DECENTRALIZED ORCHESTRATION OF COMPOSITE OGC WEB PROCESSING SERVICES IN THE CLOUD

Fei Xiao [a], Geoffrey Y.K. Shea[a, *], Jiannong Cao [a]

[a] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong - (phil.xiao, geoffrey.shea)@polyu.edu.hk, csjcao@comp.polyu.edu.hk

KEY WORDS: Geospatial Cloud Computing, Sensor Web, P2P Networking, Hypercube topology, HyperCGSF

ABSTRACT:

Current web-based GIS or RS applications generally rely on centralized structure, which has inherent drawbacks such as single points of failure, network congestion, and data inconsistency, etc. The inherent disadvantages of traditional GISs need to be solved for new applications on Internet or Web. Decentralized orchestration offers performance improvements in terms of increased throughput and scalability and lower response time. This paper investigates build time and runtime issues related to decentralized orchestration of composite geospatial processing services based on OGC WPS standard specification. A case study of dust storm detection was demonstrated to evaluate the proposed method and the experimental results indicate that the method proposed in this study is effective for its ability to produce the high quality solution at a low cost of communications for geospatial processing service composition problem.

1. INTRODUCTION

The evolutions in computing science and Web technology offer the geoscience community with continuously expanding resources for geospatial data collection and processing. Geoscience theories and technologies have improved over the last three decades, and Geographic Information Science (GIS) has evolved from traditional desktop to a network based, multiple-tier and service-oriented architecture (SOA) (David, 2005). Web services are geographically dispersed and are used by a wide array of companies and government organizations because of the reusability, flexibility, and platform independence (Shen et al., 2007). An increasing amount of geospatial resources and processing functions are available in the form of online Web services (Castronova et al., 2013; Hofer 2014). One of the most important values of employing web service technologies is the composition of web services to create value added service (Tong et al., 2011). Web service composition refers to the mechanisms that promote the collaboration and interoperability of individual web service to create software applications with a functionality that is the result of integration of the individual functionalities provided by each participating service, which has the potential to reduce development time and effort for new applications (Rao and Su, 2004). Chaining interoperable geospatial models and data resources is particularly interesting because such a chain can potentially answer more questions than the individual models alone, allowing users to achieve complex tasks in a variety of different contexts (Dubois, 2013). Web-based distributed geospatial computing and large networks of collaborating applications are the next step on the evolution of web-based geoprocessing platforms (Kiehle et al., 2007).

Many attempts to integrate OGC services into service-based geospatial workflows have been proposed within the geospatial and environmental domains (Granell et al., 2010; Goodall et al., 2011; De Jesus et al., 2012; Mullerm et al., 2013; Wang et al., 2013). However, using web services to build an asynchronous geospatial workflow is still a challenge. First, the traditional approach for composing web services is mainly constructed in a centralized manner, whereby an orchestrator component running on a single server is responsible for the execution of all process instances, while all relevant data are maintained at a single location. These centralized web service composition approaches suffer from performance bottleneck (high communication burden and high computational load) Furthermore, both the geospatial sciences and the cloud computing environment are spatiotemporal intensive. Earth science phenomena are complex processes and Earth science applications often take a variety of data as input with a long and complex workflow. It becomes then a critical challenge to deliver such complex applications to cloud as a transparent service to support massive numbers of users.

Thus, a new geospatial model service composition method is needed to address these challenges. In this paper, a Hypercube Geospatial Service Framework (HyperCGSF) is proposed to solve the handling large volume of data and dynamic and complex interaction problems in geospatial model services. This article is organized as follows: section 1 introduces the background and literatures of this study. Section 2 introduces the study area and data used. Section 3 discusses the design of the proposed geospatial model service composition approach. The model validation and case study are introduced in Section 4. The conclusions and recommendations outlook are summarized in Section 5.

---

* Corresponding author

2. RELATED WORK

2.1 Geospatial process service and service composition

OGC has issued a series of service specifications ranging from data service and processing service to catalogue service, making great progress on the use of Web services to publish and access geospatial resources. For data processing over the internet, the Web Processing Services (WPS) specification was released on 2005 by Open Geospatial Consortium (OGC) (Schut, 2007), in order to provide spatial processes through a standardized service interface based on the Hypertext Transfer Protocol (HTTP) (Foerster and Stoter, 2006). Ideally, a Cloud-enabled system where any geospatial process could be exposed and executed through a WPS implementation and the involved spatial data conform to the standards of a Spatial Data Infrastructure (SDI) cloud satisfy on demand provision of valuable geoinformation. In the first OGC compliant Cloud service ever presented (Baranski et al., 2009), the scalability feature of Cloud Computing is evaluated through a comparison of a WPS implemented in the Cloud with one implemented locally. Also, significant effort is being spent regarding SDI integration with Cloud Computing: a thorough analysis of the basic concepts along with the design and test of a Cloud-enabled SDI is examined by Schäffer et al. (2010) while Baranski et al. (2011) made use of a hybrid Cloud by combining local public IT-infrastructure in order to meet Quality of Service requirements set by INSPIRE directive (European Commission, 2007). The term Composite-WPS that is used to invoke all other services involved is introduced in the case of a bomb threat scenario (Stollberg and Zipf, 2007).

However, current web-based GIS or RS applications generally rely on centralized structure, where the geospatial data is stored on one single server. To get the required geospatial information, it is necessary to collect data and processing resources from multiple service nodes spreading over the network, composite these services as a workflow, and execute the workflow on a centralized controller. This approach has inherent drawbacks such as single points of failure, network congestion, and data inconsistency, etc. The inherent disadvantages of traditional GISs need to be solved for new applications on Internet or Web.

2.2 P2P-based technologies and their application in Geospatial science

P2P networking is a paradigm where a set of user machines at the edge of the Internet communicates with one another to share resources without the help of any central authority (Sukumar, 2014). For a P2P network, the geographical boundaries become irrelevant, and the failure of any central authority promises spontaneous growth, as well as freedom form censorship. Peers include friends, collaborators and competitors, and the resource sharing has to be implemented through decentralized protocols. Scalability is an integral part of this concept, which means that no P2P system is worth looking at unless it scales to millions of machines around the globe. Regardless of the legal ramifications of ethical issues, P2P has led to users to a new form of freedom in collaborative resource sharing. One typical application of P2P network is the generation of genomic data about newly discovered proteins by collaborating hundreds of small laboratories all over the world. In addition, Facebook and Twitter also started using P2P technologies for content distribution.

Several researches have been conducted to apply P2P technologies to construct distributed GIS and RS systems. Guan et al. (2004) explored the techniques of enabling GIS services in a P2P environment to overcome the inherent shortcomings of current GISs and presented an implementation called BP-GService. Puppin et al. (2005) applied Globus package to develop a grid information service based on P2P network, which offers a fast propagation of information and has high scalability and reliability following the OGSA standard. Lee et al. (2006) proposed a method of applying P2P network to collaborative GIS environment, particularly targeting exploratory spatial data analysis for small-group brainstorming. Gianluigi et al. (2010) proposed a grid portal for solving geoscience problems using distributed knowledge discovery services by integrating workflow technologies with data mining resources and a portal framework in unique work environment called MOSÈ.

3. THE HYPERCUBE GEOSPATIAL SERVICE FRAMEWORK

The HyperCGSF consists of a multifunctional geospatial service provider agent model, an underlying networking topology called 'hypercube', and a set of distributed algorithms to support efficient publishing, sharing, managing, and accessing the geospatial service resources (data and processes) distributed over the cloud and orchestration of geospatial processing services in a decentralized manner with the features of security, load balancing, and fault tolerance.

3.1 The Geospatial Service Provider Agent (GeoSPA) Model

GeoSPA was designed as a geospatial services hub through which the geospatial model developer and data producer can deploy standard-based geospatial services onto cloud computing system. GeoSPA supplies a series of algorithms for managing and discovering the geospatial services, as well as orchestrating the service composition execution. To achieve these functionalities, three GeoSPA service models were defined including Earth Observation (EO) data service model, processing service model, and computing service model. These three models make GeoSPA as a one-stop solution for building SDI in cloud computing environment. Figure 1 illustrates the internal structure of GeoSPA.
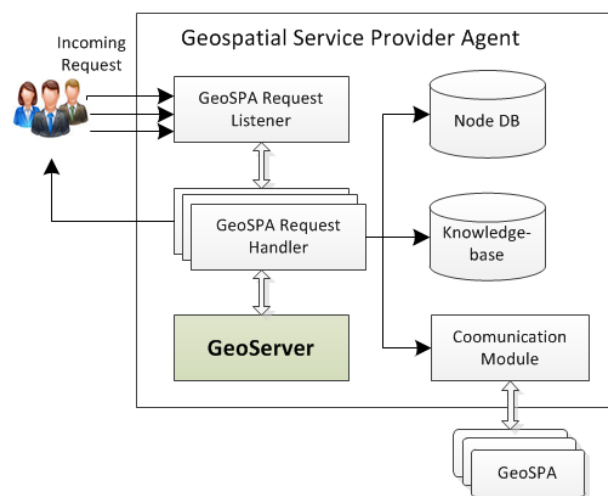

Figure 1. Internal structure of GeoSPA

As shown in Figure 1, several functional components are embedded in GeoSPA, which are introduced below. First, the *GeoSPA Request Listener* acts as the entry point of GeoSPA for processing the incoming requests sent by service consumers or other service agents. The GeoSPA Request Handler processes the incoming request in a simultaneous manner. Each GeoSPA request handler corresponds to a user-specified request. Second, a GeoSPA is equipped with a *Knowledge-base*, through which the GeoSPA can determine which kind of service model needs to be used to handle the user-specified request. Furthermore, the embedded *GeoServer* component is responsible to offer the actual web-based geospatial processing services based on the OGC WPS specification. The geoscience problems are always complex and several geospatial services need to be cooperated and coordinated to achieve complex tasks. To support this functionality, each GeoSPA was equipped with an *Agent Communication Module*, which is responsible for communicating with other GeoSPAs to exchange information. Finally, each GeoSPA is embedded a *Node Database* which is responsible for storing all information that is needed for service agent communications and geospatial workflow execution.

3.2 The Hypercube-based Network Topology

Schlosser et al (2002) proposed a network topology called 'hypercube' to manage the peers in a P2P network. Hypercube is one of the most important structures which is regular, symmetric, shorter diameter, good fault-tolerant properties and so on. A complete hypercube graph consists of $N = b^d$ nodes, where b is the number of nodes in each dimension and d is number of dimensions spanned by the cube. Figure 2 depicts the one example of hypercube topology drawn in 3D with the base b=2. Based on Schlosser, essentially every node can perform as the root node of a tree which spans all nodes in the hypercube. A complete hypercube topology has N=b^(L_max+1) nodes and has a Δ equals to log_b N, where L_max+1 is the number of dimensions spanned by the cube. There are (b-1)·(L_max+1) neighbours for each node in the hypercube.

Based on the hypercube network topology introduced above and GeoSPA model introduced in Section 3.1, this research proposed the Hypercube Geospatial Computing Framework (HyperCGSF) as a scalable and elastic geospatial service framework. The HyperCGSF contains a scalable architecture and a set of distributed algorithms to enable the efficient discovery, composition and execution of geospatial processes persisted by multiple GeoSPAs. Figure 2 illustrates the architecture of HyperCGSF.

Figure 2. Illustration of Hypercube topology with b=2
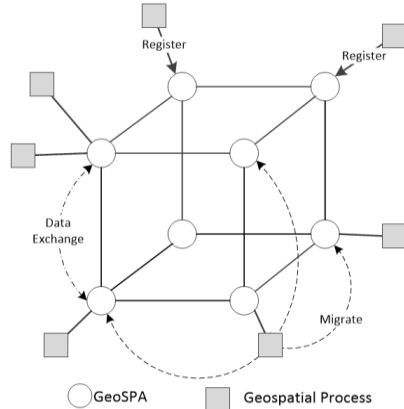
As shown in Figure 2, central to this framework is the binary hypercube topology for organizing an arbitrary number of available GeoSPA nodes (represented by circle in Figure 2). Based on the hypercube network topology introduced above, there is no centralized execution engine in HyperCGSF, and the output of geospatial process (GP, represented by square in Figure 2) is directly transferred amongst distributed service nodes. The GP can migrate between different HyperCGSF nodes for execution through the GeoSPA processing service and GeoSPA computing service. One of the most important functionalities of HyperCGSF is that it can map the static job workflow specification to the dynamic cloud computing resources on the fly for distributed job workflow execution, and cooperate with the distributed GeoSPAs to achieve complex geospatial processing tasks.

3.3 Decentralized Execution of Geospatial Service Composition

A service composition consists of a set of geospatial processes as well as their execution sequence. Figure 3 illustrates the procedure of the decentralized execution of geospatial service composition. In this case, there are five processes in the service composition. The GeoSPA$_{101}$ plays the role of geospatial process provider agent, while GeoSPA$_{100}$, GeoSPA$_{110}$, GeoSPA$_{010}$, GeoSPA$_{011}$, and GeoSPA$_{111}$ play the role of process worker agents (Figure 3(a)).
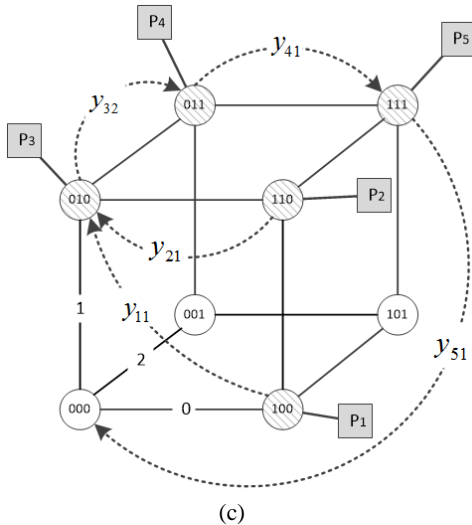

(a)

(b)

(c)

Figure 3. Illustration of decentralized plan execution

At the beginning of service composition execution, every worker agent will be dispatched a geospatial process for performing. The work agents need to request to $GeoSPA_{101}$ for the migration of geospatial process (Figure 3(b)). Upon receipt of the geospatial processing migration request, the $GeoSPA_{101}$ creates a new instance of the requested geospatial processes, which then migrates to the worker agent for execution. Then, every worker agent executes the geospatial process and cooperates with other worker agents to exchange process results. If a process execution completes successfully, the worker agent transmits the desired output variable to the process's immediate successor (Figure 3(c)). For example, in Figure 3(c), the $GeoSPA_{100}$ performs p_1 and returns the desired output parameter y_{11} to its immediate successor $GeoSPA_{010}$, where p_3 is executed using y_{11} as one of its input parameters.

4. EXPERIMENT AND DISSCUSION

4.1 The Integrated Dust Storm Detection Model

An integrated dust storm detection model (IDDM) was designed in this study as a study case to evaluate the efficiency of the HyperCGSF. Dust storms are known to have adverse effects on human health and significant impact on weather, air quality, hydrological cycle, and ecosystem. Five models were involved in IDDM: (1) The Reverse Absorption Technique (RAT) which uses the Brightness Temperature Difference (BTD) of two or more wavelengths for retrieving dust storm region (Ackerman 1997; Zhao et al., 2008). (2) The Infrared Difference Dust Index (IDDI) model developed by Legrand (2001) to detect the presence of desert dusts over Africa. (3) The Radiative Transfer Model (RTM) has been widely used to retrieve the aerosol optical thickness (AOT) effective radius (Reff) and altitude of dust layer (Shao et al., 2006) by means of look up table (LUT) based inversion calculations. (4) The land surface temperature (LST) model was taken based on a reference image at 11μm band, which synthesized from 2 previous week's clear sky 11μm in the same study area. (Zhang et al., 2006). (5) A 72-hour forward trajectory analysis was performed using the NOAA HYSPLIT model with inputs from the National Centers for Environmental Prediction/the National Center for Atmospheric Research (NECP/NCAR) global reanalysis meteorological data. Figure 4 illustrates the workflow of the IDDM.
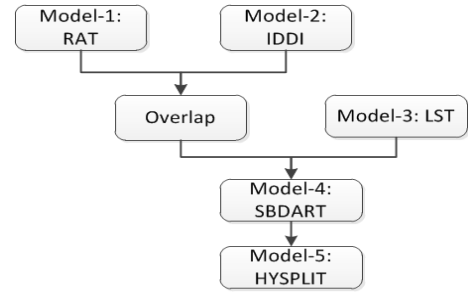

Figure 4. Workflow of the integrated dust storm detection model

4.2 Experiment Environment

Performance tests were conducted to evaluate the potential computational costs introduced by the HyperCGSF. A prototype system was implemented using the Google Compute Engine (hereafter GCE). GCE is an infrastructure service provided as part of the Google Cloud Platform. GCE is made up of three major components: virtual machines, persistent disks, and networks. It is available at several Google data centers worldwide and is provided exclusively on an on-demand basis. GCE provides worldwide Cloud services, such as IaaS, PaaS, and SaaS. Five GCE virtual machine (VM) instances (instance-1 to 8) were purchased in this research and deployed the models of IDDM introduced in Section 4.1. Five Windows Server 2008 Cloud VMs were created from one Cloud instance image. Each VM has one virtual CPU of 2.50 GHz, with 3.75 GB of RAM, a 50-GB disk, and bandwidth of 4 Gb/s.

### 4.3 Response Time for Varying Request

To evaluate the efficiency of the algorithms in the presence of many simultaneous accesses, the average execution time of IDDM workflow was recorded and compared using both the HyperCGSF and the traditional BEPL-based WPS service composition (BPEL-WPS) approach (Meng et al., 2009). The objective of this experiment was to evaluate how the average execution time varies with the increment of domain size and request rate. The request rate is the number of incoming service composition requests every minute. Figure 6 shows the experiment results using different request rates from 1 to 60, over geographical scope of $10° \times 10°$ (Figure 6 (a)), $20° \times 20°$ (Figure 6 (b)), and $30° \times 30°$ (Figure 6 (c)) with 4km spatial resolution.
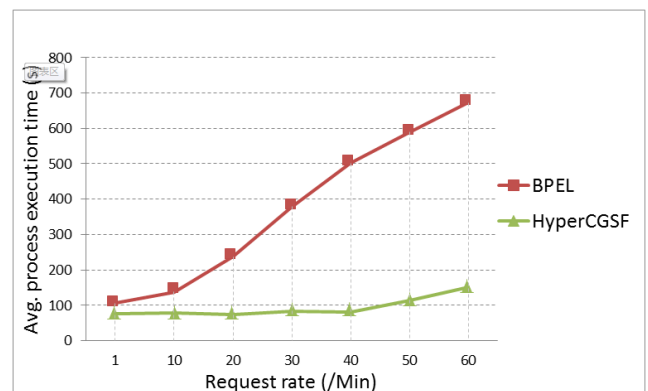
(a) Domain size: $10° \times 10°$
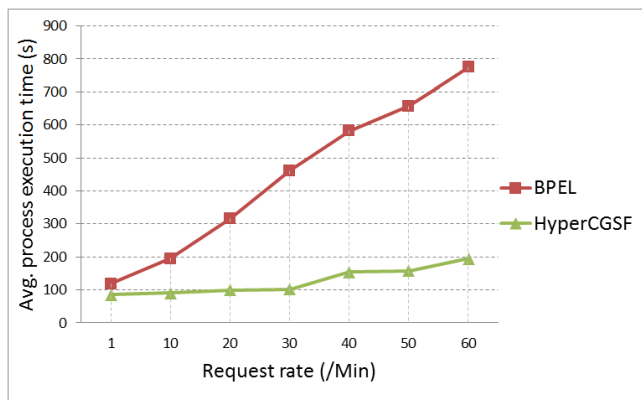
(b) Domain size: 20 °✕ 20 °
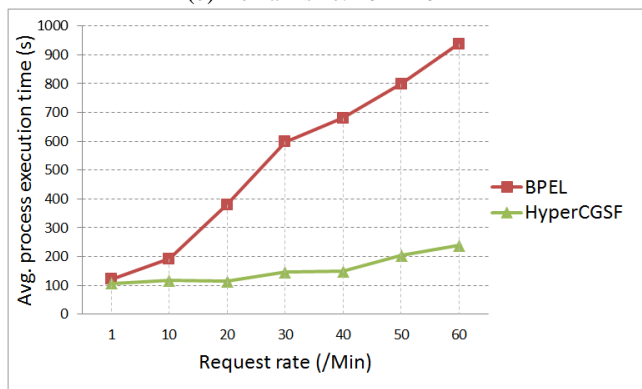
(c) Domain size: 30 °✕30 °

Figure 6: Comparison of average process execution time using different service composition methods for the IDDM.

Several conclusions can be drawn from Figure 6. First, the average execution time of BPEL-WPS and HyperCGSF increases dramatically with the number of current requests. For example, the response time for these two methods increases from about several minutes to approximately one hour when the number of requests per minute increases from 1 to 60. That is because before model execution, the service agent needs to read geospatial data with large volume from remote sites as the input parameters.

Second, the response time of HyperCGSF is less than the traditional BPEL-WPS approach for every request number, and the increasing rate of execution time for HyperCGSF is also lower than BPEL-WPS. The test result is reasonable. The BPEL-WPS approach applies the centralized manner that the interaction and data exchange movements are conducted through the orchestrator, or workflow execution engine. The geospatial processes can generate a lot of data that is irrelevant to the composite service, yet this data will be transferred to the coordinator node where it is discarded, thereby putting an unnecessary load on the network. Different from BPEL-WPS, the HyperCGSF applies the decentralized architecture in the way the service agents can communicate directly with each other to exchange processing results on demand.

One of the most advanced features of HyperCGSF is that it supports the migration of geospatial processes among various GeoSPAs. This feature is extremely useful for geoscience because the geospatial data is always Big Data and distributed on remote sites. Considering that the geoscience applications always need to process large volumes of geospatial data, transferring the geospatial processes rather than the geospatial

data over a cloud computing environment is significant in that it can dramatically decrease the volume of data transmission and increase the computing efficiency. Several studies have shown the advantages of applying the migration of the service agent in geospatial model services (Tan et al., 2015). However, some security issues must be taken into consideration before migrating a geospatial process from one GeoSPA node to another.

5. CONCLUSION AND FUTURE WORK

This paper presents a decentralized P2P-based geospatial model service composition framework (HyperCGSF) for multiple geospatial model sharing and operation through Internet. The service composition has provided a promising computing paradigm for the automatic model service composition. Based on the OGC WPS standard, the HyperCGSF was developed to provide an explicit service and interactive interface for sharing/accessing geospatial models through model services. The open standards help reduce the interoperability problems that may be encountered when using closed standards, such as commercial private standards. The HyperCGSF developed in this study can handle issues related to establishing the geospatial model workflow, allowing modellers to implement the programming interface without directly developing model services and, therefore, focus on model algorithm implementation. The experiment on the detection of dust storm presence demonstrates the feasibility, efficiency, and effectiveness of the proposed framework and its better fulfilment of geospatial model service.

Future work will be focused on the following aspects: (i) efforts will be made to improve the mobile capability of the agent in terms of movement and communication, and meanwhile make full consideration about the security issues. (ii) The High-performance Computing (HPC) capability provided by Cloud computing platform will be investigated to develop high performance agent-based geospatial service chain.

REFERENCES

Ackerman, S.A., 1997. Remote sensing aerosols using satellite infrared observations. Journal of Geophysical Research, 102(D14), 17069–17079.

Baranski, B., Schäffer, B., Redweik, R., 2009. Geoprocessing in the Clouds. In: Proceedings of the Free and Open Source Software for Geospatial (FOSS4G) Conference, Sydney, Australia, pp. 17–22.

Castronova, A.M., Goodall, J.L., 2013. Simulating watersheds using loosely integrated model components: evaluation of computational scaling using OpenMI. Environ. Model. Softw. 39, 304-313.

David, J.M., 2005. Towards a GIS platform for spatial analysis and modeling. In: Maguire, D.J., Batty, M., Goodchild, M.F. (Eds.), GIS, Spatial Analysis, and Modeling. ESRI Press, Redlands, C.A., U.S., pp. 19-39.

De Jesus, J., P. Walker, M. Grant and S. Groom (2012). "WPS orchestration using the Taverna workbench: The eScience approach." Computers & Geosciences 47(0): 75-86

Dubois, G., Schulz, M., Skøien, J., Bastin, L., Peedell, S., 2013. eHabitat, a multipurpose Web Processing Service for ecological modeling. Environ. Model. Softw. 41 (3), 123-133.

Foerster, T., Brühl, A., Schäffer, B., 2011. RESTful web processing Service. In: Proceedings of the AGILE 2011 Conference. Utrecht, The Netherlands.

Gianluigi Folino, Agostino Forestiero, Giuseppe Papuzzo, Giandomenico Spezzano, 2010. A grid portal for solving geoscience problems using distributed knowledge discovery services, Future Generation Computer Systems,26,87-96

Goodall, J.L., Robinson, B.F., Castronova, A.M., 2011. Modelling water resource systems using a service-oriented computing paradigm. Environmental Modelling and Software 26, 573-582.

Granell, C., Diaz, L., Gould, M., 2010. Service-oriented applications for environmental models: reusable geospatial services. Environ. Model. Softw. 25 (2), 182-198.

Guan, J., Wang, L., Zhou, S., 2004. Enabling GIS services in a P2P environment. In: Das, G., Gulati, V.P. (eds.) CIT 2004. LNCS, vol. 3356, pp. 776–781. Springer, Heidelberg.

Hofer, B. (2014). "Uses of online geoprocessing technology in analyses and case studies: a systematic analysis of literature." International Journal of Digital Earth: 1-17.

Kiehle, C., Greve, K., Heier, C., 2007. Requirements for next generation spatial Data Infrastructures - standardized Web based geoprocessing and Web Service orchestration. TransactionsinGIS11(6),819–834.

Lee, Y-W., Park, H.-H., Shibasaki, R., 2006. Collaborative GIS Environment for Exploratory Spatial Data Analysis Based on Hybrid P2P Network. Z. Pan et al. (Eds.) Edutainment 2006, LNCS 3942, 330-333.

Legrand, M., Plana-Fattori, A., & N'doumé, C., 2001. Satellite detection of dust using the IR imagery of Meteosat: 1. Infrared difference dust index. J. Geophys. Res., 106(D16), 18251–18274.

McKendry, I.G., Hacker, J.P., Stull, R., 2001. Long-range transport of Asian dust to the Lower Fraser Valley, British Columbia, Canada. J. Geophys. Res., 106(D16), 18361–18370.

Muller, I., Kowalczyk, R., & Braun, P., 2006. Towards agent-based coalition formation for service composition. In Proceedings of of IEEE/WIC/ACM international conference on intelligent agent technology, 73–80.

Puppin, D., Moncelli, S., R. Baraglia, N. Tonelotto, Silvestri, F., 2005. A Grid information service based on Peer-to-Peer, in: Proc. 11th Euro-Par Conf., Euro-Par 2005, in: LNCS, vol. 3648, Springer, pp. 454–464.

Rao, A. S. and Georgeff, M. P. (1991). Modeling Rational Agents within a BDIArchitecture. In Allen, J., Fikes, R., and Sandewall, E., editors, Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR 1991), pages 473{484. Morgan Kaufmann.

Schaeffer, B., 2008. Towards a transactional web processing service. In: Proceedings of the Sixth Geographic Information Days, Münster.

Schlosser, M., Sintek, M., Decker, S., Nejdl, W., (2002). A scalable and ontology-based p2p infrastructure for semantic Web Services, in: P2P, pp. 104–111.

Schut, P., 2007. Open Geospatial Consortium Inc. OpenGIS Web Processing Service. Open Geospatial Consortium, pp. 1e87.

Shen, W., Q. Hao, S. Wang, et al., An agent-based service-oriented integration architecture for collaborative intelligent manufacturing, Robotics and Computer Integrated Manufacturing 23 (2007) 315–325.

Stonebraker, M., 2010. SQL databases v. NoSQL databases, Communications of the ACM 53 (4) (April 2010) 10–11.

Tan, X.C., Di, L.P., et al., 2015. Cloud- and Agent-Based Geospatial Service Chain:A Case Study of Submerged Crops Analysis During. IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING VOL. 8, NO. 3

Tong, H., Cao, J., Zhang, S., Li, M., 2011. A distributed algorithm for Web Service composition based on service agent model. IEEE Trans. Parall. Distr. Syst. 22 (12), 2008–2021.

Wang, S., Anselin, L., Bhaduri, B., Crosby, C., Goodchild, M. F., Liu, Y., et al. (2013). CyberGIS software: A synthetic review and integration roadmap. International Journal of Geographical Information Science, 27, 2122–2145.

Whiteside and J. Evans. OGC Implementation Specification 07-067r5: Web Coverage Service (WCS) Implementation Standard. Technical report, Open Geospatial Consortium, 2008.

Zhang, P., Lu, N.M., Hu, X.Q., Dong, C.H., (2006). Identification and physical retrieval of dust storm using three MODIS thermal IR channel. Global Planet. Change 52, 197–206.

Zhao, T.L., Gong, S.L., Zhang, X.Y., Jaffe, D.A., 2008. Asian dust storm influence on North American ambient PM levels: observational evidence and controlling factors. Atmos. Chem. & Phys., 8, 2717-2728.