International Symposium for Next Generation Infrastructure
October 1-4, 2013, Wollongong, Australia

Bio-inspired cost-effective access to big data

Lijuan Wang[a]
Jun Shen[a*]

Abstract: With the rapid proliferation of services and cloud computing, Big Data has become a significant phenomenon across many scientific disciplines and sectors of society, wherever huge amounts of data are generated and processed daily. End users will always seek higher-quality data access at lower prices. This demand poses challenges to service composers, service providers and data providers, who should maintain their service and data provision as cost-effectively as possible. This paper will apply bio-inspired approaches to achieving equilibrium among the otherwise competitive stakeholders. In addition to novel models of cost for Big Data provision, bio-inspired algorithms will be developed and validated for dynamic optimisation. Furthermore, the optimised algorithms will also be applied in the data-mining research on the Alpha Magnetic Spectrometer (AMS) experiment, which is aiming to find dark matter in the universe. This experiment typically receives 200G and generates 700G data daily.

Key words: Bio-inspired algorithm; Big data; Data-intensive service provision; Cloud computing

I. Introduction

In recent years, the data generated by scientific activities, social networking, social media, as well as commercial applications has exponentially increased. This explosion of digital data and the dependence on data-intensive services are main characteristics of the IT tread in this decade. Big data is used to describe the exponential growth, availability and use of information, both structured and unstructured. Cloud computing has become a viable, mainstream solution for data processing, storage and distribution. It provides unlimited resources on demand.

Considering Big Data and the cloud together, we see a practical and economical way to deal with Big Data, which will accelerate the availability and acceptability of analysis of the data. To put Big Data to work, increasing numbers of companies are starting to use the cloud to publish Big Data as a data service. Many data services in the area of Big Data analytics have now become available.

The development of cloud computing and the delivery of Internet-scale data-intensive services have brought a number of challenges, such as the maintenance of quality of service

[a] School of Information Systems and Technology, University of Wollongong, NSW, Australia
[*]Corresponding author: Jun Shen jshen@uow.edu.au
http://dx.doi.org/10.14453/isngi2013.proc.42

(QoS) as well as excellent opportunities for businesses to gain market share, and for scientific programs to save on cost and energy as well as on space for data management. Because the main motivation for moving to the cloud is to minimise cost (or maximise earnings), that is to say, for economic reasons, many studies attempt to find approaches to supporting cost-effective service provision from holistic perspectives.

In this paper, we address the cost-effective service provision issues and propose a bio-inspired model for data-intensive service provision. After giving the economic model, the problem and the three optimization phases of data-intensive service provision will be presented. Then, the advantages of applying bio-inspired algorithms to solve data-intensive service provision is presented, as well as some headway of our research in applying bio-inspired algorithms to tackling the problems. The case study is also described. Finally, we conclude this paper and propose future work.

II. Problem statement

A. An economic model of data-intensive service provision

In general, data-intensive service composition will be supported cooperatively by data providers, service providers and service composers. Service providers and service composers become "service requesters" when they request data from data providers and services from other service providers. The various providers need a standardised way to regulate and price their resources[1].The composite service should be constructed by achieving the Pareto optimum under the Nash equilibrium for the data-provider utility, the service-provider utility and the service-composer utility.

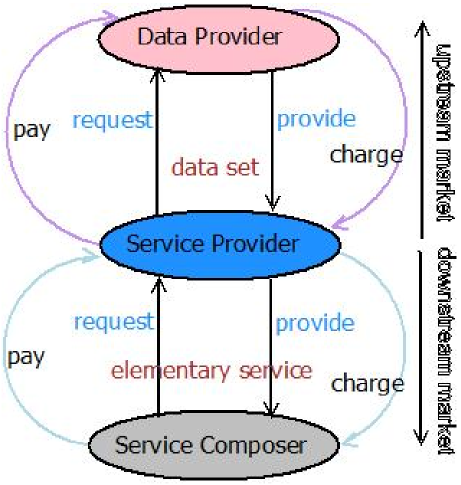

Figure 1. Service and data usage and the charging relationship.

An economic model of this service provision is assumed to be an accurate representation of the reality and to offer a suitable way to regulate the interaction among these providers. As shown in Fig. 1, in the downstream market, the service composer seeks optimal strategies to select elementary services provided by multiple service providers, who compete on the basis of price and quality of services. From the service composer's point of view, it is important to be able to assess the value of the needed services and how much it wants to pay for them to satisfy its users' requirements as well as maximising its own profit. From the service provider's perspective, it is important to be able to analyse its competitive position and improve its offers if it is to win contracts with the service composer. In the upstream market,

the service provider requests the data from the data provider. The price and location of the data may affect the total cost and the price of services. Therefore the prices of service and data have a crucial impact on the service composer's and the service provider's profits.

B. Data-intensive service composition

The data-intensive service composition problem is modelled as a directed graph, denoted as G= (V, E, D, start, end), where V= {AS1, AS2,…, ASn} and E represent the vertices and edges of the graph respectively, D= {d1, d2,…, dz} represents a set of z data servers. Fig. 2 gives an example of a directed graph for data-intensive service composition, in which data sets, as the inputs of services, are incorporated.

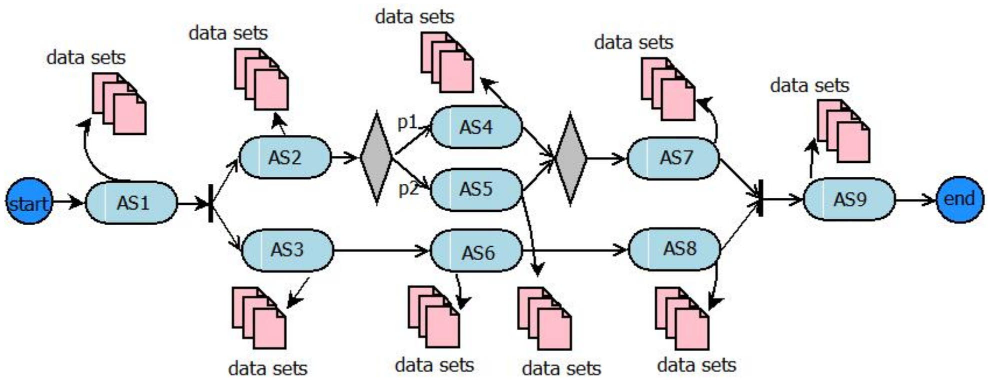

Figure 2. Example of directed graph for data-intensive service composition.

There are only two virtual vertices, the start vertex which has no predecessors, and the end vertex which has no successors. Each abstract service ASi has its own service candidate set $cs_i = \{cs_{i,1}, cs_{i,2}, \ldots, cs_{i,m}\}$, $i \in \{1,\ldots,n\}$, which includes all concrete services to execute ASi. Each abstract service ASi requires a set of k data sets, denoted by DTi, that are distributed on a subset of D. A binary decision variable xi,j is the constraint used to represent only one concrete service is selected to replace each abstract service during the process of service composition, where $x_{i,j}$ is set to 1 if $cs_{i,j}$ is selected to replace abstract service ASi and 0 otherwise.

C. Optimizations in data-intensive service composition

Data-intensive services raise new challenges for service composition. They need to access large numbers of datasets, which may be replicated at different data centres[2-4]. The economic cost of each data replica itself can never be overlooked because cloud computing is actually a model of business computing[5]. For one dataset, the access costs for a replica on one data center may be different from that for other data centres[6]. For this reason, it becomes very important to optimize data-replica selection[5,7]. In Ref. 8, it was pointed out that, although some research has considered the effect of data intensity on service composition[9,10], the communication cost of mass data transfer and its effects on business processes with different structures was overlooked. Therefore, the optimizations will be performed at two points during the lifetime of a data-intensive service composition, as shown in Figure. 3.

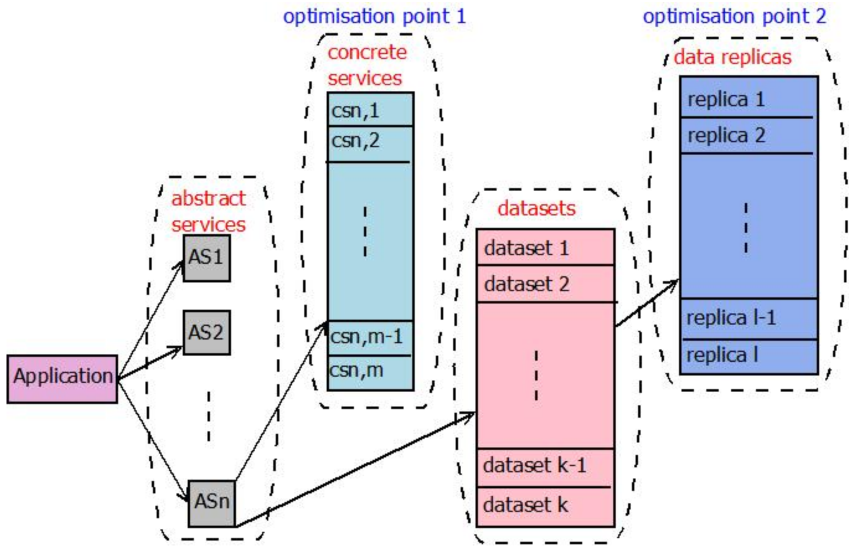

Figure 3. Service and data selection in data-intensive service composition.

The first optimisation phase occurs when the system uses a late-binding mechanism to choose concrete services. Because there may be many candidate services of similar functionality but different QoS that might be mapped to an abstract service, we need to select a set of candidate services. We refer to this phase as dynamic concrete-service selection. The second optimisation phase is dynamic data-replica selection. Because there may be multiple copies of each dataset, the services need to find the best available data replica. After these optimisations, a third optimisation phase is the dynamic optimisation and integration of candidate solutions, which globalises the two optimisations, achieved locally and will usually involve negotiations.

II. Bio-inspired cost-effective to access big data

A. Bio-inspired algorithms to optimize data-intensive service provision

As explained above, it is important to develop mechanisms for selecting concrete services and data replicas to achieve a cost-effective solution for data-intensive service composition. The location of users, service composers, service providers and data providers will affect the total cost of service provision. Different providers will need to make decisions about how to price and pay for resources. Each of them wants to maximize its profit as well as retain its position in the marketplace. Dynamic service-price-setting models using non-standard pricing mechanisms, such as auctions and negotiations, are expected to appear for composing data-intensive services. Meanwhile, the QoS-based service-composition problem is regarded as a multi-objective or constrained combinatorial optimization problem. It is well known that the computational complexity of service selection and negotiation is NP-hard. Data-intensive service provision faces challenges and constraints such as autonomy, scalability, adaptability and robustness.

Biological systems can satisfy these requirements optimally via evolution. They are autonomous entities and are often self-organized, without a central controller. Moreover, bio-inspired concepts and mechanisms have already been applied successfully to service-oriented systems by a few researchers, including some of our work in applying bio-inspired algorithms to tackling the problems related to data-intensive service provision[11-18]. Based on these

existing studies, we are confident that bio-inspired algorithms present many advantages in dealing with data-intensive service-provision problems.

Bio-inspired optimization algorithms are proposed to solve the service-provision problem because of the simplicity of the algorithms and their rapid convergence to optimal or near-optimal solutions. Bio-inspired algorithms will also demonstrate their strengths in optimizing dynamic negotiations among service composers and service or data providers. Biological entities can learn from their environment. They can sense the surrounding conditions and adaptively invoke suitable behavior. We focus on service composition involving large amounts of data transfer, data placement and data storage, which make the cost issues for the whole composition more complex.

The optimization problem of service composition will be more critical when big data is present and involved in future services and cloud-based systems. To be able to allocate, aggregate[13], duplicate and de-duplicate[15] unstructured data stored virtually everywhere, the configuration and management of data should be optimized, not only for reasons of performance and efficiency of data access, but also for lower cost and even the implicit energy and space savings in the maintenance of data storage.

B. Case study

After construction of the system model and the design of the optimisation algorithm, the optimised algorithms will also be applied in the data-mining research on the Alpha Magnetic Spectrometer (AMS) experiment which uses cloud computing to process huge amounts of data. The AMS experiment is a large-scale international collaborative project, and the only large physics experiment on the International Space Station. The purpose of the AMS experiment is to study the universe and its origin by searching for antimatter and dark matter while performing precision measurements of cosmic rays composition and flux. The AMS lab is based at CERN in Switzerland. The key technology for accessing the data remotely collected from AMS relies on data services based on the cloud computing. This is actually supported by IBM Cloud Computing Center located at Southeast University in China. Typically, the center receives 200G bytes data from AMS and generates 700G bytes data after processing them, on each single day.

III. Conclusion

In this paper, we present an economical model of data-intensive service provision and also give the problem and the three optimization phases during the lifetime of a data-intensive service composition. Our bio-inspired cost model uses all of the data-provision, service-provision and service-composition perspectives. In order to minimize the total cost of the data-intensive service provision and to maximize each player's profit, we also need to design negotiation processes among different players. The design of the negotiation processes is currently under way.

References

[1]Armbrust, M., Fox, A., Joseph, R.G.A., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., and Zaharia, M., "Above the clouds: A Berkeley view of cloud computing," EECS Department, University of California, Berkeley, *Tech. Rep. UCB/EECS-2009-28*, February 10, 2009.

[2]Milenkovic, M., Castro-Leon, E., and Blakley, J.R., "Power-aware management in cloud data centers," *Cloud Computing*, *LNCS 5931*, edited by M. Jaatun, G. Zhao, and C. Rong, Springer Heidelberg, 2009, pp. 668-673.

[3]Song, Y., Wang, H., Li, Y., Feng, B., and Sun, Y., "Multi-tiered on-demand resources scheduling for VM-based data centre," *Proceedings of 9th IEEE/ACM International Symposium on Cluster Computing and the Grid*, 2009, pp. 148-155.

[4]Winter, M., "Data center consolidation: A step towards infrastructure clouds," edited by M. Jaatun, G. Zhao, & C. Rong, *Cloud Computing, LNCS 5931*, Springer, Heidelberg, 2009, pp. 190-199.

[5]Liu, W., Shi, F., Du, W., and Li, H., "A cost-aware resource selection for data intensive applications in cloud-oriented Data Centers," *International Journal of Information Technology and Computer Science*, Vol. 1, 2011, pp. 10-17. http://dx.doi.org/10.5815/ijitcs.2011.01.02

[6]Venugopal, S., and Buyya, R., "An SCP-based heuristic approach for scheduling distributed data-intensive applications on global grids," *Journal of Parallel and Distributed Computing*, 2008, Vol. 68, No. 4, pp. 471-487. http://dx.doi.org/10.1016/j.jpdc.2007.07.004

[7]Bell, W.H., Cameron, D.G., Carvajal-Schiaffino, R., Millar, A.P., Stockinger, K., and Zini, F., "Evaluation of an economy-based file replication strategy for a data Grid," *Proceedings of the 3rd IEEE/ACM International Symposium on Cluster Computing and the Grid*, 2003, pp. 661-668.

[8]Li, Y., and Lin, C., "QoS-aware service composition for workflow-based data-intensive applications," *IEEE International Conference on Web Services*, 2011, pp. 452-459.

[9]Bucchiarone, A., and Presti, L., "QoS composition of services for data-intensive application," *Second International Conference on Internet and Web Applications and Services*, 2007, pp. 46-46. http://dx.doi.org/10.1109/ICIW.2007.50

[10]Zhang, Y., Zhou, X., and Gao, Y., "Optimizing the data intensive mediator-based web services composition," edited by X. Zhou, J. Li, H. Shen, M. Kitsuregawa, and Y. Zhang, *Frontiers of WWW Research and Development - APWeb*, LNCS 3841, Springer, Heidelberg, 2006, pp. 708-713.

[11]Yuan, S., Shen, J., and Krishna, A., "Ant inspired scalable peer selection in ontology-based service composition," *Proceedings of IEEE World Congress on Services*, United States: IEEE, 2009, pp. 95-102.

[12]Shen, J., Beydoun, G., Yuan, S., and Low, G., "Comparison of bio-inspired algorithms for peer selection in services composition," *Proceedings of IEEE International Conference on Services Computing*, 2011, pp. 250-257.

[13]Dong, F., Luo, J., Song, A., Cao, J., and Shen, J., "An effective data aggregation based adaptive long term CPU load predictions mechanism on computational grid," *Future Generation Computer Systems: the international journal of grid computing: theory, methods and applications*, Vol. 28, No. 7, 2012, pp. 1030-1044.

[14]Wang, L., and Shen, J., "Towards bio-inspired cost minimisation for data-intensive service provision," *Proceedings of IEEE International Conference on Services Economics*, United States: IEEE, 2012, pp. 16-23.

[15]Sun, Z., Shen, J., and Yong, J., "A novel approach to data deduplication over the engineering-oriented cloud systems," *Integrated Computer Aided Engineering*, Vol. 20, No. 1, 2013, pp. 45-57.

[16]Wang, L., Luo, J., Shen, J., and Dong, F., "Cost and time aware ant colony algorithm for data replica in alpha magnetic spectrometer experiment," *Proceedings of IEEE 2nd International Congress on Big Data*, United States: IEEE, 2013, pp. 254-261.

[17]Wang, L., Shen, J., and Beydoun, G., "Enhanced ant colony algorithm for cost-aware data-intensive service provision," *Proceedings of IEEE Ninth World Congress on Services,* United States: IEEE, 2013, pp. 227-234.
[18]Wang, L., and Shen, J., "Economical data-intensive service provision supported with a modified genetic algorithm," *Proceedings of In IEEE 2nd International Congress on Big Data*, United States: IEEE, 2013, pp. 358-365.