國立臺灣海洋大學

資訊工程學系

碩士學位論文

指導教授：丁培毅

基於數位浮水印技術的可證明著作權保護機制
Provable Watermark-Based Copyright Protection Scheme

研究生：黃少達 撰

中華民國 104 年 6 月

基於數位浮水印技術的可證明著作權保護機制

Provable Watermark-Based Copyright Protection Scheme

研 究 生：黃少達　　Student：Shao-Da Huang

指導教授：丁培毅　　Advisor：Pei-Yih Ting

國 立 臺 灣 海 洋 大 學
資 訊 工 程 學 系
碩 士 論 文

A Thesis
Submitted to the Department of Computer Science and Engineering
College of Electrical Engineering and Computer Science
National Taiwan Ocean University
in partial fulfillment of the requirements
for the Degree of
Master of Science
in
Computer Science and Engineering
June 2015
Keelung, Taiwan, Republic of China

中華民國 104 年 6 月

誌謝

這兩年的研究所生活令我收穫相當豐碩，首先得要感謝丁培毅老師一直以來的提攜教導，老師提供了一個頻繁互動、深度討論的學習環境，並且也以實質的行動為我們示範「發現問題、定義問題、解決問題」與「整理知識、組織知識、表達知識」的方法，令我不僅在密碼學這門直通計算機科學核心的學門中拓展了視野，也學到了學習新事物與做學問的基本功夫，讓我未來面對任何事情都能有所依歸。同時也很感謝吳宗杉老師、林韓禹老師在平時 Seminar 以及課堂之中對我的指導以及建議，我們的 Seminar 對學生來說真是很難得的磨練機會，通過這兩年來在 Seminar 報告的經驗讓我的表達能力與心理素質進步了許多。此外也很感謝長庚大學的許建隆老師在口試時對這篇論文的建議，讓我能從不一樣的角度思考、理解並釐清問題，讓論文內容更完整。

很幸運我來到了資訊安全實驗室，在這邊也得到了許多來自學長姐、同學與學弟的協助，讓我能夠更快地累積相關知識。感謝建偉、秉輝、姝儀跟昱彤在我一年級時帶領我加速理解這塊領域的內容，也為我們立下了良好的學習風範，感謝同學俊宏、鈺博平時的討論、玩樂與陪伴，一起度過這段學習的日子，感謝學弟建豪、柏嶽平時討論的意見與回饋，這些都是我一路成長的養分。

感謝我的家人，這麼多年來持續給予我物質與精神上的支持及鼓勵，也給我足夠多的空間讓我感受這個世界、享受大學及研究所生涯，讓我無後顧之憂順利完成學業。

感謝海洋資工系壘球隊，這一個充滿靈魂的地方。在這裡我真正認識了自己，不斷地自我對話，找到屬於自己的定位，知道了我該前進的方向，學到了對我一生影響重大的態度和精神。這裡是我生命中第二個家、第二個歸屬，伴隨我度過這六年大學及研究所生涯，真的很感謝在系壘相遇的學長姊、同學、學弟妹們，沒有你們就沒有今天的我。

摘要

利用數位浮水印技術提供數位資料的所有權證明 (Proof of ownership) 機制以進行著作權保護，是近二十年來備受關注的議題，本論文接續過往研究，探討強化浮水印的證據性所需要的條件，並且引入密碼學方法，提出一個在數位影像上應用於所有權歸屬爭議 (Ownership dispute resolution) 的數位浮水印方法，將被保護的原始影像進行數位簽章後，再利用此簽章計算出虛擬亂數序列，一部份作為浮水印，一部分作為嵌入浮水印的密鑰，將浮水印以低密度方式嵌入影像中，此方法除了建立浮水印與合法持有者的唯一關聯、建立浮水印與原始影像的唯一關聯之外，不可預測的虛擬亂數序列使得移除浮水印的困難度提高，縱使部份浮水印被破壞，簽章形式轉變後仍然具有不可偽造且可以驗證的特性。本文提出形式化的證明，證明了如果在有抄襲嫌疑的影像中偵測出一定比例的浮水印，這個影像不是源自於原作者所公開使用的已嵌入浮水印影像的機率為計算上可以忽略的。

關鍵詞：數位浮水印、數位版權保護、數位簽章、虛擬亂數序列。

Abstract

Watermark-based copyright protection techniques have been investigated for more than two decades in the signal processing and the digital rights management communities. In this paper, following the previous works, we discuss the requirements of a watermark scheme for providing proof of ownerships, and build our scheme based on previous well developed signal processing techniques but focus on how to employ unpredictable signature-seeded pseudorandom bit sequence to not only establish the unique relation between the watermark, the identity of true owner, and the original cover work, but also make the false positive watermark detection rate computationally negligible. We formally prove that if a valid watermark can be found in a disputed image, the probability that this image is not derived from the true owner's published image is computationally negligible.

Keywords : digital watermark, copyright protection, digital signature, pseudorandom bit sequence.

目次

圖次

第一章 前言

隨著電腦科技和網際網路的快速發展，以及在數位影像、音訊、視訊等訊號處理技術的進步，生產、處理、並且散布數位媒體變得越來越容易而快速，相對地，複製與修改這些數位媒體資料也是非常輕鬆的事情。在這樣的環境下，有越來越多的多媒體資料以數位的形式儲存起來，對於許多藝術創作者來說，更是直接使用電腦與智慧型裝置進行作品的創作與發行，而無須藉由實體管道包裝銷售、再透過特定實體裝置如 DVD 播放器被人們觀賞使用。但也正因為如此的便利與低成本，如何確保作者對這些以數位形式儲存之媒體資料的著作權，自然成為一個十分重要的議題。儘管就法律的層面而言，以中華民國著作權法為例，著作人於著作完成時即享有著作權，然而因為沒有登記註冊的制度，若發生著作權的爭議時，著作權人必須對於自己是著作權人及著作權存在及其存續時間等事項負起舉證責任。一個可能的解決方案是尋求可信賴的第三方進行作品的登記註冊，如此當著作權爭議發生時，能夠藉由第三方所提供之資訊以獲得排解紛爭的保證，然而此方案自然不是一個方便的方法，對於創作者來說也不會是一個經濟的方法，因為無法保證登記註冊的資料會有被使用到的一天，即便成功在爭議之中獲得了損害賠償，也難以保證可與支出的成本相平衡。

人們希望尋求一種簡單、有效的方法，證明自己對於某作品的合法所有權，當作品遭未經授權地複製、修改、使用而產生著作權爭議時，能夠釐清作品來源歸屬，決斷剽竊與否。通常之手段為在作品之中安插可視之文字或圖像訊息，表示出作者或合法擁有者之身分資訊，此方法在作品被複製或公開發表時，很有可能被有意或者無意地移除掉，過於不可靠，因而尋求基於資訊科技的所有權保護機制便成了資訊科學各領域研究者的目標。不可察覺的數位浮水印技術，即為近年來因應此需求而蓬勃發展的可能解決方案之一。

數位浮水印的歷史可追溯到西元 1954 年 [9]，幾十年來被廣泛使用於各種應用方向，例如廣播監控 (Broadcast monitoring)、辨認擁有者 (Owner identification)、所有權證明 (Proof of ownership)、交易追蹤 (Transaction tracking)、資料來源驗證 (Content authentication)、拷貝管控 (Copy control)、裝置控管 (Device control) 等 [10]。近二十年來由於電腦技術和網路的興起，版權物遭到非法盜用的案例越來越普遍，所有權證明 (Proof of ownership) 機制成為了數位浮水印的研究者們試圖解決的一個主要議題，浮水印技術也被大量地應用於數位媒體如影像、音訊、視訊的著作權保護上 [8][10][13][25][26][27][29]。以影像為例，浮水印技術將一些資訊緊密地嵌入一張圖片中，使得被加入的浮水印具有下列兩個重要的基本特性：(1) 不可視性：浮水印在視覺上是不會被察覺的，對於不知情的人而言，無法看出加入浮水印的圖與原圖的差異，以達成資訊隱藏的目的，也因此不會破壞圖片本身的美感。(2) 不可分離性 (又稱強健性)：即便將加

過浮水印的圖片經過各種數位訊號處理、數位影像處理，其中的浮水印不會被完全清除掉，依然能夠擷取出足以識別的資訊。大部分的數位浮水印方法運用數位訊號處理技術，針對數位影像空間域和頻率域的性質、以及編碼方法的性質，配合人類的視覺模型來達成上述的目標。

針對所有權證明機制，本文討論的問題情境如下：我們希望讓數位影像的原作者或合法擁有者，在產生 (獲得) 屬於他的原始圖片以後，可以使用浮水印方法加入專門的浮水印到原始圖片中，得到含有浮水印的一張經過後製的圖片，接著公開使用這張後製過的圖片，比如放置於個人的部落格上。在未來任何時刻，如果在一個未授權的地方發現有人使用一張視覺上和自己原先公開的圖片很相似的圖片，而且能夠從該圖片中擷取出自己當初嵌入的浮水印，那麼他應該要可以向法院提出侵權訴訟，並以其中的浮水印作為證據，證明對方使用的圖片是由他原先公開的圖片修改、重製而成的。

也就是說，被嵌入的浮水印相當於是一個由擁有人對該圖片所附加的所有權標記，我們不只期望能夠透過這個標記來辨認出圖片的合法擁有者是誰，也期望爭議發生時能透過它來證明此版權所有之圖片遭到不法盜用的事實。

因為數位浮水印具有前述「資訊隱藏」的特性，如果憑藉感官、光學或是數位訊號處理機制無法覺察到某張圖片中嵌入了浮水印，在公開管道取得這張圖片的人就沒有嘗試偵測、或者移除所藏之浮水印的動機。因此傳統利用數位浮水印系統進行所有權宣告的最簡單方法，便是以一張標示個人身分的圖片，或者是各種組織或商業機構用以識別的圖騰或註冊商標，作為浮水印嵌入到欲保護的圖片中。只要能夠提供足夠證據顯示所使用的嵌入演算法足以「抵抗」各類訊號處理及影像處理，也就是說，嵌入浮水印的圖片在被他人取得、或傳送的過程中經過這些處理之後，依然能夠透過擷取浮水印的演算法擷取出足以識別的份量，那麼這被擷取出的浮水印就足以表示特定擁有者對此圖片之所有權。

然而如果沒有較為精巧的系統設計，這樣簡單的數位浮水印方法並不足以提供充足的證據，在我們設想的情境之下用以證明合法擁有人的所有權。以下將詳述各項理由。

(1) 基於上述，浮水印方法會設法隱藏所嵌入的資訊，也就是說，使用者期望並假設這個額外附加於圖片的訊息是根本不會被他人所覺察的，當然也不會提供「如何嵌入」的相關資訊。因此傳統的數位浮水印系統，目標不在於證明敵人在不影響影像品質的前提下無法移除浮水印；也不在於證明即便所使用的浮水印方法是公開的，敵人無法偽造一張具有特定浮水印的圖片。這樣的機制在本質上並不是為了抵抗惡意攻擊而設計的，如此必然存在浮水印可被偵測、可被移除 (將於第 2 點詳述)、或可被假造 (將於第 3 點詳述) 的疑慮，再加上浮水印可以被多次加入，且加入的先後順序無法判斷，導致這樣的方法使用在數位媒體的所有權界定爭議時，僅能被視為輔助工具，「協助」擁有者或執法人員判斷一個有爭議物件的來源。

(2) 由於預設了浮水印的存在不會被發覺的立場，許多過往的文獻便把關注的焦點放在如何增進浮水印的強健性上，也就是研究如何更妥善地運用數位訊號處理、影像處理技術，搭配數學工具來藏入浮水印，儘量使得被藏入的浮水印能夠承受更多種類、更高強度的破壞而依然能夠被擷取出來。但是這個研究方向對於所有權的證明機制並無顯著的幫助，因為強健性始終是無法被證明的，我們頂多只能倚靠實驗來模擬盡可能多的破壞方法，並以統計的數據來分析該浮水印嵌入方法的表現，卻不太可能把所有種類的破壞方法都測試完畢並宣稱某個嵌入方法足以承受任意的處理方式，證明被嵌入的浮水印是不可能被移除的。尤其當我們將惡意的攻擊者納入考量時，浮水印的不可移除性將會和被保護的影像內容 (即浮水印的載體) 及攻擊者的智慧有密切的關聯，因此不論嵌入的方法為何，都必然能夠在特定的狀況下被移除，例如將浮水印嵌入在全黑或全白的圖片中。

(3) 承第 2 點所述，即便假設存在一個浮水印方法，可以被證明是絕對強健的，即被嵌入的浮水印無論如何都不可能被移除，事實上依然無法證明擁有人之所有權。原因有二，首先因為在上述的浮水印方法中，被嵌入的浮水印是圖片或註冊商標等資訊，這些圖片本身並非秘密，因此無法保證使用這些圖片的人就是合法擁有人本尊，無法和擁有者緊密地關聯在一起，無法被證明是獨一無二、只有擁有人才能生產出的資訊，甚至由於浮水印圖片很容易取得，自然也有可能被惡意者嵌入在有負面意涵的圖片中，例如具有色情、暴力、誹謗、中傷等內容的圖片，用以陷害特定個人或組織機構。其二，應用此技術的目標是達成數位資料所有權保護，但是這類浮水印的產生過程卻與欲保護之圖片內容是不相干的，那麼倘若圖片中的浮水印被成功地擷取出來，它可以被嵌入另一張完全無關的圖片之中 [17]，依舊表述了和前一張圖片中相同的識別資訊。如此即便被嵌入的浮水印可證明專屬於特定擁有者，浮水印的存在依然不足以被信任為所有權歸屬的證明，因為惡意陷害他人的可能性依舊存在。

本文的目標希望利用數位浮水印方法，從技術上提供所有權證明機制，讓數位浮水印成為一個具有法律效力的工具，足以在法庭上被視為決定性的證據。根據上述分析，想要設計出一個足以解決所有權歸屬爭議的數位浮水印方法，是必須滿足一些條件的。

首先，根據第 1 點，在我們的應用中，必須考慮最底限的狀況，以密碼學的觀點來衡量浮水印的安全性。浮水印的存在可能是眾人皆知的事，嵌入和擷取的演算法也可能 (或者根本就應該) 是公開的，只有嵌入時和擷取時所使用的鑰匙會是秘密，並且浮水印不只可能遭受到非惡意攻擊，如各類訊號處理，也可能會遭受到惡意攻擊，如惡意移除既有浮水印、惡意嵌入他人浮水印。

其次，根據第 2 點，我們必須專注於被嵌入之浮水印本身的「證據力」而非

強健性，也就是說，我們不去強調所嵌入的浮水印是否特別難以被移除，而是針對「如果已經在某張圖片裡面偵測到某浮水印，那麼這件事能夠代表什麼？它的存在能夠證明什麼?」這個問題方向去提供解決方案。

而由上述的第 3 點，我們也可以知道，相較於嵌入方式，浮水印的產生方式更會是一個關鍵步驟。浮水印產生的演算法必須具有下列兩項重要限制：(1) 浮水印必須根據擁有人的秘密資訊來產生。同時具有唯一性，只有擁有人有能力產生此浮水印。(2) 浮水印必須根據欲保護的圖片內容來產生。

想要擁有計算上可證明的效果，又必須同時兼顧浮水印產生時的這兩項要求，那麼利用密碼學的數位簽章當然是一個最適切的提案，我們可以運用擁有者的私鑰對欲保護的圖片進行簽署，再將獲得的數位簽章當作浮水印嵌入到圖片中，此時數位簽章的不可偽造性確保只有圖片的擁有者能夠產生浮水印，同時這個浮水印是針對欲保護的原圖產生的，即便攻擊者成功地從公開使用的圖片中擷取出來並嵌入到另一張不同的圖片之中，也無法用來證明原圖的擁有人對於另一張圖的所有權。然而，由於驗證資料來源和數位版權保護這兩類問題之間有極大的差異，導致上述直接援引數位簽章到數位浮水印方法中的作法無法得到證明所有權的效果。

一般使用數位簽章驗證資料來源時，希望同時保證資料來源的正確性與資料的完整性，倘若透過公開管道傳送的資料本身或對應的數位簽章有任何一點的毀損，即便只是一個位元的錯誤，都將導致簽章驗證的失敗。數位簽章本質上是一種易碎的、沒有彈性的密碼學物件，正是這種非黑即白的特性為我們保證了資料的完整性與來源的可驗證性。以上述直接運用數位簽章作為浮水印的方法而言，只要嵌入浮水印的圖片經過一個有損的壓縮程序，擷取出來的浮水印必然無法通過數位簽章的驗證程序。這種方法被稱為易碎浮水印，當我們的應用目標是利用浮水印方法來進行資料來源驗證時，十分有效，但是與本文討論的所有權證明機制有相當差異。人們的視覺允許誤差的存在，我們希望添加浮水印的圖片在視覺上與原圖難以分辨，也接受經過輕微影像處理的圖片在視覺上難以分辨的事實。因此我們仍然希望被嵌入的浮水印是強健的，只要影像沒有被破壞的太嚴重，在視覺上與原作的差異不大，應該要能夠從被修改過的圖片中擷取出最初嵌入的浮水印。換句話說，我們的目標就是希望做出一種「強健的數位簽章」當作浮水印，使之不但擁有計算上可驗證來源的特性，也同時兼具對於浮水印部分破壞的忍耐度。

如果想要使得浮水印存在於圖片中的意義變得如此重大，我們當然也需要小心地定義怎樣算是「在一張圖中偵測到某浮水印」。以往傳統的浮水印方法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，運用統計的結果來訂出門檻值，以超過門檻值表示有浮水印，低於則沒有。然而這樣做的結果是，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，而成為一個浮動的數字。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。我們希望能夠在偵測演算法中訂出一個絕對的門檻，來解決這種模糊的、模稜兩可的偵測標準問題。同時，需要避免在一張沒有嵌入浮水印的圖片中卻能偵測到自己的浮水印，讓所謂的偽陽性偵測率 (False-positive detection rate) 降低到可忽略的大小。作為證據力的一環，若能夠在一張圖中偵測到某浮水印，其應該要具有的意義是原擁有人確實曾經在該圖中添加過該浮水印。

1.1 文獻回顧

最早針對所有權證明機制與爭議解決進行討論的是 Craver 等人 [11]，他們提出了一種攻擊模式，稱為逆向攻擊 (Inversion attack)，點出了既有浮水印方法無法用於解決所有權爭議的缺陷所在。方法是假設有一名攻擊者 Bob，得到了 Alice 添加了浮水印並公開使用的圖片，他可以隨機產生一份參考圖像 (Reference pattern) 作為自己的浮水印，並從 Alice 公開的圖片中減去這個浮水印，得到一張新的、和 Alice 公開的圖片很像的、屬於 Bob 的原圖，然後就可以開始盜用 Alice 公開的圖片了。雖然 Bob 沒有移除 Alice 所添加的浮水印，但是 Bob 卻可以宣稱 Alice 公開的圖片中存在 Bob 自己的浮水印，並且當兩人把各自的原圖拿出來嘗試比對誰的原圖比較「乾淨」時，Alice 當然可以在 Bob 的原圖中偵測到她的浮水印，但 Bob 的作法將使得 Alice 的原圖中也可以偵測到他的浮水印，造成一個僵局，致使無法區辨真正的所有權歸屬。Craver 等人指出這是因為浮水印的嵌入演算法是可逆的 (Invertible)，也就是敵人可以先決定新的浮水印 w，接著執行嵌入演算法的逆向操作，從特定的一張圖片 I 中彷彿「拿掉」w ，得到 I′，並宣稱 I′ 是原圖，而 I 是加入了浮水印的圖片，並且能夠通過偵測演算法偵測到 w 的存在。Craver 等人在 [12] 中提出了一個解決方案，利用單向雜湊函式計算得到原圖的雜湊值，再將此雜湊值作為虛擬亂數產生器的種子，產生均勻分布的虛擬亂數當作浮水印進行嵌入。也就是設法讓浮水印必須單向地根據原圖而產生，那麼攻擊者就無法任意自行決定浮水印，再從別人公開使用的圖片中去衍生出他自己的原圖。Craver 等人並未在文中強調所使用的單向雜湊函式與虛擬亂數產生器是密碼學的工具，並且沒有給出形式化的證明。

而 Craver 等人的討論，引領了對於所有權歸屬爭議的研究朝向兩個主要方向發展，其中之一跟隨 Craver 等人的看法，專注於設計不可逆 (Non-invertible) 浮水印方法，並試圖憑藉這樣的方法解決所有權的爭議。他們的主要方法是基於既有的強健但可逆的浮水印方法 (如 [8])，並利用密碼學的工具如單向函式、加密系統或數位簽章系統來進行浮水印的產生或者是嵌入(擷取)密鑰的產生，以建構出一個不可逆的浮水印方法。例如 Ramkumar 等人 [22]，提出了一個針對

Craver 等人在 [12] 中方法的攻擊，並且提出了一個改進的方法。Qiao 等人 [21] 使用了加密系統來進行浮水印的產生，他們將原圖轉換到頻率域後的係數作為訊息，以擁有人事先決定好的鑰匙進行對稱式加密得到浮水印，而在擷取並驗證所有權時使用原圖及該密鑰。他們同時也證明了所有沒在驗證所有權時使用原圖的方法都會是可逆的。

Adelsbach 等人 [1][2] 對於逆向攻擊給出了形式化的定義，同時也形式化了一個更一般化的攻擊稱為模糊攻擊 (Ambiguity attack)，指的是攻擊者得到任意的一張圖片 I，不論其中是否已經添加了浮水印，只要能夠計算得到新的浮水印以及新的原圖，透過偵測演算法在 I 中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，因而 Adelsbach 等人以及後續研究者都嘗試設計足以抵抗模糊攻擊的浮水印方法。Adelsbach 等人在 [1] 針對這第一個研究路線 (即努力設計出不可逆的數位浮水印方法) 進行分析，強調了浮水印方法的偽陽性偵測率 (False-positive detection rate) 在利用不可逆浮水印方法解決所有權爭議時的影響力，並且指出若偽陽性偵測率不能降低到可忽略的大小，則模糊攻擊將會是可能的，進而逆向攻擊的風險亦將存在。

Kutter 等人提出了所謂的浮水印拷貝攻擊 (Watermark copy attack) [17]，指的是攻擊者得到一張已經嵌入浮水印的圖片，儘管他不知道浮水印如何被嵌入，仍然可以透過分析該圖片的統計特性，將其中的浮水印移植到另一張可能完全不相關的圖片之中。拷貝攻擊和模糊攻擊被合稱為協定攻擊 (Protocol attack)，它們的共通點為，攻擊者都不需要知道浮水印是如何被嵌入的，目標也不是移除被嵌入的浮水印，以得到一張乾淨的圖片來作為非法用途，而是想辦法造成所有權歸屬判定上的障礙，使得浮水印的證據性受到質疑即可。這類攻擊的存在也表示了，即便是非常強健的數位浮水印方法，仍然有可能遭受到這樣的攻擊，若沒有小心避免這樣的可能，是不足以提供充足的證據性的。

Li 等人 [18] 使用了密碼學的虛擬亂數產生器來產生浮水印，並且提出形式化的證明，證明了他們的浮水印方法是不可逆的。同時，他們的方法要求產生虛擬亂數的種子不能根據原圖計算而得，並且指出，以這樣的序列作為浮水印，浮水印和原圖在統計上是不相關的，因此，想要達成不可逆性，浮水印必須根據原圖產生不是一個必要條件。

第二個研究路線則是引入可信賴的第三方，運用類似登記註冊的方式或者借助時戳伺服器的幫助來釐清所有權爭議。Adelsbach 等人在 [2] 中，設計由擁有人向第三方申請嵌入浮水印，第三方運用他的私鑰對於包含原圖與擁有人指定的鑰匙等等的一串訊息進行數位簽章，作為浮水印的主體，他們並且證明了該方法能抵抗模糊攻擊與拷貝攻擊。Adelsbach 等人在 [3] 中首次提出了一個先前未被考慮的問題，那就是如果一張圖片引發了爭議，但參與爭取所有權的人都不是真正的合法擁有人，那麼按照以往將所有權判給其中最具說服力的競爭者的解決方案，將無法真正解決所有權歸屬問題。因此引入第三方的一個重要目的是確保爭議發生時，必然能讓真正的擁有人涉入其中。然而，如果在實務中需要引入高容

量、高計算能力的可信賴第三方，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，那麼更根本的問題會是：我們還需要數位浮水印來協助解決所有權爭議嗎？

1.2 本文貢獻

本文提出一個密碼學上可證明的數位浮水印方法，提供數位影像所有權的明確證明機制，使得數位影像在沒有合法授權的情況下遭到剽竊、修改、重製時，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所有權，釐清圖片之抄襲關係，捍衛原擁有者之各項著作權。

我們對合法擁有人持有的原始影像進行數位簽章，接著將此無法偽造的數位簽章轉化為無法預測的虛擬亂數序列，並且將一部份的序列當作浮水印本體，再將一部份的序列當作嵌入鑰匙來決定浮水印嵌入的位置，這樣的浮水印不但具有強健性能夠承受訊號處理與影像處理，也能夠抵抗惡意敵人的破壞與移除攻擊。密碼學的虛擬亂數序列已經在單向函式存在的假設下被證明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布且隨機的 λ 個位元之序列當作種子，那麼所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，輸出一張圖片，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，並且此爭議圖片對應於原作者之原圖的 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。

此方法與以往數位浮水印的主要差異包括：第一，本文的方法利用原圖以及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，僅需要透過可信賴第三方來偵測浮水印。我們並且證明了若能偵測出特定浮水印，則此浮水印是唯一地對應於特定擁有人及某張特定原圖。第二，以往的方法運用圖形辨識的精神來判定圖片中是否存在某一浮水印，也就是藉由相似度的比較來回答「是否偵測到某一浮水印？」或「偵測到的浮水印比較像浮水印 A 還是浮水印 B ？」等問題。當偵測到的浮水印同時與兩個浮水印有接近的相似度時，這樣的比較機制可能造成公信力的疑義。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原始的浮水印，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演算法也是公開的，並在安全性證明的過程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數位浮水印系統，讓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。

1.3 章節內容介紹

本文第二章中敘述了相關的密碼學和訊號處理的背景知識，第三章中描述了本文的數位浮水印方法以及所有權歸屬判斷協定，第四章中提出本文方法的正式安全性定義，並且證明第三章之建構方法滿足我們定義的安全性，第五章則呈現實驗結果與相關討論，第六章是結論。

第二章　背景知識

本章介紹所使用的訊號處理與密碼學方法之相關背景知識。

2.1 小波轉換

小波轉換是空間-頻率域數位訊號分解轉換方法，其基底涵括局部頻率資訊以及局部空間資訊。此種技術廣泛應用於訊號壓縮、偵測、通訊、與辨識。將一維訊號表示為長度 2^n 的向量 $\mathbf{x}=(x_1,x_2,\ldots,x_{2^n})$，其 m 階 Haar 離散小波轉換如下：首先將此向量分為移動平均 (Running average) $c^{m-1}=(c_1^{m-1},c_2^{m-1},\ldots,c_{2^{n-1}}^{m-1})$ 與移動差值 (Running difference) $d^{m-1}=(d_1^{m-1},d_2^{m-1},\ldots,d_{2^{n-1}}^{m-1})$，其中 $c_j^{m-1}=(x_{2j-1}+x_{2j})/\sqrt{2}$ 代表低頻的部分，$d_j^{m-1}=(x_{2j-1}-x_{2j})/\sqrt{2}$ 代表高頻的部分。不斷反覆將 c^{m-i} 分解成大小為 2^{n-i-1} 的向量 c^{m-i-1} 與 d^{m-i-1}，直到得到大小為 2^{n-m} 的向量 c^0 與 d^0；最後 $\mathbf{x}$ 可分解為 $(c^0\|d^0\|d^1\|\ldots\|d^{m-1})$。定義 $\{\mathbf{v}_j^{m-i}\}_{j=1,\ldots,2^{n-i}}$ 為第 i 階的尺度函數 (Scaling function)，$\{\mathbf{w}_j^{m-i}\}_{j=1,\ldots,2^{n-i}}$ 為小波函數 (Wavelet function)，此兩類向量的長度均為 2^n，表示如下：$\mathbf{v}_j^{m-i}=(\ldots,v_{j,k}^{m-i},\ldots)$ 與 $\mathbf{w}_j^{m-i}=(\ldots,w_{j,k}^{m-i},\ldots)$，其中 $k=1,\ldots,2^n$，

$$v_{j,k}^{m-i}=\begin{cases}1/\sqrt{2},\text{if } 2^i j-1\le k<2^i j+2^i-1\\0,\quad \text{otherwise}\end{cases}\quad \text{且}$$

$$w_{j,k}^{m-i}=\begin{cases}1/\sqrt{2},\text{if } 2^i j-1\le k<2^i j+2^{i-1}-1\\-1/\sqrt{2},\text{if } 2^i j+2^{i-1}-1\le k<2^i j+2^i-1\\0,\quad \text{otherwise}\end{cases}。$$

訊號向量 $\mathbf{x}$ 可分解為下列基底的線性組合：

$$\sum_{j=1,\ldots,2^{n-m}} c_j^0\mathbf{v}_j^0+\sum_{i=1,\ldots,m}\sum_{j=1,\ldots,2^{n-i}} d_j^{m-i}\mathbf{w}_j^{m-i}\quad 。$$

2.2 數位簽章及其不可偽造性

一個數位簽章系統包含以下三個演算法：(1) $KeyGen(1^\lambda)$：金鑰產生演算法，輸入安全參數 λ，輸出金鑰對 (PK,SK)。(2) $Sign(SK,m)$：簽署演算法，輸入密鑰 SK 以及訊息 m，輸出對應 m 的簽章 σ。(3) $Verify(PK,m,\sigma)$：簽章驗證演算法，輸入公鑰 PK、訊息 m 以及簽章 σ，當 σ 是對應 m 的合法簽章時輸出 1，否則輸出 0。

數位簽章的安全性要求在選擇訊息攻擊之下不得偽造出得以通過驗證程序之合法簽章 (Existential unforgeability under chosen message attack, EUF-CMA)

[16]。如下列的賽局所示，一個挑戰者 C 和一個惡意的攻擊者 $\mathcal{A}$ 進行互動，M 表示訊息空間：

- 起始階段：C 執行 $KeyGen(1^{\lambda})$ 以產生 (PK,SK) ，並將 PK 交給 $\mathcal{A}$。
- 詢問階段：$\mathcal{A}$ 可以一次一次地詢問（共 q_s 次）任意訊息 $m^{(j)} \in M$ 的簽章，C 則執行 $Sign(SK,m^{(j)})$ 並回傳對應的簽章 $\sigma^{(j)}$ 給 $\mathcal{A}$。
- 輸出階段：$\mathcal{A}$ 輸出一組訊息與簽章 (m^*,σ^*)。若 $m^* \notin \left\{m^{(j)}\right\}_{j=1,\ldots,q_s}$ 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:

$$Adv_{\mathcal{A}}^{EUF-CMA}(1^{\lambda}) = \Pr\left[Verify\left(PK,m^*,\sigma^*\right)=1 \text{ and } m^* \notin \left\{m^{(j)}\right\}_{j=1,\ldots,q_s}\right]。$$

一個數位簽章系統如果在上述的賽局之中，使得任意機率式多項式時間的敵人 $\mathcal{A}$ 贏得賽局的優勢 $Adv_{\mathcal{A}}^{EUF-CMA}(1^{\lambda})$ 為可忽略之值，則此簽章系統是 EUF-CMA 安全的。

2.3 虛擬亂數產生器及其不可預測性

一個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布、λ 個位元的種子 s 後，產生一串 $\ell(\lambda)$ 個位元的序列，此序列與真正均勻隨機分布的 $\ell(\lambda)$ 個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。正式的安全性描述如下：對於任意機率式多項式時間的分辨演算法 D、任意正多項式 $p(\cdot)$、及任意足夠大的整數 λ 而言，$\left|\Pr\left[D\left(G\left(U_{\lambda}\right)\right)=1\right]-\Pr\left[D\left(U_{\ell(\lambda)}\right)=1\right]\right| < \frac{1}{p(\lambda)}$，其中 U_{λ} 和 $U_{\ell(\lambda)}$ 分別為均勻分布的 λ 位元序列和 $\ell(\lambda)$ 位元序列的隨機變數。

除此之外，一個虛擬亂數產生器的輸出是不可預測的，亦即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，$\Pr\left[\mathcal{A}\left(G\left(U_{\lambda}\right)\right)=\text{next}_{\mathcal{A}}\left(G\left(U_{\lambda}\right)\right)\right] < \frac{1}{2}+\frac{1}{p(\lambda)}$，其中 $\text{next}_{\mathcal{A}}(\cdot)$ 是一個定義如下的函式：當 $\mathcal{A}$ 讀入 $G\left(U_{\lambda}\right)$ 的前 k 個位元，$\text{next}_{\mathcal{A}}(\cdot)$ 會輸出 $G\left(U_{\lambda}\right)$ 的第 $k+1$ 個位元；當 $\mathcal{A}$ 讀入全部 $\ell(\lambda)$ 個位元，$\text{next}_{\mathcal{A}}(\cdot)$ 會輸出均勻隨機挑選的一個位元 [15，章節 3.3.5]。

第三章　系統建構及所有權歸屬判斷協定

3.1 系統建構

本文所使用的基於浮水印技術之可證明著作權保護機制，是以因數分解的單向函數性質為基礎的三個演算法所構成：$(WSetup, Embed, Detect)$。$WSetup(1^{\lambda})$ 是機率式的系統初始化演算法，藉由給定一安全參數 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 後，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 是否含有與原圖 I 相關之浮水印。
上述三個演算法詳細步驟如下：

- $WSetup(1^{\lambda})$：
  (1) 首先用 RSA 簽章機制產生參數，隨機選擇兩個長度為 $\lambda/2$ 位元的質數 p_1 、 q_1，算出模數 $N_1 = p_1 \cdot q_1$ 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，接著算出簽章用的指數 $d \equiv e^{-1} (\mathrm{mod}\, \phi(N_1))$。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) 使用的參數：挑選一個 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 $p_2 \equiv q_2 \equiv 3 \,(\mathrm{mod}\, 4)$ 的隨機質數。N_2 定義單向 Rabin 函式 $f_{N_2} : QR_{N_2} \to QR_{N_2}$，$f_{N_2}(x) = x^2 \,(\mathrm{mod}\, N_2)$，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 $G_{f_{N_2}}(s) = LSB(f_{N_2}(s)) \| LSB(f_{N_2}{}^2(s)) \| ... \| LSB(f_{N_2}{}^{k\lambda-1}(s))$，$LSB(\cdot)$ 是取出最低位元 (Least significant bit) 的函式，$x \| y$ 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) 選擇一個可抵抗碰撞的雜湊函式 $H(\cdot)$。
  (4) N_1、e、$G_{f_{N_2}}(\cdot)$、$H(\cdot)$ 組成 PK，而 d 是嵌入浮水印的鑰匙 EK。演算法最後輸出 (PK, EK)。

- $Embed(PK, EK, I)$：
  (1) 首先算出原圖 I 的數位簽章 $\sigma_I = H(I)^d \,(\mathrm{mod}\, N_1)$。
  (2) 接著以這個簽章作為亂數產生器的種子產生 $k\lambda$ 位元長度的虛擬亂數 $w_I[1, k\lambda] = G_{f_{N_2}}(\sigma_I)$，其中前 λ 位元 $w_I[1, \lambda]$ 為浮水印，剩下的 $(k-1)\lambda$ 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 (每一區塊都有超過 λ 個像素)，嵌入鑰匙 $w_I[\lambda+1,k\lambda]$ 則分成 λ 個段落，每個段落 $k-1$ 個位元，用來指定 λ 位元浮水印 $w_I[1,\lambda]$ 中的第 i 個位元 $w_I[i]$ 要隱藏在 ℓ 個區塊中的哪一個。步驟如下：對浮水印中每一位元 $w_I[i]$，$i=1,...,\lambda$，根據 $k-1$ 個位元的序列 $w_I[\lambda+i]\|w_I[2\lambda+i]\|...\|w_I[(k-1)\lambda+i]$ 當作區塊資訊，從 ℓ 個像素中選出一個像素，然後用浮水印的第 i 位元 $w_I[i]$ 取代該像素的第 β 位元，待 λ 個位元全數嵌入完畢，即得到嵌入浮水印後的低頻部分 LL'。此步驟以簡化過的類 C 語言虛擬碼表示如下：

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

上述演算法中，$LL[pos][i]$ 的第一個索引代表 ℓ 個區塊中的第 pos 個區塊，第二個索引則代表該區塊中的第 i 個像素。

如果影像 I 是 512×512 大小的八位元灰階圖片，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 $1\le\beta\le8$。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影像處理對浮水印的影響，故由實驗決定 β。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，也許品質不會變那麼差，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) 嵌入浮水印得到 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 $\left(I_w, XK_I=(I,\sigma_I)\right)$。$I_w$ 即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) 執行 $G_{f_{N_2}}(\cdot)$ 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 $w_{I_\alpha}[1,\lambda]$ 的過程以簡化過的類 C 語言虛擬碼表示如下：

```
Extract(w_I, LL)
{
    for(i=1; i<=λ; i++)
    {
        pos = w_I[λ+i] || w_I[2λ+i] || ... || w_I[(k−1)λ+i];
        selectedPixel = LL[pos][i];
        extractedWatermark[i]=selectedPixel[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。

(6) 如果 $\left|NCC(w_I[1,\lambda], w_{I_\alpha}[1,\lambda]) - \frac{1}{2}\right| \geq \frac{2}{\ell}$ 則演算法輸出 1，否則輸出 0。

3.2 所有權歸屬判斷協定

設想以下情形：O 為圖片 I 的原作者，圖片 I 並未公開，而是公開使用嵌入浮水印後的圖 I_w。如果 O 發現 P 疑似未經授權公開使用與 I 及 I_w 相似度極高的圖 I_d，O 可針對 P 未獲授權使用以及傳播 I_d 的行徑提起訴訟。本方法中使用以下步驟來判斷 I_d 的所有權歸屬，其中 (PK_O, EK_O) 和 (PK_P, EK_P) 分別為 O 和 P 的浮水印系統公開參數和嵌入浮水印的私鑰，$\sigma_I^{(O)}$ 表示 O 以其簽章私鑰 (即 EK_O) 對於圖片 I 的簽章，$\sigma_{I'}^{(P)}$ 則是 P 以其私鑰 EK_P 對於圖片 I' 的簽章。

1) 原告 O 將圖片 I 和 O 針對 I 簽署的簽章 $\sigma_I^{(O)}$ 提供給可信任的第三方 T，T 可藉由 $Detect(PK_O, I, \sigma_I^{(O)}, I_d)$ 是否等於 1 來確認圖 I_d 是否含有用圖 I 和嵌入私鑰 EK_O 做出的浮水印。並用 $PSNR(I, I_d)$ 是否大於 30 做為圖 I_d 與 I 的相似度的客觀指標。若以上兩個測試有一不成立，則駁回控訴。

2) P 可藉由提供另一張圖 I' 和它的簽章 $\sigma_{I'}^{(P)}$ 給 T 提出反駁。當 $Detect(PK_P, I', \sigma_{I'}^{(P)}, I_d)$ 等於 1，表示圖 I_d 中確實含有用圖 I' 和嵌入私

鑰 EK_P 做出的浮水印，則進行下一步驟的驗證。

3) T 透過 $Detect(PK_O, I, \sigma_I^{(O)}, I')$ 來確認圖 I' 是否含有用圖 I 和嵌入私鑰 EK_O 作出的浮水印。若結果為 1，則證實 P 確實盜用 O 的圖片。若結果為 0 則 O 敗訴。接下來的定理說明如果 P 盜用 I_w 亦即 I' 是由 I_w 衍生而得到的，I' 與 I 的 $PSNR$ 值、I' 與 I_w 的 $PSNR$ 值皆高於 30，則此步驟的結果為 0 的機率極低，是計算上可忽略的。

假設浮水印的嵌入機制使浮水印有足夠的強健性而難以完全移除，則以下定理可以作為殘存一定比例浮水印之影像的明確所有權的基礎。

定理一：

若 $Detect(PK_O, I, \sigma_I^{(O)}, I') = 1$ ，則 I' 不是由 I_w 衍生得到的機率是計算上可以忽略的。

此定理的證明是建立在虛擬亂數序列的不可預測性上。在第四章中將有完整的定理描述以及證明。

在上述的情境中，若 P 想要在未經 O 授權的情況下使用 O 的圖片，則 P 必須試著將 I_w 裡的浮水印移除，得到 I'。再加入其個人的浮水印至 I' 中以得到 I_d 。若他成功地完全移除 I_w 裡的浮水印，亦即 I' 不含有 O 的浮水印，$\left| NCC - \frac{1}{2} \right| < \frac{2}{\ell}$，那麼在步驟二的 I_d 亦不含有 O 的浮水印。反之，若 P 未完整移除浮水印，T 就會在 I' 中找到 O 的浮水印，如此可證實 P 未經授權的使用。如果 I' 和 I_d 由圖 I_w 製作得到，即使無法證明浮水印加入方法的不可移除性，我們也讓剽竊者陷入風險極高的窘境，由於此種浮水印對於任一計算能力有限的攻擊者而言，都是和真正亂數序列不可區辨的，我們的方法使得他自己無法驗證是否完全移除浮水印－也就是殘存的浮水印必須滿足 $\left| NCC(w_I[1,\lambda], w_{I_d}[1,\lambda]) - \frac{1}{2} \right| < \frac{2}{\ell}$。若 P 無法完全移除 I_w 中的浮水印，就留下了侵犯他人智慧財產權的證據。

第四章　安全性證明

4.1 安全性定義

本文方法的安全性是基於「不可偽造包含特定使用者針對特定圖片產生之浮水印的圖片」來定義，簡稱為 WUF-CIA (Watermark unforgeability under chosen image attack)。

定義一 (WUF-CIA 安全性)：

如圖 1 所示，這是一個挑戰者 C 及一個攻擊者 $\mathcal{A}$ 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，並將 PK 及 I 傳送給 $\mathcal{A}$，嵌入密鑰 EK 只有 C 知道。
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 $Embed(PK, EK, I^{(j)})$ 得到添加了針對 $I^{(j)}$ 產生的浮水印的圖片 ${I^{(j)}}_w$ 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以驗證嵌入結果。
3. 輸出階段：
   $\mathcal{A}$ 輸出一張圖片 I 以及對應的 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。

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

如果一個浮水印方法是 WUF-CIA 安全的，那麼我們可以做出以下解讀：針對一位使用此方法的數位影像創作人 C 來說，任意計算能力有限的攻擊者偽造出一張包含 C 針對特定圖片 I 產生之浮水印 w_I 的圖片的可能性是計算上可忽略的，因此，如果確實在任意的圖片中偵測到 w_I，那麼這張圖片勢必就是 C 針對 I 製作的公開圖片 I_w 本身，或者是經由 I_w 修改衍生而成的圖片。

以集合的角度觀察此安全性，若我們定義 S 是由 I_w、以及所有經由 I_w 修改衍生而得的圖片之集合，那麼儘管 S 中的圖片不見得都確保能偵測到 w_I 的存在，然而對任意一張不屬於 S 的圖片來說，偵測到 w_I 的機率都是可忽略的。因此我們也可以看出，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將會是可忽略的，也就是在不知道嵌入密鑰 EK 的情況下，

幾乎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。

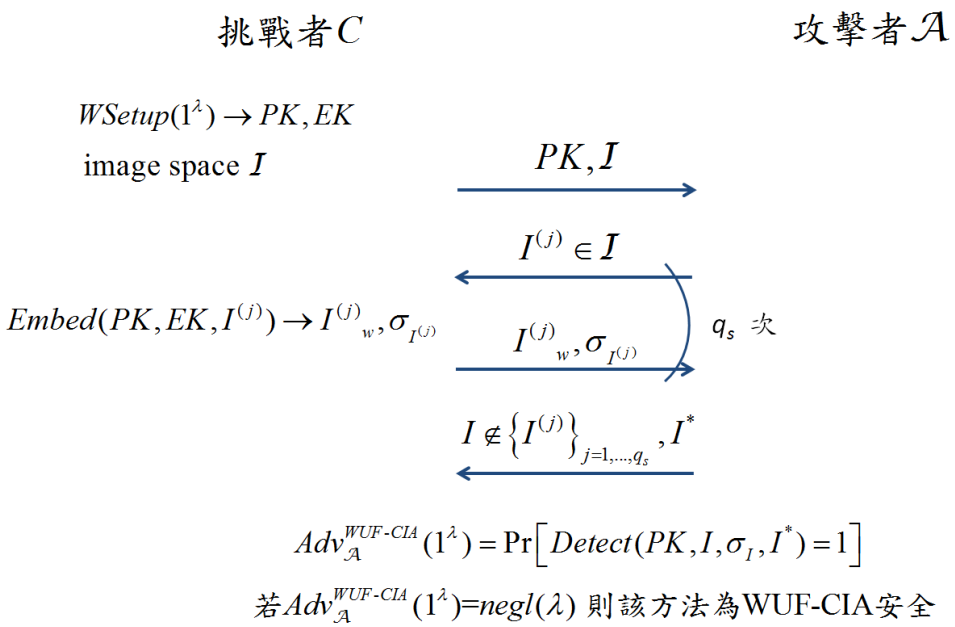


圖 1　WUF-CIA 安全性定義

4.2 安全性證明

在本節中，我們完整地證明了 3.2 節提到的定理一。

定理一：

如果下列三項前提條件成立，則 3.1 節所提出的數位浮水印方法是 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) [24]，(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 2.3 節所定義的虛擬亂數性。

由於本文使用 RSA 簽章機制以及 BBS 虛擬亂數產生器，此二者皆已經被證明安全，因此只有第 (1) 點是本文必須定義的假設，如下：

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 D、任意輸入 x 而言，$\left|\Pr[D(H(x)) = 1] - \Pr[D(U_\lambda) = 1]\right| = negl(\lambda)$，其中 U_λ 為均勻隨機分布的 λ 位元序列的隨機變數，$negl(\cdot)$ 是一個可忽略函數。

本節證明分為兩個部分，第一部分基於假設一，證明使用由雜湊並簽章 (Hash-and-sign) 方法得到的數位簽章當作種子而獲得的虛擬亂數序列具有虛擬亂數性。此延伸的虛擬亂數性質簡稱為 EP-SSPS (Extended pseudorandomness for signature seeded pseudorandom sequence)，定義如下。

定義二 (EP-SSPS 虛擬亂數性)：

在以下賽局中，一個挑戰者 C 和一個惡意的攻擊者 $\mathcal{A}$ 進行互動：

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 $PK=\left(N_1,e,G_{f_{N_2}}(\cdot),H(\cdot)\right)$ 以及嵌入密鑰 $EK=d$，同時決定影像空間 I，並將 PK 及 I 傳送給 $\mathcal{A}$，嵌入密鑰 EK 只有 C 知道。

2. 詢問階段：

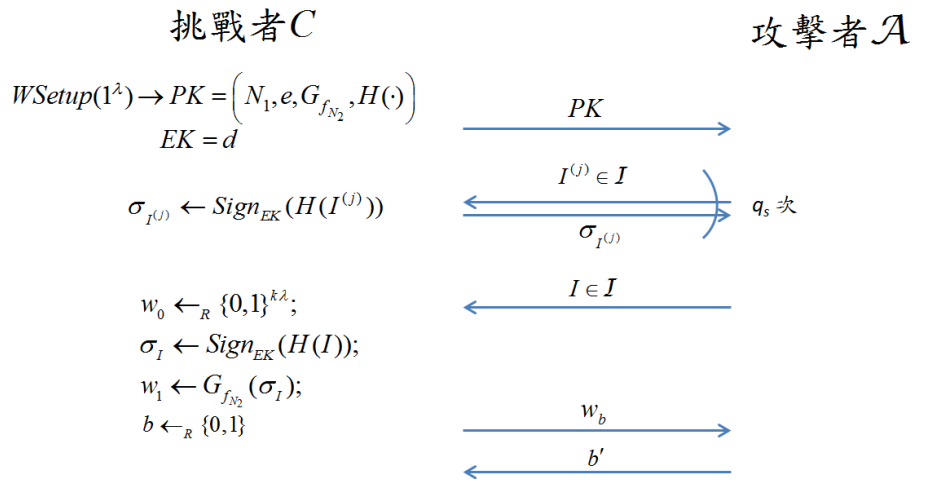
   C 提供簽署引擎，讓 $\mathcal{A}$ 可多次 (共 q_s 次) 任意選擇影像空間中的影像 $I^{(j)}$ 交給 C，C 用 N_1 和 d 執行簽章演算法 $Sign\left(EK,H(I^{(j)})\right)$ 得到針對 $I^{(j)}$ 產生的數位簽章 $\sigma_{I^{(j)}}$，並將 $\sigma_{I^{(j)}}$ 送回給 $\mathcal{A}$。

3. 質疑階段：$\mathcal{A}$ 選擇一張圖片 I 交給 C，C 隨機挑選一個位元的亂數 b，當 b 是 0 時，均勻隨機地挑選 $k\lambda$ 位元長的亂數 w_0，當 b 是 1 時，則執行 $Sign\left(EK,H(I)\right)$ 得到 σ_I，緊接著將 σ_I 當作種子執行 $G_{f_{N_2}}(\sigma_I)$ 得到序列 w_1。C 根據 b 將製作出來的 w_b 傳送給 $\mathcal{A}$。

4. 猜測階段：

   $\mathcal{A}$ 輸出一個位元的數字 b'，C 比對 b' 和 b，當 $b'=b$ 時，$\mathcal{A}$ 贏得這個賽局。

攻擊者的優勢定義為 $Adv_{\mathcal{A}}^{EP-SSPS}(1^{\lambda})=\Pr[b'=b]$。如果對於任意機率式多項式時間的敵人 $\mathcal{A}$ 而言，贏得賽局的優勢 $Adv_{\mathcal{A}}^{EP-SSPS}(1^{\lambda})=\frac{1}{2}+negl(\lambda)$，則稱呼此種產生虛擬亂數序列的方法是 EP-SSPS 安全的。表示用這樣雜湊並簽章再產生虛擬亂數序列的方法，得到的序列將會是與均勻隨機分布的亂數在計算上不可分辨的。此定義如圖 2 所示。



$$Adv_{\mathcal{A}}^{EP-SSPS}(1^{\lambda})=\Pr[b'=b]$$

若 $Adv_{\mathcal{A}}^{EP-SSPS}(1^{\lambda})=\frac{1}{2}+negl(\lambda)$ 則該種虛擬亂數為 EP-SSPS 安全

圖 2 EP-SSPS 虛擬亂數性定義

引理一：

如果使用的雜湊函式滿足假設一，則本文使用 $G_{f_{N_2}}\left(Sign\left(EK,H(I)\right)\right)$ 作為亂數的方法，滿足上述 EP-SSPS 的定義。

證明：

這裡我們使用一系列賽局替換 (Sequence of games) 的方式 [24] 來證明，如圖 3 所示。

Game 0：上述原始定義的 EP-SSPS 賽局。我們定義 $G_0 = Adv_{\mathcal{A}}^{\text{Game 0}}(1^{\lambda})$。

Game 1：在 EP-SSPS 賽局的第 3 步驟質疑階段中，當 C 挑選到的 b 是 1 時，他改為挑選 λ 個位元的均勻隨機亂數 $h \leftarrow_R \{0,1\}^{\lambda}$，並執行 $Sign\left(EK, h\right)$ 得到 s_1，緊接著將 s_1 當作種子執行 $G_{f_{N_2}}(s_1)$ 得到序列 w_1。我們定義 $G_1 = Adv_{\mathcal{A}}^{\text{Game 1}}(1^{\lambda})$。

Game 2：在 EP-SSPS 賽局的第 3 步驟質疑階段中，當 C 挑選到的 b 是 1 時，他挑選 λ 個位元的均勻隨機亂數 $s_1 \leftarrow_R \{0,1\}^{\lambda}$，接著將 s_1 當作種子執行 $G_{f_{N_2}}(s_1)$ 得到序列 w_1。我們定義 $G_2 = Adv_{\mathcal{A}}^{\text{Game 2}}(1^{\lambda})$。

機率分析如下：

在 Game 0 與 Game 1 之間，差別只在於簽署演算法輸入的訊息是 $H(I)$ 或者是均勻隨機的 λ 位元亂數。基於假設一，這兩者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演算法 $\mathcal{A}$ 在 Game 0 的優勢與在 Game 1 的優勢非常接近，差異是可忽略的，也就是 $\left|G_0 - G_1\right| = negl_1(\lambda)$。

在 Game 1 與 Game 2 之間，差別只在於亂數產生器的種子是使用 $Sign(EK, U_{\lambda})$ 或者是直接使用 U_{λ}。由於 RSA 簽章演算法是一個排列函數 (Permutation function)，同時簽署的訊息是均勻隨機分布的，因此產生的簽章也會是均勻隨機分布的數字，因此任意機率式多項式時間的演算法 $\mathcal{A}$ 在 Game 1 的優勢與在 Game 2 的優勢是完全相等的，也就是 $G_1 = G_2$。

最後，因為在 Game 2 之中，亂數產生器的種子是使用真正均勻隨機分布的 λ 位元亂數，根據 2.3 節中虛擬亂數產生器的定義，任意機率式多項式時間的演算法 $\mathcal{A}$ 在 Game 2 的優勢為 $G_2 = \frac{1}{2} + negl_2(\lambda)$。

由以上，我們可以推得 $\left|G_0 - G_2\right| = \left|G_0 - G_1\right| = \left|G_0 - \left(\frac{1}{2} + negl_2(\lambda)\right)\right| = negl_1(\lambda)$，

因此 $G_0 = \frac{1}{2} + negl_1(\lambda) + negl_2(\lambda) = \frac{1}{2} + negl_3(\lambda)$，證明完畢。■

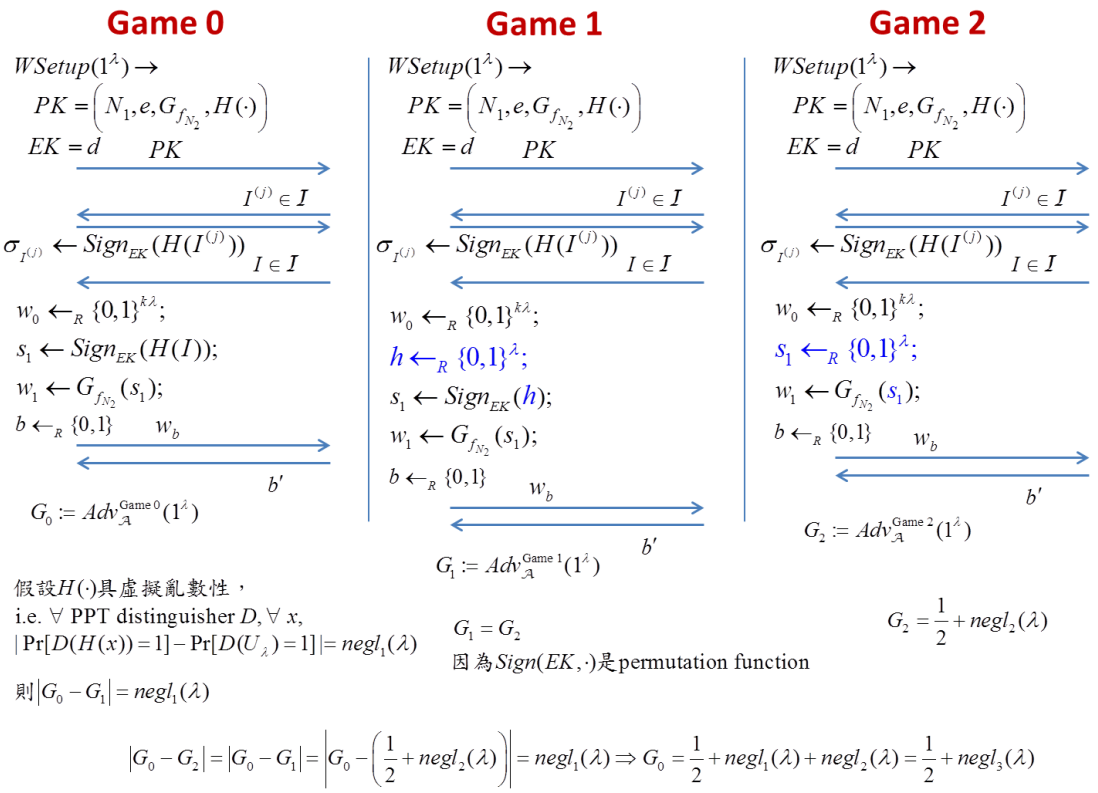


圖 3　EP-SSPS 虛擬亂數性證明

本節的第二部分證明，則基於引理一來證明定理一。證明手法為假設本文方法不是 WUF-CIA 安全，然後推論得到引理一將不會成立。

證明：

假設存在一個機率式多項式時間的演算法 $\mathcal{A}$，在 WUF-CIA 賽局中具有不可忽略的優勢，即存在多項式 $p(\cdot)$，使得 $Adv_{\mathcal{A}}^{WUF-CIA}(1^\lambda) > \frac{1}{p(\lambda)}$，則可以利用 $\mathcal{A}$ 建構一個具有不可忽略優勢可贏得 EP-SSPS 賽局的演算法 Sim。建構方式如下：

首先 EP-SSPS 賽局的挑戰者會執行 $WSetup(1^\lambda)$ 得到 $PK = \left(N_1, e, G_{f_{N_2}}, H(\cdot)\right)$，$EK = d$，並且決定影像空間 $\mathcal{I}$，將 $(PK, \mathcal{I})$ 傳送給 Sim。Sim 開始執行演算法 $\mathcal{A}$，將 $(PK, \mathcal{I})$ 傳送給 $\mathcal{A}$，接下來在 $\mathcal{A}$ 進行的一連串嵌入詢問過程中，每當 $\mathcal{A}$ 傳送一張圖片 $I^{(j)}$ 給 Sim，Sim 便將 $I^{(j)}$ 轉送給 EP-SSPS 的挑戰者，然後獲得 $\sigma_{I^{(j)}}$，接著 Sim 執行 $Embed$ 演算法的 (2)、(3)、(4)、(5)、(6) 步驟，得到嵌入浮水印的圖片 $I^{(j)}{}_w$，並將 $\left(I^{(j)}{}_w, \sigma_{I^{(j)}}\right)$ 傳送給 $\mathcal{A}$。結束詢問階段後，$\mathcal{A}$ 會把它偽造出的 I 和 I^* 交給 Sim，這時候 Sim 將 I 轉交給 EP-SSPS 的挑戰者，得到序列 w_b，然後 Sim 就可以利用 w_b 執行 $Detect$ 演算法的 (3)、(4)、(5)、(6) 步驟，執行 $DWT(\cdot)$ 把 I^* 分成 (LL, LH, HL, HH)，接著根據 $w_b[\lambda+1, k\lambda]$ 所決定的位置，由 LL 中取出

$w_{I^*}[1,\lambda]$，計算 $NCC(w_b[1,\lambda],\ w_{I^*}[1,\lambda])$，當 $\left|NCC(w_b[1,\lambda],w_{I^*}[1,\lambda])-\frac{1}{2}\right|\geq\frac{2}{\ell}$ 時，*Sim* 輸出 1，反之輸出 0。此演算法運作流程如圖 4 所示。

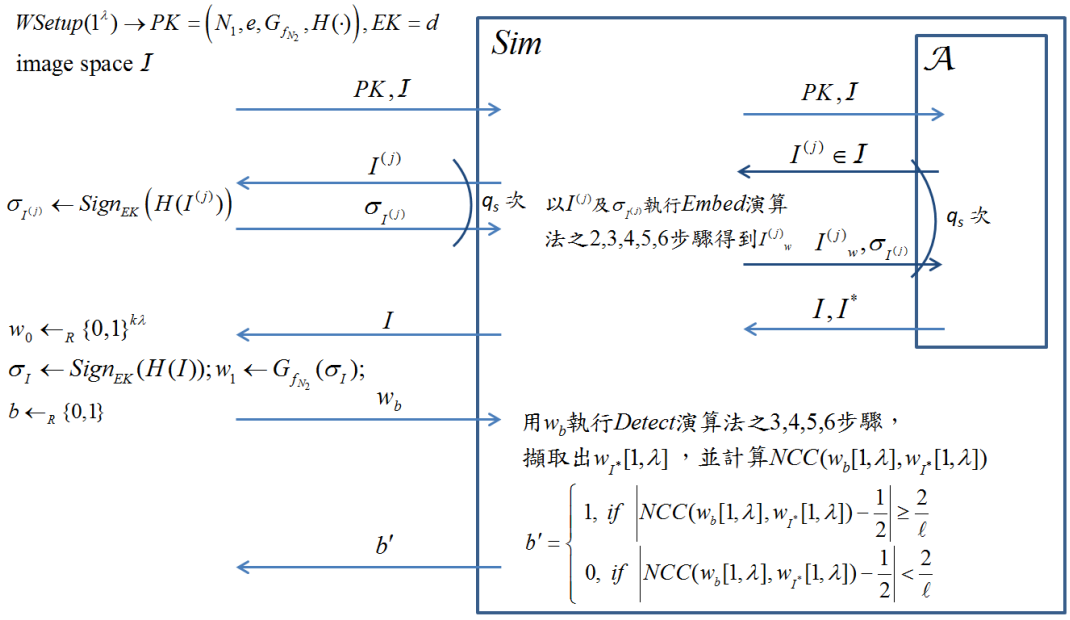


圖 4　*Sim* 演算法之建構方式

機率分析如下：

$$Adv_{Sim}^{EP-SSPS}(\lambda)=\Pr[b'=b]$$

$$=\frac{1}{2}+\frac{1}{2}\left(\left|\Pr\left[Sim^{Sign_{EK}(\cdot)}(PK,w_1)=1\right]-\Pr\left[Sim^{Sign_{EK}(\cdot)}(PK,U_{k\lambda})=1\right]\right|\right)$$

$$=\frac{1}{2}+\frac{1}{2}\left(\left|\Pr\left[\left|NCC(w_1[1,\lambda],w_{I^*}[1,\lambda])-\frac{1}{2}\right|\geq\frac{2}{\ell}\right]-\Pr\left[\left|NCC(U_{k\lambda}[1,\lambda],w_{I^*}[1,\lambda])-\frac{1}{2}\right|\geq\frac{2}{\ell}\right]\right|\right)$$

上式中 $\Pr\left[\left|NCC(U_{k\lambda}[1,\lambda],w_{I^*}[1,\lambda])-\frac{1}{2}\right|\geq\frac{2}{\ell}\right]$ 的數值可以估計其上限如下：

$$\Pr\left[\left|NCC(U_{k\lambda}[1,\lambda],w_{I^*}[1,\lambda])-\frac{1}{2}\right|\geq\frac{2}{\ell}\right]=\Pr\left[\left|\frac{\sum_{i=1}^{\lambda}\left(U_{k\lambda}[i]=w_{I^*}[i]\right)}{\lambda}-\frac{1}{2}\right|\geq\frac{2}{\ell}\right]\quad,$$

根據 Chernoff 不等式 [7]，$\Pr\left[\left|\frac{\sum_{i=1}^{n}X_i}{n}-p\right|>\delta\right]<2e^{-\frac{\delta^2}{2p(1-p)}n}$，其中 X_i 是二元隨

機變數，$p=\Pr[X_i=1]$，因此，$\Pr\left[\left|\frac{\sum_{i=1}^{\lambda}\left(U_{k\lambda}[i]=w_{I^*}[i]\right)}{\lambda}-\frac{1}{2}\right|>\frac{2}{\ell}\right]<2e^{-\frac{\left(\frac{2}{\ell}\right)^2}{2\cdot\frac{1}{2}\cdot\frac{1}{2}}\lambda}=2e^{-\frac{8}{\ell^2}\lambda}$ 。

假設 $\ell=16$，λ=4096，即我們使用 4096 位元之 RSA 簽章、以及由 4096 位元拉長為 5·4096 位元之 BBS 虛擬亂數產生器，則上述機率低於 $2e^{-128}$，可知使用足夠安全之系統參數的情況下，均勻分布的亂數序列 $U_{k\lambda}[1,\lambda]$ 與 $w_{I^*}[1,\lambda]$ 比對的 NCC 值分布的尾端機率將會是可忽略的數值。該式可繼續推演如下：

$$\frac{1}{2}+\frac{1}{2}\left(\left|\Pr\left[\left|NCC(w_1[1,\lambda],w_{I^*}[1,\lambda])-\frac{1}{2}\right|\geq\frac{2}{\ell}\right]-\Pr\left[\left|NCC(U_{k\lambda}[1,\lambda],w_{I^*}[1,\lambda])-\frac{1}{2}\right|\geq\frac{2}{\ell}\right]\right|\right)$$

$$=\frac{1}{2}+\frac{1}{2}\left(\left|\Pr\left[Detect(PK,I,\sigma_I,I^*)=1\right]-negl(\lambda)\right|\right)$$

$$=\frac{1}{2}+\frac{1}{2}\left(Adv_{\mathcal{A}}^{WUF-CIA}(1^\lambda)-negl(\lambda)\right)>\frac{1}{2}+\frac{1}{2}\left(\frac{1}{p(\lambda)}-negl(\lambda)\right)>\frac{1}{2}+\frac{1}{2}\left(\frac{1}{2p(\lambda)}\right)$$ 。

因此 *Sim* 將具有不可忽略的優勢能夠分辨由合法簽章做為種子而產生的亂數或者是真正均勻隨機分布的亂數，故 EP-SSPS 的安全性不會滿足，引理一將不成立，得到矛盾。證明完畢。■

4.3 針對協定攻擊之分析

本文朝向以數位浮水印方法解決所有權歸屬爭議之目標而努力，期望能提供充足證據性，而根據前言所述，模糊攻擊與浮水印拷貝攻擊是使得浮水印方法用於解決所有權爭議的重大阻礙，因而在此處我們分析本文方法達成的安全性與此二類攻擊之關聯性。

我們的核心解決理念為使用密碼學上具備不可偽造性的數位簽章，將之轉化為不可預測、與均勻隨機亂數不可分辨的虛擬亂數序列以作為浮水印主體，使得這樣的簽章具有一定程度的容錯能力但依然可以驗證，藉以成功將浮水印、擁有人以及被保護的數位影像三者緊密結合，讓我們在一張圖片中成功偵測到一個浮水印時，能夠唯一地對應到特定擁有者及特定圖片。

這樣的浮水印方法，明顯是不需要擔心遭受浮水印拷貝攻擊的。縱使攻擊成功了，擁有人根據原圖 A 嵌入的浮水印被近乎無損地複製到不相關的圖片 B 之中，由於這個浮水印必須以擁有人之公鑰以及圖片 A 來進行偵測方可偵測得到，是一個針對圖片 A 的所有權宣告，因此這個浮水印在 B 之中已不具絲毫意義，而不會發生某個使用者的所有權宣告被敵人惡意濫用的情形。

至於模糊攻擊，在本文的方法底下則如是：敵人擁有自己的一組浮水印系統參數，當他拿到了任意的一張圖片 I，若能有效率地計算出一張新的原圖 I' 及

他針對 I' 製作的浮水印 (即簽章 $\sigma_{I'}$)，使得在 I 中能偵測到這個新的浮水印 (即 $Detect(PK, I', \sigma_{I'}, I) = 1$)，則攻擊成功。我們的方法設計亦能夠抵擋這樣的攻擊，但詳細的安全性定義及其證明不在本文中描述，敬請關注後續文章。

第五章 實驗結果

我們用每個像素為 8 個位元、512×512 大小的灰階圖片「Lenna」測試。本實驗的主要目的為決定第三章 *Embed* 演算法中提到的雜訊強度參數 β 並驗證此浮水印方法在一般傳輸及影像處理下的強健性。

圖 5 顯示運用各個 β 值加入浮水印後，「暴力移除浮水印攻擊」時 PSNR 值的變化。暴力移除係指攻擊者把已嵌入浮水印之圖變換到轉換域，接著將 *LL* 中每個像素相對應的位元清除，此位元由嵌入演算法中的參數 β 所決定；最後轉換回空間域，得到沒有原作者浮水印的圖。

在肉眼無法辨識的前提下，浮水印應該要嵌入在圖片中比較重要的地方。此實驗結果顯示，$\beta=6$ (雜訊振幅為 $\{32,0,-32\}$) 是有效的嵌入浮水印強度。接下來的實驗皆是以 $\beta=6$ 作為參數之結果。

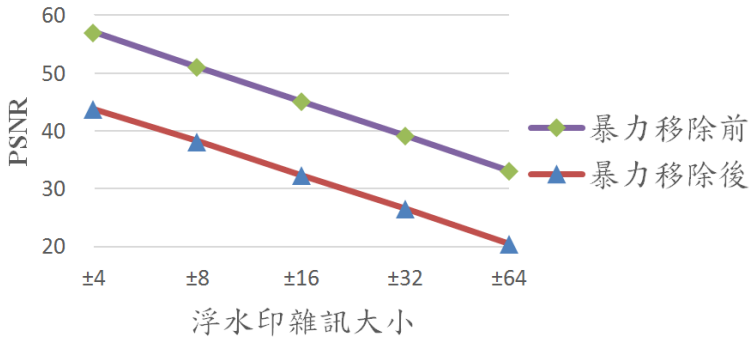


圖 5 在不同浮水印強度下，嵌入浮水印之圖經由暴力移除浮水印攻擊後 PSNR 之變化

為了釐清某圖片的所有權，有爭議的圖片必須和原作者的圖片非常相似，也就是 PSNR＞30。若有爭議的圖片已經過空間轉換的處理，則在提取浮水印、計算 PSNR 以及殘留浮水印比例前，須先將此圖轉換回符合 PSNR＞30。例如將圖片平移或旋轉回與原圖儘可能接近的狀況。

圖 6 為圖片經過 JPEG 壓縮 80% 至 40%，PSNR 的變化及殘留浮水印比例。

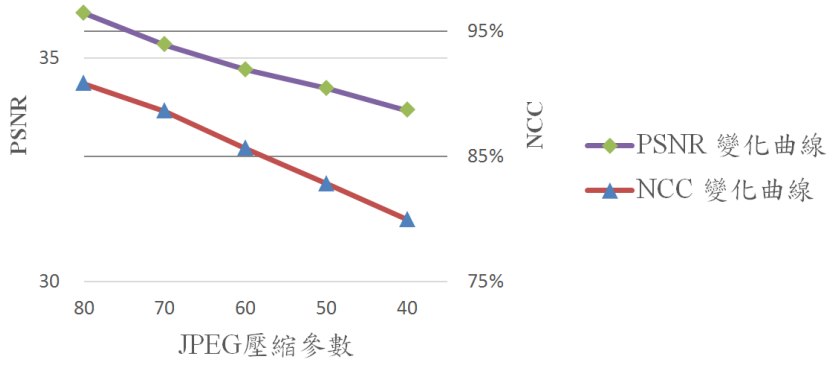


圖 6 JPEG 壓縮對圖片品質及殘存浮水印的影響

圖 7 顯示圖片經過放大/縮小 150% 至 25%，並經由原圖校正後，PSNR 的變化及殘留浮水印比例。

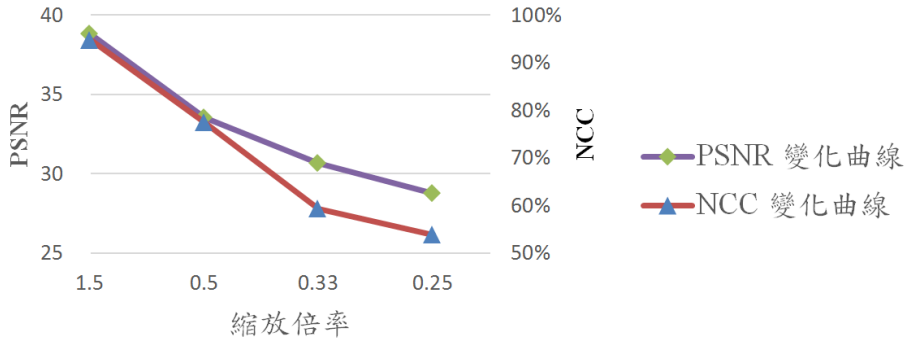


圖 7 放大/縮小不同倍率對圖片品質及殘存浮水印的影響

圖 8 為圖片經過原圖校正後，旋轉角度誤差在 0.1 度至 0.5 度時，PSNR 的變化及殘留浮水印比例。

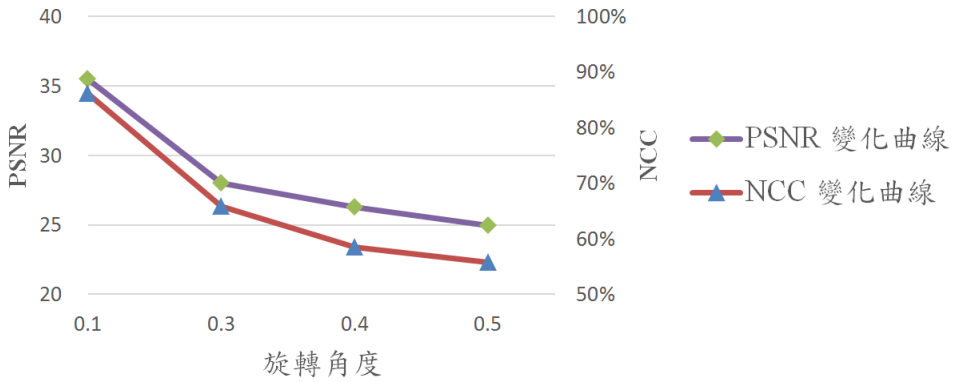


圖 8 旋轉角度對圖片品質及殘存浮水印的影響

圖 9 為圖片經由原圖校正後，水平平移誤差 0.2 至 0.8 像素時，PSNR 的變化及殘留浮水印比例。

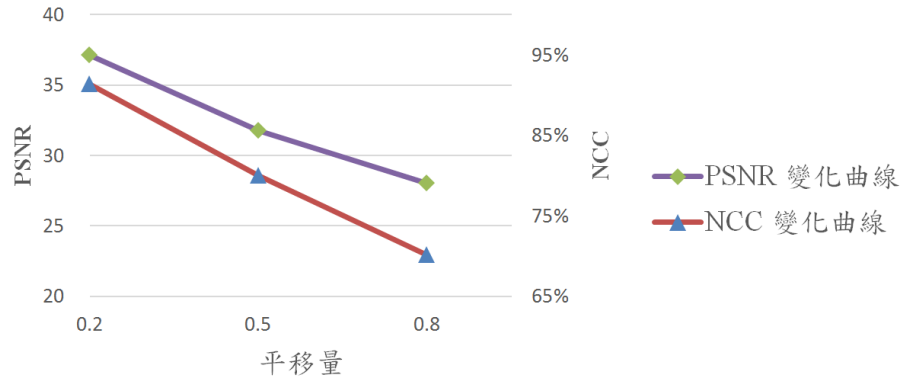


圖 9 平移對圖片品質及殘存浮水印的影響

圖 10 為圖片加入密度為 0.005 至 0.025 的胡椒鹽雜訊後，PSNR 的變化及殘留浮水印比例。其中胡椒鹽雜訊的密度是指所有像素中，受到影響的像素比例。

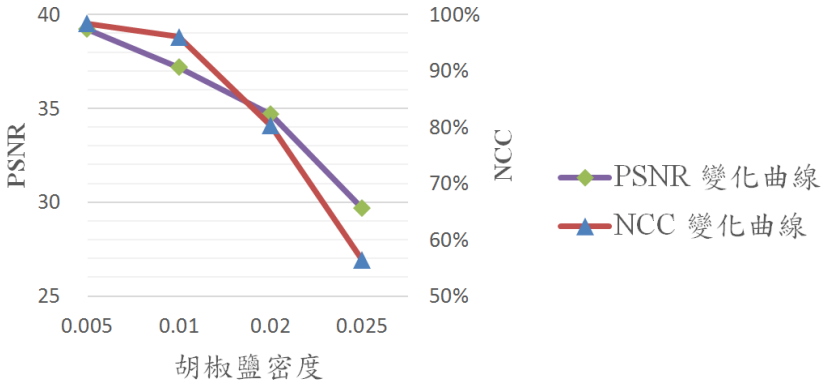


圖 10　雜訊密度對圖片品質及殘存浮水印的影響

圖 11 為將圖片經由離散餘弦轉換 (DCT) 轉至頻率域，再把高頻的部分係數清除為 0 後轉換回空間域，得到處理後圖片的 PSNR 變化及殘留浮水印比例。假設轉換域最左上角 (最低頻) 之點為原點，距離原點一定半徑長度以外的係數值將被設定為 0，圖中橫軸的濾除門檻參數即為此半徑距離。

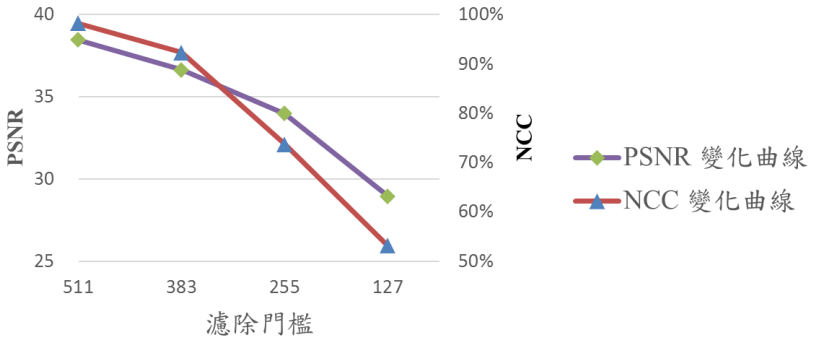


圖 11　濾掉高頻部分對圖片品質及殘存浮水印的影響

可以看出嵌入浮水印的圖片在上述實驗中即使經過各式的影像處理，但只要 PSNR 仍大於 30，大部分的浮水印依舊會殘留在圖片上，浮水印被完全移除的機率極低。

在我們的實驗過程中也發現，本文低密度而不均勻的嵌入浮水印方式，將導致使用於保護非常平滑的影像時，容易使用中位數濾波處理將浮水印近乎完全清除。實驗結果如圖 12 所示，為圖片經由三乘三、五乘五、七乘七的中位數濾波處理後，PSNR 的變化及殘留浮水印比例。

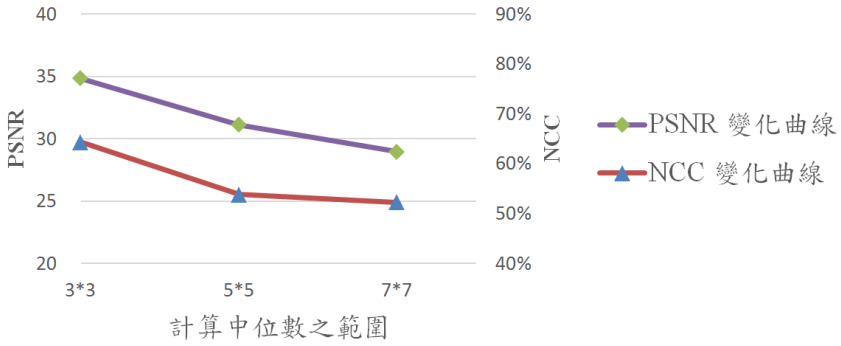


圖 12　中位數濾波處理對圖片品質及殘存浮水印的影響

雖然浮水印的不可移除性並非本文的主要目標，卻是先前大部分數位浮水印論文所欲達成的目標，強化了證據力的浮水印仍須仰賴不容易被移除的嵌入方式，方能發揮實務價值。如果將本文的系統中直接修改某個像素中一個位元的作法改為架構於展頻浮水印技術 [8] 之上，並以密碼學上不可預測的亂數序列做為展頻系統中的 PN 序列，如此可大幅強化浮水印的不可移除性。

第六章 結論

當公開使用的數位影像疑似未經授權抄襲或修改，著作權擁有者可提出侵權的訴訟，本文以過去所提出的訊號處理機制為基礎，並著重在如何使用以數位簽章做為種子產生不可預測的虛擬亂數序列，建立浮水印和擁有人及原始圖片的唯一連結，降低浮水印被惡意移除的機率，並使得浮水印的偽陽性偵測率降低到計算上可忽略的大小，即對任何使用者來說，近乎不可能在沒有嵌入過自己浮水印的圖片之中，找到能夠以自身公鑰成功偵測的浮水印。在此機制之下，透過具公信力的第三方進行所有權爭議解決步驟，便足以證明爭議圖片來源歸屬，進而捍衛擁有者之各項著作權。第五章所提出的實驗決定了浮水印雜訊密度 β 並驗證此浮水印在一般傳輸及影像處理下的強健性。其中最重要的是殘留一定比例的浮水印即可證明所有權的歸屬。

參考文獻

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

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

[3] 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.

[4] A. Adelsbach, and A. R. Sadeghi, “Advanced Techniques for Dispute Resolving and Authorship Proofs on Digital Works,” Proc. of SPIE: Security and Watermarking of Multimedia Contents V, Vol. 5020, 2003.

[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] D. Boneh and J. Shaw, “Collusion-Secure Fingerprinting for Digital Data,” Advances in Cryptology-Crypto'95, LNCS 963, Springer, 452–465, 1995.

[7] 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.

[8] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, “Secure Spread Spectrum Watermarking for Multimedia,” IEEE Trans. on Image Processing, 6, 12, 1673–1687, 1997.

[9] I. J. Cox and M. L. Miller, “The First 50 Years of Electronic Watermarking,” EURASIP Journal on Advances in Signal Processing, 2002(2), 126–132, 2002.

[10] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, Digital Watermarking and Steganography, 2nd Ed., Morgan Kaufmann Publishers Inc., 2008.

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

[12] 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.

[13] C. Fei, D. Kundur, and R. H. Kwong, “Analysis and Design of Secure Watermark-Based Authentication Systems,” IEEE Trans. on Information Forensics and Security 1(1), 43–55, 2006.

[14] J. Fridrich, “Robust Bit Extraction from Images,” Proc. IEEE Intern. Conf. Multimedia Computing and Systems (ICMCS'99), 1999.
[15] O. Goldreich, Foundations of Cryptography: Volume 1, Basic Tools, Cambridge University Press, 2000.
[16] 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.
[17] M. Kutter, S. Voloshynovskiy, and A. Herrigel, “The Watermark Copy Attack,” Proc. of SPIE: Security and Watermarking of Multimedia Contents II, Vol. 3971, 2000.
[18] 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.
[19] C.-Y. Lin and S.-F. Chang, “Generating Robust Digital Signature for Image/Video Authentication,” Multimedia and Security Workshop at ACM Multimedia'98, 1998.
[20] B. Pfitzmann and M. Schunter, “Asymmetric Fingerprinting,” Advances in Cryptology - Eurocrypt'96, LNCS 1070, Springer, 84–95, 1996.
[21] L. Qiao and K. Nahrstedt, “Watermark Methods for MPEG Encoded Video: Towards Resolving Rightful Ownership,” Proc. of ICMCS, Vol. 9, 194–210, 1998.
[22] 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.
[23] H. T. Sencar and N. Memon, “Watermarking and Ownership Problem: A Revisit,” Proc. of the ACM Workshop on Digital Rights Management, 2005.
[24] V. Shoup, “Sequences of Games: A Tool for Taming Complexity in Security Proofs,” Cryptology ePrint Archive: Report 2004/332.
[25] K. SriSwathi and S. G. Krishna, “Secure Digital Signature Scheme for Image Authentication over Wireless Channels,” Int. J. Comp. Tech Aool. 2 (5): 1472–1479, 2011.
[26] L. F. Turner, “Digital Data Security System,” Patent IPN WO 89/08915, 1989.
[27] X.-G. Xia, C. G. Boncelet, and G. R. Arce, “Wavelet Transform Based Watermark for Digital Images,” Comm. ACM 57(3): 86–95, 2014.
[28] W. Zeng and B. Liu, “On Resolving Rightful Ownerships of Digital Images by Invisible Watermarks,” Proc. of the International Conference on Image Processing, 552–555, 1997.
[29] E. Zielinska, W. Mazurczyk, and K. Szczypiorski, “Trends in Steganography,” Comm. ACM 57(3): 86–95, 2014.