基於浮水印技術且具不可偽造性及不模糊性的著作權保護機制

黃少達　　丁培毅　　吳宗杉

國立臺灣海洋大學

10257040@mail.ntou.edu.tw　pyting@mail.ntou.edu.tw　ibox456@gmail.com

***摘要*—本文改進過往運用浮水印技術保護數位影像著作權的方案，提出可抵抗惡意偽造、惡意移除及協定攻擊的著作權保護機制，浮水印嵌入時運用原始圖片的數位簽章計算出不可預測的虛擬亂數序列，其中一部分作為浮水印，剩餘部分作為嵌入及擷取浮水印的鑰匙。此方法除了建立浮水印與著作權所有人以及被保護的圖片之間的唯一關聯，具有不可偽造的性質，即便浮水印遭受惡意破壞，仍然能夠通過驗證，提供有效的版權證據。本文分析以浮水印作為著作權證明工具時所需要滿足的安全性質，提出浮水印不可偽造性(WUF-CIA)以及不模糊性(NA)兩個安全性定義以建立著作權保護所需要的證據性，並且證明所提出的方法滿足這兩項安全性，如此數位浮水印才能夠成為具有法律效力的數位證據。**

***關鍵詞*—數位浮水印、著作權保護、數位簽章、虛擬亂數序列**

一、前言

隨著電腦科技以及網際網路的快速發展，大量的影像、音訊、視訊等多媒體資料以數位的形式發佈與儲存，同時因為訊號處理技術的進步，這些數位媒體的修改與重製也越來越容易，在這樣的環境下，數位媒體的著作權保護一直是非常重要的議題。其中以不可察覺的數位浮水印技術保護多媒體資料的著作權已經研究多年，大部分研究成果著重於如何加強數位浮水印的強健性：使被嵌入的浮水印在經過多重或是多種訊號處理程序後，依然能夠穩定地擷取與驗證。然而自從 Craver 等人[8]提出逆向攻擊(Inversion attack)、Kutter 等人[12]提出浮水印拷貝攻擊(Watermark copy attack)以後，如果沒有妥善設計的協定，再強健的浮水印嵌入方法在遭受此等攻擊時，所偵測到的浮水印將不足以被採信為合法擁有者的著作權證明。

Craver 等人在[8]中指出，如果使用的浮水印方法是可逆的(Invertible)，就存在有效率的攻擊，使得一張數位影像可以同時偵測出兩個浮水印，攻擊者能夠和真正著作權擁有人一樣提供有說服力的證據，導致著作權歸屬無法判定的僵局。因此既有的浮水印方法，例如 Cox 等人的方法[7]，即便以實驗展示優越的強健性，要作為證明著作權的工具還有相當的距離。Craver 等人雖然提出一個方案試圖解決這樣的問題：使用單向雜湊函式計算原始圖片的雜湊值作為虛擬亂數產生器的種子，產生均勻分佈的序列作為嵌入的浮水印，此機制的浮水印是單向地由原圖產生的，因此浮水印產生方法是不可逆(Noninvertible)的，但是 Craver 等人並沒有提出正規的證明。

在 Craver 等人的研究之後，以浮水印技術解決著作權歸屬爭議的研究主要朝向兩個方向發展：其一跟隨 Craver 等人的腳步，將目標放在設計不可逆的浮水印嵌入機制，主要是修改既有強健的浮水印方法，在產生與嵌入過程中加入密碼學的元件如單向函式、加密系統或數位簽章系統。例如 Ramkumar 等人[15]攻擊 Craver 等人在[9]中提出的方法，並提出了一個改進的方法。Qiao 等人[14]使用加密系統來產生浮水印，將原圖轉換到頻率域後的係數作為訊息，以擁有人的密鑰進行對稱式加密得到浮水印，密鑰及原圖也用來擷取浮水印及驗證所有權，他們更進一步證明所有在驗證著作權歸屬時不使用原圖的方法都是可逆的。

Adelsbach 等人[2][3]正式地定義逆向攻擊，也定義更一般化的模糊攻擊(Ambiguity attack)：攻擊者得到任意的一張圖片 I，不論其中是否已經含有浮水印，只要能夠計算得到新的浮水印以及新的原圖，使得在 I 中可以偵測到這個新的浮水印則攻擊成功，逆向攻擊是模糊攻擊的一個特例，Adelsbach 等人以及後續研究者都以設計抵抗模糊攻擊的浮水印嵌入方法為目標。Adelsbach 等人在[2]中指出浮水印方法的偽陽性偵測率(False-positive detection rate)對利用不可逆浮水印方法解決著作權爭議時的負面影響 -- 若偽陽性偵測率不能降低到可忽略的大小，則模糊攻擊是可能的，逆向攻擊的風險亦隨之存在。

Kutter 等人提出浮水印拷貝攻擊[12]：攻擊者得到一張已經嵌入浮水印的圖片，即便不知道如何嵌入浮水印，仍然可以分析該圖片的統計特性，將其中的浮水印移植到另一張完全不相關的圖片中。拷貝攻擊和模糊攻擊被合稱為協定攻擊(Protocol attack)，它們的共通點在於攻擊者不需要知道如何嵌入浮水印，目標不是移除嵌入的浮水印，得到一張乾淨的圖片來作為非法用途，而是阻礙著作權歸屬的判定，使得浮水印的證據性受到質疑。

Li 等人[13]使用密碼學的虛擬亂數產生器來產生浮水印，並證明該方法足以抵抗模糊攻擊。他們的方法要求虛擬亂數的種子不能由原圖計算得出，如此浮水印和原圖在統計上是不相關的，因此「浮水印必須根據原圖產生」並

不是達成不可逆性的必要條件。

第二個研究路線則是引入公正可信賴的第三方，運用登記註冊的方式或者借助時戳伺服器的幫助來釐清著作權爭議。Adelsbach 等人在[3]中提出由原圖擁有人向第三方申請嵌入浮水印，第三方運用其私鑰對於原圖及擁有人指定的鑰匙等訊息進行數位簽章，作為浮水印的主體，並證明該方法能抵抗模糊攻擊與拷貝攻擊。然而這種方法需要假設高頻寬、高容量、高運算能力的可信賴第三方的存在，能夠儲存並確保所有原擁有人之註冊資料的安全，並且能夠進行有效率的查詢，在解決爭議時仰賴它提供充分的資訊，實務成本相當高而使得可行性低，甚至會衝擊以浮水印解決著作權爭議的根本動機。另外由於數位簽章本身是脆弱的，任何一位元的改變都使得數位簽章無法驗證，因此將數位簽章本身作為浮水印就使得攻擊者有機可乘，能夠輕易地破壞浮水印。

上述這兩條研究路線的成果，都還沒有完全解決問題，還沒能使數位浮水印具備足夠的證據力，在法庭中作為有效的證據。關鍵原因在於從圖片中偵測到的浮水印，其本身具有的意義不足，無法藉由它的存在指向唯一的擁有者以及被保護的影像。就如同 Adelsbach 等人在[4]中所提到的困境，如果一張圖片引發了著作權爭議，但浮出檯面爭取著作權的人都不是真正的著作權擁有人，那麼按照以往將著作權判給其中最具說服力的競爭者的解決方法，將無法真正解決著作權歸屬問題。Adelsbach 等人因此引入可信賴的第三方來解決這個困境，確保爭議發生時，必然能讓真正的擁有人涉入其中。

本文提出一個可以用密碼學方法證明的數位浮水印機制，針對浮水印應用於著作權爭議所需要的證據性提出兩個嚴謹的定義：浮水印不可偽造性 (WUF-CIA) 以及不模糊性 (NA)，並證明所提出的方法滿足這些安全性定義。如此可提供數位影像著作權歸屬的明確證明機制，使得數位影像如在未授權的情況下遭到修改、重製並非法使用時，只要與原作品難以區辨，原始擁有者可以在法庭上提供直接有效的證據以證明其著作權，釐清合法與非法之使用。

本機制中合法擁有人對原始影像進行數位簽章，接著將此不可偽造的數位簽章轉化為無法預測的虛擬亂數序列，一部份序列當作浮水印本體，其餘序列當作嵌入鑰匙來決定浮水印本體嵌入的位置。密碼學的虛擬亂數序列在單向函式存在的假設下是計算上不可預測的，如果一個虛擬亂數產生器以一串均勻分佈且隨機的 λ 個位元之序列當作種子，得到的虛擬亂數序列將具有如下的不可預測性：對任意機率式多項式時間的敵人 $\mathcal{A}$ 而言，在 $\mathcal{A}$ 看到輸出序列的前 i 個位元之後，成功預測第 i+1 個位元是 0 或 1 的機率只比 1/2 高出一個可忽略的函數值 $\varepsilon(\lambda)$。利用足夠長的虛擬亂數序列作為浮水印以及嵌入密鑰，再搭配固定且公開的浮水印嵌入演算法，我們可證明對任意機率式多項式時間、沒有看過該浮水印序列的敵人而言，偽造一張圖片，其中可擷取出和該浮水印序列有超過 50%一定比例(如 60%)的相同位元的機率是可忽略的。因此驗證者有接近 100%的信心得到下述推論：如果能夠從一張爭議圖片中擷取出和原作者嵌入其原始圖片中，由數位簽章衍生出的虛擬亂數序列浮水印 δ 百分比相同的位元，並且此爭議圖片相對原作者之原圖的 PSNR 值夠高，那麼此爭議圖片必然是由作者所公開使用、具有浮水印之原圖修改而來，是一個未授權的、侵犯原作者著作權的非法使用案例。不模糊性的安全性定義一方面可以說是針對模糊攻擊而來的，然而也補足了上面的安全性保證，最主要希望保證攻擊者不但無法偽造出任何原始擁有者的浮水印，也無法聲稱任何一張圖片中具有攻擊者的浮水印，不會因此造成著作權歸屬判斷的阻礙。

本文方法與過往方法的主要差異包括：第一，本文的方法利用原圖以及著作權擁有人的私鑰來產生浮水印並由擁有人自行嵌入，毋須可信賴第三方協助進行嵌入或維持登錄資訊，僅需要假設公開金鑰基礎建設(Public key infrastructure)，以確保參與爭取著作權的人必定包含真正的著作權擁有人，並透過可信賴第三方來偵測浮水印。此機制保證若能偵測出特定浮水印，則此浮水印唯一地對應於特定擁有人及特定原始圖片。第二，傳統的浮水印偵測標準是去計算擷取出來的浮水印和原本所添加的浮水印之間的相似度，運用例如相關係數(Correlation coefficient)等方法來比較相似度，並且經由大量的實驗測試，運用統計的結果訂出門檻值，以超過門檻值表示有浮水印，低於則沒有。然而，這門檻值會因為被嵌入的浮水印不同，還有嵌入的載體圖片(實驗時使用的測試圖片)本身內容不同，而成為一個浮動的標準。也就是說，對於不同的原始圖片擁有人而言，「自己的浮水印出現在某圖片中」這件事的評斷標準都不一樣，即便是同一個人，當所嵌入的浮水印不同時，標準同樣會隨之變動。相對地，本文的方法運用固定的門檻值來判定圖片中是否存在浮水印，並且使得發生誤判的機率降到計算上可以忽略的大小。第三，我們由密碼學的角度來建構本文的系統並分析系統安全性，數位影像中包含浮水印的狀態是公開的，浮水印的產生及嵌入、偵測演算法也是公開的，並在安全性定義與證明中提供敵人浮水印嵌入引擎，以期建立一個可證明的、足以作為執法者判決依據的著作權保護系統，讓此種數位浮水印如同數位簽章一般成為一個具有法律效力的工具。

本文第二節敘述相關的密碼學背景知識，第三節描述所提出的著作權保護方法，第四節提出正式的安全性定義並證明本文的方法滿足這樣的安全性，第五節則是結論。

二、背景知識

本節簡介所使用之密碼學相關背景。

2.1 數位簽章及其不可偽造性

數位簽章系統包含三個演算法：(1)金鑰產生演算法 $KeyGen(1^{\lambda})$ ：輸入安全參數 λ ，輸出金鑰對 (PK, SK) ；(2)簽署演算法 $Sign(SK, m)$ ：輸入密鑰 SK 以及訊息 m ，輸出 m 的簽章 σ ；(3)驗證演算法 $Verify(PK, m, \sigma)$ ：輸入公鑰 PK 、訊息 m 以及簽章 σ ，當 σ 是 m 的合法簽章時輸出 1，否則輸出 0。

安全的數位簽章要求在選擇訊息攻擊下，任何敵人不得偽造出可通過驗證程序之簽章(Existentially unforgeable under chosen message attack, EUF-CMA) [11]。以下為挑戰者 C 和惡意的攻擊者 $\mathcal{A}$ 互動的賽局，$\mathcal{M}$ 為訊息空間：

- 起始階段：C 執行 $KeyGen(1^{\lambda})$ 產生 (PK, SK)，並將 PK 交給 $\mathcal{A}$。
- 詢問階段：$\mathcal{A}$ 可以多次地詢問(共 q_s 次)任意訊息 $m^{(j)} \in \mathcal{M}$ 的簽章，C 則執行 $Sign(SK, m^{(j)})$ 將對應的簽章 $\sigma^{(j)}$ 給 $\mathcal{A}$。
- 輸出階段：$\mathcal{A}$ 輸出一組訊息與簽章 (m^*, σ^*)。若 $m^* \notin \{m^{(j)}\}_{j=1,...,q_s}$ 且 $Verify(PK, m^*, \sigma^*) = 1$，則 $\mathcal{A}$ 贏得此賽局。

$\mathcal{A}$ 之優勢 $Adv_{\mathcal{A}}^{EUF-CMA}$ 定義為：$Adv_{\mathcal{A}}^{EUF-CMA}(1^{\lambda}) = \Pr\left[Verify(PK, m^*, \sigma^*) = 1 \text{ 且 } m^* \notin \{m^{(j)}\}_{j=1,...,q_s}\right]$。對於一個數位簽章系統，如果任意機率式多項式時間的敵人 $\mathcal{A}$ 贏得上述賽局的優勢是可忽略的，則此簽章系統是 EUF-CMA 安全的。

2.2 虛擬亂數產生器及其不可預測性

密碼學的虛擬亂數產生器 $G(s)$ 是一個確定式多項式時間的演算法，輸入一個均勻隨機分佈、λ 位元的種子 s，產生一串 $\ell(\lambda)$ 位元的序列，此序列與真正均勻分佈的 $\ell(\lambda)$ 位元的序列是計算上不可分辨的，其中 λ 是安全參數，多項式函數 $\ell(\lambda)$ 大於 λ。正式的安全性定義如下：對於任意機率式多項式時間的分辨演算法 $\mathcal{D}$、任意正多項式 $p(\cdot)$、及任意足夠大的整數 λ 而言，$\left|\Pr\left[\mathcal{D}(G(U_{\lambda})) = 1\right] - \Pr\left[\mathcal{D}(U_{\ell(\lambda)}) = 1\right]\right| < \frac{1}{p(\lambda)}$，其中 U_{λ} 和 $U_{\ell(\lambda)}$ 分別為均勻分佈的 λ 位元序列和 $\ell(\lambda)$ 位元序列的隨機變數。

此外虛擬亂數產生器的輸出是不可預測的，對於任意機率式多項式時間的預測演算法 $\mathcal{A}$、任意正多項式 $p(\cdot)$、及任意足夠大的整數 λ 而言，$\Pr[\mathcal{A}(G(U_{\lambda}))_{1..k} = \text{next}_{\mathcal{A}}(G(U_{\lambda}))] < \frac{1}{2} + \frac{1}{p(\lambda)}$，其中函數 $\text{next}_{\mathcal{A}}(\cdot)$ 定義如下：當 $\mathcal{A}$ 讀入 $G(U_{\lambda})$ 的前 k 個位元時，$\text{next}_{\mathcal{A}}(\cdot)$ 為 $G(U_{\lambda})$ 的第 $k+1$ 個位元；當 $\mathcal{A}$ 讀入全部 $\ell(\lambda)$ 個位元後，$\text{next}_{\mathcal{A}}(\cdot)$ 為均勻隨機挑選的一個位元[10，章節 3.3.5]。

三、系統建構

本文提出「基於浮水印技術之可證明著作權保護機制」，包含以基於因數分解的單向函數為基礎的三個演算法：$(WSetup, Embed, Detect)$。$WSetup(1^{\lambda})$ 是機率式的系統初始化演算法，輸入安全參數 λ，得到可公開的參數 PK 以及嵌入浮水印用的秘密鑰匙 EK；$Embed(PK, EK, I)$ 為嵌入浮水印的演算法，先用 PK 與 EK 產生唯一的浮水印，再將該浮水印嵌入至原圖 I，得到含有浮水印的新圖像 I_w 與提取浮水印所需的秘密鑰匙 $XK_I = (I, \sigma_I)$，其中 σ_I 是運用 EK 針對原圖 I 產生的數位簽章；$Detect(PK, XK_I, I_{\alpha})$ 運用 PK 及 XK_I 來偵測圖片 I_{α} 是否含有原圖 I 之浮水印，詳述如下：

***WSetup*(1^{λ})**：

(1) 首先產生RSA簽章機制之參數：隨機選擇兩個長度為 $\lambda/2$ 位元的質數 p_1、q_1，計算模數 $N_1 = p_1 \cdot q_1$ 以及 $\phi(N_1) = (p_1 - 1)(q_1 - 1)$，並選出和 $\phi(N_1)$ 互質的驗證指數 e，計算簽章用的指數 $d \equiv e^{-1} (\text{mod } \phi(N_1))$。得到驗證用的公鑰 (N_1, e) 與簽章私鑰 d。

(2) 挑選BBS[5]虛擬亂數產生器(PRG)之參數：挑選Blum整數 $N_2 = p_2 \cdot q_2$，其中 p_2 與 q_2 是 $\lambda/2$ 位元且滿足 $p_2 \equiv q_2 \equiv 3 \pmod 4$ 的隨機質數。N_2 定義單向Rabin函式 $f_{N_2}(x) = x^2 \pmod{N_2}$，$f_{N_2} : QR_{N_2} \to QR_{N_2}$，其中 QR_{N_2} 是在 $Z_{N_2}^*$ 中二次剩餘數(平方數)的集合，$G_{f_{N_2}} : \{0,1\}^{\lambda} \to \{0,1\}^{k\lambda}$ 定義為 $G_{f_{N_2}}(s) = LSB(f_{N_2}(s)) \| LSB(f_{N_2}^{\ 2}(s)) \| ... \| LSB(f_{N_2}^{\ k\lambda-1}(s))$，$LSB(\cdot)$ 表示最低位元(Least significant bit)，$x \| y$ 表示字串序列 x 串接字串序列 y，s 為 λ 位元的亂數種子。

(3) 選擇一個抗碰撞的雜湊函式 $H(\cdot)$。

(4) N_1、e、$G_{f_{N_2}}(\cdot)$、$H(\cdot)$ 組成 PK，d 是嵌入浮水印的鑰匙 EK，演算法輸出 (PK, EK)。

***Embed*(PK, EK, I)**：

(1) 首先計算原圖 I 的數位簽章 $\sigma_I = H(I)^d (\text{mod } N_1)$。

(2) 以 σ_I 作為亂數產生器的種子，產生 $k\lambda$ 位元的虛擬亂數 $w_I[1, k\lambda] = G_{f_{N_2}}(\sigma_I)$，其中前 λ 位元 $w_I[1, \lambda]$ 為浮水印，其它 $(k-1)\lambda$ 位元 $w_I[\lambda+1, k\lambda]$ 為嵌入浮水印的秘密鑰匙。

(3) 對原圖 I 進行一階離散小波轉換，分成四個部分 $DWT(I) = (LL, LH, HL, HH)$，由於低頻的 LL 中存放著圖片中較為重要的資訊，故將浮水印 $w_I[1, \lambda]$ 嵌入 LL，使得浮水印受到暴力移除破壞時，圖片品質得以顯著下降。

(4) 將 LL 分割成 $\ell = 2^{k-1}$ 個區塊(每一區塊至少包含 λ 個像素)，將嵌入鑰匙 $w_I[\lambda+1, k\lambda]$ 分成 λ 個段落，每個段落 $k-1$ 個位元，用來指定浮水印 $w_I[1, \lambda]$ 中的第 i 個位元 $w_I[i]$ 要隱藏在 ℓ 個區塊中的哪一個區塊。步驟如下：對浮水印中每一位元 $w_I[i]$, $i = 1, ..., \lambda$，以 $k-1$ 個位元的序列 $w_I[\lambda+i] \| w_I[2\lambda+i] \| ... \| w_I[(k-1)\lambda+i]$ 表

示的二進位值blk為區塊資訊，用浮水印的第 i 位元 $w_I[i]$ 取代像素 $LL[blk][i]$ 的第 β 位元，得到嵌入浮水印後的低頻部分 LL'。以虛擬碼表示如下：

```
Insert(w_I, LL) {
    for (i=1; i<=λ; i++) {
      blk = w_I[λ+i] || w_I[2λ+i] || ... || w_I[(k-1)λ+i];
      LL[blk][i]_β = w_I[i];
    }
    return LL;
}
```

上述演算法中，$LL[blk][i]_\beta$ 的第一個索引代表 ℓ 個區塊中的第 blk 個區塊，第二個索引則代表該區塊中的第 i 個像素，下標 β 則代表該像素中的第 β 位元。如果影像 I 是八位元灰階圖片，放在第 β 位元的浮水印相當於振幅 $\{2^{\beta-1}, 0, -2^{\beta-1}\}$ 的雜訊，其中 $1 \le \beta \le 8$。在肉眼無法辨識的前提下，浮水印應該要嵌入在圖片比較重要的地方以抵抗各種影像處理，故由實驗決定 β，在我們的實驗中，$\beta = 6$ 是一個效果較好的選項。由於只在 ℓ 個像素中挑選其一放入浮水印，攻擊者若欲完全移除浮水印，就需將所有 ℓ 個像素的第 β 位元清除，如此勢必造成圖片品質大幅下降。如果運用比較智慧的方法，品質可能不會變那麼差，但是只要影像不是空白無意義的，由於敵人沒有原始圖片也沒有簽章和由簽章導出的浮水印，要完全移除乾淨是非常困難的。

(5) 得到 LL' 之後，進行反離散小波轉換 $IDWT(LL', LH, HL, HH)$ 轉回空間域，得到嵌入浮水印之圖片 I_w。

(6) 演算法輸出 $(I_w, XK_I = (I, \sigma_I))$。$I_w$ 為最後可公開使用的圖片，往後若有爭議，原圖 I、簽章 σ_I 就是偵測浮水印的私鑰 XK_I。

***Detect*$(PK, I, \sigma_I, I_\alpha)$**：

(1) 首先執行簽章驗證演算法驗證 $H(I)$ 是否等於 $\sigma_I^e \pmod{N_1}$。通過驗證則繼續執行下一步驟，否則輸出 0。

(2) 產生 $k\lambda$ 位元的虛擬亂數 $w_I[1, k\lambda] = G_{f_{N_2}}(\sigma_I)$。

(3) 使用 $DWT(\cdot)$ 進行一階離散小波轉換把 I_α 轉換成 (LL, LH, HL, HH)。

(4) 根據 $w_I[\lambda+1, k\lambda]$ 所決定的位置，由 LL 中取出 $w_{I_\alpha}[1, \lambda]$ 準備隨後與 $w_I[1, \lambda]$ 進行比對。取出 $w_{I_\alpha}[1, \lambda]$ 的過程以虛擬碼表示如下：

```
Extract(w_I, LL) {
    for (i=1; i<=λ; i++) {
      blk = w_I[λ+i] || w_I[2λ+i] || ... || w_I[(k-1)λ+i];
      extractedWatermark[i]=LL[blk][i]_β;
    }
    return extractedWatermark;
}
```

(5) 以漢明距離(Hamming distance)計算代表位元相似度的正規化交叉相關函數(Normalized cross correlation)，$NCC(w_I[1,\lambda], w_{I_\alpha}[1,\lambda]) = 1 - \frac{1}{\lambda}\sum_{i=1}^{\lambda}\left(w_I[i] \oplus w_{I_\alpha}[i]\right)$，相關值高代表兩張圖所提取出的浮水印位元串相似性高。如果 $\left|NCC(w_I[1,\lambda], w_{I_\alpha}[1,\lambda]) - \frac{1}{2}\right| \ge \frac{2}{\ell}$ 則演算法輸出 1，否則輸出 0。

四、安全性定義與證明

4.1 浮水印不可偽造性

這是基於「不可偽造對應特定使用者以及特定圖片之浮水印」的概念來定義，簡稱為 WUF-CIA(Watermark unforgeability under chosen image attack)。

定義一 (WUF-CIA 安全性)：

圖 1 是一個挑戰者 C 及攻擊者 $\mathcal{A}$ 之間的賽局：

1. **環境設置階段**：C 以安全參數 λ 執行 *WSetup* 演算法產生公開參數 PK 以及嵌入密鑰 EK，同時決定影像空間 I，將 PK 及 I 傳送給 $\mathcal{A}$，嵌入密鑰 EK 為 C 的秘密。
2. **詢問階段**：C 提供浮水印嵌入引擎，讓 $\mathcal{A}$ 可多次(共 q_s 次)任意選擇影像空間中的影像 $I^{(j)}$ 交給 C，C 執行 $Embed(PK, EK, I^{(j)})$ 得到添加浮水印 $w_{I^{(j)}}$ 的圖片 ${I^{(j)}}_w$ 和對應的簽章 $\sigma_{I^{(j)}}$，並將 $({I^{(j)}}_w, \sigma_{I^{(j)}})$ 送回給 $\mathcal{A}$，讓 $\mathcal{A}$ 能夠自行擷取浮水印驗證。
3. **輸出階段**：$\mathcal{A}$ 輸出一張圖片 I 以及具有偽造水印的圖片 I^*，C 確認 $I \notin \{I^{(j)}\}_{j=1,\ldots,q_s}$ 後，以密鑰 EK 計算簽章 σ_I，執行 $Detect(PK, I, \sigma_I, I^*)$ 以檢查 I^* 中是否存在 C 針對 I 產生的浮水印，當 *Detect* 演算法輸出 1 時，$\mathcal{A}$ 贏得這個賽局。

攻擊者優勢為 $Adv_{\mathcal{A}}^{\text{WUF-CIA}}(1^\lambda) = \Pr[Detect(PK, I, \sigma_I, I^*) = 1]$。如果一個浮水印方法在上述的賽局中，使得任意機率式多項式時間的敵人 $\mathcal{A}$ 贏得賽局的優勢 $Adv_{\mathcal{A}}^{\text{WUF-CIA}}(1^\lambda) = negl(\lambda)$，$negl(\cdot)$ 表示一個可忽略的函數，則稱這個浮水印方法是 WUF-CIA 安全的。

當一個浮水印方法滿足 WUF-CIA 安全性，則針對一位使用此方法的數位版權所有者 C 來說，任意計算能力有限、不知道嵌入密鑰 EK 的攻擊者偽造出一張可偵測出 C 嵌入在圖片 I 之浮水印 w_I 的可能性是計算上可忽略的，也就是說針對 I 而言，浮水印 w_I 只有 C 有能力製作出來，因此如果在某張圖片中偵測到 w_I，即可明確且唯一地指出著作權爭議的參與對象是 C 而宣稱被保護的原圖正是 I，並且也能夠確實地對應這個參與對象 C 在現實世界的身分。此外如果一個數位浮水印方法滿足此安全性定義，則該方法的偽陽性偵測率是可忽略的，也就是計

算上不可能在一張沒有嵌入特定浮水印的圖片中偵測到該浮水印。

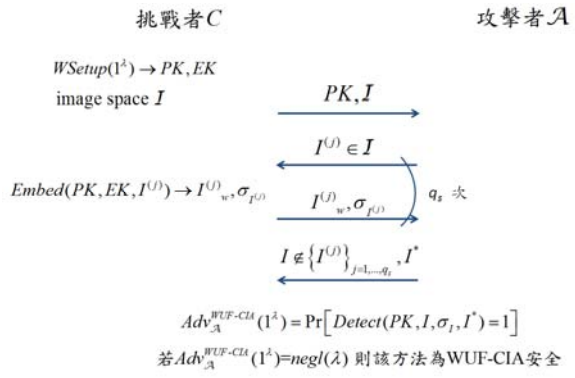

圖 1　WUF-CIA 安全性賽局與定義

定理一：

如果下列三項條件成立，則第三節所提出的方法是 WUF-CIA 安全的：(1) 所使用的雜湊函式 $H(\cdot)$ 具有亂度平滑性 (Entropy smoothing) [16]，(2) 所使用的數位簽章系統是 EUF-CMA 安全的，(3) 所使用的虛擬亂數產生器滿足 2.2 節所定義的虛擬亂數性質。

由於本文使用 RSA 簽章機制以及 BBS 虛擬亂數產生器，此二者在因數分解的假設下皆已證明安全，因此只有第(1)點是本方法需要的額外假設：

假設一 (雜湊函式之亂度平滑性)：

雜湊函式 $H(\cdot):\{0,1\}^* \to \{0,1\}^\lambda$ 在滿足下列條件時被稱為具有亂數平滑性：對於任意機率式多項式時間的分辨演算法 $\mathcal{D}$、任意輸入 x 而言，$\left|\Pr\left[\mathcal{D}\left(H\left(x\right)\right)=1\right]-\Pr\left[\mathcal{D}\left(U_\lambda\right)=1\right]\right| = negl(\lambda)$，其中 U_λ 為均勻隨機分佈的 λ 位元序列的隨機變數，$negl(\cdot)$ 是一個可忽略函數。

證明. 由於篇幅限制，詳細證明請見[1]。 □

請注意，滿足 WUF-CIA 尚不足以提供完整的浮水印「證據性」，還缺少一個重要的性質 -- 就是沒有保證經由 *Detect* 演算法偵測到的浮水印，都是由 *Embed* 演算法嵌入的。這個性質將在下一小節不模糊性的定義與證明中顯現出來，藉由下一小節的證明，能夠保證只有透過 *Embed* 演算法才能產生可成功偵測的浮水印。

4.2 不模糊性

本節將證明第三節的方法足以抵抗模糊攻擊。模糊攻擊是指攻擊者能夠宣稱在一張圖片中偵測到一個從未被嵌入過的浮水印。過往研究都試圖要防制這個關鍵性的攻擊，以得到可以順利排解著作權糾紛的協定。以下參考 Adelsbach 等人[3]的定義，並根據本文提出的方法稍作修改，提出「不模糊性」安全性定義，簡稱 NA (Non-ambiguity) 如下：

定義三 (NA 安全性)：

圖 2 是一個挑戰者 C 及攻擊者 $\mathcal{A}$ 之間的賽局：

1. 環境設置階段：$\mathcal{A}$ 輸入安全參數 λ 執行 *WSetup* 演算法，產生公開參數 *PK* 以及嵌入密鑰 *EK*，同時決定影像空間 I，將 *PK* 及 I 傳送給 C，嵌入密鑰 *EK* 只有 $\mathcal{A}$ 知道。
2. 影像決定階段：C 任意選擇影像空間 I 中的圖片 I 交給 $\mathcal{A}$，指定此圖片為稍後 $\mathcal{A}$ 必須宣稱為其所有之「已嵌入 $\mathcal{A}$ 之浮水印」的圖片。
3. 輸出階段：$\mathcal{A}$ 輸出一張圖片 I^* 以及對應的簽章 σ_{I^*} 當作是 I 的原圖和被嵌入的浮水印，C 則執行 $Detect(PK, I^*, \sigma_{I^*}, I)$ 以檢查 I 中是否存在 $\mathcal{A}$ 針對 I^* 產生的浮水印，當 *Detect* 演算法輸出 1 時，$\mathcal{A}$ 贏得這個賽局。

攻擊者的優勢 $Adv_{\mathcal{A}}^{\mathrm{NA}}(1^\lambda) = \Pr\left[Detect(PK, I^*, \sigma_{I^*}, I) = 1\right]$。如果一個數位浮水印方法在上述賽局之中，使得任意機率式多項式時間的敵人 $\mathcal{A}$ 所具有的優勢 $Adv_{\mathrm{A}}^{\mathrm{NA}}(1^\lambda) = negl(\lambda)$，則稱此浮水印方法是 NA 安全的。

定理二：

如果所使用的雜湊函式滿足假設一之亂度平滑性，同時所使用的虛擬亂數產生器滿足 2.3 節所定義之虛擬亂數性，則第三節提出的方法是 NA 安全的。

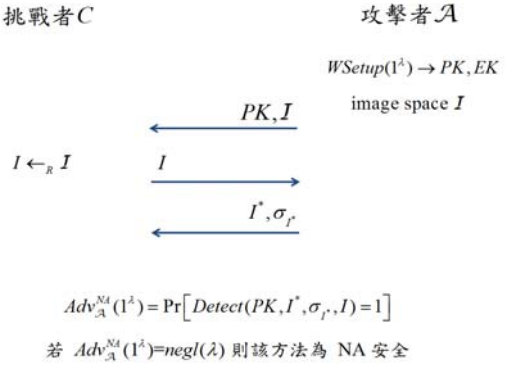

圖 2　NA 安全性賽局與定義

證明. 以下藉由一系列賽局(Sequence of games)替換的方式[16]來證明。

Game 0：上述原始定義的 NA 賽局。我們定義敵人 $\mathcal{A}$ 贏得此賽局之事件的機率 $G_0 = Adv_{\mathcal{A}}^{\mathrm{NA,\ Game\ 0}}(1^\lambda)$。

Game 1：在 NA 賽局的輸出階段中，C 並不直接執行 $Detect(PK, I^*, \sigma_{I^*}, I)$，而是在 *Detect* 演算法的第(2)步驟中，選擇均勻隨機分佈的 λ 個位元亂數 $s \leftarrow_R \{0,1\}^\lambda$ 來當作種子執行 $G_{f_{N_2}}(s)$ 得到序列 R，接著使用 R 繼續執行原本 *Detect* 演算法的第(3)、(4)、(5)步驟：執行 $DWT(\cdot)$ 把 I 分成 (LL, LH, HL, HH)，接著根據 $R[\lambda+1, k\lambda]$ 所決定的區塊，由 LL 中取出 $w_I[1,\lambda]$，計算 $NCC(R[1,\lambda], w_I[1,\lambda])$ 來判定偵測結果。我們定義 $G_1 = Adv_{\mathcal{A}}^{\mathrm{NA,\ Game\ 1}}(1^\lambda)$。

Game 2：在 NA 賽局的第(3)步驟輸出階段中，C 並不直接執行 $Detect(PK, I^*, \sigma_{I^*}, I)$，而是在 *Detect* 演算法的第(2)步驟時，直接選擇均勻隨機分佈的 $k\lambda$ 個位元亂數

$R \leftarrow_R \{0,1\}^{k\lambda}$ 來取代原本經由虛擬亂數產生器 $G_{f_{N_2}}(\cdot)$ 得到的序列，接著使用 R 繼續執行原本 *Detect* 演算法的第(3)、(4)、(5)步驟：執行 $DWT(\cdot)$ 把 I 分成 (LL, LH, HL, HH)，接著根據 $R[\lambda+1, k\lambda]$ 所決定的區塊，由 LL 中取出 $w_I[1,\lambda]$，計算 $NCC(R[1,\lambda], w_I[1,\lambda])$ 來決定偵測結果。我們定義 $G_2 = Adv_{\mathcal{A}}^{\text{NA, Game 2}}(1^{\lambda})$。

機率分析

在 Game 0 與 Game 1 之間，差別只在於虛擬亂數產生器 $G_{f_{N_2}}(\cdot)$ 使用的種子是 $\mathcal{A}$ 對於其輸出圖片 I^* 所簽署的簽章 σ_{I^*}，抑或是均勻隨機分佈的 λ 位元亂數。基於假設一，$H(I^*)$ 和均勻隨機分佈的 λ 位元亂數對於任意機率式多項式時間的演算法而言是不可分辨的，同時 RSA 簽章演算法是一個排列函數，因此可知 σ_{I^*} 和均勻隨機分佈的 λ 個位元亂數對於任意機率式多項式時間的演算法來說是不可分辨的。是故，任意機率式多項式時間的敵人 $\mathcal{A}$ 在 Game 0 的優勢與在 Game 1 的優勢之差異是可忽略的，也就是 $|G_0 - G_1| = negl_1(\lambda)$。

在 Game 1 與 Game 2 之間，差別只在於 *Detect* 的(3)、(4)、(5)步驟中被使用至圖片 I 中擷取浮水印以進行比較的序列 R 究竟是經由虛擬亂數產生器產生的，抑或是真正均勻隨機分佈的 $k\lambda$ 個位元亂數。基於 2.3 節虛擬亂數產生器之安全性定義，虛擬亂數產生器的輸出和均勻隨機分佈的亂數對於任意機率式多項式時間的演算法而言是不可分辨的，因此任意機率式多項式時間的敵人 $\mathcal{A}$ 在 Game 1 的優勢與在 Game 2 的優勢之差異是可忽略的，也就是 $|G_1 - G_2| = negl_2(\lambda)$。

最後在 Game 2 之中，由於在偵測時使用了真正均勻隨機分佈的亂數 R 到圖片 I 中進行擷取並計算 $NCC(R[1,\lambda], w_I[1,\lambda])$ 的值，根據 Chernoff 不等式[6]可以推出對於固定的 ℓ，$\Pr\left[\left|NCC(U_{k\lambda}[1,\lambda], w_I[1,\lambda]) - \frac{1}{2}\right| \geq \frac{2}{\ell}\right]$ 是計算上可忽略的，因此任意機率式多項式時間的敵人 $\mathcal{A}$ 在 Game 2 中的優勢是可忽略的，也就是 $G_2 = negl_3(\lambda)$。

綜合以上分析我們可知，$|G_0 - G_2| \leq |G_0 - G_1| + |G_1 - G_2| = negl_1(\lambda) + negl_2(\lambda)$，因此 $|G_0 - negl_3(\lambda)| \leq negl_1(\lambda) + negl_2(\lambda)$，也就是任意機率式多項式時間的敵人 $\mathcal{A}$ 在 Game 0 中的優勢 $G_0 \leq negl_1(\lambda) + negl_2(\lambda) + negl_3(\lambda)$ 是計算上可忽略的。 □

五、結論

本文運用數位簽章為種子產生不可預測的虛擬亂數序列，製作出與著作權擁有人以及欲保護的原始圖片唯一關聯的數位浮水印，基於虛擬亂數序列的特性，設計可允許破壞的簽章物件，排除浮水印被惡意移除的可能性，並且訂出絕對的偵測門檻，令浮水印的偽陽性偵測率降低到計算上可忽略的大小。本文證明此方法滿足 WUF-CIA 以及 NA 兩安全性，將此兩安全性合併考量，可確保當浮水印以指定使用者的公鑰偵測到的時候，唯一連結到該使用者以及特定的原圖，同時只可能在有嵌入過這浮水印的圖片(也就是嵌入到原圖後得到的圖片以及其衍生物)中偵測到它，在此機制之下，配合具公信力有裁量權的第三方進行著作權爭議解決步驟，便可有效地保障真實擁有者之著作權。

六、參考文獻

[1] 黃少達,"基於數位浮水印技術的可證明著作權保護機制,"國立臺灣海洋大學資訊工程學系碩士學位論文, 國立臺灣海洋大學博碩士論文系統, http://ethesys.lib.ntou.edu.tw, 2015.

[2] A. Adelsbach, S. Katzenbeisser, and A. Sadeghi, "On the Insecurity of Non-invertible Watermarking Schemes for Dispute Resolving," Proc. of IWDW, 2003.

[3] A. Adelsbach, S. Katzenbeisser, and H. Veith, "Watermarking Schemes Provably Secure Against Copy and Ambiguity Attacks," Proc. of the ACM Workshop on Digital Rights Management, 2003.

[4] A. Adelsbach, B. Pfitzmann, and A. R. Sadeghi, "Proving Ownership of Digital Content," Proc. of IH'99, Lecture Notes in Computer Science, Vol. 1768, 117–133, 2000.

[5] L. Blum, M. Blum, and M. Shub, "A Simple Unpredictable Pseudo-Random Number Generator," SIAM Journal on Computing 15 (2): 364–383, 1986.

[6] H. Chernoff, "A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the Sum of Observations," The Annals of Mathematical Statistics, Vol. 23, No. 4, 493–507, 1952.

[7] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure Spread Spectrum Watermarking for Multimedia," IEEE Trans. on Image Processing, Vol. 6, No. 12, 1673–1687, 1997.

[8] S. Craver, N. Memon, B. Yeo, and M. Yeung, "Can Invisible Watermarks Resolve Rightful Ownerships," Technical Report RC 20509, IBM Research Institute, 1997.

[9] S. Craver, N. Memon, B. Yeo, and M. Yeung, "Resolving Rightful Ownership with Invisible Watermarking Techniques: Limitations, Attacks, and Implications," IEEE Journal on Selected Areas in Communications, Vol. 16, No. 4, 573–586, 1998.

[10] O. Goldreich, Foundations of Cryptography: Volume 1, Basic Tools, Cambridge University Press, 2000.

[11] S. Goldwasser, S. Micali, and R. Rivest, "A Digital Signature Scheme Secure against Adaptive Chosen-Message Attacks," SIAM J. Computing, Vol. 17, No. 2, 281–308, 1988.

[12] M. Kutter, S. Voloshynovskiy, and A. Herrigel, "The Watermark Copy Attack," Proc. of SPIE: Security and Watermarking of Multimedia Contents II, Vol. 3971, 2000.

[13] Q. Li and E. Chang, "On the Possibility of Non-invertible Watermarking Schemes," Proc. of IHW'04, Lecture Notes in Computer Science, Vol. 3200, 13–24, 2004.

[14] L. Qiao and K. Nahrstedt, "Watermark Methods for MPEG Encoded Video: Towards Resolving Rightful Ownership," Proc. of ICMCS, Vol. 9, 194–210, 1998.

[15] M. Ramkumar and A. Akansu, "Image Watermarks and Counterfeit Attacks: Some Problems and Solutions," Symposium on Content Security and Data Hiding in Digital Media, 102–112, 1999.