have an equal ability to provide the service, and that we wish to load-balance the usage equally across them [7]. In realistic deployments however, it is unlikely that this will be the case. Leaving aside any consideration of differences in the overall capacity of service providing nodes - that is to say the total amount of \( \pi \) which a single node may provide at a given time - here we consider a heterogeneity of the ability of nodes in the population to provide the resource. One node, for example, may be able to perform twice as much work as another, perhaps delivering the resource faster.

Such heterogeneities of service providing nodes’ abilities may be represented by selling agent firstly taking account of a notion of the cost of provision of the resource, \( \pi \), and hence different relative valuations of \( \pi \) when compared with the notion of money.

This valuation may be built in to our existing model by modifying the payoff function of the selling agents to take account of the cost of provision, and a preference weight on the price.

In the original model, seller \( s_i \)'s utility gain, or payoff, \( P_{s_i} \), was equivalent to its revenue, \( R_{s_i} \), given by

\[
P_{s_i} = R_{s_i} = \sum_{j=1}^{|B|} p_{s_i} q_{ij}.
\]

Introducing a node’s cost of providing \( \pi \), \( c_{s_i} \), and taking account of a utility preference weight, \( w_{s_i} \), we may instead propose that \( s_i \)'s payoff,

\[
P_{s_i} = \sum_{j=1}^{|B|} q_{ij} (w_{s_i} p_{s_i} - c_{s_i}).
\]

Let us explore this by means of the above two-node example. For the spread buyers used in [7], the quantity of \( \pi \) purchased by buyer \( b_j \) from seller \( s_i \),

\[
q_{ij} = \frac{(v_{b_j} - p_{s_i})}{(v_{b_j} - p_{s_1}) + (v_{b_j} - p_{s_2})},
\]

where \( v_{b_j} \) is the valuation, and hence maximum price payable by each buyer.

For a population of spread buyers, the payoff functions (equation 3) for \( s_1 \) and \( s_2 \) are therefore

\[
P_{s_1} = \sum_{j=1}^{|B|} \frac{v_{b_j} - p_{s_1}}{(v_{b_j} - p_{s_1}) + (v_{b_j} - p_{s_2})} (w_{s_1} p_{s_1} - c_{s_1}), \tag{5}
\]

and

\[
P_{s_2} = \sum_{j=1}^{|B|} \frac{v_{b_j} - p_{s_2}}{(v_{b_j} - p_{s_1}) + (v_{b_j} - p_{s_2})} (w_{s_2} p_{s_2} - c_{s_2}), \tag{6}
\]

Sellers \( s_1 \) and \( s_2 \) will then attempt to maximise their respective payoff function as before. The outcome resource allocation occurs when the system is at equilibrium.

In our example, if \( s_1 \) represents a node able to perform twice as much work as \( s_2 \), then we might say that \( w_{s_1} = 1.0 \) and \( w_{s_2} = 0.5 \), since \( s_2 \) values its work twice as much, when compared to money. For the purposes of the example, we initially assume that both sellers bear a fixed cost; \( c_{s_1} = c_{s_2} = 100 \), and that \( v_{b_j} = 300 \), to give us a reasonable range of prices.

Using an iterative procedure, we calculate that the system described is at equilibrium when \( p_{s_1}^e = 243.1 \) and \( p_{s_2}^e = 262.4 \). At this point, neither \( s_1 \) nor \( s_2 \) may increase its respective payoff by unilaterally choosing a different price.

The allocation at these prices, therefore, is our expected outcome allocation. This may be calculated for each service provider by substituting equation 4 into equation 1:

\[
l_{s_i} = \sum_{j=1}^{|B|} \frac{(v_{b_j} - p_{s_i}^e)}{(v_{b_j} - p_{s_1}^e) + (v_{b_j} - p_{s_2}^e)}.
\]

Solving this using the sellers’ weights, our determined equilibrium price values, and assumed costs in the example,

\[
l_{s_1} = 0.6 \times |B|,
\]

and

\[
l_{s_2} = 0.4 \times |B|.
\]

In other words, we can expect \( s_1 \) to take 60\% of the load, and \( s_2 \) to take 40\%, given the above valuations and costs.

This result may be validated in simulation. Figure 1 shows the normalised outcome loads for the two-node scenario with parameters as described above. The predicted allocation is quickly achieved, and remains stable.

The above approach may be generalised to other buyer behaviours, such as those found in [17], by replacing equation 4 with one which captures the appropriate buyer behaviour.
Mean and standard deviation over 30 independent runs.

We have, therefore, a method for predicting the expected outcome resource allocation, given the service providers’ costs and valuations, and known buyer behaviour. Furthermore, we are able to validate these predictions in simulation.

III. ACHIEVING DESIRED OUTCOME ALLOCATIONS

The ability to predict an outcome allocation, the allocation at equilibrium, is in itself useful. However, in resource allocation tasks, we often have a particular objective in mind. Can we determine, therefore, what seller valuation and cost values should be used in order for the system to evolve to a particular desired outcome?

Let us explore this also by means of some examples; firstly a trivial one, and then a more complex objective.

A. A Trivial Example

Suppose initially that we wish in fact to achieve an even load-balance between two providers, equivalent to the original model in [7]. In this instance, we wish for the following outcome:

\[ l_{s_1} = l_{s_2} = 0.5 \times |B|. \]

Since the population of buyers is homogeneous, for each seller, \( s_i \in S \) it must be that

\[ \sum_{j=1}^{B} q_{ij} = q_{ik}, \forall b_k \in B, \]

where \( q_{ik} \) is the quantity of \( \pi \) bought by buyer \( b_k \) from seller \( s_i \).

Clearly therefore, \( l_{s_1} = 0.5 \times |B| \) if and only if \( q_{ij} = 0.5 \), for any and all \( b_j \in B \). Hence we now know that the payoff functions for \( s_1 \) and \( s_2 \) will be as follows:

\[ P_{s_1} = \frac{|B| \times (w_{s_1} p_{s_1} - c_{s_1})}{2} \]

\[ P_{s_2} = \frac{|B| \times (w_{s_2} p_{s_2} - c_{s_2})}{2} \]

Now, given the buyer decision function for spread buyers (from equation 4 above), we also know that we want

\[ \frac{(v_{b_j} - p_{s_1})}{(v_{b_j} - p_{s_1}) + (v_{b_j} - p_{s_2})} = 0.5, \]

which when solved for a given \( v_{b_j} \), 300 in the example above, gives

\[ p_{s_1} = p_{s_2}. \]

In other words, an even load-balance will be the outcome allocation when both sellers quote the same price for \( \pi \) at equilibrium, as we already know. We also know from previous work [7] that this equilibrium may be achieved with zero cost values, \( c_{s_1}^\pi \) and when \( w_{s_1} = w_{s_2} = 1 \).

B. A More Complex Example

Now let us consider a more complex desired outcome allocation; \( l_{s_1} = \frac{2}{3} \times |B|, l_{s_2} = \frac{1}{3} \times |B| \). We wish for \( s_1 \) to provide twice the load as \( s_2 \).

Following the method above, for a given homogeneous buyer valuation, \( v_{b_j}^\pi \), we may calculate the required relationship between \( p_{s_1}^\pi \) and \( p_{s_2}^\pi \) at equilibrium to achieve our desired outcome:

\[ \frac{(v_{b_j} - p_{s_1}^\pi)}{(v_{b_j} - p_{s_1}^\pi) + (v_{b_j} - p_{s_2}^\pi)} = \frac{2}{3}, \]

which simplifies to

\[ p_{s_2}^\pi = \frac{v_{b_j} + p_{s_1}^\pi}{2}. \] (8)

In order to achieve our desired outcome allocation, the prices \( p_{s_1}^\pi \) and \( p_{s_2}^\pi \) at equilibrium, must conform to this relationship.

The question then arises of what valuation and cost values, \( c_{s_1}^\pi, c_{s_2}^\pi, w_{s_1} \) and \( w_{s_2} \), can be chosen in order to satisfy this constraint. In order to begin to answer this, we firstly take a look at how the equilibrium prices vary with respect to the differences in seller valuations.

Figure 2 shows two distinct regions within the space of heterogeneous seller valuations. Firstly, we note the region when \( w_{s_2} < \frac{1}{3} \), where \( s_2 \) has no equilibrium price. This may be explained as follows. Since \( P_{s_i} = \sum_{j=1}^{B} q_{ij} (w_{s_i} p_{s_i} - c_{s_i}) \), and \( l_{s_i} = \sum_{j=1}^{B} q_{ij} \), in order to achieve any positive load \( l_{s_i} \), with a non-negative payoff \( P_{s_i} \), it follows that we must have

\[ w_{s_i} \leq \frac{c_{s_i}^\pi}{p_{s_i}^\pi}. \] (9)

Now, the maximum possible price \( p_{s_i}^\pi \) which would lead to any transaction occurring will be the buyers’ reserve price, or valuation, \( v_{b_j}^\pi \), which given homogeneous buyers is, as we have already seen, the same for all \( b_j \in B \). Since, given a fixed cost \( c_{s_i}^\pi \), this is the price which places a lower bound on the constraint on \( w_{s_i} \), it follows that there exists no price for which the seller \( s_i \) has a non-negative payoff when
$w_{s1} < \frac{c_{s1}}{v_{b1}^*}$. In the example illustrated above, this is indeed \(\frac{1}{3}\). Below this weight value, no transactions will take place, since they would require either buyer or seller to accept a negative payoff.

However, given that $s_2$ is unable to trade when $w_{s1} < \frac{1}{3}$ in the above example, in this region $s_1$ exists in monopoly. As we would expect, it is therefore able to charge the maximum acceptable price for the buyers, (i.e. $p_{s1}^* = v_{b1}^*$) and be confident of securing a normalised load of 1.

In the competitive region of the weight-space however, the equilibrium prices follow a more complex pattern. Recall that in order to achieve our desired resource allocation, we require a ratio between the sellers’ prices as described in equation 8. In our example, when $v_{b1}^* = 300$, $\forall b_j \in B$, we require that

$$p_{s2}^* = 150 + \frac{p_{s1}^*}{2}.$$ 

In our example, this occurs when $w_{s1} = 1.0$ and $w_{s2} = 0.42105$. Therefore, by fixing values for $c_{s1}^*$ and $c_{s2}^*$, we have identified weight values which we predict will achieve our desired outcome allocation of $l_{s1} = \frac{2}{3} \times |B|, l_{s2} = \frac{1}{3} \times |B|$. We now validate this prediction in simulation. Figure 3 shows the evolution of the load on each service providing node over time, where $w_{s1} = 1.0$ and $w_{s2} = 0.42105$ for 30 independent runs; mean and standard deviation are shown. The system achieves the desired allocation, where $s_1$ is providing twice the level of load as $s_2$, and clearly, as with the original results in [7], a rough approximation to the predicted outcome is achieved quickly, while the final allocation remains stable.

We are, therefore, able to predict the outcome resource allocation for given sets of buyers and sellers, or conversely to parameterise them in order to achieve a desired outcome. Sellers taking account of the cost of provision, and their valuation of money when compared with this cost, leads to a load-balance which reflects the varied ability of service providers. This is achieved transparently, from the buyers’ perspective.

IV. FURTHER SIMULATION RESULTS

We have already seen how the equilibrium prices vary with respect to differences in seller valuations (see figure 2 above). But how does this translate into outcome resource allocations? Figure 4 illustrates the example scenario’s corresponding predicted outcome loads between $s_1$ and $s_2$, normalised by the size of the buyer population.

Here, the monopolistic region is clearly visible, when $w_{s2} < \frac{1}{7}$, as $l_{s1} = 1$ and $l_{s2} = 0$. The relationship between the valuation differences and the outcome allocation in the competitive region is, however, more complex. This is due to the non-linearity of the spread buyer decision function being employed here, and the shape of the curves in this region are dependent upon this behaviour.

We now validate this prediction, as before, in simulation. Figure 5 shows experimental results compared with the predictions in figure 4. Mean and standard deviation results are shown, in each case at the allocation reached after 1000 iterations.
Fig. 5. Predicted and experimental outcome loads with respect to seller valuations, \( w_{s2} = 1.0, w_{s2} \) varies along the x axis. Experimental results are mean and standard deviation over 30 independent runs.

Firstly, we note that in the competitive region of the valuation space, our experiments follow the predicted outcome loads to a high degree of accuracy. The stochastic nature of the evolutionary market agents’ algorithms ensures that there remains some variation at equilibrium, due to ongoing mutation.

During the monopolistic phase however, when \( w_{s2} < \frac{1}{3} \), we note not only that the experimental result for \( l_s \) only loosely follows its predicted path, but also that there is a high standard deviation between the simulation runs. However, further observation reveals that for the experimental results in this region, \( l_s + l_{s2} \) is often less than 1, meaning that the users’ desired load is not being met in all cases. This is in fact also due to ongoing mutation in the \( s_1 \)'s population at equilibrium.

The results in figure 6, from a typical run when \( w_{s1} = 1.0 \) and \( w_{s2} = 0.1 \), illustrate this effect. As the price mutates, there is a probability that it will increase above the buyers’ maximum acceptable price. When it does so, no transactions occur, as the seller, \( s_1 \) in this case, has priced itself out of the market. The result is temporarily a load - and payoff - of zero. Of course, this individual will quickly be selected out, and the price will return to a lower value, however, the presence of this step function in the fitness landscape of the evolutionary market agent means that this behaviour will continue. Nevertheless, this is not of great concern, since the price’s proximity to the step in the fitness landscape only occurs in monopoly, and will not affect regular operation of the algorithm.

Here we have considered decentralised resource allocation in the presence of service providing nodes with heterogeneous abilities; the amount of work a node is able to perform. In this scenario, service providing nodes may be said to value their resources differently, when compared with a common notion of money.

We described how the market-based approach in [7] may be extended and generalised in order to take account of these concepts. We demonstrated how this heterogeneity of seller valuations leads to a stable and uneven outcome load-balance. This is achieved despite a lack of any central control or cooperation between nodes; only self-interest is relied upon. However, the possibility of complex buyer decision functions means that there may not be a straightforward mapping between valuations and outcome allocation. We therefore provided a method for its prediction and validated our predictions in simulation.

Further to this, we described how a desired outcome resource allocation of our choosing may be reached, by determining suitable valuations for the service providing nodes, given the buyers’ decision functions. The achievement of a desired, uneven load-balance is achieved transparently to service users, as no modification to the buyers’ behaviour is required.

Our future work will consider the results obtained here in the presence of larger numbers of service providing nodes, though our intuition suggests that this will involve a straightforward scaling up of the method presented in this paper. We also aim to be able to quantify how quickly the desired allocation is reached, especially in the presence of more dynamic environments.

Finally, market-based control has been applied to a variety of different real-world scenarios [3], many of which are not concerned with purely computational resources, such as air-conditioning control and transport logistics [18]. We believe that our approach will be equally applicable in such scenarios, where the characteristics of self-interest and decentralisation apply.

V. CONCLUSIONS AND FUTURE WORK
ACKNOWLEDGEMENTS

This work was partially supported by an UK EPSRC/BT CASE project on "Autonomic Solutions for Virtualised ICT Systems" and an EPSRC grant (No. GR/T10671/01) on "Market-Based Control of Complex Computational Systems".

REFERENCES


