



Cambridge
Conference

www.cambridgeconference2003.com

Ordnance Survey
Romsey Road
SOUTHAMPTON
SO16 4GU
United Kingdom

Evaluation tests performed over a proposed anti-piracy system for digital vector datasets

Colonel A Bacci
Director
Military Geographic Service
Uruguay

Dr C Lopez
TheDigitalMap Ltd
Uruguay

Paper 5B.2

Evaluation tests performed over a proposed anti-piracy system for digital vector datasets

Colonel A Bacci
Director
Military Geographic Service
Uruguay

Dr C Lopez
TheDigitalMap Ltd
Uruguay

Recent achievements at the SGM in Uruguay includes the completion of the digitalization of urban areas for the whole country (~300 units). Licenses to use the new dataset will be offered to companies and institutions under a cost recovery strategy. To be successful, some means to prevent piracy are being considered.

The most promising is based upon watermarking, a technique which inserts a unique, imperceptible serial number (SN) in each vector dataset identifying the original customer. The SN is expected to survive in the dataset even after format exchange (like DGN=>DWG), minor editions, deletions, etc.

This paper describes the test procedures applied to the proposed system. To test its capabilities, we provided the supplier with a dataset from a city and received from him five watermarked instances corresponding to five different customers, each with a different SN. We chose at random one of them, and applied a sequence of modifications described in the paper. The resulting file was returned to the supplier without further information. The supplier was able to correctly identify the original customer, proving the suitability of the approach for tracking back the dataset to its source.

The whole process was repeated twice, using maps corresponding to urban and rural areas, and in both cases the success was equivalent.

The experiment is expected to be representative of a real situation, with some dozens of customers all having formally identical datasets. If we find an illegal copy, and if we are able to link it to a particular customer, we will be able to apply the terms of the contract signed in occasion of the sale. The previous situation was that there was no possibility to prove in court that the illegal map came from a particular customer, so contracts were useless for such purpose.

It is believed that this technology will discourage the customers from making copies of their datasets, thus limiting the piracy threat.

Motivation

The deployment of the Internet infrastructure provides a number of benefits, including easy exchange of data, ideas and information. However, it has created new problems and exacerbated old ones. Now, the exchange of data in digital format is simple, fast and cost efficient, at least for sizeable files. Once transferred, there are popular tools like the CD copier which allows producing locally identical copies as the original at a very low cost. Only digital content can be copied and transferred so easily; other analog media cannot be copied without severe quality effects, or at a significant cost, or involving costly equipment, or even all together. This technical barrier was very useful to protect cartography, because (for many applications) a photocopy cannot be confused to the original printed map, while a full size printer is outside the reach of most individuals.

There are also specific and generic legal protections which apply to maps. Considered jointly with the technical barriers they explain why many cartographic services in all countries have had a business model based on the cost recovery of map production through the sales of many individual paper copies of the same original. Things change once the map is in digital form, mostly because the old technical barriers simply do not hold. Overall, the business model is now under controversy for many reasons. In the policy front, there exists an international movement that proposes to follow the lead of countries like the EEUU that state that the best results from information (in general, not only cartography) is to disseminate it at the lowest possible cost. Discussing this trend is outside the goal of this paper, but it has been recently considered by Lemmens (2003).

The other front is technical. The working hypothesis for a number of cartographic organizations in Europe, Latin America and probably in other places is that the cartography investment in general, and the digital cartography in particular, should produce revenues by sales in order to cover developing costs and provide funds for new accomplishments. This business model (relatively new in some countries) have substituted in whole or in part previous practices, based upon the fact that the national bodies in charge of cartography production have had their operating costs mostly from the state budget. Under such hypothesis, the revenues through direct sales were a minor part of the overall budget, and sales themselves were not the core business, which was in turn the long term plans of cartography production.

The new need to have substantial revenues through sales, plus the technological differences raised by the digital cartography have put the new business model under controversy. In one hand, society requests information in digital format, which can be easily manipulated by a computer enabling a more ample and sophisticated use for the same information. In the other hand, and disclosing specific countermeasures, delivering the information in digital format makes easier the job of creating illegal copies because no technical barriers are applied. Since copying is easy, cheap, and the resulting material is identical as the original, the hypothesis of cost recovery through a significant number of sales quickly should be abandoned. This paper deals with the tests performed over a new technology based upon Steganography, specifically developed for vector maps.

The paper is organized as follows: under *Technical alternatives* we report briefly some copy protection schemes intended to inhibit the copy operation itself. In *Brief introduction to Steganography* we describe the basics of the approach, and how to use it. *The Experiment* describes the tests performed in Uruguay over a particular technology, describing the data available, the underlying assumptions and the procedures applied in the test. The last three chapters summarize the *Results* of the experiment, the *Conclusions* that could be extracted from it and at the end we put the *References* quoted in the text.

Technical alternatives

There exist a number of technical approaches against piracy; they can be classified in three groups: a) those which attempt to inhibit the copy operation itself; b) those which attempt to inhibit the use of the copied dataset; and c) those which try to discourage the copy operation. The differences might be subtle. Once the PC in the 80's and the CD copier in the late 90's became popular, many practical barriers have been lowered dramatically. Some CD's have hidden marks that confuse the device for copying but not for playing; SONY has recently deployed a system based upon such an idea. The system was quickly broken. The second option is very common with software. You are allowed to make as many copies as you want, but in order to use it you must provide an activation key, which could be keyed manually or read from a physical device named *dongle*. In the case of digital information, this scheme has been used in the past by some GIS packages. The whole process is based upon cryptography. The data is kept encrypted in the discs, and only when open by the software the key is read and the information unencrypted to be displayed or printed. The main problem is that the information is protected only if the original format could not be overridden through an export operation. It also forces the customer to keep using the same GIS software, and limits the format that the cartography producer might offer.

The third approach is somewhat indirect. By copying analog information the quality degrades, diminishing the usefulness and the value of the copy for a would-be pirate. Another alternative is to coordinate the behavior of the GIS software with the data, by displaying information bundled with the data about the legal customer. This information could be included in specific fields in the data file (different from the geographic information itself), or be included in a companion file. Displaying a notice about the legal owner might have a significant dissuasive effect. However, it requires the collaboration of the GIS software in order to look for, decode and display the notice attached to the data, a service which might not add value to the GIS package itself. Another problem is that such information is usually included as a comment in the file, and those fields are easily stripped out in any format conversion procedure (either deliberate or not).

We could include in this category the Steganographic approach. It attempts to include in the file information that links it to a particular customer (the one who paid for the file). Unlike the earlier approaches, such information is not visible and is not located in a particular place in the dataset, but bundled with the data itself. Even attempting to do so, the would-be pirate will not be sure that the secret information has been removed because it is not always visible. If the legitimate owner of the first sample knows that such secret information exists, and if he does not know how to remove it or to verify that has already been removed, he will be a lot more careful at the moment of letting someone make a copy. The reason is simple: if starting from an illegal copy he or his organization could be trace back as first owner, through any standard written contract signed in occasion of the sale he will be held responsible by piracy.

Brief introduction to Steganography

Cryptography is an ancient science devoted to secret communication between parties. It is assumed that there exists a third party, which has access to the exchanged message; the trick is to make the content incomprehensible to him, while being able to decode it through some previously agreed procedure. The analysis of the procedure, its properties, its strengths and weakness, etc. is the topic of Cryptography. The goal is to protect the secret message during the communication; once decoded, the protection no longer applies. In the context of digital information, with Cryptography we can assure that only the legitimate client could decode the file, but offers no protection afterwards.

The problem we are considering here is a little bit different: we need to include information in the file in order to prove that an illegal copy found in the computer of Mr. C (which is not a customer) was provided directly or indirectly by Mr. A, a legal customer. Notice that Mr. A could have received the information through a secure transaction, which precludes anyone (including Mr. C) to steal the file at that stage. Once Mr. A received the file, he could have easily made as many copies as he wished, and eventually, sell, lend, or provide one of them as a gift to Mr. C. If there exists more than a customer A, like A1, A2, ..AN, the problem of Steganography is how to distinguish between all the possible providers of illegal copies to Mr. C, and identify the single one involved.

This problem is different to the one solved by Cryptography, and thus requires different techniques. The information (identifying the original customer) should travel with the data file without raising suspicious. The general idea is denoted as Steganography (from the greek words *steganos*=hide and *grafia*=writing) and in many aspects is a close relative of Cryptography. Here the message is inserted into another, but without been noticed. For example, if we want to hide a message within a text, the compound message should make sense. If the message is within a partiture, the music should be reasonable. Some authors say that Steganography operates while the information is in use, in opposition to Cryptography which operates only during the transport or transmission stage. Both techniques are complementary, and there exist cases that they should be used together. The secret message is also known in the literature as *the watermark*.

This ancient technique has been revisited recently, and is now a hot area of research. A general overview has been presented by Bender *et al.* (1996) for audio, video, still images, text, etc. For its application in vector cartography see López (2002, 2003a, 2003b). In this particular case, the watermark is a serial number that can be associated to a specific client. It is inserted by the data producer, or by a third party. The carrier of the message is the vector map itself, and the information should reach a judge or an expert who, after analyzing the content and extracting the watermark, will be able to identify the legal customer of the dataset.

The experiment

We have designed an experiment in order to test the feasibility of the scheme. The provider of the watermarking service supplied five samples of an urban cartography (to be described afterwards) belonging to five different fictitious customers. Each sample had a unique watermark only known by the provider. The SGM received the five samples, selected one at random and renamed it. Some operations were performed over such dataset (to be described later) and the resulting file was delivered back to the provider. The overall scheme would be successful if the provider was able to identify through the watermark the original customer, owner, of the particular file selected by the SGM.

Data description

Urban data

The SGM selected the area of José E. Rodó as representative of the whole dataset. It comprises an area of 456 ha., and the file has a nominal scale of 1:5000. It is represented in Fig. 1. The file is organized in 60 layers, holding information in plant and in height. Its native format is DGN (the dataset was created with Microstation 95/SE) but for the experiment it has been supplied in AutoCAD DWG format.

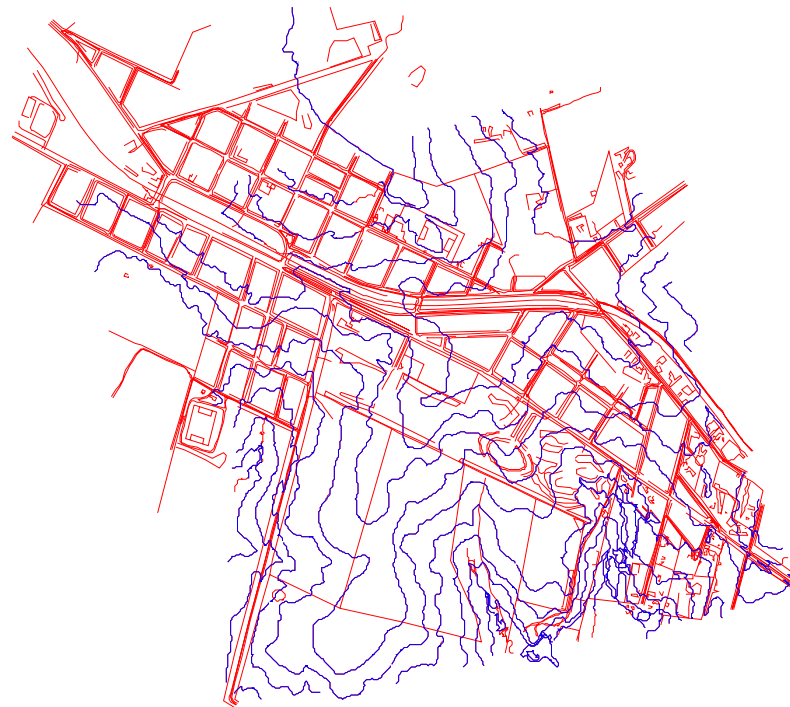


Fig 1 Sample of José E. Rodó. The contour lines are included

Rural data

We have used a cartography describing only hydrography and contour lines of an area 30 km² located near the Cauca Valley, near Cali, Colombia. The heights represented are between 1000 and 2200 m, and the area is traversed by two main rivers. The original scale of the information is 1:10.000 and was developed in ArcINFO 7.1. In Fig. 2 we show only the contour lines, which illustrates the overall complexity of the area.

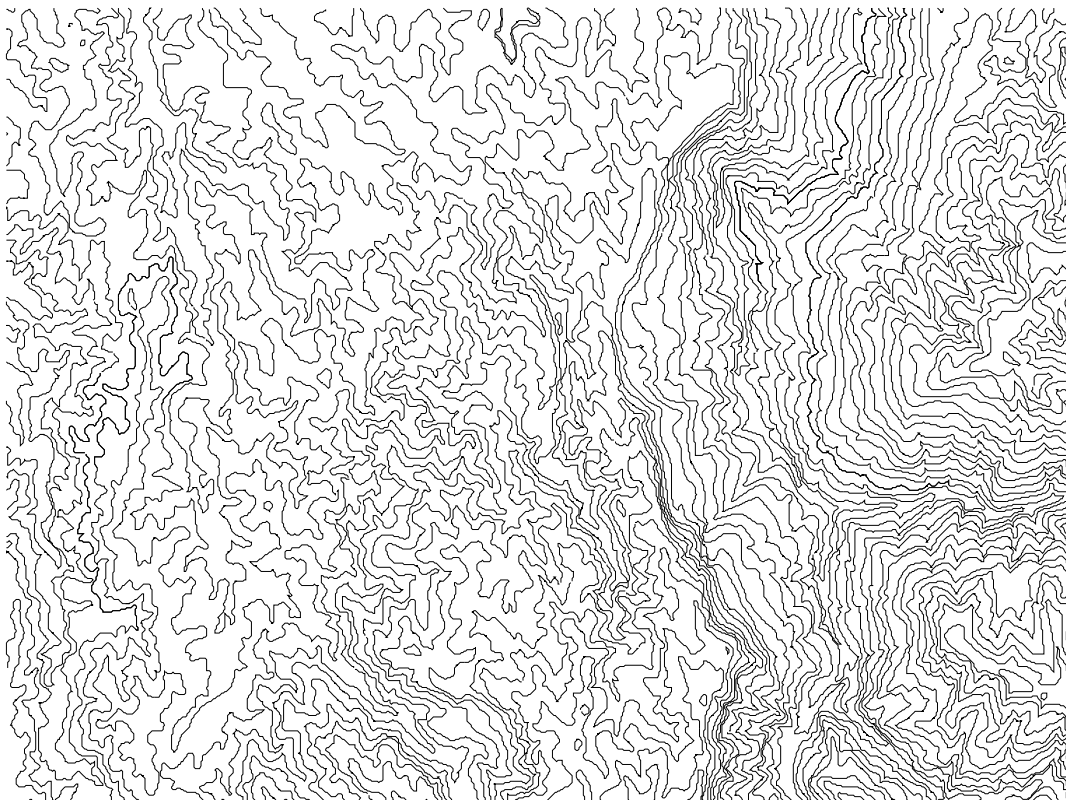


Fig. 2. Contour lines of the rural test area

Hypothesis assumed

Normal users require digital charts with different goals: thematic cartography, location of features in either rural or urban areas (like hospitals, railway stations, etc.), locate themselves or provide others with precise locations through coordinates, analyze the topography to derive slope, erosion, etc. With these goals in mind we restrict ourselves about the likely procedures that a typical user might apply to a map. We assume that:

The modified file should be functionally equivalent to the original

The commercial value of the dataset is tightly connected with its planimetric accuracy. Thus, we discarded as feasible any transformation or manipulation that significantly affects or degrade it.

Editions performed

For either the rural or urban area, all the samples delivered to the fictitious customers were very similar in size, differing at most in a few Kb. For the urban case, a transformation from DWG to DXF were performed, followed from importation to Microstation 95/SE and an internal transformation from 3D to 2D, and stored in DGN format. While in this format, it has been analyzed by EDG in order to detect internal errors. Afterwards, the files were again loaded within Microstation in order to check visually that all of them look the same. One of them was selected at random, exported back to DWG format and given back to the provider.

In the rural case the starting point was DGN. Within Microstation, the dataset was compacted, and the resulting file were analyzed with EDG. Some minor editions and deletions were applied, as well as a non isometric georeferentiation. The file was then converted to DXF and given back to the provider without any further information, including any parameter of the non linear transformation applied.

Results

Both for the urban and rural case, the watermarking provider received back AutoCAD files. In the rural case, it was necessary to guess inverse transformations in order to analyze it. In both cases the original customer (one out of five) was detected without error.

Conclusions

In many countries the map production organizations are at present facing conflictive demands. Society claims for better and more products, to be delivered in digital format for its use within computer systems. Such efforts require substantial investment in technology, human resources and time. However, in many countries specially in the Third world the governments need to diminish to a minimum the investments in the national budget, forcing or stimulating each of its branches to adopt a strategy directed to have direct income from its products. The goal is to fund a significant part of the budget through sales.

The production of digital cartography is shared as a common goal for society, but adopting it raises immediately the concern about piracy, a problem almost inexistent in paper format. Once delivered the second digital sample of a cartography, it is necessary to identify who was the source for any illegal dataset once located. Without the use of specific technologies, all the samples delivered to customers are identical, raising to a dilemma at the moment to assign responsibility to the customer or customers involved. Being anonymous facilitates and stimulates the irresponsibility of the legitimate customers, with severe damage to the data producer.

Steganography provides a framework where to develop a solution to identify the pirate. The trick is to insert a serial number in each sample of the cartography, unique for each customer. Unlike other solutions, the serial number is hidden making more difficult to erase it. In addition, it is not stored in any specific location of the file but embedded with the geometric information itself. This particular fact allows the information to survive changes of format (a legitimate transformation) like converting from DGN to DXF, a process that might ignore comments and other accessory information included in the original file.

In order to test the technology, the SGM of Uruguay performed two tests. In each of them, a set of five customers for urban and rural cartography were simulated by producing five samples with different serial numbers. At random, one of them was selected and some transformations applied like those expected from the legitimate clients. The resulting files (one for rural and one for urban areas) were delivered back to the technology provider. This stage resembles the situation when an illegal copy of the dataset has been located. With the files, and without having any information about the process applied to the datasets, the provider was able to identify without error both fictitious customers, proving in this limited context the usefulness of the watermarking solution. Adding this extra technical protection to the existing copyright protection legal body, the SGM will be able to sell the new digital urban cartography recently developed for the whole country.

References

- Bender, W. Gruhl, D., Morimoto, N and Lu, A., 1996. "Techniques for data hiding". IBM Systems Journal, 35, 3&4, 313-336
- Lemmens, M., 2003. "Free of Charge or Revenue Generation", GIM International, 17, 2, 40-49
- López, C., 2003a. "Digital Rights Managements of Geo-Datasets: Protection against Map Piracy in the Digital Era", GIM International, 17, 2, 51-53
- López, C., 2003b. "Intellectual Property Protection through Watermarking". (last visited may 2003). <http://www.thedigitalmap.com/curso/lpintro.pps>
- López, C., 2002. "Watermarking of Digital Geospatial datasets: a review of Technical, Legal and Copyright issues", International Journal of Geographic Information Science, 16, 6, 589-607.