Ain Shams University
Faculty of Engineering
Computer and Systems Engineering Department

Protecting GIS Data Using a Distributed Transparent Mechanism

Dissertation Submitted for the
Partial Fulfillment of PhD Degree

Submitted By
Eng. Ashraf Farouk Tammam
B.Sc. in Computer Engineering, Military Technical College, 1994
M.Sc. in Computer and Systems Engineering, Ain Shams University, Faculty of Engineering, 2004

Supervised By
Dr. Yasser Hesham Dakroury
Professor of Computer Engineering
Computer and Systems Engineering Department,
Faculty of Engineering, Ain Shams University

Dr. Ismail Abdel Ghafar Farag
Professor of Computer Engineering
Computer and Operation Research Department,
Military Technical College

Cairo 2011

جامعة عين شمس
كلية الهندسة
قسم هندسة الحاسبات والنظم

حماية بيانات نظم المعلومات الجغرافية بإستخدام آلية شفافة موزعة

رسالة مقدمة للحصول على درجة دكتوراه الفلسفة فى هندسة الحاسبات والنظم

مقدمة من

مهندس / أشرف فاروق تمام

بكالوريوس هندسة الحاسبات - الكلية الفنية العسكرية - 1994
ماجستير هندسة الحاسبات والنظم - كلية الهندسة - جامعة عين شمس - 2004

تحت إشراف

دكتور / ياسر هشام دكرورى
أستاذ بقسم هندسة الحاسبات والنظم
كلية الهندسة - جامعة عين شمس

دكتور / إسماعيل عبد الغفار فرج
أستاذ بقسم هندسة الحاسبات و بحوث العمليات
الكلية الفنية العسكرية

القاهرة 2011

ABSTRACT

Ashraf Farouk Tammam, Protecting GIS Data Using a Distributed Transparent Mechanism, PhD Dissertation, Ain Shams University, Faculty of Engineering, Computer and Systems Engineering Department, 2011.

Geographic Information System (GIS) became much more than a mapping tool. It became a mainstream within the enterprise and organizations serving their geographical data on the networks. GIS involves the use of more complex explorations, analysis and visualizations of the earth we are living on its surface. Today, it is used in many military and commercial applications.

GIS data is one of the most important GIS components. It is very expensive and sometimes contains confidential information. Protecting the GIS data is a very critical task especially when this data is related to military and confidential applications.

In this research, we propose the building of a framework to protect the GIS data from being copied or accessed by unauthorized users. The proposed framework combines symmetric cipher, asymmetric cipher and digital watermarking to achieve its goal. Using this framework, the GIS data is encrypted using symmetric cipher and distributed to many servers. Only authorized users are able to access this data and see the information inside it. The symmetric cipher key is exchanged between the authorized users' desktops and the servers containing the encrypted GIS data using asymmetric cipher and digital watermarking.

The effectiveness and advantages of the proposed framework are objectively evaluated using two data sets of Environmental Systems Research Institute (ESRI) shapefile format. The effect of using cryptography and digital watermarking on the time required to display the shapefiles is studied. We also study all types of failures that can occur and introduces solutions for them.

At the end, we show that the proposed framework can be the basis for building many GIS applications by applying it to the Automatic Vehicle Location (AVL) application.

TABLE OF CONTENTS

CHAPTER (3)

CRYPTOGRAPHY

CHAPTER (4)

GIS DATA PROTECTING FRAMEWORK

CHAPTER (5)

IMPLEMENTATION AND TESTING OF DTEDPM

CHAPTER (6)

APPLYING DTEDPM TO AUTOMATIC VEHICLE LOCATION

CHAPTER (7)

CONCLUSIONS AND FUTURE WORK

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

GIS plays an important role as an area of Information Technology (IT). It is used in many fields and applications. One of the most important GIS components is the data. GIS data is expensive and sometimes contains confidential information. GIS data providers have a problem protecting their data from being copied or accessed by unauthorized users.

GIS data providers have two scenarios to provide their data to users. In the first one, the data providers put the GIS data on their servers and build web applications for users to be able to access this data. Many techniques can be used to secure the network infrastructure and protect servers containing the GIS data from attackers. However, these techniques alone cannot achieve the full protection. If some one succeeds to get access to these servers, he will be able to get a copy of this data and distribute it.

In the second one, the data providers build standalone applications and provide the application with the GIS data to their users. In this case, users will be able to copy this data and distribute it without taking any permission from its owner. Most of the researches in this area focus on how to copyright the GIS data using the concept of digital watermarking. However, watermarking has a lot of limitations and if it is used alone, it will not able to achieve the goal.

Both the previous techniques share an important limitation. Whether the GIS data is stored on the servers of the GIS data providers or watermarked and distributed to users, the data is stored in a non encrypted format. Any one succeeds in getting a copy of this data will be able to access all the information stored in it using any GIS application. This limitation is not accepted when this data is related to military and confidential applications.

1.2 Proposed Approach

In this dissertation, we propose a framework or to protect the GIS data from being accessed by unauthorized users. It also protects the GIS data from being copied by authorized and non authorized users.

The proposed framework can protect the GIS data used by both desktop and web GIS applications. It can also be applied to any raster and vector GIS spatial data formats. However, we will focus on protecting the vector GIS data used by the desktop GIS applications.

The proposed framework uses a combination of Advanced Encryption Standard (AES) symmetric cipher, Rivest Shamir Adleman (RSA) asymmetric cipher and digital watermarking to protect the GIS data. It also uses a Database Management Systems (DBMS) to store the information of all the users and desktops authorized to access the encrypted GIS data.

The proposed framework performs its goal using two components. The first component is the offline component and it encrypts the GIS data on a standalone desktop before being stored on many network connected servers or distributed to users. The second component is the online component and it gives the authorized users the ability to access the encrypted GIS data by decrypting it to the memory of their desktops and displayed it from there. If both authorized and unauthorized users try to copy the GIS data, they will take an encrypted version of it. This encrypted version cannot be used in any GIS application.

We try to achieve the high availability of the proposed framework by avoiding a single point of failure. This can be done by distributing the different modules of its online component and encrypted GIS data to many servers. We also try to make it applicable to be used in a combination of many computing environments at the same time.

We put into consideration making the proposed framework applicable to both single user and multiuser GIS desktop applications. In the single user applications, all the modules of the online component are installed on a single desktop while in the multiusers applications, the client/server computing model is applied.

We study all types of failures that can occur when the client server computing model is applied and introduces solutions for them.

1.3 Dissertation Outline

The rest of this dissertation is organized as follows:

Chapter 2, **GEOGRAPHIC INFORMATION SYSTEM**, gives an overview of GIS; including GIS definition and components, spatial data types used by GIS, why it is important to georeferene the GIS data, real world coordinate systems, three dimensional GIS, remote sensing and GIS applications.

Chapter 3, **CRYPTOGRAPHY**, describes some terms and techniques related to cryptography; including cryptography basics, symmetric and symmetric ciphers with a detailed example for each type and the different methods used to authenticate messages.

Chapter 4, **GIS DATA PROTECTING FRAMEWORK**, describes the proposed framework and how it protects the GIS data; including the importance of protecting the GIS data and the related work in this area, our assumptions, considerations and the details of the different modules of its components.

Chapter 5, **IMPLEMENTATION AND TESTING OF DTEDPM**, presents the full details of the proposed framework implementation; including implementation overview, full description of its four programs, database architecture, dealing with various types of failures, results of testing it using two GIS data sets and the analysis of these results.

Chapter 6, **APPLYING DTEDPM TO AUTOMATIC VEHICLE LOCATION SYSTEM**, shows how the proposed framework can be applied to AVL application; including AVL components and the implementation details of the modules built on it to provide the tracking service.

Chapter 7, **CONCLUSIONS AND FUTURE WORK**, summarizes the dissertation and outlines ideas for the future work in protecting GIS data area.

CHAPTER 2

GEOGRAPHIC INFORMATION SYSTEM

In this chapter we will show the importance of using GIS and its main components. We will study the two types of spatial data used in GIS (raster and vector), their common formats and the sources of getting such data. We will study the different real world coordinate systems used to georeference the GIS data. We will take a look at the remote sensing technology and how it is integrated with GIS. Finally, we will mention some examples of the fields where the GIS are applicable.

2.1 What Is GIS?

Maps and geographic information are essential to how we know the world. The endless complexity of the world around us presents us with a multitude of choices about what to represent and how to represent that complexity in the form of maps and as geographic information [27]. GIS involves many issues and choices for doing this task.

GIS is an integrated software package specifically designed for use with geographic data that performs a comprehensive range of data handling tasks. These tasks include data input, storage, retrieval and output, in addition to a wide variety of descriptive and analytical processes [65].

Data input is the most important and difficult task. It is the process of converting a real world area to a set of layers (raster and vector) capable to be managed by GIS as shown in figure 2.1.

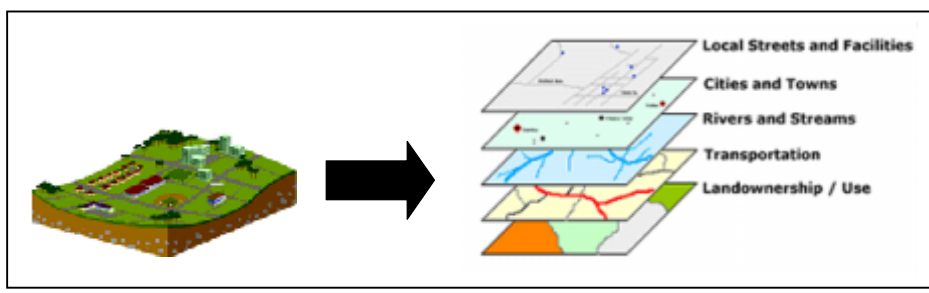

Figure 2.1: Converting Real World Area to Set of Layers

2.2 GIS Components

Enterprise GIS has five main components as shown in figure 2.2. These components are [44]:

- People: the users of the system.
- Applications: the processes and programs they use to do their work.
- Data: the information needed to support those applications.
- Software: the core GIS software.
- Hardware: the physical components on which the system runs.

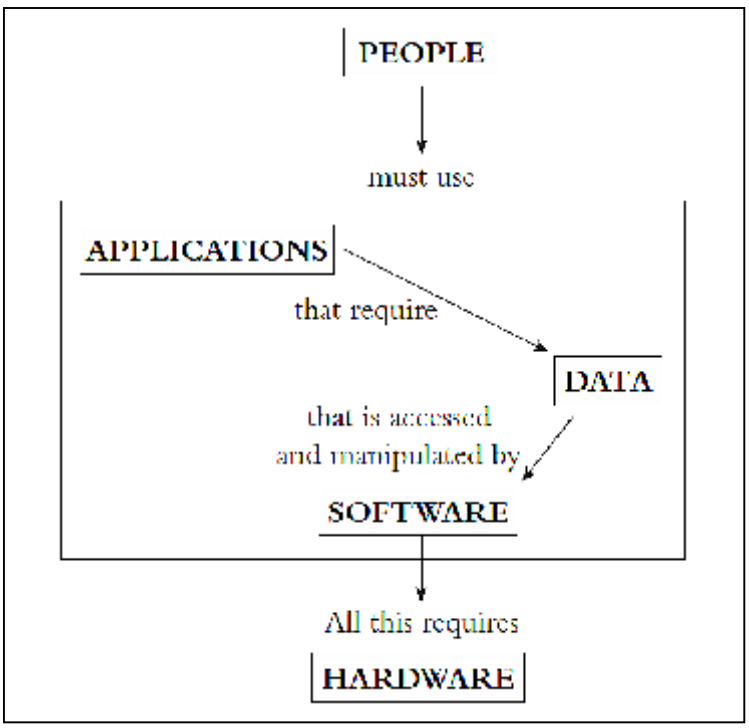

Figure 2.2: Components of an Enterprise GIS [44]

The triad in the center of figure 2.2 of applications/data/software represents the core of the information system. Ideally, it should work regardless of who the people are (if the application is well designed) and be flexible enough to work on whatever innovations in hardware come along, even in the absence of any hardware. That is why hardware is at the bottom, the least important element of the information system.

2.2.1 People

The people are the most important component, although some would argue for the data. Information systems, geographic or not, spring from the needs of people in organizations to do work, answer questions and generally interact with the world and the people and organizations in it. An information system is supposed to support the work, to make it quicker to do with more consistent results and to provide high levels of confidence in the output. The process of design and implementation of a GIS begins with people and their needs and ends up with applications in the hands of people who do the work. The entire system exists to support them and their tasks.

2.2.2 Applications

The applications come next in the hierarchy because they define the work that needs to be done. In organizations people need to create all kinds of reports, make all sorts of decisions and generally apply their skills so the work gets done. The processes they develop to do these things are the applications. Some applications are routine and get done multiple times a day, whereas others are less routine but get done with some regularity and then there are specific analytical applications that might have to be accomplished only rarely or even just once. The applications arise out of the mission and goals of the organization. In any information system you need to know what applications the system will be expected to support.

2.2.3 Data

Applications require data to work. You cannot generate a map of sales potential or customer locations without the appropriate data tables necessary to create that type of output. These tables will reside in a database (possibly more than one) and the system will require software to access, manage and manipulate the data so that the application can generate a useful product.

2.2.4 Software

The software is suite of products that provides specialized toolsets for organizations to share and discover spatial resources. Today, there are many GIS products available in the market. Some of these products can be used by any kind of people while others required to be used by programmers to develop applications on the top of it. One of the companies that have a large variety of GIS products is ESRI [22].

Another source of GIS software is the open source software and libraries. There are many open source GIS software covering both desktop [53] and web based [14-71] technologies. Developers can use this software as it is or customize it to build their GIS applications with a minimum effort and free of charge.

2.2.5 Hardware

The hardware required to run the GIS application depends on the complexity of application and its number of users. Also, it depends on the size and types of the used GIS data. It can be ranged from a normal desktop to large servers.

2.3 GIS Data Types

GIS uses two data types to translate a real world area or a paper map into digital form. These types are raster and vector. Most GIS software today handles both types of data.

2.3.1 Raster Data Type

The raster data [56] is a rectangular array of equally spaced cells, which taken as a whole represent thematic, spectral, or picture data. It can represent every thing from qualities of a land surface such as elevation or vegetation, to satellite images, scanned maps and photographs. Figure 2.3 shows some examples of the raster data.