



## **Spatial Web Services Mining Privacy**

**A.V. Krishna Prasad**

Associate Professor

Department Computer Science  
Matrusri Institute of P.G.Studies

India

kpvambati@gmail.com

**Dr S.Ramakrishna**

Professor

Department Computer Science S.V.U campus

India

drramakrishna@yahoo.com

**Dr.D.SravanKumar**

Principal and Professor

CSE KITE Women's College of Professional Engg Sciences

India

dasojusravan@yahoo.co.in

### **Abstract:**

The Web Services paradigm promises to enable rich flexible and dynamic interoperation of highly distributed, heterogeneous network enabled services. The Idea of Web Services mining that it makes use of the findings in the field of data mining and applies them to the world of Web Services. Web Services Security represents a key requirement for today's distributed interconnected digital world and for the new generations and Web 2.0 and Semantic Web. To date, the problem of security has been investigated very much in the context of standardization efforts. Personal judgments are made usually based on the sensitivity of the information and the reputation of the party to which the information is to be disclosed. On the privacy front, this means that privacy invasion would let more quality and sensitive personal information. In this paper, we had implemented a case study on integrated privacy issues of Spatial Web Services Mining. Initially we improved privacy of Web Services Mining using P3P. Finally, we validated this privacy issues by implementing a Web 2.0 Mapping Microsoft .NET application using a GPS receiver system.

### **1. Introduction**

Web Services are becoming more and more complex, involving numerous interacting business objects within considerable processes. Web Services are self-contained, self-describing, modular applications that can be published, located, and dynamically invoked across the Web [1]. In order to fully explore Web Service business opportunities while ensuring a correct and reliable modeling and execution, analyzing and tracking Web Services interactions will enable them to be well understood and controlled. The Web Services paradigm promises to enable rich, flexible and dynamic interoperation of highly distributed, heterogenous network enabled services. The idea of Web Services Mining that makes use of the findings in the field of data mining and apply them to



the world of Web Services and Service Oriented Architectures.

With the advent of evolution of WWW from Web searching to Web Mining, many business applications that utilize data mining techniques are extracting useful business information on the web. The web is transforming from a web of data to a web of both semantic data and services. This trend is providing us with increasing opportunities to compose potentially interesting and useful services from existing services. Using open web API's like Google (Spatial Google Map API version 2), Amazon and eBay we can implement data mining algorithms like Apriori for various business applications like CRM with ease. The emerging concept of Semantic Web Services aims at more sophisticated Web Services technologies: on basis of Semantic Description Frameworks, Intelligent mechanisms are envisioned for Discovery, Composition, and contracting of Web Services. The Semantic Web Services on basis of the Web Services Modeling Ontology (WSMO) and related initiatives. Ontology is a representation of a set of concepts within a domain and the relationships between those concepts. In other words, ontology defines the vocabulary of a domain.

Web Service Security represents a key requirement for today's distributed interconnected digital world and for the new generations, Web 2.0 and Semantic Web. To date, the problem of security has been investigated very much in the context of standardization efforts; these efforts however, have dealt mainly with adopting existing security techniques such as encryption for use in Web Services. The standards also focused on addressing the problem of security interoperability thought the development of standard formats for security assertions, tokens and credentials. Interoperability is certainly an important issue for Web Services in that easy and flexible service composition requires that security relevant information be seamlessly transmitted across different services. Often sensitive information is exchanged among users, Web Services, and software agents. This exchange of information has highlighted the problem of privacy. A large number of strategies employed to preserve people's privacy require users to define their respective privacy requirements and make decisions about the disclosure of their information. Personal judgments are made usually based on the sensitivity of the information and the reputation of the party to which the information is to be disclosed. In the absence of a comprehensive privacy preserving mechanism, no guarantees about the information disclosure can be made. On the privacy front, this means that privacy invasion would net more quality and sensitive personal information [2].

## 2. Web Services Mining

The main purpose of Web Services Mining is to apply some mining and conformation techniques as a feedback loop to improve the Web Services concepts in particular to abstract process and the orchestration model. [3] Generally, formal traditional approaches develop using their Web Service modeling formalisms, a set of techniques to analyze the composition model and check some properties. Figure 1. (a), proposes a formal framework for modeling, specifying and analyzing the global behavior of Web Services compositions. This approach models web services by Mealy machines (finite state machines with input and output). Figure 1. (b) Proposes a log based approach that starts from the execution log to mine and analyze web services behavior. This mining step will be used in future to improve the web services modeling. [4]

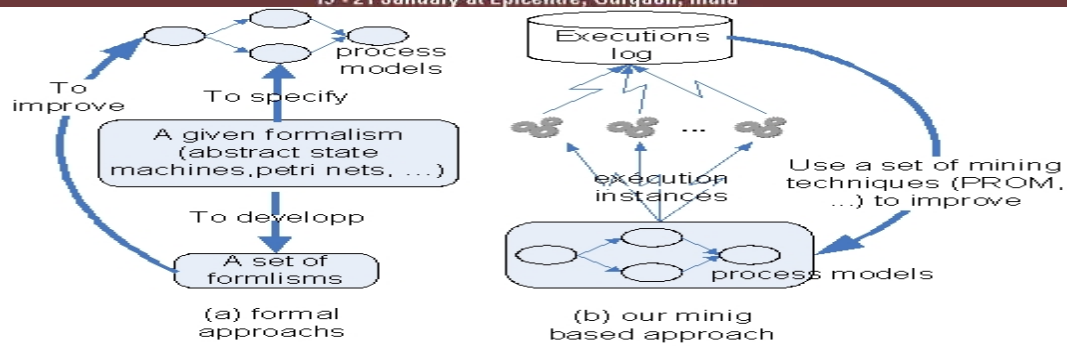


FIGURE 1. LOG BASED APPROACH COMPARED TO BASED FORMALISM APPROACHES

### 3. Privacy Issues

#### 3.1 Overview of Web Services Privacy

Privacy- Preserving Data Mining (PPDM) refers to data mining techniques developed to protect sensitive data while allowing useful information to be discovered from the Web.[5] Any solution to privacy of Web Service should support two important functions that are complementary and very crucial: How to communicate the privacy policies of an organization to interested parties, such as the individuals whose data are being collected by the organization; and How to enforce such privacy policies within the organization. The most well-known approach to the communication of privacy policies is the W3C's Platform for Privacy Preferences Project (P3P). P3P enables Web sites to encode their data collection and data use practices in a machine-readable XML format, known as P3P Policies. In short, P3P and APPEL (A P3P Preference Exchange Language) are designed to enable users to play an active role in controlling their private information. Refer to TABLE 1 which explains Three Layer Privacy Responsibility Framework and Engineering Issues.[6]

TABLE 1. THREE LAYER PRIVACY RESPONSIBILITY FRAMEWORK AND ENGINEERING ISSUES

Privacy Spheres	Where data is stored	Engineer's responsibility	Engineering Issues
User Sphere	Users desktop Personal, computers, Laptop, Mobile phones, RFID chips	Give users control over access to themselves (in terms of access to data and attention)	What data is transferred from the client to a data recipient? Is the user explicitly involved in the transfer? Is the user aware of remote/or local application storing data on his system? Is data storage is transient or persistence?
Joint	Web	Give users	Is the user fully aware of how his data is



Sphere	Service providers, servers and databases	some control over access to themselves (in terms of access to data and attention) Minimize users future privacy risks	used and how he controls this?
Recipient Sphere	Any data recipients, servers and databases of network providers, service providers or other parties with whom data recipient shares data	Minimize users future privacy risks	What data is being shared by the data recipient with other parties? Can the user expect or anticipate a transfer of his data by the recipient? Is personal data adequately secured? Is data storage transient or persistence? Can the processing of personal data be foreseen by the user? Are there secondary uses of data that may not be foreseen by the user? Is there a way to minimize processing (e.g. by delegating some pre-processing to User Sphere)?

Each P3P policy is specified by one POLICY element that includes the following major elements:  
As an example, consider the following P3P Statement in Box 1:

**BOX 1. EXAMPLE OF A P3P STATEMENT**

```
<STATEMENT>
<PURPOSE> <admin required="opt-in"/> </PURPOSE>
<RECIPIENT> <public /> </RECIPIENT?
<RETENTION> <indefinitely/> </RETENTION>
<DATA-GROUP>
<DATA ref="user.home-info.postal"> </DATA>
<DATA-GROUP>
<STATEMENT>
```

Often, sensitive information is exchanged among users, Web Services and software agents. This exchange of information has highlighted the problem of privacy[7]. A large number of strategies employed to preserve people’s privacy requires users to define their respective privacy requirements and make decisions about the disclosure of their information. Personal judgments are usually made based on the sensitivity of the information and the reputation of the party to which the information is to be disclosed. On the privacy front, this means that privacy invasion would net more quality and sensitivity personal information.

**4. Implementations and Validations**

**4.1.1 Securing Privacy for Web Services Mining**

A factor that limits the application of web mining is the need to protect privacy of users.[8] Web Mining uses data that are available on the web anyway but the use of Data Mining makes it possible to induce general patterns that can be applied to personal data to inductively infer data that should remain private. Privacy has become a major concern because the news contains numerous stories of personal information misuse. One of the misuses of personal information is identity theft, but that's by no means the largest misuse. Many users also feel that gathering personal information for marketing purposes without permissions and full disclosure of the requestor will use the information is also a major misuse of personal information. The proliferation of spy ware blockers shows that users are becoming aware of covert attempts by some Web sites (including pop-up advertisements) to steal data from their systems. Personal information covers a range of topics today. Most developers recognize that name, address, telephone number, and other personally identifying information is private. However, users don't want developers to know a lot of information that some developers see as belonging to the public domain.

Web Services Mining using open Web Google API, can help developers and designers, and address privacy issues by providing privacy policy.[9] Many users are beginning to request this functionality (privacy policy) and they may not visit our site more than once if we don't provide it. In addition it's also important to personalize the user's experience with our site so we don't have to ask the same questions every time the users visit the site. Personalization helps users have a better experience and improves user efficiency. Refer to Box 2 which contains the Contents of P3P.xml file.

#### BOX 2. CONTENTS OF P3P.XML

```
<META omens="http://www.w3.org/2002/01/P3Pv1">
<POLICY-REFERENCES>
  <POLICY-REF about="http://www.mipgs.ac.in/PrivPolicy.p3p#GeneralPolicy">
    <INCLUDE>/*</INCLUDE>
    <COOKIE-INCLUDE/>
  </POLICY-REF>
</POLICY-REFERENCES>
</META>
```

Figure 2 displays the Privacy Demonstration with user name displayed, with options for save cookie and read cookie.

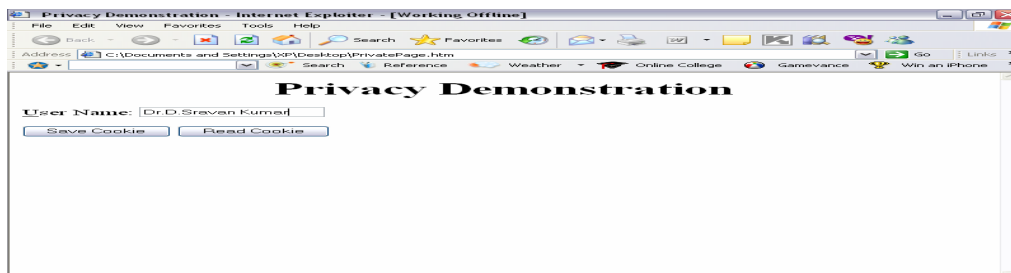


FIGURE 2. PRIVACY DEMONSTRATION USING P3P



#### **4.1.2 Implementation of Spatial Web Mapping .NET application using GPS Receiver and its Privacy issues**

We implemented a case study: communication is one of the oldest mechanisms for devices to communicate with each other. Starting with the IBM PC and compatible computers, almost all computers are equipped with one or more serial ports and one parallel port. [10 -14] As the name implies, a serial port sends and receives data serially, one time at a time. In contrast, a parallel port sends and receives data eight bits at a time, using eight separate wires. Despite the comparatively slower transfer speed of serial ports over parallel ports, serial communication remains a popular connectivity option for devices because of its simplicity and cost-effectiveness. Although consumer products today are using USB connections in place of serial connections, still a lot of devices use serial ports as their sole connections to the outside world. This tool communicates with other serial devices using the new SerialPort class available in the .NET Framework 2.0 and the .NET Compact Framework 2.0 application implementing communicate with a GPS receiver and then extract the useful data for displaying the current location on a map.

#### **Creating a Mapping Application using a GPS Receiver and Microsoft Virtual Earth:**

This application builds a mapping application using Microsoft Virtual Earth and a GPS receiver. Microsoft Virtual Earth (VE) is a map and search system comprising maps, aerial images, business directories, and so on. Using VE, we can search for businesses and addresses, as well as ask for directions.(<http://local.live.com>) To developers, Microsoft exposes the VE Map control that allows us to embed VE maps into our own application. We can then build our own custom solution using the mapping services provided by VE. The VE Map control consists of a JavaScript page and a style sheet. The VE Mapcontrol is hosted at <http://dev.virtualearth.net/mapcontrol/v3/mapcontrol.js> The CSS is located at <http://local.live.com/css/MapControl.css>.

In this tool, we will build a windows application that allows us to track our location in real time by getting positioning information from our GPS receiver and then feeding the data to VE. For this to work, having Internet connectivity is a must, possibly through wireless hotspots or General Packet Radio Service (GPRS). In addition this application is also able to plot traveled paths by extracting GPS data from a file. For example, we can track our travelled path by saving all the positioning information using GPS receiver (a pocket PC or notebook). When we return to the office, we can upload the saved data to the application, and it will display the path we have travelled using VE.

This application can use a Bluetooth-enables GPS receiver (such as Holux GPSlim 236) that connects to our computer (such as Pocket PC or computer) via Bluetooth (through a serial communication). We can view the maps of a particular location as well as pan and zoom into specific areas. Besides viewing the roads of an area, we can also choose to view the aerial images. Depending on the locations we have specified, we can also zoom into specific areas and view the highly detailed map. Refer to Figure 3, which contains the detailed sequence diagram of the entire application. Refer to Figure 4 and Figure 5, which contains the screen shots of the application.



**Map India 2010**

19 - 21 January at Epicentre, Gurgaon, India

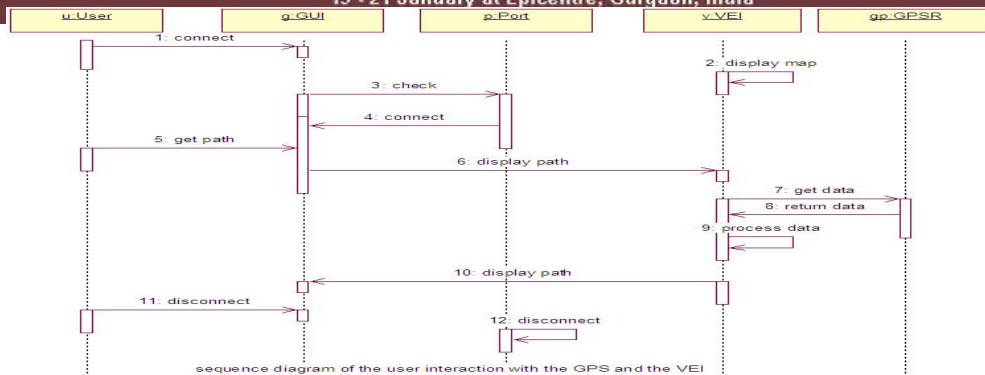


FIGURE 3. DETAILED SEQUENCE DIAGRAM OF THE APPLICATION

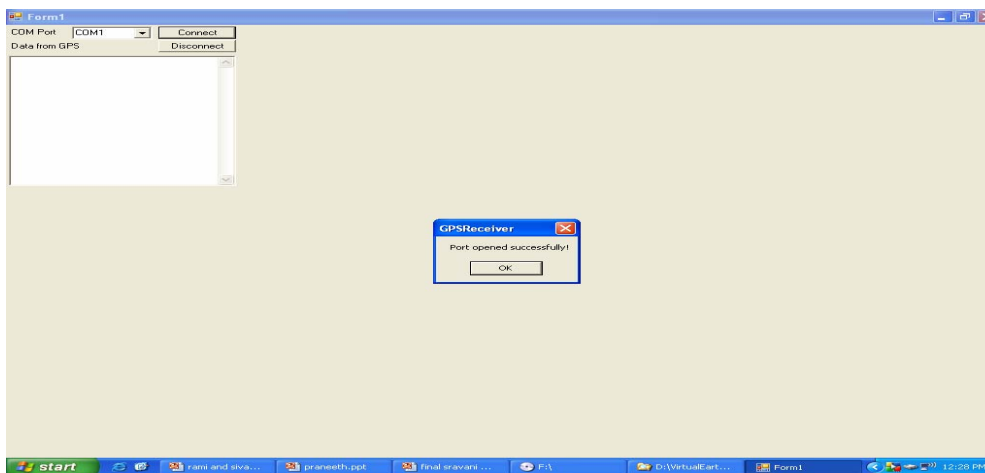


FIGURE 4. GPS Receiver Simulator connecting to application through Port1

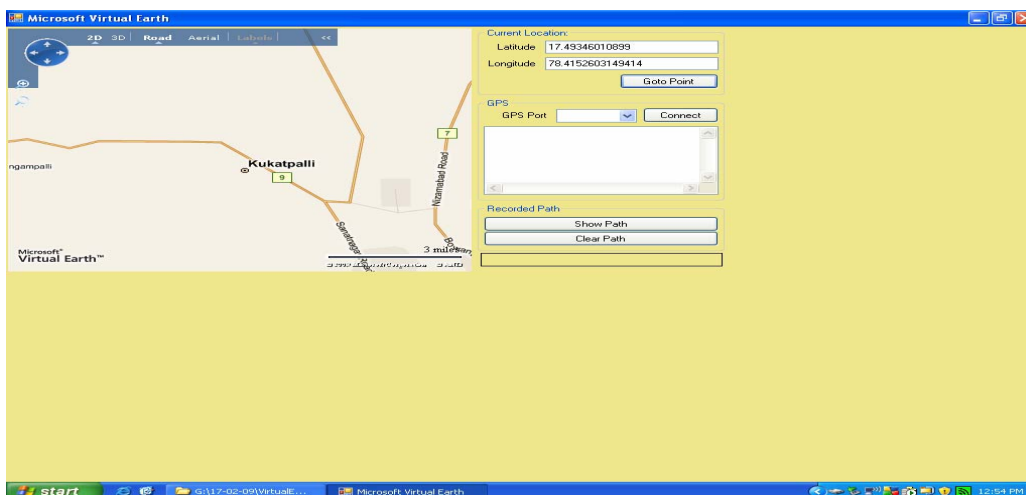




FIGURE 5. Mapping Application with JNTU Virtual Earth working with GPS Receiver

## 5. Conclusion

In this paper, we discussed various privacy issues integrating web services mining, geospatial web by implementing a case study. Further work pertaining to Privacy Preserving data mining includes data perturbation and SMC (Secure Multiparty Computation). Privacy enforcement in web services and semantic web using decentralized models like distributed, registry-based and peer to peer reputation models. Finally geospatial privacy includes geoRBAC. The widespread deployment of location based services and mobile applications, as well as the increased concern for the management and sharing of geographical information in strategic applications like environmental protection and homeland security, have resulted in strong demand for spatially aware access control systems. These application domains impose interesting requirements on access control systems. In particular, the permissions assigned to users depend on their position in a reference space; users often belong to well-defined categories; objects to which permissions must be granted are located in that space; and access control policies must grant permissions based on object locations and user positions.

## 6. Acknowledgements

The authors wish to thank the following members for implementing these concepts: D.Shravani, P.Praneeth and Ch.Venkatabhilash.

## 7. References

- [1] Hsinchun Chen, Xin Li, Michael Chau, Yi-Jen Ho and Chunju Tseng, "Using Open Web APIs in Teaching Web Mining," *IEEE Transactions on Education*, vol. 52 no. 4, pp. 482–490, November 2009.
- [2] Ravi Kothuri, Albert Godfrind and Euro Beinat, *Pro Oracle Spatial for Oracle Database 11g* Apress publishers, 2007.
- [3] Sarah Spiekermann and Lorrie Faith Cranor "Engineering Privacy," *IEEE Transactions on Software Engineering*, vol. 35 no. 1, pp. 67—82, January-February 2009.
- [4] Lixin Fu, Hamid Nemati and Fereidoon Sadri *Privacy-Preserving Data Mining and the Need for Confluence of Research and Practice* IGI Global, 2009.
- [5] Shinsaku Kiyomoto, Keith M.Martin, and Kazuhide Fukushima," Privacy-Aware Location Database Service for Granular Queries", *WISTP LNCS 5746*, pp. 28—37, IFIP 2009.
- [6] Rahul katarya and Dr.Satbir Jain, "Web Search with mobile incorporation Geographic information system" ; *CSI Communications*, Vol No 33, Issue No.8, November 2009.
- [7] Neil daswari, Christoph kers and Anita kesavan, "Foundations of Security, What every Programmer needs to know" A press publishers, 2007.
- [8] Christopher C.Miller, "A Beast in the Field: The Google Maps Mashups as GIS/2", *Cartographica, The International Journal for Geographic Information and Geo Visualization*, Volume 41, issue 3, pp. 187 – 199, 2005.
- [9] Laszlo Zentai, "Application of Web 2.0 in Cartographic education: Is it time for Cartography 2.0?" 2006.
- [10] Johnny Long, "Google Hacking for penetration testers", *SYNGRESS*, 2005
- [11] Zao Liu et.al, "Implementing a Caching and Tiling Map Server: a Web 2.0 Case Study", 2006.
- [12] Ravi Kothuri, et.al., "Pro Oracle Spatial for Oracle Database 11 g", *Apress*, 2007.
- [13] Jon Arking, "Professional Windows Live Programming", *Wrox Publication*, 2008.
- [14] Yan Huiwu et.al., "The Research and Application of the Web Map Issuing Technology",



**Map India 2010**  
19 - 21 January at Epicentre, Gurgaon, India

Research center for Environmental Engineering & Management, Shenzhen, 2008