基于中国剩余定理的 GIS 数字水印算法

周 旭 毕笃彦
(空军工程大学工程学院四系信号与信息处理实验室,西安 710038)

摘 要 利用数字水印技术对数据进行保护是一种行之有效的方法,但目前提出的许多算法都是针对多媒体数据的。为了将该技术应用于地理信息系统(GIS)数据保护,首先介绍了数字水印的基本体系,并分析了 GIS 数据的特点,然后利用基于中国剩余定理的数字水印算法,构造了适用于 GIS 数据的水印嵌入和提取的方法,以用于实现 GIS 数据的版权标识和篡改提示。该算法符合 Asmuth-Bloom 体系,它的优点在于可以根据部分数据恢复出水印信息,其不仅有较好的抗攻击能力,而且在恢复时不需要原水印信息。由于该算法是基于点的坐标来进行水印的嵌入,因此在 GIS 数据中有较好的推广价值,其不仅适用于关系型数据库,且可以应用到三角网格模型中。

关键词 数字水印 地理信息系统 版权保护 中国剩余定理 Asmuth-Bloom 体系
中图法分类号: TP309.7 **文献标识码**: A **文章编号**: 1006-8961(2004)05-0611-05

Use Digital Watermarking to Protect GIS Data by Chinese Remaindering

ZHOU Xu, BI Du-yan
(*Signal and Information Processing Lab, fourth department, college of engineering, Air Force University of engineering, Xi'an* 710038)

Abstract Along with the popularization of Internet and the development of multimedia techniques, internet has become a main way for information communication. People can get a lot of information through the Internet, such as text, image, graphics, audio and video etc. Unfortunately however: internet and multimedia also afford virtually unprecedented opportunities to pirate copyrighted material. This risk is same to the distribution of GIS data through internet. As a result, digital Watermarking has been presented as an efficiently solution to copyright protection. Dozens of schemes and algorithms have been proposed, but most of those are for multimedia data. In this paper, a new watermarking method is proposed for the protection of GIS data. This approach applies the Chinese Remaindering to construct embedding watermarking in GIS data. The influence of embedding process on GIS data would ignore under the lower measurement precision. It can be used as tamper proof and authentication of GIS data . This algorithm accord with Asmuth-Bloom system, so we can extract watermarking use only a part of GIS data. The extraction process doesn't need the original watermarking, because the embedding procedure is based on the states that a positive integer is uniquely specified by its remainder modulo relatively prime integers. The algorithm bases on the coordinate of the points, so it can be widely used in GIS data protection such as relation database, as well as triangular mesh. The results show that the algorithm can work well. Finally, some open problem is discussed.

Keywords digital watermarking, geographic information system, copyright protection, chinese remaindering, asmuth-bloom system

1 引 言

随着因特网的日益普及,信息的交流已达到了前所未有的深度和广度,其发布形式也愈加丰富了,但是随之出现的问题也十分严重,如侵权更加容易,篡改也更加方便。在这种情况下,如何既能充分利用因特网的便利,又能有效地保护知识产权,已受到人们的高度重视。现在,地理信息系统的应用越来越广泛,已经不再局限于地学领域。其结果是,地理信息

基金项目:全国骨干教师资助计划项目(GG-020-90039-3243)
收稿日期:2003-06-05;**改回日期**:2003-11-14

系统中的数据生产者和数据使用者之间已经截然分开了。这样二者之间将面临着信任危机，即从生产者来说，希望在数据发布中能避免未经授权的拷贝制作和发行，同时能防止未授权的用户使用；从使用者来说，则希望能得到未经篡改的正确数据，并能确定它们的内容是否被修改、伪造或特殊处理过。要解决上述问题，必须要提出一个 GIS 数据的版权保护方案。由于采用传统密码方法并不能完全解决版权保护问题[1]，于是人们开始研究通过永久性数字水印来解决这一难题。

2 数字水印技术的主要机制

数字水印(Digital Watermarking)技术是指用信号处理的方法在其他数据(宿主数据)中永久镶嵌具有可鉴别性的数字信号或模式，而且并不影响宿主数据可用性的技术。由于数字水印是信息隐藏技术的最重要的一个分支[2]，它可为计算机网络上的多媒体数据产品的版权保护提供一个有效的解决方法，因此数字水印的应用前景广阔，它可以广泛用于未来数字作品的版权保护、隐蔽通讯、电子商务等领域。由于在绝大多数情况下，人们希望添加的水印信息是不可察觉的或不可见的，并且希望攻击者在不破坏数据本身质量的情况下无法将水印去掉，因此数字水印应具有如下特点：

(1) 顽健性　即数字水印难以被清除，如果使水印破坏，则数据质量也会严重下降，可见，数字水印对恶意攻击(如宿主数据被分割、篡改、滤波等)应具有稳健性，但是目前能抵抗所有恶意攻击的数字水印算法还没有出现。

(2) 安全性　即数字水印本身是安全的，难以被篡改或伪造。

(3) 保护性　即数字水印能为版权提供保护，能对数据产品的归属提供完全和可靠的证据，并能监视被保护数据的传播和鉴别其真伪，以及能控制非法拷贝等。

(4) 隐蔽性　即数字水印对数据的正常使用没有影响，也就是对用户来说是不可感知的；对攻击者来说，即使是对大量嵌入水印后的数据进行统计分析也不能提取水印或确定水印的存在。

数字水印的加载和检测过程如图 1、图 2 所示。

其中，水印可由多种模型构成，如随机数字序列、数字标识、文本及图像等。从顽健性及安全性考

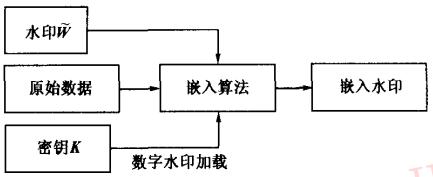

图 1

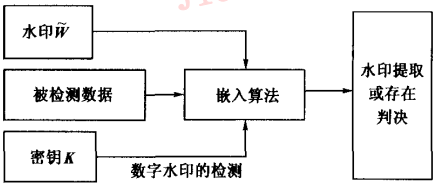

图 2

虑，通常需要对水印进行伪随机化以及加密处理。检测水印时，根据不同算法，有的不需要原始水印。

设 $\boldsymbol{I}$ 为原始数据，$\boldsymbol{W}$ 为水印信号，$\boldsymbol{K}$ 为密钥，处理函数为 F，那么处理后的水印为

$$\tilde{\boldsymbol{W}} = F(\boldsymbol{I},\boldsymbol{W},\boldsymbol{K})$$

设嵌入水印后的数据为 $\boldsymbol{I}_{\mathrm{w}}$，编码函数为 E，则嵌入算法可表示为 $\boldsymbol{I}_{\mathrm{w}}=E(\boldsymbol{I},\tilde{\boldsymbol{W}})$。与之类似，设水印提取的过程为解码函数 D，则提取出的水印为 $\hat{\boldsymbol{W}}=D(\boldsymbol{I}_{\mathrm{w}},\boldsymbol{I})$。由于受攻击的原因，$\tilde{\boldsymbol{W}}$ 和 $\hat{\boldsymbol{W}}$ 有可能不同。

在目前的许多数字水印算法中，大部分抵抗数据压缩和噪音攻击的能力较强，对于数据遭受剪裁攻击等几何变换的抵御能力较差[3]。而在本文采用的算法中，在数据缺失不严重的情况下，仍然能较好的提取水印，顽健性更强，而且在数字水印的提取过程中，并不需要原始水印。

3 基于中国剩余定理的数字水印算法[4]

中国剩余定理最早见于《孙子算经》，是数论中最重要的基本定理之一，其实质上是刻画了剩余系的结构。它要解决的问题的一般形式是：已知 m_1，$m_2,\cdots,m_s$ 是两两互质的正整数，求最小正整数 x，使它被 $m_1,m_2,\cdots,m_s$ 除所得余数分别为 $a_1,a_2,\cdots,a_s$。上述问题可等同为求以下一次同余方程组：

$$x \equiv a_j \bmod m_j (j = 1,2,\cdots,s)$$

的最小正整数解。

中国剩余定理证明该方程组必有解，并给出同余方程组的解为

$$x \equiv M_1 \times a_1 + M_2 \times a_2 + \cdots + M_s \times a_s (\bmod m)$$

其中,$m=m_1\times m_2\times\cdots m_s$;$M_j\equiv l \bmod m_j$,$j=1,2,\cdots,s$,且 M_j 能被 $m_i(i=1,2,\cdots,s$ 且 $i\neq j)$整除。

基于中国剩余定理构造了如下秘密分存(n,t)门限方案:设 k 是一个要分存的秘密整数,选一素数 $p>k$,另选一组两两互素的整数 $m_1,m_2,\cdots,m_n$,并满足与 p 互素的条件,不妨设 $m_1<m_2<\cdots<m_n$,此外 $m_1,m_2,\cdots,m_n$ 还要同时满足如下条件:

$$m_1\times m_2\times\cdots\times m_t>p\times m_n\times m_{n-1}\times\cdots\times m_{n-t+2}$$

令 $l<[m_1\times m_2\times\cdots m_t/p]$,作

$$L=k+l\times p<l\times p=(l+1)\times p\leqslant m_1\times m_2\times\cdots\times m_t \quad (1)$$

令 $L\equiv k_i \bmod m_i$,$i=1,2,\cdots,n$。其中 k_i 即是分存的结果。

可以证明:若是得到其中上面分存结果中任意 t 个 k_i 便可恢复出 k。根据中国剩余定理,方程组

$$L\equiv k_{i_j} \bmod m_{i_j},\ j=1,2,\cdots,t \quad (2)$$

在模 $m_{i_1},m_{i_2},\cdots,m_{i_t}$ 下有唯一解 $k=L-p\times l$。

如果宿主数据完整,则可以从中提取出全部(n 个)k_i,显然可以恢复出 L,从而得到被分存的整数 k。如果宿主数据不完整,则只要从中可以提取出 k_i 的数目大于或等于 t,就仍然可根据上述定理,从同余方程组中得到唯一解 L,进而可确定 k 的值,但是,如果只能从残余数据中提取出 $k_{i_1},k_{i_2},\cdots,k_{i_{t-1}}$,即少于 t 个 k_i,则无法准确恢复 L,也就无法恢复出 k。满足这种条件的方案,在密码学上称为 Asmuth-Bloom 体系。

4　GIS 数据的特点及对数据保护的要求

针对不同的应用,GIS 数据的数学模型有所不同,但总的来说,有两种人们普遍认可的基本地理数据模型,它们是连续域模型和确定对象模型,简称为域模型和对象模型。

域模型通常用于表达具有连续变化的地理现象,这些地理现象通常可以用平滑的数学函数表示,如数字地形模型。在域模型下,每个属性假定在空间连续和平滑地变化,其变化可以用一个平滑的数学函数有效地描述。这些域通常可分离为给定分辨率下的规则格网或 Delauney 三角网,其适宜用关系数据库或格网数据结构表示,如 TIN 不规则三角网,并且多使用栅格模型。

对象模型适用于便于抽象的地理实体,它们不仅具有清晰定义的边界和定义好的一系列属性,而且是当前应用最为广泛的模型,其典型的例子如在土地利用调查中,每一地块的边界已准确界定,且每一地块所具有的相同属性都已定义好,另外,对象模型还强调空间对象的可区分性和标识性。在对象模型中,地理实体由清晰而确定的点、线、面、和体等几种基本几何元素表示,其中点表示如居民地、城市等抽象点的地理坐标;线连接一系列确定的已知坐标点,可以表示河流、道路;面由确定的线围绕而成,可以表示宗地或行政区域;体由光滑的面组合而成,可以表示楼群等实体,另外,还可用属性数据来进一步描述点、线、面和体,因它是常量值,故可在采集地理信息数据时获得。由于对象模型的形式简单,同时为了便于表示属性和便于保持拓扑关系,因此对象模型通常用关系型数据库实现。

在当前 GIS 应用日益广泛的情况下,对 GIS 数据的保护主要有以下要求:

首先,是保护生产者的权益。数据的生产者可先将水印嵌入原始 GIS 数据,然后发布他的水印版本作品。这样当该作品被盗版或出现版权纠纷时,生产者即可从盗版作品或水印版作品中获取水印信号作为依据,以保护生产者的权益。同时,为避免未经授权的拷贝制作和发行,可以将不同用户的识别号(ID)作为水印(指纹)嵌入作品的合法拷贝中。这样一旦发现未经授权的拷贝,就可以根据此拷贝所恢复出的水印(指纹)来确定它的来源。

其次,对数据的使用者而言,都希望能得到篡改提示。为实现该目的,通常可先将原始数据分成多个独立块,然后对每个独立块按照一定的算法计算出一个校验值,再将校验值作为水印加入每个块中。这样使用者就可通过检测每个数据块中的水印信号与校验值比较来确定作品的完整性,但要求校验算法保密。

5　GIS 数据中数字水印嵌入和提取方案

嵌入的水印信息可以是图像、文字、用户识别码,或其他编码的形式,在水印嵌入的过程中,都将其视为二进制比特流。为了得到篡改提示,在水印中应该加入校验值;为了增强水印算法的顽健性,应对水印信息进行随机化,其目的是使水印信息接近白噪声。因为中国剩余定理处理的是正整数,所以应将加密后的水印信息比特流进行按固定长度的 bit 进行切割,以转化为十进制数。

按 Asmuth-Bloom 体系选取合适的参数:m_i,

p,t,l,n，若随机化后的水印信息为 $\boldsymbol{k}$，按照式 1 和式 2 对水印进行分存，则分存后的水印总长为原来大小的 n 倍，设分存后的水印为：$\boldsymbol{k}_1,\boldsymbol{k}_2,\cdots,\boldsymbol{k}_n$，且它们都应视作比特流，设它们都是 q bit 长。n 值的选取要参考原始数据和水印数据的比例来考虑水印的容量，以防止水印嵌入时发生越界的问题，而且 m_i,p,t,l,n 等参数应该保密。

前文所述的两种 GIS 数学模型在目前的 GIS 平台中应用非常广泛，并且主要利用关系型数据库实现。为了嵌入水印，可以在数据库中增加一个容纳水印信息的属性表，其优点是容易实现，且不用改变原有的数据，但是这种方法抗攻击能力太差，攻击者只需简单地将该属性表去掉，就可以去掉水印。当然也可以将水印信息和校验信息结合在一起，将二者进行伪随机化，再放置到属性表中。这样当攻击者在剥离掉水印的同时，也会丢失校验信息，这虽然增强了顽健性，但仍嫌不足。若要更进一步增强水印的抗攻击性，可以将水印隐藏在 GIS 数据中“较重要位处”，具体方法可根据实际数据来确定。这种方法的好处是显而易见的，因为一旦较重要位处被破坏，数据品质也会严重地下降。文献[4,5]中提出的基于变换域的、扩频的水印隐藏方法，虽较适用于图像数据，但应用于 GIS 数据中则比较困难。

针对 GIS 数据的特点，作如下分析：点的位置信息在 GIS 数据中是非常重要的，其在域模型和对象模型中都普遍存在，通常用经纬度来表示，并用 32bit 长整型数记录。因受字长的限制，经纬度精度最高可取 1″/1 000，一般为 1″/100，其对一个 1∶50 000的地图来说，该 32 bit 长整型数的最后 1 bit 的变化对数据的影响是微不足道的，仅相当于对数据施加了扰动，因此这里就是存放水印信息的最佳位置。嵌入水印时，首先将要加水印的数据按预定的分割模式分割为大小相当的 n 个块，在每个块中有 r 个点，由于每个点有两个坐标，因此要求 $2r>q$，然后将每个点的坐标的最后 1 bit 依顺序换为相应的 $\boldsymbol{k}_i(i=1,\cdots,n)$的 bit。这样水印嵌入后，就对原有数据结构没有影响，对使用者来说，由于该 bit 位的变化对数据的影响远远低于测量精度，因此对实际使用没有不良影响。

水印的提取算法实际是水印分存嵌入算法的逆向算法。由于算法符合 Asmuth-Bloom 体系，其恢复水印时，并不需要获得全部 n 份分存的水印信息，而只需要获得其中的 t 份就可以，因此恢复水印时，可首先将携带水印信息的数据按分割模式分为 n 块，再从中任意选取 t 块，然后通过将每个点的坐标的最后一个 bit 顺序取出来得到 $\boldsymbol{k}_i$，再结合所对应的 m_i，根据中国剩余定理恢复出伪随机化后的水印 $\boldsymbol{k}$，即可由 $\boldsymbol{k}$ 得到原始的水印。

6 实验及结论

按如上所述的数字水印嵌入算法来完成以下试验：

嵌入的水印信息为如图 3 所示的图标，嵌入水印时，先用 RAR 软件进行压缩，这里对图标进行压缩是为了避免生成的密文太长，以至于要求的宿主文件太大；然后再使用 gost 加密算法（密钥是字符串 air force university of engineering），来得到 1 000Byte的密文（如图 4 所示）；接着再一次对密文进行压缩，目的是使最终获得的正整数较小；最后用 BCD 码将密文转为正整数。使用上述水印算法，向一大小为 1.58 MByte 的等高线数据文件中嵌入水印。该文件的数据形式是由一系列的点组成的等高线。试验证明，在处理完成的等高线文件中不仅能完整提取出所嵌入的图标，而且在将该数据文件去掉部分数据后，所提取出的图标与图 3 完全一样。由于该等高线文件的图幅范围大致为960 000km^2，因此，由于水印嵌入带来的误差可忽略不计。

Face02.ico

图 3 试验中嵌入的水印信息

12c9590b5af4ea457c89c75e3236691716e4b57800064622975642c8103
5201f42d5a664fd9d4b9f852b11d11fc8e252ed51106d1fbcda662be3ef
60cb33b574233eaa2be8bbbe479b3a8952bda9a5e0de48f5d19d3686d94
ab7ecd6ceaf1ddf666832c9f9e95a68496ccd5f5252bb6a356e25bb00f6
d2c6bb414e2fea25f3ccbf1379b0316f0825c8246d932ba163d429e833f
dc018e61d4dfd5fcd2918d22aeebf348080032ee0a3cf69848cd389e81a
9a78fdcd02135884a70ae20acb26ed1cad5fd94772d46fe14425708af3a
0ee0a9cdcca5eccecfc35c7a75813c3083aa505761667e0a19f1de69cc3
0e38c722b82d3541178c510f48e3baade85356cce64bd41a0bc6c2d608d
f20d469d2a8b3305cea5c03154fea91adb84bc1c7ca8bddaca9e6b5a538
dc2a31fb9e186b44cd07c2f525766490720b60f24858233c78648e22863
fab19b6cad02b5ecb04f2113f59423774ec9c59580d198fac2cf5294487
ed7cf73b9794774e5cdc2b82e3f8890876418253e1ba0e5f76418253e1b
0e5f76418253e1ba0e5f76418253e1ba0e5f76418253e1ba0e5f7641825
e1ba0e5f76418253e1ba0e5f76418253e1ba0e5f76418253e1ba0e5f764
8253e1ba0e5f76418253e1ba0e5f76418253e1ba0e5f76418253e1ba0e5
76418253e1ba0e5f76418253e1ba0e5f76418253e1ba0e5fb262d8edb08
057a

图 4 对压缩后的图标使用 gost 加密算法得到的密文

7 结　论

本算法的优点是：①对数据的剪裁攻击有很好的抵抗能力，并可以依靠部分数据恢复数字水印；②提取水印时不需要原始水印信息，隐蔽性也比较好，但顽健性还需依靠于一个好的伪随机方法和一个健壮的数据分块算法。另外，隐蔽性的好坏是否还决定于水印信息的隐藏位置，还有待作进一步的研究。针对 GIS 数据的实际情况，数字水印也可隐藏在属性数据中，并都要求隐蔽算法保密。对于篡改提示，较全面的解决方案可以采用易碎水印（Fragile Watermarking）[6]。在目前 GIS 平台各自相对封闭的情况下，可以将校验值的提取算法放在平台内部，分发给使用者，这样对使用者来说也可以得到比较满意的篡改提示。由于本文提出的方法是基于点的坐标来嵌入水印，且点的信息是 GIS 系统的基础，因此基本上在任何类型的 GIS 数据中都可使用本算法，如线数据、面数据、体数据，并且不仅仅适用于关系型数据库，而且对于三角网格数据[7]（如 TIN 模型、规则网格模型等）也适合。

参 考 文 献

1 Gong L, Qian X. The complexity and composability of secure interoperation [A] In: Proceedings of the IEEE Symposium on Research in Security and Privacy [C], Oakland, California, USA, 1994:190～200.

2 刘瑞祯，谭铁牛. 数字图像水印研究综述[J]. 通信学报，2000，**21**(8)：39～48.

3 Matheson L R, Mitchell S G, Shamoon T G, *et al*. Robustness and Security of Digital Watermarks [A]. In: Proceedings of Financial Cryptography Secod International Conference '98, of Lecture Notes in Computer Science[C], Anguilla, British West Indies, 1998,1465:227～240.

4 杨义先，钮心忻，任金强. 信息安全新技术[M]. 北京：北京邮电大学出版社，2002. 12 .

5 周利军，周源华. 数字图像水印的扩频实现[J]. 红外与激光工程，2000，**29**(5)：27～31.

6 Ping Wah Wong. A public key watermark for image verification and authentication[A]. In: International Conference on Image Processing[C], Chicago Illinais, USA, 1998:445～459.

7 尹康康，潘志庚，石教英. 一种强壮的网格水印算法[J]. 计算机辅助设计与图形学学报，2001，**13**(2)：102～107.

周　旭　1975 年生，空军上尉，2003 年获空军工程大学空军工程学院信号与信息处理专业硕士学位，现为西安空军工程大学工程学院电子工程学院通信与信息系统博士研究生。研究兴趣为地理信息系统、信息加密、免疫算法。

毕笃彦　1962 年生，博士生导师，1986 年获国防科技大学电子系硕士学位，1997 年获法国图尔大学博士学位，现为空军工程大学工程学院电子工程系教授。研究方向为图像处理、图像分析、数据压缩。