Distributed Geospatial Analysis through Web Processing Service: A Case Study of Earthquake Disaster Assessment

Xiaoliang Meng\(^1\)
\(^1\)Research Center of Spatial Information and Digital Engineering, Wuhan University
Wuhan, Hubei 430079, China
Email: mengxiaoliang2000@126.com

Yichun Xie\(^2\) and Fuling Bian\(^1\)
\(^2\)Institute for Geospatial Research and Education, Eastern Michigan University
Ypsilanti, MI 48197, USA
Email: yxie@emich.edu, flbian@yahoo.com

Abstract—Web Processing Service (WPS) is a new standard approved by Open Geospatial Consortium (OGC), which is aimed to define a standardized interface facilitating the publishing of geospatial processes, and the discovery of and binding to those processes by clients. WPS also acts as a middleware service that obtains geospatial data from an external resource in order to run a process on a local implementation. Consequently, it can be used to wrap other existing OGC geospatial standards designed to provide geospatial services, such as Web Map Service (WMS), Web Feature Service (WFS), and Web Coverage Service (WCS) so as to achieve distributed geospatial analyses. This paper demonstrates three approaches of using WPS to chain geospatial services. The methods are illustrated through a prototype online system for conducting Earthquake Disaster Assessment.

Index Terms—Distributed geospatial analysis, Geospatial service, Web Processing Service, Services chain, Earthquake Disaster Assessment System

I. INTRODUCTION

Web-based GIS (also known as distributed GIS or Internet-based GIS) signifies a new type of GIS platform, which is easily accessible, easy-to-use, and easy-to-share. However, the development of Web-based GIS is facing many challenges, both classic and new. Interoperability is a classic challenge because each GIS system usually has its own proprietary data format. New challenges come from several fronts. The first one is how to incorporate GIS analytical functions with input data of spatial features and descriptive characteristics across Internet [1]. The second one is how to take advantage of large volumes of distributed data and alleviate bottlenecks of massive user interactions and Internet communication [2]. The third one concerns the adoption of development tool kits and development standards that are either open source or proprietary [3, 4]. These challenges have led to rapid growth of various solutions to Web-based GIS software and systems. For instance, enterprise information systems are evolving toward Service Oriented Architecture (SOA), while geospatial technologies are moving along the same direction [5]. The open and interoperable SOA is taking place of the traditional monolithic GIS.

Geospatial analysis (or process) using the technology of distributed geospatial database system becomes more and more important because of its capacity of supporting data interoperability in transmission [6]. Distributed geospatial analysis (DGA) has long been a technical challenge for several disciplines involved with GIS development. DGA can be defined as an execution of geographic data processing and analysis with geospatial data physically located in a varied number of spatial databases, not necessarily homogeneous, interconnected by a computer network [7]. Many efforts have been reported in the literature on how to address the practical issues of distributed geospatial analysis [4, 7, 8, 9]. Geospatial data are present in different formats and with numerous specifications, such as scale, projection, spatial reference, representation type, thematic, DBMS type, date, etc. The heterogeneity and complexity of the geospatial data complicates geo-
information retrieval across a network which is an important feature of DGA [10]. Therefore, interoperability embedded in DGA becomes a popular research area that is significant to geographic information science in recent decades, because it reduces the cost of operating geospatial databases by sharing resources and exchanging among different systems in a distributed computing environment [7].

In order to interoperate a multi-vendor and distributed GIS, there exists a need for publishing, sharing, and accessing spatial data from distributed geospatial databases interconnected through a communication network. A comprehensive review of current significant efforts that have been made to address this research challenge has been done by Peng [11]. There are plenty of active efforts of searching for solutions to improve accessibility to distributed geospatial data and publish the geospatial data over the World Wide Web among researchers and vendors [4, 11, 12]. In short, the technical problems can be induced to three main aspects: (1) Finding localization of the required geospatial data, (2) Enabling remote access to the geospatial data, and (3) Transforming different geospatial data representations [10]. Besides these technical issues, data sharing is also hindered by institutional barriers, which are not discussed in the paper. In order to solve the technical challenges, the conformance to standards is essential for distributed geospatial database systems. Complying with corporate and industry standards is a fundamental principle of good geospatial data management [7].

The atomic element of the SOA paradigm is the Services [13], which can be considered as the functions coupled with certain service standards and protocols that describe the pre- and post-conditions. Web services’ nativity provides a new mechanism for achieving interoperability among distributed data and applications. Many geospatial services were also designed and created during the transmission process, such as Web Map Service (WMS), Web Feature Service (WFS), Web Coverage Service (WCS) and the latest Web Processing Service (WPS). Today, these Open Geospatial Consortium (OGC) web services become the standardized interfaces that are used to publish geospatial processes including the capture, modeling, manipulation and analysis of geospatial data [14]. In general, “Geospatial Service” in this paper is defined as the web service that provides geospatial data access or geospatial data process which use XML as the standard data format and HTTP as the transfer protocol.

Each service interface standard has its own characteristic. WMS, which provides GIS mapping services, dynamically produces spatially referenced maps of client-specified ground rectangle from the geographic dataset, and returns pre-defined pictorial renderings of the maps in an image or graphics format (PNG, JPG, GIF, SVG, etc.). WFS and WCS, which provide GIS data services, retrieve geospatial features (GML) and coverage dataset (satellite imagery, DEM, TIN, etc.) that meet client-specified selection criteria [15]. These standardized service interfaces focus on the geospatial data accessing capabilities. Geospatial data processing capabilities have not been made normative until WPS was adopted as a standard by the OGC. The novelty of this WPS service is capable of resolving the key interoperability problem of chaining these services together.

In this study, we discuss how to use WPS to establish geospatial services chains so as to achieve a distributed geospatial analysis in three different approaches. A prototype Earthquake Disaster Assessment System is designed and developed as the illustration. The remainder of the paper is organized as follows: Section 2 gives an overview of WPS and the related natures, while Section 3 describes the proposed three approaches to create geospatial services chains with WPS. In Section 4, a case study of earthquake disaster assessment was developed to demonstrate the application of chaining geospatial services. Finally, Section 5 concludes the paper and discusses future work.

II. WEB PROCESSING SERVICE

Open Geospatial Consortium (OGC) launched Version 1.0 of the Web Processing Service (WPS) Interface Standard on February 22, 2008, which was first released in September 2005. This new standard aims to define a standardized interface that facilitates the publishing of geospatial processes and the discovery of and binding to those processes by clients [16]. This WPS interface standard allows executing geospatial processes on the web on the basis of XML/GML communication encoding.

A “geospatial process” means an algorithm, a model or a calculation that is made available at a service instance operating on spatially referenced raster or vector data. This process is not a chain unless it is created as an opaque chain. This point of view can be described as [17]:

Process P1 produces results R1-RN from input I1-Ii

There are three mandatory operations specified by WPS interface, namely GetCapabilities, DescribeProcess and Execute [16]. The response to a GetCapabilities request generally describes the names and abilities of the processes in the form of XML-based metadata. The response to a DescribeProcess
request returns the detailed information about the processes containing input and output parameters as well as their available formats. An Execute operation is used to run specified processes, which are also based on XML/GML.

Figure 1 shows a simple class diagram summarizing the WPS interface. The WPS interface class inherits the getCapabilities operation from the OGCWebService interface class [16]. Because WPS offers a generic interface, it does not limit itself to certain specific processes that are supported. Instead, each implementation of WPS defines a process it supports, as well as the associated inputs and outputs. WPS allows for the provision of input data in two different methods. Data can either be embedded in an Execute request, or referenced as a web accessible resource. In the former approach, WPS acts as a stand-alone service. In the latter method, WPS acts as middleware service by obtaining data from an external resource in order to run a process on the local implementation. Owing to this generic nature, WPS can be used to wrap other existing OGC geospatial standards (such as WMS, WFS and WCS) that are designed to provide geospatial services.

III. THREE APPROACHES FOR CREATING GEOSPATIAL SERVICE CHAINS WITH WPS

A service chain is a sequence of services where, for each adjacent pair of services, an occurrence of the first action is necessary for the occurrence of the second action [17, 22]. When no single service could meet a task requirement, we should combine several services together to do the work. Composing service chains could solve complex problems, which is regarded as the central concept of SOA framework. This possibility is also perceived as the solution to deal with geospatial data accessing and processing tasks by composing and reusing several simple geospatial web services [23].

A WPS process normally performs a specific geospatial service. Chaining of WPS processes facilitates the creation of repeatable workflows. Currently there are three methods for incorporating existing geospatial services (such as WMS, WFS and WCS), including WPS itself, into service chains.

A. Using BPEL to orchestrate a service chain that includes one or more WPS processes

WPS is compatible with Simple Object Access Protocol (SOAP) and Web Service Description Language (WSDL) [16], which is used by Business Process Execution Language (BPEL) for the purpose of describing web services. WPS can offers more sophisticated service chaining capabilities since it uses BPEL to orchestrate a chain including other services and WPS services themselves.

BPEL is an XML-based standard for defining process flows. This open standard makes it interoperable and portable across many environments. Today, IBM and Microsoft work jointly on making the BPEL and its supporting tools popular and standard [24]. BPEL is widely used in addressing the automation and orchestration of business processes in business networks. It can be used for describing a key aspect for a geospatial web service chain. It also provides a core of process description concepts that are
needed to define interactions among distributed geospatial processes [25]. This core of concepts is used both for defining the internal processes of a participant to an interaction and for publishing the external protocol that defines the interaction behavior of a participant among several distributed servers without revealing its internal behavior.

Figure 2 shows the collaboration diagram of using BPEL to orchestrate a service chain, and uses a data reduction process as the example. A client requests a geospatial data process by deploying a BPEL to describe it. WPS executes the request with the BPEL script calling a geospatial data process with reference to other web services as input parameters. Response from the WPS process only contains references to the results. And the results that are GML documents are stored on the WPS server that performed this process. This reference could then be forwarded to other WPS processes. The BPEL script also could send requests to other WPS processes. It passes the reference to the results from the former WPS process. Response from the BPEL script tells the client where to get the processed results. The client retrieves the GML documents from the WPS service.

To illustrate this method of building the chain using BPEL, Figure 3 shows a simplified snippet of the BPEL code, adapted from the example of Figure 2. Parameters have been omitted.

Figure 3. BPEL Snippet

Figure 4 shows a sample flow tree of the BPEL process. Each BPEL process can specify a tree. The leaves of the tree are the basic activities while the internal nodes correspond to the structured activities [26]. The lined arrows indicate links. Figure 4 is the process tree that illustrates Figure 3.
B. Using WPS Interface to design a sequence of web services including other services

The BPEL approach has its limitations. For instance, there is a problem with the transfer of binary data that is served in response to a WMS GetMap or WCS GetCoverage request, which cannot be orchestrated using the BPEL approach [27]. This difficulty can be overcome by using a WPS interface to develop a chain of geospatial services compliant with WSDL interface. WPS supports to use WSDL when it is required to help and support discovery and binding. WPS can be used in orchestrating a service chain, for the reason that a WPS service can be constructed to call other geospatial web services and WPS services as well, acting as the services chaining engine.

Figure 5 shows the collaboration diagram of using a WPS interface to orchestrate a service chain, and uses a buffer process as the example. A client requests the buffer process by calling WPS GetCapabilities operation, and this request is based on HTTP Get and Post. The WPS Interface executes the request with reference to other WPS services as input parameters. The response from other WPS processes only contains a GML document as the reference to the result. The WPS Interface can also request other geospatial web services, such as WFS, WMS, etc., by calling GET operations of these services. Response from the WPS interface tells the client to retrieve GML document as the result.

C. Simple cascading service chains created via WPS GET operation

The former method we discussed is commonly treated as the centralized chaining approach using WPS service as the central service to invoke other geospatial services. Furthermore, simple service chains can be orchestrated even via the WPS GET operation, since WPS also offers a service discovery mechanism that can be used without the overhead and complexity of WSDL [18]. In other word, the WPS interface also supports cascading chaining. This method enables geospatial services to exchange data directly because each individual service communicates with one another.

Figure 6 shows the collaboration diagram of a simple cascading service chain orchestrated via WPS GET operation, and uses a feature fusion process as the example. Similar as the former method, a client requests the fusion process by calling the WPS Get operation. This WPS service aims at aggregating more than one heterogeneous GML feature collections into one homogeneous GML feature collections by communicating with other geospatial services, such as WFS, for fusing GML features. The aggregated features can be requested as following:

IV. IMPLEMENTATION

Earthquake Disaster Assessment System is a research project in Institute for Geospatial Research and Education (IGRE) at Eastern Michigan University, in cooperation with China Data Center at University of Michigan and Spatial Information & Digital Engineering Research Center at Wuhan University. The goal of this project is to pilot a distributed GIS system capable of geospatial process and visualization in order to quickly provide critical assistance, assessment, and decision support after the earthquake disaster of magnitude 7.9 struck eastern Sichuan in China in May 2008 [28, 29].

The system aims at mapping, investigating, and assessing disastrous losses through the distributed data sharing and management [30]. This system is capable of supporting multimedia output functions, such as fast automated generation of assessment reports, which integrate maps, charts, tables, and texts, management documents, field videos and photos, etc. As members of this research group, we are mainly responsible for developing an open web service platform for the administration, application, visualization and interoperation of geospatial data (including both vector and raster data) using web services (WMS, WFS and WCS) chaining with WPS within the system.

The system prototype has been implemented based on the geospatial services described in the previous section. All the service chains are implemented in Java using the existing 52° North WPS framework. Figure 7 presents the architecture of Earthquake Disaster Assessment System. We developed a module called dynamic service discovery engine (DSDE) to locate the required data and to perform the selected spatial analyses, which is the key component of the system. DSDE obtains the XML description (or metadata) to identify if a particular data source or geospatial service is relevant to a particular spatial analysis. DSDE will search over all accessible servers starting at a specified web link to find available geospatial services to compose service chains. Distributed databases provide the required features and images data. Web Servers provide the GIS functionalities for clients to execute geospatial processes through the browser based platform which acts as a workbench for demonstration and research purposes.

Figure 8 shows a simplified snippet of WSDL code, which defines a WPS buffer process service.

Figure 9 shows the geospatial process application, which allows users to perform the buffer function in order to determine the earthquake affected region and calculate the number of deaths and missing persons hit by the earthquake. The death information is extracted through clipping with the earthquake affected region by using the buffer operation, and then summing up the population information in the buffer zone by using the aggregation process.

Figure 7. Architecture of Earthquake Disaster Assessment System
Figure 8. a WSDL Snippet

```xml
<!-- Concrete Binding Over HTTP -->
<wSDL:binding name="EarthquakeProcessHttpGet" type="tns:EarthquakeProcessHttpGet">
  <http:binding verb="GET"/>
  <<wSDL:operation name="GMLBuffer">
    <http:operation location="GMLBuffer"/>
    <wSDL:input name="GMLBufferRequest"/>
    <wSDL:output name="GMLBufferResponse"/>
  </wSDL:operation>
</wSDL:binding>
```

Figure 9 Buffer Process in Earthquake Disaster Assessment System

<table>
<thead>
<tr>
<th>Tab 1 Input Parameters</th>
<th>Tab 2 Result: Deaths from 12 May to 29 May, 2008 in the Buffer Zone</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>CNTY.PY</th>
<th>CNTY.HE</th>
<th>DEATH_TOTAL</th>
<th>%_2000.Pop</th>
<th>AMM.00</th>
<th>PROV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wenchuan Xian</td>
<td>21.45</td>
<td>11939</td>
<td>11935</td>
<td>14.2196</td>
<td>Sichuan</td>
</tr>
<tr>
<td>Ya'an</td>
<td>21.45</td>
<td>11939</td>
<td>11935</td>
<td>14.2196</td>
<td>Sichuan</td>
</tr>
<tr>
<td>Mianyang</td>
<td>11939</td>
<td>11935</td>
<td>14.2196</td>
<td>Sichuan</td>
<td></td>
</tr>
<tr>
<td>Liangshan</td>
<td>11939</td>
<td>11935</td>
<td>14.2196</td>
<td>Sichuan</td>
<td></td>
</tr>
<tr>
<td>Li County</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinasu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pengzhou</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yibin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meishan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

© 2010 ACADEMY PUBLISHER
V. CONCLUSIONS AND FUTURE WORK

This paper looks into how to make geospatial services chains using WPS in order to achieve the distributed geospatial analysis. This study comes up with three methods to orchestrate the chains and successfully developed a prototype web platform for deploy the services chains in an Earthquake Disaster Assessment System. Using BPEL in accompany with WPS is dependent on WSDL documents. Although a WSDL document seems redundant for the reason that the WPS DescribeProcess operation response contains some same information as a WSDL description, reusing services can be better supported by applications like BPEL designer that make it possible to orchestrate single services using graphical tools. The use of the WPS interface to orchestrate a service chain can cover the shortage when facing the problem with the transfer of binary data, which cannot be orchestrated using the BPEL approach. Furthermore, some simple processes can be chained by using cascading services via WPS GET operation. The geospatial services chain architecture could be centralized or cascaded depending on which process the geospatial services deal with.

Since the assessment of earthquake disaster needs a wealth of data from different areas, including population, land use, environment, etc., we designed the assessment platform, which relies to multiple distributed computing environments and data sources. One goal of OGC geospatial data services is to offer the standard interface for access to distributed data sources based on SOA principle. BPEL is an important tool enabling services interoperable and portable across many environments. The Earthquake Disaster Assessment System offers access to many distributed data sources using geospatial services chains orchestrated by standardized service interfaces centered on WPS. Through the geospatial services chains we discussed, combining data sources and data processes can be tailored to the users’ needs. This project, in fact, provides the middleware to bridge the gap between data providers and data users.

The future work of our study will concentrate on realizing more sophisticated geospatial analyses through the methods discussed above and complete the Earthquake Disaster Assessment System which is still in development.

REFERENCES


Xiaoliang Meng was born in 1981, Wuhan, China. He got his Bachelor of Engineering from Wuhan University in 2004. He now is a candidate of Ph.D in State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing in Wuhan University.

From 2008 to 2009, he was a visiting scholar and staff in Institute for Geospatial Research and Education, Eastern Michigan University. His main research interests include WebGIS, Spatial-Informatics and Digitalized Technology.

Yichun Xie got his diploma in Geography from Anhui Normal University in 1978. He received Master Degree of Arts in Urban Studies / Urban Planning from The University of Akron in 1991 and Ph.D in Geography from State University of New York at Buffalo in 1994.

He is a professor in Department of Geography and Geology and the director of Institute for Geospatial Research and Education, Eastern Michigan University. Concentrations: GIS, Urban Modeling in GIS Environment, Spatial Statistics.

Fuling Bian received her diploma from Wuhan Technical University of Surveying and Mapping in 1964.

She is a professor and the director of Research Center of Spatial Information and Digital Engineering in the International School of Software, Wuhan University, Director of GIS (China) Association Training & Education Committee. Research Areas: Spatial Information & Digital Techniques.