Chapter 5

Multimedia Security

In this chapter, previous work concerning multimedia security will be discussed. Firstly, traditional techniques used to ensure the security of multimedia data will be briefly discussed. Next, the recent main research focus with regards to multimedia security, watermarks, will be discussed. Finally, other research material regarding the security of multimedia will be reported.

5.1 Securing multimedia using traditional techniques

Some traditional techniques can be used to protect multimedia from unauthorised access (including viewing, copying, modifying, distributing, etc.). The more common techniques include encryption, checksums, hash-functions, and time stamps. Encryption when used in this context converts the original multimedia data into a form that cannot be interpreted without correctly decrypting it, therefore enforcing confidentiality, although it can be used to enforce non-repudiation when public key encryption is used. Checksums or hash-functions can be used to ensure the integrity of the multimedia data, specifically when the data is transmitted over a network. Time stamps are used to derive the owner, as well as the time at which the multimedia data was generated. Ownership of the multimedia data can be verified by owning the earliest reliable time stamp for it (and if the multimedia data is time stamped at each change of ownership, we can determine its chain of custody).

The above traditional techniques are usually used in various combinations, depending on their application, as components of a multimedia security system. In most cases, some form of authentication is integrated into the multimedia security system, which in this context is used to prove who created the multimedia data (e.g. for watermarks), or to verify the user requesting the multimedia data.

5.2 Watermarks

The duplication of digital image, video, and audio multimedia data has increased rapidly in the recent years due to the decrease in quality of analogue duplication, and also because it is becoming easier to duplicate and transmit large amounts of digital multimedia data. The availability of digital multimedia is rapidly increasing with the continual growth of multimedia services such as e-commerce, video-on-demand, digital newspapers, pay-per-view, video conferencing, image/audio/video databases, digital libraries, digital cameras, etc. A need has therefore been created for techniques that can be used for copyright protection of digital multimedia data. Copyright protection concerns the authentication of multimedia

content and/or ownership of it, which can be used to identify whether the concerned multimedia data was illegally copied or forged.

5.2.1 The definition of a watermark

A developing method to enforce copyright protection has been the introduction of *digital watermarks*. A digital watermark is a secret code which is embedded into multimedia data that allows for the verification of the owner and/or content of that data. Watermarking is a special case of a technique called *steganography*, which involves taking a piece of information and hiding it within another. Steganography takes advantage of unused or insignificant areas of data (often found within images, sound recordings, and even disks), replacing them with information.

Generally, a watermarking algorithm consists of three parts. The first part is the actual *watermark*; each owner has a unique watermark which is used for identification. The *marking algorithm* is the second part, which is used to embed the watermark into the multimedia data. The third part is the *verification algorithm*, which allows us to authenticate the multimedia data using the watermark embedded within it (Wolfgang & Delp, 1997). The diagram below illustrates the watermarking process:

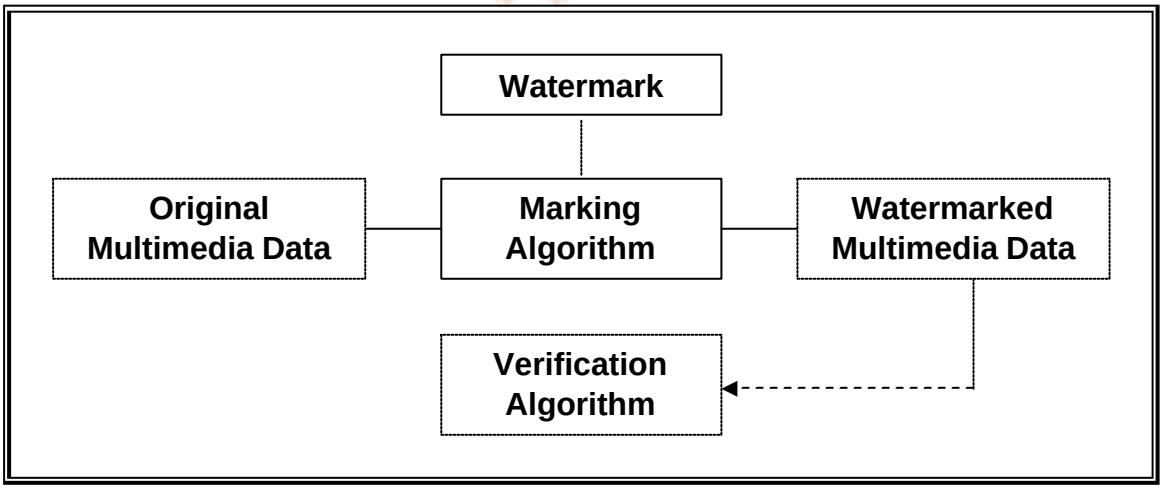

Figure 5.1: The watermarking process (Adapted from Wolfgang & Delp, 1997)

5.2.2 Watermark requirements

A watermark must satisfy a number of properties in order for it to be effective. Firstly, the watermark must be robust. Although any watermark can in theory be removed with knowledge of the process of insertion, it should be virtually

impossible to remove the watermark without this knowledge (without first degrading the multimedia data to the point where it has no practical use). The watermark should in particular be robust to common geometric distortions (scaling, translation, etc.), common signal processing (analogue to digital, digital to analogue conversion, re-sampling, etc.), collusion, and forgery. Secondly, the watermark should not be perceptually visible, or at least should not interfere with the multimedia data (transparency requirement). Ideally, the watermarking process should apply to all three multimedia types (image, audio, and video). Thirdly, the marking algorithm should allow multiple watermarks to be embedded into the multimedia data in a way that will allow each watermark to be independently verified (capacity requirement). Lastly, the verification algorithm must indisputably identify the owner of the multimedia data, and the accuracy of the identification should lessen gracefully in the event of an attack (Cox, Kilian, Leighton & Shamoon, 1996).

5.2.3 Perceptually based watermarks

The most common form of digital watermark is the *perceptually based watermark*, which is specifically designed to exploit certain aspects of the human visual system in order to decrease the chance of it being detected. Other types of watermarks include visible watermarks used to embed information into some form of multimedia, e.g. a company trademark may be visibly embedded into the background of an image. The main focus of this section will however be on perceptually based watermarks. In theory, any watermarking algorithm that tries to embed an invisible mark into multimedia data is said to be perceptually based. To be truly transparent and robust, however, the watermarking process must make more practical use of perceptual information.

Cox, Kilian, Leighton & Shamoon (1996) argue that a watermark should be placed in perceptually significant regions of the multimedia data in order for it to be robust. Most techniques prior to this one specifically avoid inserting a watermark into perceptually significant regions. The main reason for this is that changes made to perceptually significant regions result in perceptual distortions more easily than the addition of a watermark to perceptually insignificant regions. The common strategy has therefore been to place a watermark into the high frequency components of a signal's spectrum i.e. in the areas of that multimedia type where some property such as colour changes quickly and by large amounts. These prior techniques have however not been robust toward many intentional and even some unintentional attacks.

The main argument in Cox, Kilian, Leighton & Shamoon (1996) is that some common processing operations performed on multimedia data degrade the perceptually insignificant regions. Although most commonly performed operations are information-lossless, there are some operations that need to be performed in such a way as to reduce the amount of multimedia data (e.g. compression schemes such as JPEG, DIVX AVI, MP3, etc.). These operations often degrade the quality of the data by eliminating perceptually insignificant components of the multimedia data. In order to preserve the watermark in such cases, it must be placed in the *perceptually significant* components.

Besides the above mentioned operations that are often performed to reduce the amount of data that needs to be transmitted over a network, intentional attacks may be performed to deliberately remove or forge the watermark, as well as other unintentional operations that can be harmful. Operations such as low pass filtering can easily remove watermarks that have been inserted into high frequency components of a signal's spectrum. It is therefore best to insert the watermark into perceptually significant components, because if these harmful operations were to be performed over the entire spectrum (including perceptually significant areas), the multimedia data would lose its practical value (assuming that knowledge of the exact locations of insertion is not know to the attacker).

The above argument proves that a watermark is more robust when inserted into the perceptually significant components of the multimedia data. One problem with this approach is that we still need to insert the watermark in a way that does not visually interfere with the multimedia data (i.e. that the changes are not visibly noticeable). A common method in use is to make a large amount of very small changes to any single frequency/intensity coefficient, rather than a small amount of considerable change to any coefficient (in the perceptually significant regions of the multimedia data) (Cox, Kilian, Leighton & Shamoon, 1996).

A watermark's requirements of transparency and robustness therefore conflict with each other; placing the watermark in perceptually insignificant components provides transparency with little robustness, while placing the watermark in perceptually significant components provides more robustness with less transparency (depending on the technique used). A number of methods have been proposed which include a visual model as part of the watermarking process. These digital watermarking techniques often take advantage of frequency selectivity and weighing in order to provide some perceptual rules or principles into the watermarking process. Some of these techniques not only take advantage of frequency selectivity, they are also designed to adapt to local image or frame

characteristics, resulting in a very robust and transparent technique. Watermarks that are created using these techniques are called *image adaptive watermarks* (Wolfgang, Podilchuk & Delp, 1999).

5.3 Other multimedia security research

The majority of the research community's effort regarding multimedia security has gone into digital watermarks. This is due to the increasing demand for copyright protection of multimedia data. Only a small amount of research has been dedicated to other areas of the multimedia security domain. The following briefly discusses some significant multimedia security research work.

In the first research paper that I discuss, Lou & Yin (2001) propose a method of integrating a spatial database processing technique (using bintree encoding), together with a technique that provides (hierarchical) access control, to enhance the security of spatial databases. According to Lou & Yin (2001), "*spatial data* consists of spatial objects made up of points, lines, regions, rectangles, surfaces, and volumes". The proposed method is called camouflage hierarchy (CH) and consists of three phases.

In the first phase, the spatial data image is converted into a bit-stream using linear bintree coding. Bintree coding is based on the theory of recursive decomposition, and is mainly used to reduce the amount of spatial data. The root of the bintree represents the entire original image. The root's two child nodes each represent an equal half of the original image (divided in the x direction). The next level of the bintree divides each of the root's children further into equal halves (divided in the y direction). This process is recursively repeated. A leaf node is reached if the node's entire region consists of the same colour. The resulting linear bintree is represented using two bit-streams B and L. B represents the internal and external nodes of the tree when traversed using breadth first search (internal nodes are represented by 0, and leaf nodes by 1). L represents the colour of the leaf nodes when traversed using breadth first search (0 represents black, and 1 represents white). Finally, B and L are co-joined to form the bit-stream D.

The second phase involves encrypting the bit-stream D using a stream cipher system that allows for access control in a user hierarchy structure. This stream cipher system allows the secret keys used for each user to be divided into classes corresponding to the defined user hierarchy structure. The bit-stream D is then encrypted with the user's key to form the encrypted bit-stream E. Using this

stream cipher system, a user can decrypt data that was encrypted by another user with the same or lower security classification.

The third phase uses a data hiding algorithm (i.e. steganography) to embed the encrypted spatial data E into a cover image. The discrete cosine transform (DCT) based algorithm used embeds the data into the middle frequency coefficients of the cover image, allowing the hidden data to be more imperceptible. This also ensures that data loss is kept at a minimum during compression or modification (e.g. zooming, rotation, etc.), which mainly affects high frequency areas of an image (Lou & Yin, 2001).
Performing the inverse of the above CH process retrieves the original data.

In the next research work that I discuss, Chun & Atluri (2000) propose an authorisation model that can provide a form of access control for geo-spatial images. The model is called Geo-Spatial Authorization Model (GSAM), and uses an enhancement to the spatial access method MX-quadtree, called the MX-RS quadtree, to enforce access control when used to index an image database. This technique provides access control in two ways. Firstly, it controls the depth a user can traverse by limiting the resolution level the user is allowed to access. Secondly, the extent a user can view is controlled, by allowing the user access to only certain regions of a higher resolution image (e.g. a satellite image of that user's property).

The MX-RS quadtree method assigns an entire region to the root node of the quadtree. The children of the root divide the region into four smaller regions, with each child referencing an image list that makes up one of the smaller regions at a higher resolution. The children of these child nodes further divide these sub-regions (at an even higher resolution), and the process continues recursively until a leaf node is reached (i.e. until the maximum resolution for that area is reached). The images in the same level of the quadtree have the same resolution and can be represented as a linked list that forms the entire region at that resolution.

The main difference that distinguishes an MX-RS quadtree from an MX quadtree is that each node in an MX-RS quadtree has a reference to an image list, whereas only the leaf nodes of an MX quadtree contain references to an image list. The MX-RS method therefore allows for efficient enforcement of privilege modes such as *view* and *zoom-in*, simply by regulating the traversal of the quadtree down toward the higher resolution images. For example, the *view* mode can be performed by traversing the tree and retrieving all images that contain the

requested region and have a resolution less than or equal to the allowed resolution level (Chun & Atluri, 2000).

The third multimedia security related research work that I discuss took place at the Fraunhofer Institute for Computer Graphics, where two systems have been developed which are used to provide multimedia security; TIE (Tool for Image Encryption) provides a form of encryption for images, and SysCoP (System for Copyright Protection), which is used to enforce intellectual property rights protection for multimedia.

TIE is used to partially encrypt an image in a manner that will allow a potential user to view the image at a lower quality (so that it is still recognisable), while at the same time the user would have to decrypt the image to its original form in order for it to be of any practical value. This will for example allow a company or individual to sell images remotely, while at the same time it will allow the buyer to 'preview' the image before agreeing to buy it. The quality of a certain area of the original image is decreased, and that area of the original image is encrypted using the standard block cipher DES. The encrypted data is then integrated into the transmitted image in a manner that does not visibly show when viewing the image (e.g. in an Application Extension Block of a GIF image). In order to restore the quality of the encrypted image to its original form, the user must receive the decryption key from the source of the image (which will be considered a purchase), allowing the user to decrypt the integrated data in the (semi-encrypted) image and replace the lower quality area of the transmitted image with the original area. The original image therefore does not need to be re-transmitted; only the decryption key needs to be transmitted upon agreement.

SysCop is used to embed an invisible watermark onto the multimedia data. The watermark contains copyright information which can be used to prove intellectual property rights to the multimedia data. Other information can also be included in the watermark, including details concerning the buyer. This information can be used to identify the source of illegal copies of the multimedia data. The SysCop watermark is robust and cannot be detected or removed without using the relevant secret key (Storck & Koch, 1997).

Although the above two systems improve the security of multimedia data, practical usage of such systems in a distributed network require a secure protocol for the identification and authentication of the user. Storck & Koch (1997) describe a method of controlling access to image data transferred via the World Wide Web, which was implemented as a plug-in for the Netscape browser. The first step

when using this method is to prepare a secure web server which will contain all the images in their original form, including associated image information such as access rights, pricing, etc. A user wishing to access these images must first register with the web server and provide a public key that will be used to encrypt the desired images. The user will install a plug-in on his/her machine, and all requests will be made through the plug-in. When an image request is made, the user is authenticated via the plug-in (to ensure access to the user's private key), after which the image request, user name and a time stamp is encrypted (using the private key) and sent to the web server. The user is authenticated when the web server correctly decrypts the message with the user's public key, and the time stamp is valid.

Once the user is authenticated, the desired image is embedded with a watermark (containing user details) using the SysCop system, and then encrypted using the TIE system (using a randomly generated key). The encrypted image is then sent to the plug-in together with access rights and cost information. At this point, the plug-in displays the semi-encrypted image, as well as any other image related information. The plug-in ensures that the user performs actions that are within his/her access rights e.g. decrypting or printing the original image. When the user wants to perform a certain action (such as decryption), the request is sent to the web server, which will check for the relevant access rights, register the purchase, and send the decryption key (encrypted with the user's public key) to the plug-in software. The plug-in will then decrypt the image to its original form and display the image (or perform any other authorised action that was requested).

Actions such as printing and saving the partially encrypted image can be illegally performed by reading the image data directly from memory (even though there may be only a limited amount of practical use in doing so). The printed or illegal image is still however watermarked with the customer information, and can later be used to prove who made the illegal copies of the image (Storck & Koch, 1997).

Lastly, Ioannou, Moschovitis, Ntalianis, Karpouzis & Kollias (2000) explain a method of user authentication that takes place between a web browser and a web server before any multimedia data can be transferred to the client/browser. A logon screen produced by the web server is presented to the browser, which contains the user name and password fields together with a randomly generated number (generated each time the initial logon screen is requested). The browser then calculates the MD5 digest of a string containing the password, user name, and random number, and sends this digest to the web server (together with the user name in plaintext). The server then calculates the same digest using the

corresponding data retrieved from its user database, and compares it with the digest calculated by the web browser. If they match, the user is authenticated, allowing any request for multimedia data from that browser to be fulfilled.

5.4 Conclusion

A more powerful multimedia security system can be modelled and developed using a combination of traditional techniques (such as encryption) together with more recent techniques (such as watermarking and authorisation models based on multimedia characteristics). Although the main research focus concerning multimedia security has been on the theory and development of watermarks, some efforts have recently been made to ensure the security of multimedia data in other areas of this domain.