Facultatea de Electronică şi Telecomunicaţii
Universitatea Politehnica Timişoara

Corina Naforniță

Creşterea securităţii reţelelor de comunicaţii de date prin autentificare bazată pe watermarking

Raport de cercetare din cadrul grantului CNCSIS, de tip TD, cod 47,
număr 33385/29.06.04, tema 15,
număr adiţional 2930/2006 tema 9

Cuprins

1. Introducere în securitatea rețelelor de comunicații de date

1.1 Introducere

Nevoia de protejare a informației digitale împotriva copierii şi manipulării ilegale a apărut o dată cu dezvoltarea pe scară largă a comunicațiilor prin intermediul *Internet*-ului. Publicarea electronică şi comerțul electronic a informațiilor digitale cresc pericolul de falsificare şi furt intelectual.

Dezvoltarea rapidă a tehnologiei digitale face foarte necesară dezvoltarea metodelor pentru protejarea produselor multimedia împotriva pirateriei. Atacurile pirat includ accesul ilegal al datelor pe *Internet*, modificări ale conținutului făcute cu rea-voință, retransmisia copiilor neautorizate. Impactul acestui gen de atacuri ar putea fi foarte mare atât pe plan financiar (pierderi financiare cauzate de accesarea şi folosirea neautorizată a datelor), precum şi în planul securității.

Când este vorba de semnale analogice, problema se rezolvă de la sine, deoarece copiile sunt de o calitate mai redusă decât originalele (casete audio şi video). În schimb, informația digitală poate fi copiată perfect şi distincția între original şi copii este dificil, dacă nu imposibil de făcut. În plus, nu există nici un mecanism pentru a depista copierea ilegală sau modificarea conținutului. Pentru a asigura protecția împotriva copierii şi protecția copyright-ului pentru date audio-video digitale, s-au dezvoltat două tehnici complementare: criptarea şi marcarea transparentă. Sistemele de criptare pot fi folosite pentru a proteja datele digitale pe durata transmisiei, de la transmițător la receptor [BajBor01]. După ce receptorul a recepționat datele decriptate, datele sunt identice cu cele originale şi nu mai sunt protejate. Tehnicile de marcare transparentă pot complementa criptarea prin plasarea unui semnal secret, imperceptibil, adică a unui marcaj transparent, direct în datele originale în aşa fel ca aceasta să rămână prezent întotdeauna [VoyPit99].

Criptografia este ştiinta scrierilor secrete. Folosirea acestei metode reduce riscul unei utilizări neautorizate a datelor, dar prin această metodă nu putem avea nici o informație cu privire la proprietarul datelor. După ce au fost decodate legal sau ilegal, datele pot fi redistribuite. În acest caz, atacurile vizează decodarea datelor.

Steganografia. Este tehnica ce se ocupă cu transmiterea datelor ascunse în imagini. Cu ajutorul acestei tehnici se pot transmite mesaje secrete, rolul ei nefiind acela de a proteja imaginea, ci de a transmite pe un canal, diverse mesaje ascunse. Atacurile în cazul steganografiei vizează descoperirea mesajului.

Amprentarea (fingerprinting). Amprenta este un fel de număr de serie ascuns. Amprentarea este folosită pentru a afla care din utilizatorii autorizați, clienți, a încălcat contractul prin furnizarea datelor unor terți neautorizați.

Marcarea transparenta. Această metodă este folosită pentru a ne da informații cu privire la proprietarul datelor. De asemenea, ne mai poate oferi şi alte tipuri de informații, în funcție de marcajul ce este înglobat în informația gazdă ce trebuie protejată.

1.2 Criptarea şi semnătura digitală

"There are two kinds of cryptography in this world: cryptography that will stop your kid sister from reading your files, and cryptography that will stop major governments from reading your files. This book is about the latter."

--Bruce Schneier, Applied Cryptography: Protocols, Algorithms, and Source Code in C--

Datele care pot fi citite şi înţelese fără măsuri speciale se numesc date clare. Metoda prin care datele clare sunt mascate în aşa fel încât să ascundă esenţa se numeşte criptare, rezultând date cifrate. Procesul invers de transformare a datelor cifrate în date clare se numeşte decriptare. Criptarea convenţională, numită şi criptare cu cheie secretă sau cheie simetrică, foloseşte o cheie atât pentru criptare cât şi pentru decriptare. Cheia este cunoscută doar de către destinatarul mesajului. Pentru toţi ceilalţi utilizatori ai reţelei cheia este secretă. Witfield Diffie şi Martin Hellman, cercetători la universitatea Stanford au pus, în anul 1976, bazele criptografiei asimetrice cu chei publice [DifHel76]. Aceasta este o metodă asimetrică care foloseşte două chei: una publică, care criptează datele, şi o cheie privată pentru decriptare. Cheia publică poate fi cunoscută de către orice utilizator al reţelei în timp ce cheia privată e secretă. Pe baza cheii publice se poate face identificarea sursei de unde soseşte un anumit mesaj. Cu alte cuvinte folosind acest algoritm se poate face şi autentificarea mesajului.

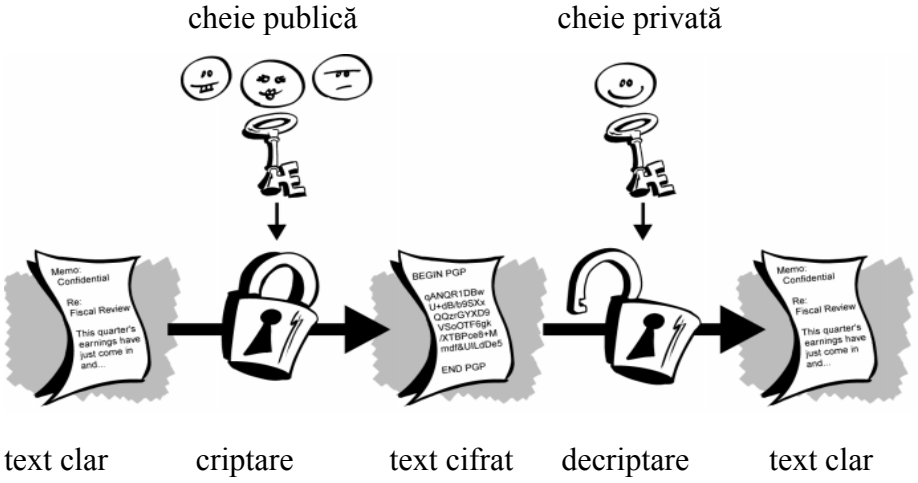

Figura 1.1. Criptarea cu cheie publică.

Unul din avantajele criptării cu cheia publică este că oferă o metodă pentru implementarea semnăturilor digitale. De aceea Guvernul S.U.A a decis elaborarea unui standard de semnătură digitală bazat pe utilizarea cheilor publice şi secrete. Standardul DSS (*Digital Signature Standard*) a fost publicat în 1991. Semnătura digitală oferă posibilitatea de verificare a autentificării originii informaţiei şi a integrităţii ei. În loc să criptăm informaţia folosind cheia publică a altei persoane, o criptăm cu cheia privată personală. Dacă informaţia poate fi decriptată cu cheia publică personală, atunci

informația provine de la tine. Metoda cea mai simplă de inserare a semnăturii digitale e arătată în figura de mai jos:

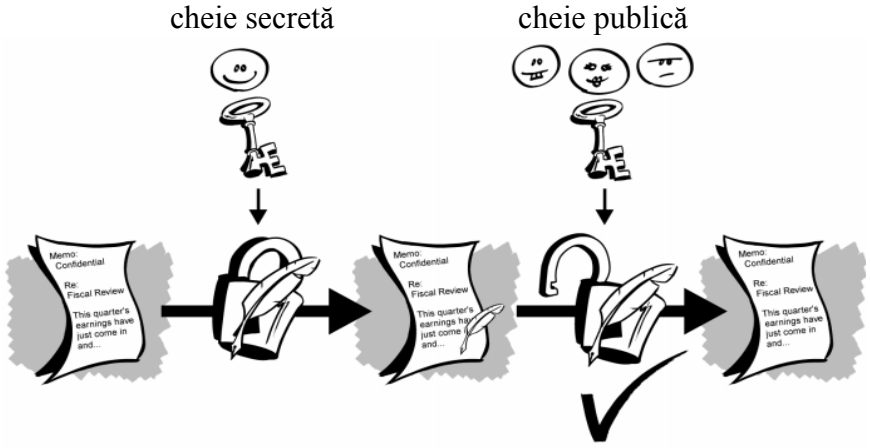

Figura 1.2. Inserarea unei semnături digitale.

O semnătură digitală are acelaşi scop ca şi o semnătură scrisă de mână. Diferența este că cea scrisă de mână e uşor de falsificat pe când cea digitală e aproape imposibil de fasificat.

1.3 Servicii de autentificare

Autentificarea exprimă ideea că unele resurse au fost puse la dispoziție pentru a garanta că entitățile sunt ceea ce susțin că sunt, sau că informația nu a fost manipulată de parți neautorizate. Autentificarea este specifică obiectivelor securității pe care încearcă să le atingă. Exemple de obiective includ controlul accesului, autentificarea entităților, autentificarea mesajelor, integritatea datelor, nerepudierea şi autentificarea cheilor.

Până la mijlocul anilor ’70 se credea că autentificarea şi confidențialitatea sunt conectate intrinsec. Odată cu descoperirea funcțiilor *hash* şi a semnăturilor digitale s-a observat că autentificarea şi confidențialitatea sunt obiective ale securității informației separate şi independente. Separarea lor nu e doar folositoare, ci şi esențială. De exemplu dacă Alice, aflată într-o țară, comunică cu Bob, aflat în altă țară, statele gazdă s-ar putea sau nu să permită confidențialitatea canalului; una dintre țări, sau ambele, ar putea dori să monitorizeze toate comunicațiile. Însă Alice şi Bob vor să fie siguri de identitatea celuilalt, de integritatea şi originea informațiilor pe care le trimit şi pe care le primesc.

Scenariul anterior relevă câteva aspecte independente ale autentificării. Dacă Alice şi Bob vor asigurări despre identitatea celuilalt sunt două posibilitați:

-Alice şi Bob comunică fără întârzieri apreciabile de timp adică comunică în timp real;

-Alice şi Bob schimbă mesaje cu o anumită întârziere adică mesajele sunt rutate prin numeroase rețele, stocate şi apoi redirijate după o perioadă.

În primul caz, Alice şi Bob vor dori să-şi verifice identitatea în timp real. Acest lucru se poate realiza dacă Alice îl provoacă pe Bob să răspundă la o întrebare la care

numai el ştie răspunsul corect. Acelaşi lucru îl poate face şi Bob pentru a o identifica pe Alice. Aceasta este o autentificare a entităţilor sau, mai simplu, o identificare.

În cel de-al doilea caz nu e indicată o întrebare şi aşteptarea unui răspuns şi, în plus, s-ar putea ca, comunicarea să se desfaşoare într-o singură direcţie. Sunt necesare alte tehnici pentru autentificarea originii mesajului. Aceasta este o noua formă de autentificare numită *autentificare a originii datelor*.

Cel mai puternic şi folosit serviciu de autentificare din lume este *Kerberos Authentication Server*, creat la MIT. Kerberos oferă un mijloc de verificare a identităţii principalelor calculatoare (de exemplu staţie de lucru sau server de reţea) dintr-o reţea. Autentificarea se face cu ajutorul unei autorităţi de încredere numită terţ de încredere (*trusted third-party*). Această sarcină (autentificarea) este îndeplinită fără a face apel la autentificarea asigurată de către sistemul de operare al calculatorului gazdă, fără a avea încredere în adresa sa, fără a fi necesară prezenţa unor dispozitive suplimentare de securitate şi pe baza ipotezei că pachetele care traversează reţeaua pot fi citite sau modificate. Totuşi numeroase funcţii ale sistemului Kerberos sunt folosite doar pentru iniţierea unei conexiuni şi presupun absenţa oricărui atac de tipul *hijacke.* O astfel de utilizare are implicit încredere în adresele calculatoarelor gazdă implicate. În aceste condiţii Kerberos realizează autentificarea, de tipul celei oferite de o a treia parte, de încredere. Se folosesc în acest scop metode convenţionale de criptare, cum ar fi algoritmii simetrici. În schema de autentificare a protocolului Kerberos sunt implicate următoarele entităţi:

- Serverul de autentificare Kerberos, AS,
- Serverul de acordare a tichetului, TGS,
- Clientul C, care tebuie autentificat, pentru a i se acorda acces la un seviciu furnizat de serverul S,
- Serverul S, la care se cere acces din partea clientului C.

Procesul de autentificare se desfăşoară după cum urmează: Un client, C, transmite o cerere server-ului de autentificare, AS, cerând acreditări pentru un anumit server, S. AS răspunde cu aceste acreditări, criptate, în cheia clientului C. Aceste acreditări constau în:

1) un "tichet" pentru server;
2) o cheie de criptare temporară (de obicei numită cheie de sesiune).

Clientul transmite "tichetul" (care conţine identitatea clientului şi o copie a cheii sesiunii, ambele criptate în cheia serverului) serverului. Cheia sesiunii (acum împărţită de client şi de server) este folosită pentru autentificarea clientului şi poate fi opţional utilizată pentru autentificarea serverului. Ea poate fi folosită şi pentru a cripta în continuare comunicaţia între cele două părţi sau pentru a permite schimbul unei chei separate de sub-sesiune care să fie folosită pentru criptarea comunicaţiei ulterioare.

Implementarea Kerberos-ului constă din unul sau mai multe servere de autentificare care rulează pe calculatoare gazdă sigure. Serverele de autentificare întreţin o bază de date de utilizatori principali (utilizatori sau servere) şi cheile lor secrete. Există biblioteci de programe pentru realizarea criptării şi implementarea protocolului Kerberos. Din dorinţa de a adăuga posibilitatea autentificării unor tranzacţii, o aplicaţie de reţea

tipică adaugă una sau două apeluri la biblioteca Kerberos, pentru a se transmite mesajele necesare pentru realizarea autentificării.

Odată obținute, acreditările trebuie folosite pentru verificarea identității utilizatorilor principali care intervin într-o anumită tranzacție, pentru a se asigura integritatea mesajelor schimbate între ei, sau pentru a asigura securitatea mesajelor. În aplicația curentă poate fi ales nivelul de protecție necesar.

Pentru a se verifica identitățile utilizatorilor principali dintr-o tranzacție, clientul trimite "tichetul" serverului. Deoarece "tichetul" poate fi interceptat şi modificat de către un atacator, se transmite informație adițională pentru a se demonstra că mesajul a fost transmis de către utilizatorul principal căruia i-a fost atribuit "tichetul". Această informație (numită autentificare) este criptată în cheia sesiunii şi include o ştampilă temporală. Aceasta dovedeşte că mesajul a fost generat recent şi că nu este un răspuns.

Criptarea autentificatorului în cheia sesiunii demostrează că aceasta a fost generată de către cineva care posedă această cheie. Deoarece nimeni în afară de utilizatorul principal, care face solicitarea şi server nu cunoaşte cheia sesiunii (aceasta nu este nici odată transmisă în clar în rețea) se obține garanția identității clientului.

Integritatea mesajelor schimbate între utilizatorii principali mai poate fi garantată şi folosind cheia sesiunii (trecută în "tichet" şi conținută în acreditări). Această abordare asigură atât detectarea atacurilor de tip răspuns cât şi a atacurilor bazate pe modificarea şirului de date al mesajului. Ea se bazează pe generarea şi transmiterea valorii unei funcții *hash* de mesajul clientului, criptată cu cheia sesiunii. Confidențialitatea şi integritatea mesajelor schimbate între principalii utilizatori pot fi realizate prin criptarea datelor de transmis folosind cheia sesiunii trecută în "tichet" şi conținută în acreditări.

Funcțiile *hash* – cunoscute şi sub denumirea de funcții de compresie, contracție sau de rezumat al mesajului ori amprentă digitală – sunt funcții care au la intrare un şir de date de lungime variabilă *n* – numit preimagine – şi la ieşire dau un şir de lungime fixă *m* (uzual 128 sau 160 biți) numit valoarea funcției hash.

2. Sisteme de marcare transparenta

2.1 Definiții

"A distinguishing mark or device impressed in the substance of a sheet of paper during manufacture, usually barely noticeable except when the sheet is held against strong light"
- Oxford English Dictionary -

Watermarking-ul adică marcarea transparentă este operația de introducere a unei informații transparente, imperceptibile pentru sistemul auditiv, respectiv vizual numită *marcaj*- watermark în informația gazdă, care poate fi text, semnal audio, imagine statică sau video. Denumirea de watermark provine de la cuvintele din limba engleză water- apă şi mark – marcaj şi desemnează un marcaj transparent, invizibil, asemănător transparenței apei.

Marcajul conține, în general, informații despre originea şi/sau destinația informației gazdă. Deşi nu e folosit direct în protecția proprietății intelectuale, el ajută la identificarea sursei şi destinatarului, fiind util în cazul disputelor privind dreptul de autor sau distribuitor al informației.

Teoretic, marcajul trebuie să protejeze informația permanent, deci trebuie să aibă calitatea de a fi robust, astfel încât să nu poată fi înlăturat din informația gazdă, fără degradarea esențială a calității acesteia. Acest marcaj este asemănător unei semnături cu observația că trebuie să fie transparent.

Procedeul de marcare transparentă pentru a putea realiza protejarea informației, constă din două operații:

- introducerea marcajului în datele gazdă, înainte de transmisie sau stocare;
- extragerea marcajului din datele recepționate şi compararea marcajului adăugat la emisie cu cel extras la recepție, pentru autentificare, în caz de dispută.

2.2 Proprietățile marcării transparente

Pentru a fi eficient, un marcaj trebuie să fie:

1. **Imperceptibil** – Marcajul ar trebui să fie perceptual invizibil, sau prezența sa nu ar trebui să interfereze cu datele care trebuiesc protejate. Cu toate acestea, chiar şi diferențe abia vizibile se accentuează dacă produsul original este direct comparat cu cel marcat. De aceea, aceste diferențe rămân neobservate pentru un observator uman deoarece produsul original este accesibil numai proprietarului legal.

2. **Robust** – Marcajul trebuie să fie dificil de înlăturat (teoretic imposibil de înlăturat). Dacă se dispune doar de o informație parțială despre marcaj (de exemplu, nu se cunoaşte cu exactitate localizarea marcajului în imagine), atunci încercările de înlăturare sau de

distrugere a marcajului ar trebui să conducă la degradarea severă a calității imaginii. Evident, un marcaj folosit cu scopul de protejare a dreptului de autor ar trebui să fie detectabil până în punctul în care calitatea produsului rămâne în limite acceptabile pentru orice tip de modificare. În particular, un marcaj ar trebui să fie robust la:

a. **Procesări obişnuite ale semnalului** – Marcajul ar trebui să poată fi încă extras, chiar dacă imaginii i se aplică procesări obişnuite. Acestea pot fi conversii D/A, A/D, reeşantionare, compresie cu pierderi, precum şi îmbunătățiri obişnuite aplicate unei imagini cum ar fi îmbunătățirea contrastului, corecția culorilor, egalizarea histogramei etc.

b. **Distorsiuni geometrice obişnuite (pentru imagini şi video)** – Marcajele din imagini şi video ar trebui să fie rezistente şi împotriva operațiilor geometrice cum ar fi rotire, translație, decupare, scalare.

c. **Atacuri de subterfugiu: prin înțelegere secretă/complotare şi falsificare** – În plus, marcajul ar trebui să fie robust împotriva mai multor indivizi care fiecare deține o copie marcată a datei. Cu alte cuvinte, marcarea ar trebui să fie robustă împotriva combinării copiilor aceluiaşi original. De asemenea, dacă un marcaj urmează să fie folosit ca probă juridică, trebuie să fie imposibil pentru atacatori să combine imaginile lor pentru a genera un marcaj diferit, valid cu intenția de a înşela o terță parte.

3. **Universal** – Acelaşi algoritm de marcare transparentă ar trebui să fie aplicabil pentru toate cele trei tipuri de date considerate. Acest lucru serveşte în marcarea produselor multimedia. De asemenea, această caracteristică este favorabilă pentru implementarea de algoritmi de marcare a imaginilor şi semnalelor video pe suport hardware comun.

4. **Neambiguu** – Recuperarea marcajului ar trebui să identifice fără îndoială proprietarul. De asemenea, acuratețea identificării proprietarului ar trebui să se degradeze progresiv în fața atacurilor.

Alte proprietăți pe care marcajele ar trebui să le aibă sunt:

5. **Complexitate** – Semnalele de marcare ar trebui să fie foarte complexe. Acest lucru este necesar pentru a se putea produce un set suficient de marcaje sesizabile. Un set foarte mare de marcaje previne refacerea unui marcaj anumit prin proceduri iterative de tip *trial and error*. În majoritatea cazurilor, complexitatea unei marcaj este direct legată de mărimea produsului în care trebuie să fie înglobată.

6. **Cheia asociată** – Marcajele ar trebui să fie asociate cu un număr de identificare numit cheia marcajului. Cheia este folosită pentru a forma, detecta şi înlătura marca. Prin urmare, cheia ar trebui să fie privată şi să caracterizeze exclusiv proprietarul legal. Orice semnal digital, extras dintr-un produs digital se presupune că este marcajul dacă şi numai dacă el este asociat cu o cheie printr-un algoritm bine stabilit. Această condiție previne crearea unor marcaje contrafăcute.

7. **Detecție/căutare automată –** Marcajele ar trebui să combine în mod facil o metodă de căutare care scanează automat orice domeniu accesibil în rețea.
8. **Detecție de încredere** – Marcajele ar trebui să constituie o dovadă suficientă şi de încredere a proprietății asupra unui anume produs. Detecția alarmelor false ar trebui să apară foarte rar (de preferință niciodată). Un anumit marcaj este o dovadă credibilă pentru a demonstra posesia dreptului de autor atunci când probabilitatea de eroare la înglobarea ei într-o imagine digitală este nesemnificativă. Totuşi, detecția cu un nivel de certitudine scăzut poate fi făcută cu scopul de a reduce probabilitatea de refuz în timpul monitorizării pe *web*.

9. **Invizibilitate statistică** – Marcajele nu ar trebui să poată să fie refăcute folosind metode statistice. De exemplu, proprietatea unui număr mare de produse digitale, marcate cu aceeaşi cheie, nu ar trebui să permită extragerea marcajului aplicând metode statistice. De aceea, marcajele ar trebui să depindă de conținutul produsului.

10. **Marcaje multiple** – Ar trebui să fim capabili să înglobăm un număr suficient de marcaje în aceeaşi imagine. Fiecare marcaj ar trebui să fie detectabil folosind cheia unică corespunzătoare. Această caracteristică pare să fie necesară deoarece nu putem preveni ca cineva să marcheze un produs deja marcat. Este de asemenea avantajos în cazurile în care dreptul de autor este transferat de la un proprietar la altul (un proces asemănător celui de amprentare). Menționăm că proprietarul legal al imaginii este singurul care poate dispune de o copie a imaginii care conține marca sa.

2.3 Aplicații posibile ale marcării transparente

Marcarea transparentă prezintă interes şi pentru aplicații care nu țin de securizarea informațională.
1. Monitorizarea transmisiilor TV – Dacă o firmă care îşi face publicitate, doreşte să afle câte din reclamele plătite au fost efectiv transmise, poate să monitorizeze transmisiile TV cu ajutorul observatorilor umani. Desigur, acest lucru se poate dovedi extrem de costisitor şi în plus nefiabil. Există de asemenea sisteme de monitorizare, care nu apelează la observatori umani. Acestea se împart în două categorii: pasive şi active. Sistemele de monitorizare pasive, încearcă să recunoască direct conținutul difuzat, la fel ca şi observatorii umani. Aceste sisteme sunt calculatoare care compară semnalul difuzat cu semnalele pe care le au în baza de date (şi anume, semnalele care reprezintă spoturile publicitare). Aceste sisteme se pot dovedi nepractice, din cauza mărimii bazelor de date. În practică, aceste sisteme nu sunt folosite pentru a verifica dacă, de exemplu, o reclamă a fost difuzată. Ele sunt folosite mai ales pentru a obține date despre competitori.

Pentru a obține acuratețea cerută de procesul de verificare, ar trebui folosite sisteme de monitorizare active, care se bazează pe informații asociate, difuzate o dată cu conținutul propriu-zis al reclamelor. Marcarea poate fi o soluție pentru monitorizarea activă a transmisiilor TV.

2. Identificarea proprietarului – Acest lucru se poate face printr-o inscripționare vizibilă a autorului. Acest tip de „marcare” a proprietății poate însă fi uşor eliminat din

semnalul multimedia respectiv. Cel mai bun exemplu în acest sens este decuparea unei porţiuni dintr-o imagine, care să nu conţină „marca" autorului. Deoarece marcajele pot fi imperceptibile şi inseparabile de semnalul original, pot reprezenta o soluţie ideală pentru identificarea autorului.

3. Dovada proprietăţii – Ar fi de dorit ca marcajele să servească nu numai pentru a „marca" proprietatea, dar chiar să o dovedească. Dacă Alice creează o imagine şi o marchează cu marca „© 2006 Alice", atunci Bob poate fura imaginea respectivă, şi folosind un program de procesare a imaginilor, poate înlocui marca cu „© 2006 Bob". Dacă Alice nu a înregistrat imaginea la o autoritate centrală, ea va trebui să demonstreze că imaginea îi aparţine. Dacă atacatorul nu dispune de un detector al marcajului, eliminarea acestuia poate fi greu de făcut. Pe de altă parte, chiar dacă marcajul nu poate fi eliminat, folosind propriul sistem de marcare, Bob poate să arate că marcajul lui ar exista în originalul lui Alice. Astfel, o terţă parte nu ar putea să îşi dea seama cui aparţine imaginea. Această problemă ar putea fi rezolvată dacă, în loc de a demonstra proprietatea prin marcare, s-ar demonstra că o imagine derivă din alta.

4. Înregistrarea operaţiilor efectuate (*Transaction tracking*) – Marcajul înregistrează una sau mai multe operaţii care au fost făcute asupra copiei unui produs multimedia. De exemplu, marcajul poate „memora" o identitate a cumpărătorului (se presupune că fiecare cumpărător are o copie diferită a originalului, marcajele nefiind aceleaşi). În acest fel se pot identifica piraţii sau chiar scurgerile de informaţii. Industria de film americană pierde anual 3 miliarde de dolari din cauza copiilor ilegale. Una dintre cauze este distribuirea pentru cei 5803 de membri cu drept de vot a filmelor nominalizate la Oscar. Prin introducerea unui marcaj individual pentru fiecare membru, s-au descoperit cei care furnizau copiile ilegale.

5. Autentificarea conţinutului – Acest lucru poate fi realizat prin înglobarea semnăturii digitale în semnalul multimedia. Această semnătură mai este cunoscută şi sub numele de marcă de autentificare. Dacă un semnal ce conţine o astfel de marcă este modificat, se poate afla cum a fost distorsionat.

6. Controlul copierii (*Copy Control*) – Prevenirea apariţiei copiilor ilegale poate fi făcută prin criptare. Există trei posibilităţi prin care un adversar poate obţine acces neautorizat la produse multimedia: dacă decriptează datele fără a avea o cheie; dacă obţine o cheie prin reverse-engineering; sau cel mai simplu dacă obţine o cheie în mod legal, făcând copii ilegale ale datelor decriptate. Marcajele însă pot rămâne în conţinut şi după decriptare. Cu toate acestea, protejarea DVD-urilor împotriva copierii nu a fost făcută încă cu succes, deoarece nu orice DVD-player conţine un detector al marcajului.

2.4 Principii de bază ale marcării transparente

După cum am văzut, principial marcarea transparentă constă din două prelucrări de bază desfăşurate la emisie, respectiv la recepţie:

-introducerea marcajului, cu respectarea cerinţelor de transparenţă perceptuală şi robusteţe, în datele gazdă ce urmează a fi marcate;

-extragerea marcajului din semnalele marcate recepționate (posibil modificate) şi compararea acestuia cu valoarea introdusă la emisie, în caz de dispută.

Pentru a îndeplini cerința de robustețe, marcajul introdus la emisie va fi guvernat de una sau mai multe chei criptografice sigure (secrete sau publice), chei necesare şi în procesul de detecție de la recepție.

Transparența perceptuală se realizează în concordanță cu un anumit criteriu de perceptibilitate care poate fi implicit sau explicit. Astfel, eşantioanele individuale ale semnalului gazdă, folosite pentru inserarea informației de marcaj vor putea fi modificate numai între anumite limite situate sub pragurile de sensibilitate ale simțurilor umane (auz, văz). În cazul prelucrării imaginilor vom obține o mască perceptuală care ne va spune cât de mult pot fi alterați anumiți pixeli, şi care sunt aceştia, fără a afecta calitatea imaginii.

Un exemplu de mască perceptuală îl putem vedea în figura următoare [PerHer99].

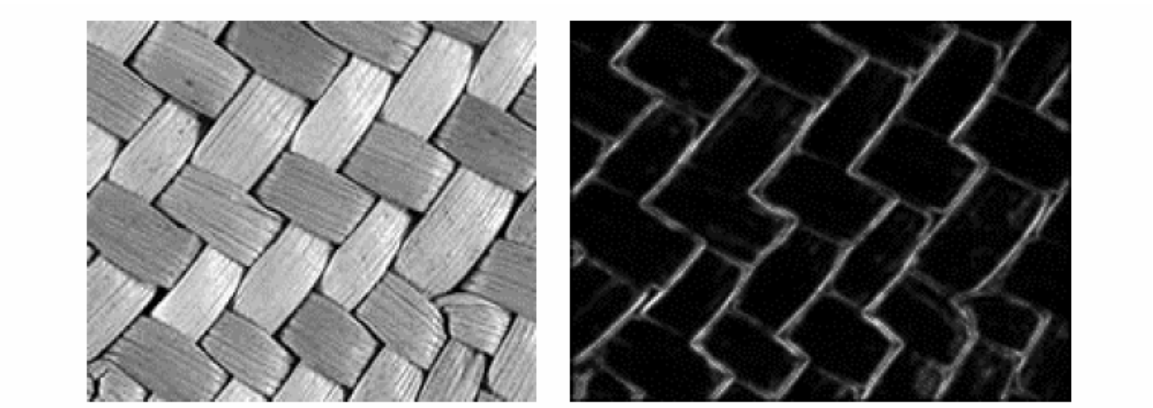

Figura 2.1. a)imagine originală a unei țesături; b)masca perceptuală a țesăturii.

Inserția transparentă a marcajului în semnalul digital gazdă este posibilă numai datorită faptului că destinatarul final este omul. Simțurile sale sunt detectoare imperfecte caracterizate de praguri de sensibilitate (intensitate sonoră respectiv nivel de contrast) minime precum şi de fenomenul de *mascare*. Mascarea se referă la faptul că o componentă, dintr-un semnal dat, audio sau video, poate deveni imperceptibilă în prezența unui alt semnal, numit semnal de mascare. Majoritatea tehnicilor de codare ale semnalelor audio şi video folosesc caracteristicile sistemului auditiv uman (HAS – *human audio system*) şi ale sistemului vizual uman (HVS – *human visual system*) direct sau indirect.

Pentru ca semnalul de marcaj să fie robust (în ciuda amplitudinii mici a acestuia cerută de condiția de transparență), el este împrăştiat pe mai multe eşantioane în conformitate cu cerințele de granularitate, ceea ce conduce la detecția lui şi din date marcate afectate de distorsiuni. Tehnica de împrăştiere este similară cu cea de întrețesere şi este folosită pentru a preveni distrugerea semnificativă, sau chiar înlăturarea marcajului în urma unor atacuri de *cropping*. Modul în care împrăştierea este efectuată depinde de cheia secretă, unică pentru fiecare marcaj în parte.

Inserția marcajului

Marcarea transparentă conține două prelucrări distincte [BajBor01]:

-*Generarea semnalului de marcaj* (watermark): notat cu W, ce va fi adăugat apoi semnalului gazdă X; în general semnalul W depinde de informația de marcaj I şi de o cheie K:

$$W=E_1(I,K),$$

unde E_1 este o funcție - de cele mai multe ori o modulație sau o împrăştiere.
Sunt aplicații în care semnalul de marcaj W poate depinde de semnalul gazdă X:

$$W=E_1(I,X,K).$$

-*Introducerea semnalului de marcaj* W în semnalul gazdă X, astfel încât să fie îndeplinite condițiile de transparență perceptuală şi de robustețe, rezultând semnalul marcat Y:

$$Y=E_2(X,W).$$

Marcarea transparentă poate avea loc în orice domeniu: spațial sau transformat. Ca urmare, înainte de inserarea marcajului sau de detecția sa, datele gazdă trebuie convertite în domeniul în care se face prelucrarea: spațial, Fourier discret, *wavelet*, transformata cosinus discretă (DCT), fractali. Fiecare din aceste domenii au proprietăți specifice. Procesul de marcare transparentă se poate realiza pentru date gazdă comprimate sau date necomprimate.

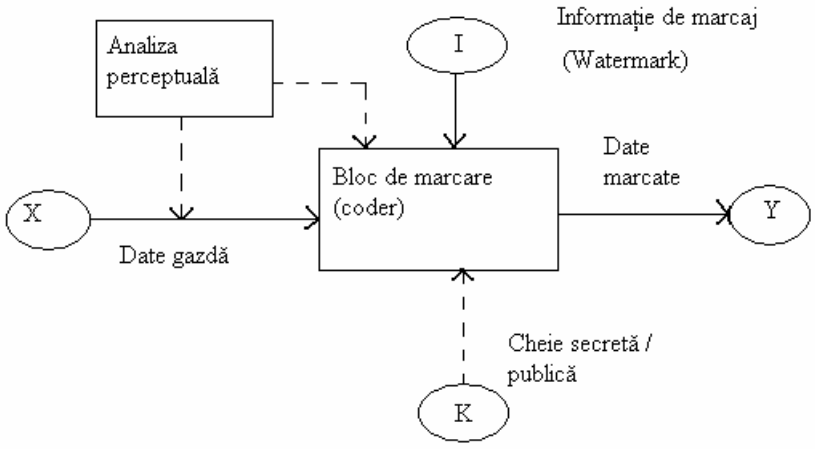

Figura 2.2. Schema de principiu pentru procesul de marcare transparentă.

Detecția marcajului

Detectorul de marcaj (D) are la intrare un semnal recepționat Y' care poate proveni dintr-un semnal marcat sau nu şi care poate să fie sau nu afectat de erori. Pentru extragerea informație de marcaj I' , semnalul original X poate să fie sau nu prezent, după cum detecția este neinversabilă sau inversabilă.

I'=D(X,Y',K) - detecție neinversabilă;
I'=D(Y',K) - detecție inversabilă.

În cazul aplicațiilor de protecție a drepturilor de proprietate, informația de marcaj detectată I' se compară cu originalul I deținut de proprietar:

$$C(I,I') = \begin{cases} 1, & \text{daca } c \geq \gamma \\ 0, & \text{in rest} \end{cases}$$

unde C reprezintă funcția de comparație prin corelație, c este valoarea funcției de corelație dintre I şi I' , iar γ este valoarea pragului de comparație.

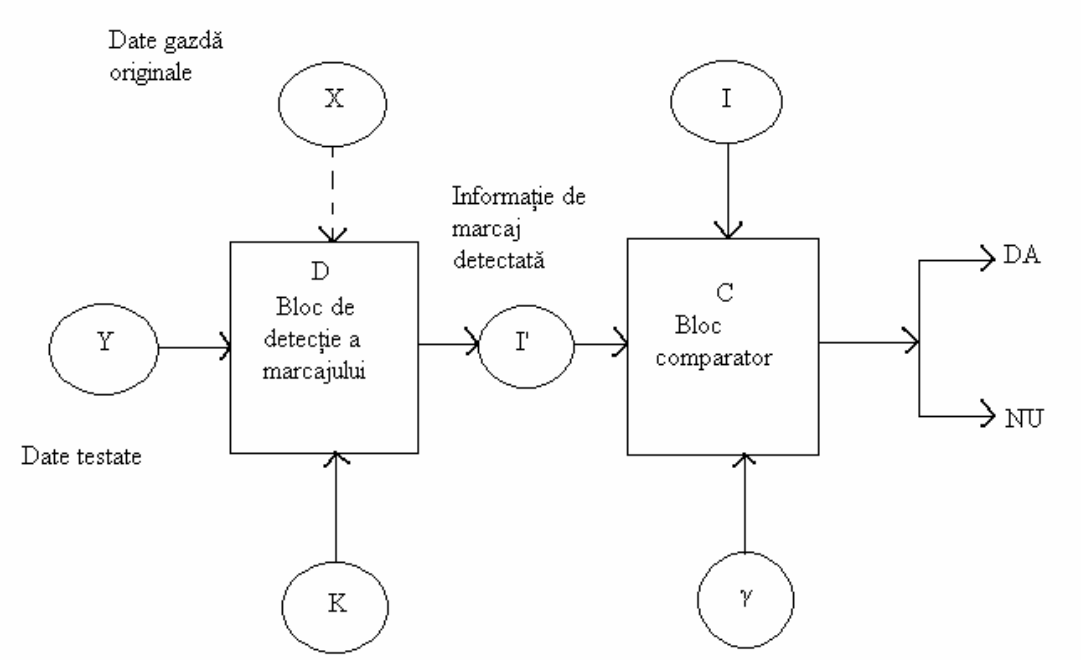

Figura 2.3. Schema de principiu pentru procesul de detecție şi comparație a marcajului transparent.

Majoritatea tehnicilor de marcare transparentă se bazează pe principiul transmisiunilor cu spectru împrăştiat (*spread spectrum*).

Principial, comunicațiile cu spectru împrăştiat constau din transmiterea unui semnal de bandă îngustă (marcajul I în acest caz) pe un canal de bandă largă cu interferențe (semnalul audio sau video - X).

Avantajele de bază ale transmisiunilor cu spectru împrăştiat sunt: reducerea efectelor interferențelor, precum şi prevenirea interceptării semnalului.

-Inserarea marcajului:

Informația de marcaj I se împrăştie prin modulare cu un zgomot pseudo-aleator (PN – *pseudo-noise*) care constituie, în acest caz, cheia K, asigurând astfel camuflarea (mascarea) zonelor din semnalul original afectate de marcaj. Folosirea tehnicii de spectru

împrăştiat protejează eficient marcajul mai ales împotriva manipulărilor neintenționate, proprii procesărilor uzuale în transmitere şi stocare (compresie, scalare).

-Detecția marcajului:

În cazul marcării bazate pe *spread spectrum*, detecția autorizată (se cunoaşte K) este uşor de făcut chiar şi în lipsa originalului X, utilizând un receptor cu corelator. Desincronizările care pot să apară vor putea fi compensate prin utilizarea unui corelator cu fereastră glisantă (*sliding corelator*), care va găsi prin alunecare valoarea maximă a funcției de corelație şi deci va detecta valoarea adevarată a informației de marcaj I.

Detecția marcajului implică decizia dacă datele au fost marcate folosind o anumită cheie sau nu. Detectorul produce o anumită ieşire binară. Pentru a analiza corectitudinea funcționării unui detector, avem doi parametri importanți: P_D=probabilitatea detecției corecte şi P_F = probabilitatea de alarmă falsă. Aceşti parametri se pot folosi de asemenea şi pentru compararea diverselor metode de marcare transparenta. Pentru ca o metodă să fie cât mai bună, ea trebuie să aibă P_D cât mai mare şi P_F cât mai mic.

În general pentru detecția este folosită o versiune de marcaj generat local. Acest marcaj este corelat cu datele recepționate. Dacă cheia folosită la receptor este cea corectă, atunci valoarea corelației este mare.

După detecția corectă a prezenței marcajului este posibilă extragerea acestuia cu ajutorul intercorelației. În acest caz se ia o decizie independentă asupra fiecărui bit în parte.

2.5 Atacuri asupra sistemelor de marcare

Cazurile care conduc la apariția erorilor în procesul de detecție al marcajului se numesc atacuri. Canalul de distribuție a documentelor multimedia de la producător la utilizatori, poate fi considerat ca un *canal de atac* asupra marcajului inserat, astfel că abordarea acestei probleme face apel la metodele din teoria transmisiunii informației. Atacurile pot genera distorsiuni întâmplătoare sau intenționate; ele pot avea loc în timpul transmisiei sau asupra mediului de memorare. Atacurile pot fi clasificate şi în funcție de obiectivele pe care încearcă să le elimine [Naf05b].

Obiectivele marcării transparente sunt: *imperceptibilitatea, robustețea, capacitatea, şi securitatea criptografică.*

Imperceptibilitatea se refera la faptul că marcajul ar trebui să fie invizibil din punct de vedere perceptual şi în plus să nu interfereze cu datele ce trebuie protejate. Deşi diferențele abia vizibile se accentuează dacă produsul original este comparat direct cu cel marcat, ele rămân practic neobservate, deoarece produsul original este accesibil doar proprietarului legal.

Robustețea se referă la posibilitatea de înlăturare a marcajului: cu cât acesta este mai greu de înlăturat, cu atât marcajul este mai robust. Atacurile asupra robusteții sunt cele care încearcă să elimine sau să estimeze marcajul prin prelucrarea semnalului primit.

Capacitatea se referă la cantitatea de informație conținută de marcajul înglobat. Atacurile asupra capacității vizează reducerea capacității marcajului. Astfel, într-o aplicație de

fingerprinting, fiecare utilizator primeşte o copie a originalului, în care este inserat un marcaj diferit. Prin formarea unei coaliții de atacatori, printr-un atac, se reduce capacitatea marcajului.
Securitatea criptografică se referă la integritatea semanticii marcajului şi identificarea sursei în aplicația de autentificare, respectiv la asigurarea confidențialității marcajului în marcarea robustă. Atacurile asupra securității criptografice sunt atacuri de falsificare ilegală. Un exemplu grăitor este fotografia lui Bill Clinton alături de soția sa Hillary, care în varianta falsă pare că este alături de Monica Levinsky. Sistemul propus de C.-Y. Lin poate reconstitui imaginea originală.

Obiectivul atacatorului este *reducerea securității sistemului de marcare*, cu alte cuvinte să reducă probabilitatea de extragere/detecție a marcajului original, respectiv să crească probabilitatea de extragere/detecție a unui marcaj care nu a fost inserat în semnalul marcat (extragere falsă).

Clasificarea atacurilor

Atacurile pot folosi *o singură copie marcată* a unui original, şi atunci aceste *atacuri* ar putea fi *neintenționate* sau *intenționate*; dacă însă sunt folosite *mai multe copii* ale documentului original, atunci ele sunt în mod clar intenționate.

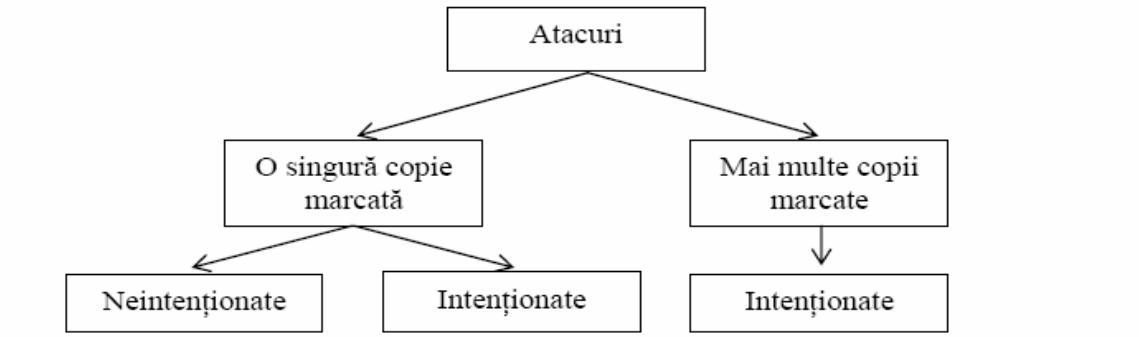

Figura 2.4. Clasificarea atacurilor în funcție de numărul de copii marcate.

Atacurile asupra unei copii marcate, pot fi clasificate după cum urmează:
- *neintenționate*: conversiile în alt format, care rezultă din compresie, cuantizarea pixelilor, modificarea ratei de bit, compensarea raportului de aspect, sau conversia tipului de fişier (analog-digitală sau digital-analogică).
- *intenționate* care se pot clasifica în două mari categorii:

- atacuri asupra unui *singur cadru* / *single-frame*: filtrare, transformări geometrice, atacuri criptografice, de protocol, de estimare a semnalului gazdă sau a marcajului, zgomot aleator – aceastea exploatează informația spațială şi/sau frecvențială a unei imagini,

- atacuri asupra *mai multor cadre* / *multiple-frames*: medierea mai multor cadre; estimarea mai multor cadre (de exemplu prin mediere ponderată) - folosesc informația temporală de la cadru la cadru într-un videoclip.

Atacurile ce acționează asupra unei copii marcate sunt de mai multe feluri:

Compresia - în categoria atacurilor *neintenționate*, cel mai întâlnit atac este compresia. Aceasta înlătură părțile nesemnificative din punct de vedere perceptual dintr-un semnal multimedia. Astfel, prin compresie, semnale perceptual asemănătoare, ajung să fie identice.

Filtrarea - modelează marcajul ca un zgomot aditiv, care este o presupunere rezonabilă în cazul marcării transparente de tip spectru împrăştiat. Înlăturarea marcajului este echivalentă cu o problemă de eliminare a zgomotului din imagine (*denoising)*, rezultatul fiind estimarea imaginii originale.

Zgomotul aleator - introduce distorsiuni imperceptibile, chiar la un raport semnal pe zgomot de 20 dB, dar impactul negativ asupra detecției marcajului este nesemnificativ.

Soluțiile posibile împotriva atacului prin zgomot AWGN sunt codurile corectoare de erori precum şi folosirea diversității şi a combinării repetițiilor marcajului în mod optim.

Atacurile geometrice

Atacurile geometrice au ca obiectiv desincronizarea detectorului, pentru ca detectorul să nu găsească marcajul. Atacurile geometrice posibile sunt: translația, rotația, redimensionare (*scaling*), respectiv combinații ale acestora; curbarea neliniară a imaginii; ştergere şi inserare de cadre (în cazul semnalelor video).

Atacurile de tip protocol

Atacurile de tip protocol periclitează întregul concept al sistemului de marcare transparentă. Exemple sunt:

- *atacul de ambiguitate sau inversiune***.** Atacatorul extrage un marcaj propriu din semnalul marcat, rezultând un semnal pseudo-gazdă, care atunci când este folosit în detecția informață, va permite detecția acestui marcaj al atacatorului. Astfel apare incertitudinea cu privire la identitatea deținătorului drepturilor de autor. Pentru protejarea copyright-ului, marcajele trebuie să fie neinversabile. Cu alte cuvinte un atacator nu ar trebui să poată extrage o marcă din semnalul multimedia marcat. O soluție la această problemă este că marcajul să fie dependent de semnalul original printr-o funcție one-way (neinversabilă).
- *atacul de copiere*, care estimează marcajul dintr-un semnal marcat şi îl inserează într-un alt semnal, numit semnal *țintă*. Acest tip de atac este aplicabil dacă poate fi produs în semnalul țintă un marcaj valid fără cunoaşterea sistemului de marcare sau a cheii. Din nou, marcajele dependente de semnalul original pot fi rezistente la atacul de copiere. Atacul este gândit pentru autentificare.
- *atacul mosaic* împarte semnalul astfel încât este afişat ca o entitate, dar detecția marcajului nu este posibilă. Acest tip de atac este un atac Stirmark.
- *atacul de remarcare* în care o imagine marcată va fi remarcată cu un alt sistem, întrebarea se pune, care a fost prima marcă introdusă. O posibilă soluție ar fi ca generarea marcajelor să depindă de moment (time-stamping).

Soluții posibile împotriva acestor tipuri de atacuri sunt stabilirea unor reguli de construire a sistemului de marcare, pentru a le combate pe cele cunoscute, cum ar fi folosirea marcajelor non-inversabile în cazul atacului de ambiguitate. O altă posibilitate este folosirea unui detector flexibil.

Atacurile de tip criptografic au ca țintă elemente criptografice din sistemul de marcare transparentă. Ele folosesc metode standarde de criptanaliză pentru atacarea sistemului:

- căutarea prin forță brută a mesajului inserat;
- estimarea cheii;
- atacul Oracle: pentru marcaje rezistente la falsificare, detectorul este disponibil astfel încât se poate deduce cum se poate înlătura marcajul. Oracolul produce un răspuns DA sau NU la întrebarea „este imaginea marcată?" şi eventual dă şi informații suplimentare, ca de exemplu coeficientul de intercorelație.

Atacurile de estimare

O altă categorie de atacuri sunt cele de estimare. În aceste cazuri, se estimează marcajul sau semnalul gazdă, fără a cunoaşte cheia secretă, dar cu informații despre statistica marcajului, respectiv a semnalului gazdă. Aceste atacuri sunt aplicabile când marcajul a fost inserat în mod redundant, sau când avem la dispoziție mai multe copii marcate.

Atacul de remodulare

Cu marcajul estimat un atacator poate să remoduleze o imagine: se extrage marca estimată din imaginea marcată (modulare negativă). În cazul unui detector cu corelație, prin aceste acțiuni se anulează corelația pozitivă, cu condiția ca marcajul estimat să fie asemănător cu cel original. Pe de altă parte, prin extragerea unei versiuni amplificate a marcajului estimat, detectorul prin corelație nu va reuşi să găsească marcajul în imaginea atacată.

Atacurile ce folosesc mai multe copii marcate (*multiple-copy*) intervin când semnalul gazdă este acelaşi, iar marcajul este diferit. În esență, aceste atacuri sunt de coliziune.

Există două mari abordări: prelucrarea semnalelor, respectiv codarea informației. Tipuri de coliziune pot fi clasificate după cum urmează:

- de estimare : liniară (prin mediere, ponderată sau nu; prin filtrare; zgomot alb) respectiv, neliniară;
- de tip statistic – min, max, median, minmax, negativ modificat, negativ aleatoriu;
- ipoteza marcării transparente (*marking assumption*) – coliziune prin copiere-şi-lipire, aleatoare, votare în majoritate, atacuri binare (and, or, xor).

Intr-un atac de coliziune, o coaliție de pirați care au versiuni diferite ale aceluiaşi produs multimedia, examinează copiile diferite în speranța creării unui nou semnal care să nu fie legat de nici unul dintre ei. Există mai multe tipuri de coliziune. O metodă este sincronizarea copiilor marcate diferit şi medierea lor, care este un exemplu simplu de coliziune liniară. Un alt atac de coliziune, numit „copiere-şi-lipire", în care atacatorii asamblează porțiuni tăiate din propriile copii, rezultând într-un semnal nou. Alte atacuri folosesc operații neliniare, cum ar fi luarea valorii maxime sau mediane a componentelor din semnale.

2.6 Clasificarea tehnicilor de marcare

Tehnicile de marcare transparentă existente pot fi clasificate după cum urmează:

- -Perceptibile
- -Imperceptibile
  - -Fragile
  - -Robuste
    - - Publice
    - - Private
    - - În domeniul spațial/temporal
    - - În domeniul transformatei.

Marcajele perceptibile creează schimbări sesizabile în semnalul original, atunci când sunt înglobate, dar nu împiedică semnalul marcat să comunice mesajul original. Deşi marcarea perceptibilă nu este atât de răspândită, din moment ce „marcarea" poate fi supărătoare pentru sistemul vizual uman, ea a fost implementată cu succes pentru imagini prin înglobarea unui *logo* vizibil, care permite ca toate detaliile imaginii să fie văzute.

Marcajele imperceptibile pot fi, în funcție de aplicație, *fragile* sau *robuste*. Un marcaj fragil este înglobat în semnalul multimedia astfel încât aproape orice transformare nedorită a semnalului marcat va duce la alterarea acesteia, furnizând astfel informații despre modificări ale semnalului făcute cu rea voință. În contrast cu marcajele fragile, marcajele robuste sunt înglobate în semnalul gazdă astfel încât eliminarea lor să fie dificil de făcut. Acestea trebuie să fie rezistente împotriva atacurilor intenționate. Un atac este orice fel de modificare a semnalului multimedia marcat care poate afecta calitatea marcajului extras. În cadrul tehnicilor de marcare robuste, se pot defini două tipuri de sisteme de marcare, *publice* respectiv, *private*. Sistemele de marcare publice folosesc pentru detecția respectiv extragerea marcajului, numai semnalul multimedia marcat, precum şi o cheie de marcare, şi nu au nevoie de semnalul multimedia original. Aceste sisteme sunt mai puțin robuste în estimarea marcajului, dar volumul de calcul este mult mai redus. Sistemele de marcare private folosesc pentru detecția şi extragerea marcajului atât semnalul multimedia marcat, cât şi cel original. Aceste sisteme, deşi sunt mai robuste, nu sunt potrivite în aplicații în care se cere viteză şi eficiență, cum ar fi căutări automate în baze de date a produselor multimedia furate.

Tehnicile de marcare robuste se mai pot clasifica în tehnici în *domeniul spațial/temporal*, respectiv, tehnici în *domeniul transformatelor*. În prima categorie, marcajul este înglobat în domeniul spațial (pentru imagini), respectiv, în domeniul temporal (pentru semnale audio), în timp ce în a doua categorie, se lucrează asupra transformatelor DCT, Fourier sau *wavelet* ale semnalului gazdă.

3. Marcarea fragila pentru diverse forme ale informației digitale

3.1 Marcarea pentru semnalele audio

În multe aplicații integritatea înregistrărilor audio trebuie stabilită fără dubii chiar înainte ca semnalul să fie folosit, unele persoane au nevoie să ştie sigur că înregistrarea nu a fost modificată neautorizat.

Când avem de a face cu discursuri, unele aplicații în care integritatea trebuie asigurată sunt: protecția mărturiilor înregistrate anterior care trebuie folosite ca dovezi în săli de judecată; protecția interviurilor, care pot fi editate în scopuri malițioase.

Legat de domeniile muzicale câteva exemple de aplicații sunt: verificarea integrității reclamelor difuzate la radio sau TV pentru a fi siguri că au fost transmise conform contractului încheiat; verificarea integrității melodiilor difuzate la radio sau distribuite pe Internet.

Există două metode pentru verificarea integrității sistemelor audio: *watermarking*-ul care ne permite să inserăm date în semnalul dorit şi *amprentarea* care constă în extragerea unei "semnături" din semnalul audio.

Marcajul introdus în semnalele audio este de asemenea un semnal audio şi conține date care pot fi recuperate din semnalul marcat. Ideal este ca marcajul să nu introducă nici o degradare perceptibilă în semnal, ceea ce înseamnă că semnalul original şi cel marcat trebuie să-i sune la fel unui ascultător.

În continuare vom defini trei clase de sisteme de verificare a integrității bazate pe marcaje:

Metode bazate pe watermarking fragil – constă în inserarea unui marcaj fragil în semnalul audio, de exemplu a unui watermark cu putere redusă. Dacă semnalul marcat e editat, watermark-ul nu trebuie să fie detectat. Prin editare înțelegem orice modificare care ar putea corupe conținutul înregistrării, ca de exemplu ştergerea sau inserția de segmente audio, care trebuie să facă marcajul nedetectabil. În schimb compresia defectuoasă sau zgomotul din canal nu trebuie să afecteze detecția de watermark.

Marcările extrem de fragile sunt folosite pentru a verifica dacă semnalul a fost manipulat în vreun fel, chiar fără distorsiuni perceptibile. De exemplu o companie de înregistrări poate marca conținutul CD-urilor sale cu un watermark foarte fragil. Dacă melodiile de pe CD sunt codate pe bit (de exemplu în format MPEG), apoi decodate şi înregistrate pe un nou CD, marcajul nu ar putea fi detectat în noua înregistrare chiar dacă al doilea semnal sună exact ca cel original. Un CD-player poate verifica apoi prezența watermark-ului iar dacă acesta nu este găsit înseamnă că înregistrarea a suferit manipulări nepermise şi CD-ul va fi refuzat. Deficiența acestei metode este inflexibilitatea: de vreme ce marcajul e extrem de fragil, proprietarul nu poate defini un set de manipulări acceptate (doar copierea semnalului audio).

Metode bazate pe watermarking semi-fragil – care sunt o variație a clasei prezentate anterior. Ideea este să se împiedice fragilitatea excesivă a marcajului prin creşterea puterii sale. Watermark-ul semi-fragil e capabil să reziste la mici modificări ale

semnalului audio dar devine nedetectabil când modificările sunt semnificative. Dificultatea întâlnită e găsirea unui prag de robustețe pentru fiecare aplicație în parte.

Metode bazate pe watermarking robust. Marcajul trebuie să poată fi detectat în ciuda oricăror modificări pe care semnalul le poate suporta.

La alegerea unei metode de ascundere a datelor trebuie să ținem cont de modalitatea de reprezentare a semnalului şi de mediul folosit la transmisie. Rata de date ce se foloseşte este dependentă de rata de eşantionare şi de tipul semnalului audio ce trebuie marcat. O valoare tipică pentru rata de date este cea de 16 bps, dar ea poate varia între 2 bps şi 128 bps.

Codarea de bit (*low bit coding*)

Codarea de bit este cea mai simplă metodă de înglobare a unor date în alte structuri de date. Prin înlocuirea celui mai puțin semnificativ bit – LSB – al fiecărui eşantion cu un bit dintr-un şir codat binar, putem ascunde o cantitate mare de date într-un semnal audio. Capacitatea ideală a unui canal este de 1kbps/1kHz. De exemplu, rata de bit de 8 kbps, într-un canal ideal, fără zgomot, corespunde cu o frecvență de eşantionare de 8 kHz, iar o rată de 44 kbps va corespunde unei frecvențe de eşantionare de 44 kHz.

În schimbul acestei capacități mari de canal, un zgomot audibil este introdus în semnal. Cât de mult afectează acest zgomot semnalul, depinde în mod direct de semnalul gazdă. De exemplu mulțimile zgomotoase din timpul transmisiilor sportive live vor masca acest zgomot, în schimb dacă semnalul gazdă este un recital de pian, acest zgomot va fi audibil. Pentru a compensa această variație, a fost folosită atenuarea adaptivă a datelor.

Dezavantajul major al acestei metode este imunitatea sa slabă la diverse manipulări. Informația codată poate fi distrusă de zgomotul din canal, de exemplu prin re-eşantionare, aceasta în cazul în care nu s-au folosit tehnici redundante de codare. Pentru a fi robuste, aceste tehnici reduc rata de date. În practică, aceste metode sunt folositoare doar atunci când mediile de transmisie sunt digitale cap – la – cap.

Codarea de fază

Metoda codării de fază presupune substituirea fazei unui semnal audio inițial, cu o fază de referință, care reprezintă de fapt datele ce trebuie ascunse, codate. Faza a două segmente adiacente este ajustată pentru a se putea păstra faza relativă dintre segmente.

Codarea de fază, atunci când poate fi folosită, este una din cele mai eficiente metode dacă ne referim la raportul semnal-zgomot perceput. Atunci când relația de fază dintre componentele de frecvență este schimbată foarte mult, va avea loc o dispersie de fază. Totuşi, atâta timp cât modificările fazei sunt suficient de fine (acest termen de suficient de fine depinde de observator – profesioniştii din radio-difuziune pot detecta modificări care pentru observatorul obişnuit sunt imperceptibile), se poate obține o codare a datelor care să fie inaudibilă.

Tehnica de împrăştiere a spectrului (spread spectrum)

Într-un canal de comunicație obişnuit este, de obicei, de dorit ca informația de fază să fie concentrată pe o regiune cât mai îngustă a spectrului de frecvențe pentru a conserva banda de frecvențe disponibilă şi pentru a reduce puterea.

Tehnica de bază de împrăştiere a spectrului, pe de altă parte, este folosită pentru a coda un şir de informații prin împrăştiere spectrului de date pe o porțiune cât mai largă

din banda de frecvențe. Cu ajutorul acestei tehnici se poate asigura o recepție corectă, chiar şi atunci când există interferențe la anumite frecvențe.

O secvență pseudo-aleatoare e generată folosind o cheie secretă. Aceasta e aplicată informației codate pentru a modula secvența transformând-o într-una cu spectru împrăştiat. Watermark-ul este scalat dupa un model psiho-acustic.

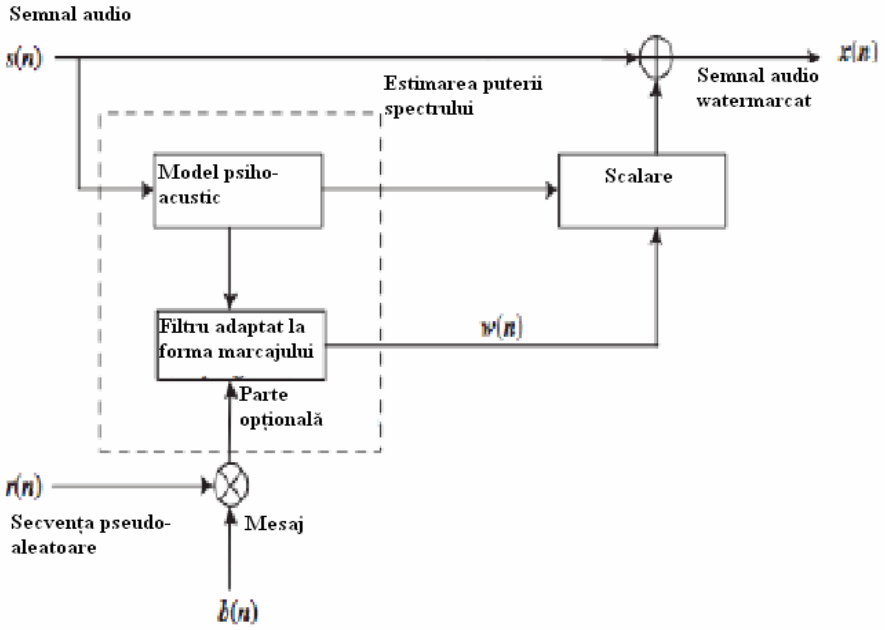

Figura 3.1. Schema de marcare cu spectru împrăştiat.

Atenuarea adaptivă a datelor

Presupune folosirea unui factor de atenuare optim ce variază cu nivelul de zgomot din semnalul gazdă. Adaptând atenuarea la schimbările pe termen scurt ale semnalului sau ale nivelului de zgomot, putem păstra zgomotul codat la un nivel extrem de scăzut în timpul intervalelor de tăcere şi putem mări nivelul atunci când avem segmente mai zgomotoase.

Codarea pentru corecția erorilor

Pentru a compensa erorile apărute datorită zgomotului din canal şi a modificărilor făcute asupra semnalului gazdă se folosesc metode de codare cu corecția erorilor.

Analiza semnalului

Posibilitatea de detectare a zgomotului alb inserat într-un semnal gazdă este dependentă de nivelul de zgomot original din acesta. Pentru a maximiza cantitatea de date ce poate fi ascunsă într-un semnal (teorema water-filling), este util ca nivelul de zgomot să fie exprimat cantitativ, în funcție de modificările ce apar între amplitudinile a două eşantioane adiacente:

$$\sigma^2_{local} = \frac{1}{|S_{\max}|} \cdot \frac{1}{N} \cdot \sum_{n=1}^{N-1} [s(n+1) - s(n)]^2$$

Se mai pot folosi şi altfel de metode pentru ascunderea datelor în semnalele audio. Una dintre aceste metode este aceea de ascundere a datelor folosind *ecoul.* Această metodă presupune introducerea unui ecou în semnalul audio gazdă. Datele se ascund prin modificarea a trei parametri ai ecoului: amplitudinea inițială, rata de descompunere (*decay rate*), şi decalajul (*offset*) dintre semnalul original şi ecou. Pe măsură ce decalajul dintre semnalul original şi ecou scade, cele două semnale tind să se îmbine. La un anumit punct, urechea umană nu mai poate face distincția între cele două semnale. Ecoul este perceput ca o rezonanță adăugată. Acest punct este foarte greu de determinat cu exactitate. El depinde de calitatea înregistrării originale, de tipul semnalului original, şi de ascultător. În general, acest punct se găseşte în jurul valorii de 1 ms.

3.2 Marcarea pentru secvențe video

Secvențele video constă dintr-o serie de imagini fixe, ce se transmit consecutiv, la intervale de timp egale. În consecință majoritatea tehnicilor de watermark-are a imaginilor statice pot fi folosite şi în cazul secvențelor video. O deosebire importantă față de imaginile statice este aceea că în cazul secvențelor de imagini, spațiul disponibil pentru inserarea watermark-ului este mult mai mare. La fel de important este şi faptul că în cazul secvențelor video, sistemul demarcare trebuie să lucreze în timp real.

Marcarea pentru secvențele video a început să se dezvolte tot mai mult, în special de la apariția DVD-urilor (*digital video disk*), care trebuie să conțină un marcaj, compatibil atât cu DVD "player"-ul cât şi cu "recorder"-ul, astfel încât să nu fie posibilă piratarea lor sau redarea de DVD-uri piratate.

O altă aplicație a marcarii pentru secvențe video este televiziunea codată (*pay-per-view*), unde semnalul video este marcat individual pentru fiecare receptor în parte. La receptor se găseşte un decodor cu acces condiționat (secvența este decriptată doar dacă cheia de marcare din decodor corespunde cu cea folosită la watermark-are).

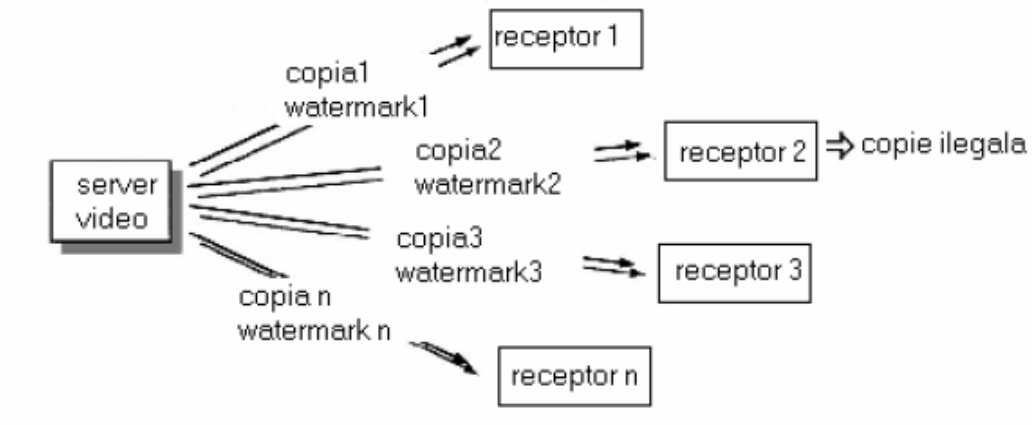

Figura 3.2. Principiul watermarking-ului individual al secvențelor video.

După cum se ştie schemele de comunicație cu spectru împrăştiat trasmit un semnal de bandă îngustă folosind un canal de bandă largă, reducându-se astfel

interferențele între simboluri (ISI). La fel ca şi în cazul imaginilor statice, o astfel de schemă poate fi adaptată pentru a fi folosită la watermark-are; watermark-ul fiind semnalul de bandă îngustă, iar secvența video semnalul de bandă largă.

În cele ce urmează vom exemplifica marcarea transparentă bazată pe principiul comunicării cu spectru împrăştiat pentru semnale video în domeniul spațial [HarGir98].

O secvență video este reprezentată de un semnal video tridimensional având două dimensiuni în plan şi cea de-a treia fiind timpul. Dacă se foloseşte sistemul de baleiaj pe linii, semnalul video poate fi considerat unidimensional.

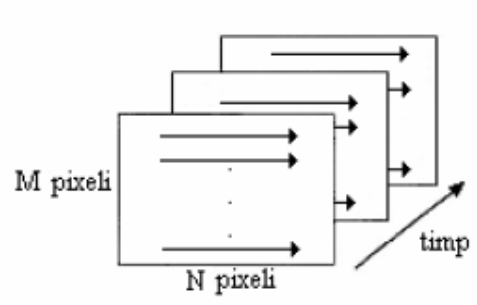

Figura 3.3. Semnal video cu baleiaj pe linii.

Inserarea marcajului în secvența video

Fie informația de marcaj I formată dintr-un număr de N_1 biți a_j:

$I=\{a_j\}$, $j=1..N_1$, $aj \in \{-1, 1\}$

N_1 reprezintă încărcătura marcajului şi valoarea sa maxima este 60÷70 biți.

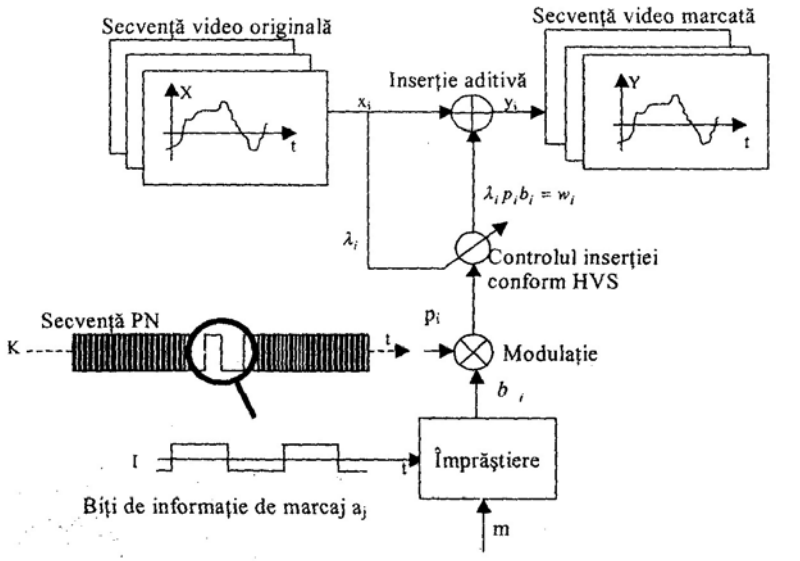

Figura 3.4. Modelul adăugării marcajului la secvența video.

Această secvență I este împrăştiată pe un număr mare de biți conform factorului de împrăştiere *m* (*chip rate, spread factor*) în scopul creşterii robusteții marcajului. Tipic, *m* ia valori în domeniul ($10^3 \div 10^6$). Secvența împrăştiată va fi:

$$b_j = a_j, \quad jm \le i < (m+1)j, \quad i \in N.$$

Secvența împrăştiată, amplificată cu un factor local $\lambda_i \geq 0$, va modula un semnal pseudo-aleator, care constituie în acest caz cheia K, formată din biți p_i:

$$p_i \in \{-1,1\}, \quad i \in N.$$

Secvența de marcaj împrăştiată $W=\{w_i\}$

$$w_i = \lambda_i b_i p_i , \quad i \in N$$

se adună la semnalul video x_i obținându-se astfel secvența video marcată.

$$Y=\{y_i\}, \quad i \in N$$
$$y_i = x_i + \lambda_i b_i p_i , \quad i \in N.$$

Secvența pseudo-aleatoare $K=\{p_i\}$ (şi în consecință şi secvența de marcaj împrăştiată $W=\{w_i\}$) sunt semnale cu caracter pseudo-aleator, deci greu de detectat, localizat şi înlăturat.

Securitatea unui algoritm de marcare trebuie să rezide în cheia K. Această cheie se alege astfel încât să asigure securitatea criptografică şi în plus, pentru detecție cu corelator, se mai cer şi bune proprietăți de corelație şi ortogonalitate. În acest caz, în acelaşi semnal gazdă, se pot insera mai multe marcaje care vor fi detectate fără ambiguitate la recepție.

Factorul de amplificare λ_i , este funcție de proprietățile locale ale semnalului video şi poate utiliza fenomenul de mascare spatială şi temporală specific sistemului vizual uman (HVS), astfel încât amplitudinea marcajului local să fie maxim posibilă, în condițiile de invizibilitate perceptuală.

Inserarea informației de marcaj w_i se face în acele zone care sunt mai puțin vizibile, de exemplu în zonele conținând detalii fine sau contururi.

Detecția marcajului din secvența video

În cazul utilizării principiului spectrului împrăştiat, detecția autorizată (deci cu cunoaşterea cheii K) se poate face cu uşurință şi fără cunoaşterea originalului X, utilizând un corelator.

Înainte de demodulare (realizată cu corelator), secvența video marcată Y' (ea poate diferi de cea originală datorită prelucrărilor din transmisie sau atacurilor) este filtrată trece sus pentru a înlătura componentele majore ale semnalului video. Filtrarea trece sus nu este obligatorie, dar îmbunătățeste performanțele întregului sistem de marcare deoarece reduce intermodulațiile dintre semnalul de marcaj şi semnalul video.

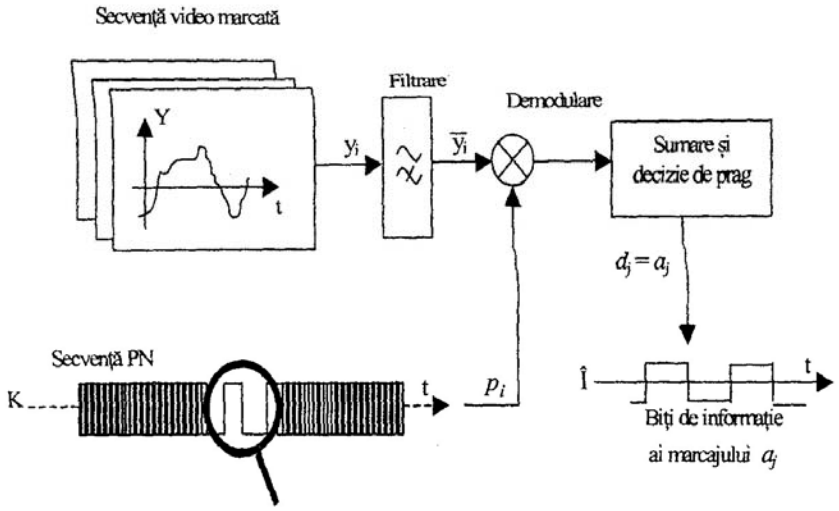

Figura 3.5. Modelul extragerii marcajului din secvența video.

Detecția constă în multiplicarea semnalului video marcat (Y') cu aceeaşi cheie K folosită la inserare, urmată de calcularea sumei de corelație d_j pentru fiecare bit inserat şi o decizie de prag pentru determinarea valorii estimate ($\hat{a}_j$) a bitului a_j.

$$d_j = \sum_{i=jm}^{(j+1)m-1} p_i \overline{y_i} = \sum_{i=jm}^{(j+1)m-1} p_i \overline{(x_i + \lambda_i b_i p_i)}$$

unde "——" indică valoarea medie iar y_i este valoarea bitului i din semnalul video marcat după filtrarea trece sus; se consideră că y_i nu este afectat de erori la recepție, deci:

$$y_i' = y_i = x_i + w_i$$
$$\underbrace{\sum_{i=jm}^{(j+1)m-1} p_i \overline{x_i}}_{S_1} + \underbrace{\sum_{i=jm}^{(j+1)m-1} p_i \overline{\lambda_i b_i p_i}}_{S_2} = S_1 + S_2$$

S_1, S_2 descriu contribuția la suma de corelație a semnalului video filtrat şi a semnalului de marcare filtrat.

În ipoteza că prin filtrare trece sus s-a îndepartat din y_i semnalul video x_i , S_1=0; se consideră de asemenea că filtrarea trece sus are o influență neglijabilă asupra semnalului de marcare:

$$\overline{\lambda_i b_i p_i} \equiv \lambda_i b_i p_i$$

În aceste ipoteze:

$$dj \equiv \sum_{i=jm}^{(j+1)m-1} p_i^2 \lambda_i b_i = \sigma_p^2 a_j m E[\lambda_i] ,$$

unde σ_p^2 este dispersia zgomotului pseudo-aleator K, iar $E[\lambda_i]$ reprezintă valoarea medie a coeficienților λ_i pe *m* pixeli.

Semnul sumei de corelație d_j va da valoarea bitului de marcare a_j, în cazul unei decizii hard:

$$sign\ d_j = sign\ (a_j\ \sigma_p^2 m\overline{\lambda_i}) = sign(\ a_j\ )$$

cu $\sigma_p^2 m\overline{\lambda_i} > 0$

Din această ecuație rezultă că bitul transmis a_j a fost +1 dacă corelația dintre semnalul video ce conține bitul inserat şi secvența pseudo-aleatoare este pozitivă, şi -1 dacă corelația este negativă.

Dacă secvența pseudo-aleatoare K utilizată la detecție nu este cea de la inserție (sau dacă nu este sincronizată cu secvența K utilizată la emisie), schema nu funcționează şi biții regăsiți sunt aleatori.

Desincronizările secvenței pseudo-aleatoare la recepție vor putea fi compensate utilizând un corelator cu fereastră glisantă (*sliding corelator*): se aplică experimental toate deplasările posibile, adevarata valoare a lui a_j fiind gasită pentru valoarea maximă a corelației d_j, comparativ cu toate celelalte deplasări.

Tipul de decoder folosit nu necesită originalul X pentru detecție, deci schema prezentată este de tip inversabil (*oblivious*). Folosirea originalului X, deci a unei scheme neinversabile (*nonoblivious*), care se va scădea din Y înainte de demodulare (în locul filtrării), va înlătura orice interferență între semnalul video Y şi marcajul W, ceea ce va face ca detecția să fie mult mai robustă.

Performanțele schemei propuse vor putea fi evaluate calculând rata erorii pe bit: BER (*Bit Error Rate*). Un bit este eronat dacă:

$$\text{sign}(\ S_1+S_2\ ) \neq \text{sign}(S_2)$$

Această situație apare când:

$$\text{sign}\ S_1 \neq \text{sign}\ S_2 \quad \text{şi} \quad |\ S_1\ | > |\ S_2\ |\ .$$

Probabilitatea de eroare este:

$$\text{BER} := \frac{1}{2}\ \text{erfc}\left[\frac{\sigma_p \sqrt{m} E\cdot(\lambda_i)}{\sqrt{2}\cdot\sqrt{(\sigma_x)^2 + (\mu_x)^2}}\right]$$

unde σ_x^2, respectiv μ_x reprezintă dispersia, respectiv valoarea medie a semnalului video. Din formulă se vede că BER scade dacă *m*, σ_p şi $E[\lambda_i]$ cresc.

Debitul informației de marcare este:

R_{WM}= *numărul pixelilor de luminanță pe secundă* / *factorul de împrăştiere*

Dacă se doreşte menţinerea unei anumite probabilităţi a erorii şi în prezenţa atacurilor, argumentul din expresia BER trebuie crescut printr-un factor de siguranţă.

Metoda expusă pentru marcarea semnalelor video în domeniul spaţial poate fi aplicată în orice domeniu transformat al imaginii. Fiecare domeniu transformat are propriile sale avantaje şi dezavantaje.

3.3 Marcare fragilă pentru imagini

Multe scheme de marcare transparentă pentru imagini au fost propuse pentru a proteja proprietatea intelectuală într-o eră în care imaginile digitale sunt uşor de modificat şi pot fi reproduse aproape la perfecţie. Într-un sistem de marcare fragilă un semnal (watermark) e inserat în imagine astfel încât probabilitatea de a detecta reproduceri ale imaginii să fie mare. Watermark-ul e practic invizibil ochiului uman. Aceste tipuri de marcări se regăsec în sisteme de autentificare a imaginilor. Majoritatea marcărilor utilizate sunt în domeniul marcării robuste însă multe aplicaţii ar putea beneficia de pe urma marcării fragile.

Aplicaţii ale marcării fragile

Un watermark fragil e un marcaj care e schimbat sau distrus când imaginea e modificată printr-o transformare liniară sau neliniară. Marcajele fragile nu sunt potrivite să întărească protecţia dreptului de autor al imaginilor (copyright); un atacator doreşte să distrugă marcajul inserat şi marcajul fragil este, prin definiţie, uşor de distrus. Sensibilitatea lor la modificări conduce la folosirea lor pentru autentificarea imaginilor. Este posibil ca cineva să fie interesat să verifice dacă o imagine a fost modificată după ce fost marcată.

Sistemele de autentificare a imaginilor işi găsesc domenii de aplicaţie în drept, comerţ, apărare naţională şi jurnalism. De vreme ce imaginile digitale sunt uşor de modificat, un sistem de autentificare sigur e folosit pentru a arăta că nu au apărut falsuri dacă credibilitatea imaginii e pusă sub semnul întrebării.

Exemple comune sunt marcarea imaginilor într-o bază de date pentru a găsi falsurile, folosirea marcajelor de către agenţiile de ştiri pentru a fi sigure că o imagine nu e modificată sau fabricată pentru a falsifica evenimente şi marcarea imaginilor în comerţ astfel încât cumpărătorul să fie sigur că imaginile cumpărate sunt autentice şi în conformitate cu contractul încheiat. Alte situaţii cuprind imaginile folosite în sălile de judecată, fotografie, jurnalistică, sau imagini folosite în spionaj.

O altă metodă de a verifica autenticitatea unei opere digitale este folosirea unui sistem de semnături digitale. Un "rezumat" (*digest*) al datelor de autentificat se obţine prin utilizarea unei funcţii *hash.* Pentru autentificare se verifică semnătura examinând rezumatul (posibil modificat) al datelor, apoi folosind un algoritm de verificare se determină dacă datele sunt autentice.

Dacă scopul marcării fragile şi a semnăturilor digitale e similar, sistemele de marcare oferă unele avantaje faţă de semnături însă dezavantajul îl constituie modificările (inserarea marcajului) efectuate asupra imaginii. Prin faptul că watermark-ul e inserat

direct imaginii nu vom avea nevoie de informații adiționale pentru verificarea autenticității. Astfel, informația critică necesară pentru autentificări e ascunsă discret şi e mai greu de îndepărtat decât semnăturile digitale. O semnatură digitală vede imaginea ca un şir arbitrar de biți şi nu exploatează structura sa unică. Semnatura e capabilă să determine dacă imaginea a fost modificată dar nu poate să caracterizeze modificările aduse ei. Multe sisteme de watermarking pot "indica" care parte a imaginii marcate a fost atacată şi să determine natura atacului.

Scheletul inserării şi detectării unui marcaj fragil e similar cu a oricărui sistem de marcare. Cheia de marcare e folosită pentru a genera watermark-ul şi e un identificator desemnat proprietarului sau imaginii respective. Imaginea originală e ținută secretă sau poate nici să nu fie disponibilă în unele aplicații, ca de exemplu camera digitală. Doar imaginea marcată poate fi transmisă, prezentată, sau distribuită. Imaginea originală e perceptual identică cu cea marcată. În figură avem un exemplu de imagine originală şi imagine marcată fragil.

Figura 3.6. Imaginea originală; Imaginea marcată.

Când un utilizator primeşte o imagine el foloseşte un detector pentru a evalua autenticitatea imaginii primite. E nevoie şi de informație suplimentară care poate fi cheia de marcare, watermark-ul, imaginea originală sau alte informații. Detectorul se bazează de obicei pe detecția statistică. E de preferat ca detectorul să afle şi locul în care imaginea a fost modificată.

Proprietăți specifice ale marcării fragile

- Detecția falsului – e cea mai importantă proprietate a marcarii fragile. Se doreşte de asemenea să se observe cât de mult a fost alterată imaginea şi unde.
- Transparența perceptuală – marcajul nu trebuie să fie vizibil la observarea normală sau să interfereze cu funcționalitatea imaginii. În cele mai multe cazuri aceasta se referă la menținerea calităților estetice ale imaginii. Din nefericire nu există prea multe informații depre cum afectează zgomotul introdus de procesul de marcare alte operații de procesare.
- Detecția nu ar trebui să utilizeze imaginea originală.

- Detectorul ar trebui să localizeze şi să caracterizeze alterările aduse imaginii marcate.
- Watermark-ul trebuie detectat şi în caz de decupare a imaginii – de exemplu o persoană poate fi interesată de o porțiune (fețe, oameni) a unei imagini mai mari.
- Watermark-urile generate cu chei diferite trebuie să fie "ortogonale" în timpul detecției – marcajul inserat într-o imagine generată folosind o cheie de marcare trebuie să fie detectată numai furnizând informația suplimentară corespunzătoare. Orice altă informație suplimentară ar trebui să nu reuşească să detecteze marcajul.
- Cheia de marcare ar trebui să fie greu de obținut din informațiile suplimentare – e deosebit de importantă pentru sistemele care au chei distincte de inserare si de detecție.
- Inserarea unui marcaj de către persoane neautorizate trebuie să fie dificil de făcut.
- Watermark-ul trebuie să poată fi inserat şi în domeniul comprimat – nu e acelaşi lucru ca şi când spunem că marcajul trebuie să reziste compresiei pentru că acesta este un atac.

Atacuri asupra marcajelor fragile

E practic imposibil să se proiecteze un sistem invulnerabil față de orice formă de atac, şi noi metode de a învinge sistemele de marcare vor fi inventate în timp. Dar cunoştințele despre atacurile comune sunt necesare pentru găsirea unor soluții mai bune.

Primul tip de atac e modificarea neintenționată a imaginii marcate (adică modificarea imaginii presupunând că nu există watermark). Acest tip de atac ar trebui recunoscut de orice marcaj fragil, însă e menționat, pentru că e cea mai răspândită formă de atac. Variații ale atacului cuprind decuparea si înlocuirea (de exemplu fața unei persoane cu alta).

Alt tip de atac e cel în care imaginea e modificată fără a afecta marcajul sau să se creeze o nou marcaj pe care detectorul să îl considere autentic. Sunt marcaje fragile care descoperă repede unele modificări ale imaginii, dar nu pot detecta modificări atent concepute. Un exemplu în acest sens e un marcaj inserat în cel mai puțin semnificativ bit al unei imagini. O încercare de modificare fără a realiza că watermark-ul e în cel mai puțin semnificativ bit va schimba marcajul şi va fi detectată. Dar dacă, un atacator încearcă alterarea imaginii fără a schimba nici unul din biții cei mai puțin semnificativi sau să înlocuiască cu un nou set de biți puțini semnificativi, detectorul va considera imaginea autentică.

Unele atacuri folosesc un marcaj valid a unei imagini marcate ca marcaj pentru o altă imagine. E uşor de realizat un astfel de atac dacă se deduce cum a fost inserat marcajul. Acest tip de atac se poate folosi asupra aceleiaşi imagini; marcajul se distruge, apoi imaginea se modifică iar în final se reintroduce marcajul.

Un atacator poate fi interesat să distrugă complet marcajul şi să nu lase nici o urmă a prezenței sale. Pentru aceasta el poate introduce un zgomot aleator în imagine, poate folosi programe concepute pentru a distruge marcaje (ca de exemplu Stirmark) sau poate folosi analiza statistică pentru a estima imaginea originală.

Exemple de marcări fragile

Tehnicile de marcare fragilă se împart în tehnici care lucrează direct în domeniul spațial şi tehnici care lucrează în domeniul transformatelor (DCT, wavelet).

Marcaje în domeniul spațial

Tehnicile timpurii de watermark-ări fragile inserau marcajul direct în domeniul spațial al unei imagini, ca de exemplu tehnicile descrise de Walton [Wal95] şi van Schyndel [SchTirOsb94]. Aceste metode inserau marcajul în cel mai puțin semnificativ bit pentru transparență perceptuală. Dezavantajul lor major era uşurința cu care puteau fi descoperite şi ocolite cât şi inabilitatea lor de a fi comprimate fără a afecta marcajul.

Wolfgang şi Delp [WolDel96] au continuat munca lui van Schyndel pentru a îmbunătăți robustețea şi localizarea marcajelor cu tehnica VW2D. Marcajul e inserat prin adăugarea unei secvențe M bipolare în domeniul spațial. Detecția se face printr-un detector corelator modificat.

P. Wong descrie o altă tehnică de marcare fragilă care obține un rezumat al imaginii utilizând o funcție *hash* [Won98]. Imaginea, dimensiunile imaginii şi cheia de marcare sunt modificate de funcția *hash* în timpul inserției şi sunt folosite pentru a schimba cel mai puțin semnificativ bit al imaginii originale. Acest lucru se face astfel încât, atunci când informația suplimentară corectă şi imaginea marcată nealterată sunt furnizate detectorului se obține o imagine binară aleasă de proprietar (de exemplu un *logo* al unei companii). Această metodă are proprietăți de localizare şi poate să identifice regiuni de pixeli modificați în imaginea marcată.

Tehnica lui Yeung şi Mintzer e mai vastă şi nu se inserează doar o valoare binară în cel mai puțin semnificativ bit. Cheia de marcare e folosită pentru a genera secvențe pseudo-aleatoare (una pentru fiecare canal sau componentă a culorii) care controlează modificările ulterioare ale pixelilor. După procesul de inserție, un proces de difuzie a erorilor e folosit pentru răspândirea efectelor pixelilor alterați, făcând ca marcajul să fie mai greu de văzut [YeuMin97].

Marcaje în domeniul transformatelor

Numeroase transformate, ca de exemplu transformata cosinus discretă (DCT) şi transformata Wavelet sunt folosite pentru compresia cu pierderi a imaginilor şi avem multe metode în care coeficienții transformatelor pot fi alterați (cuantizați) pentru a minimiza distorsiunile perceptuale. Sunt multe avantaje în utilizarea domeniului transformatelor pentru marcarea fragilă. Multe sisteme de marcare fragilă sunt adaptate pentru compresia cu pierderi (de exemplu JPEG) care au avantajul că marcajul e inserat în reprezentarea comprimată. Proprietățile transformatelor pot fi folosite pentru a caracteriza modul de alterare al imaginii. Marcajele trebuie să prezinte o anumită robustețe pentru unele modificări (strălucirea imaginii), însă trebuie să poată detecta alte tipuri de modificări (înlocuirea pixelilor).

Wu şi Liu descriu o tehnica bazată pe codarea JPEG. Watermark-ul e inserat prin schimbarea coeficienților DCT cuantizați, înainte de codarea sursei. Pentru a nu bloca coeficienții alterați nu se marchează coeficientii DC cu energie mică [WuLiu98].

Kundur şi Hatzinakos apoi Xie şi Arce descriu tehnici bazate pe transformata wavelet [KunHat98, XieArc98]. Kundur inserează marcajul prin cuantizarea coeficienților transformatei wavelet Haar în timp ce Xie et al inserează selectiv biții marcajului prin procesarea imaginii, după ce e într-o formă comprimată, folosind un algoritm SPIHT. O analiză wavelet a unei imagini conține atât informație spectrală despre frecvențe cât şi informații spațiale despre marcajele inserate.

Tehnicile care folosesc domeniul transformatelor sunt mai complexe şi mai scump de evaluat decât cele din domeniul spațial, dar oferă un grad mai ridicat de robustețe. Se poate pune în discuție de ce e importantă robustețea pentru sistemele de marcare fragile. Este utilă pentru simplul fapt că unele operații de bază de procesare a imaginilor- ca de exemplu cele care sunt folosite pentru stocarea imaginilor watermark-ate – nu trebuie să altereze marcajele inserate. În ultimii ani cel mai mare accent s-a pus pe domeniul transformatei wavelet pentru că oferă cel mai ridicat grad de robustețe la procesări de bază ale imaginilor. Transformata wavelet (WT) oferă descompunerea semnalului în benzi de frecvență înguste în timp ce păstrează spațiul semnalelor de bază limitat. Acest lucru prezintă deosebită importanță când avem de a face cu semnale reale, în special când trebuie luată în considerare localizarea spațială. În plus numărul mare de funcții de bază oferă flexibilitate analizei şi adaptare uşoară la o aplicație particulară. Toate acestea explică atenția de care se bucură WT când vorbim de procesarea imaginilor, inclusiv watermark-are digitală pentru autentificarea imaginilor.

D. Kundur şi D. Hatzinakos, în lucrarea lor "Digital watermarking for telltale tamper proofing and authentication", prezintă o tehnică de marcare fragilă în domeniul transformatei wavelet prin cuantizarea coeficienților corespunzători. Aceasta permite detecția falsurilor în regiuni spațiale sau de frecvență localizate, facând posibilă identificarea frecvențelor modificate într-o imagine. Folosirea unei Funcții de Evaluare a Falsurilor (*Tampering Assessment Function* -TAF) îngăduie utilizatorul să ia decizii dependente de aplicație cu privire la credibilitatea datelor recepționate. Autorii demonstrează cu rezultate experimentale potențialul de folosire a metodei impuse de ei pentru alte aplicații de autentificare a informațiilor multimedia.

O altă procedură de autentificare a imaginilor digitale e descrisă în [YuLuLiaShe00]. Această tehnică permite detecția falsurilor premeditate cât şi a distorsiunilor accidentale introduse de compresie. Autorii modelează cantitatea de modificări asupra coeficienților wavelet, introdusă de atacuri accidentale versus atacuri premeditate ca o distribuție Gaussiană cu variații mici versus variații mari. Ei demonstrează că integrarea unui răspuns asupra falsului la fiecare dimensiune ne permite să deosebim atacurile premeditate de cele accidentale oferind astfel un anumit grad de robustețe sistemului. Metoda e capabilă să autentifice imagini cu compresie JPEG fără nici un acces la imaginea originală nemarcată.

Ambele metode prezentate mai sus oferă imaginilor digitale protecție față de atacuri cât si stocare eficientă a datelor. Dar aceste tehnici au nevoie ca utilizatorul să determine dacă un atac e răuvoitor sau nu. În plus necesită un grad de interacțiune în procesul de inserție. Metoda propusă in [PaqWar02] afirmă că atacurile acceptate sunt predeterminate şi procedura de inserție este proiectată să folosească la maxim HVS. În plus îşi propune să îmbunătățească rezoluția frecvenței a transformatei wavelet discrete prin utilizarea pachetelor wavelet–WP. În descompunerea WP se folosesc pentru nivelul următor atât ieşirile imaginii filtrate trece-jos, cât şi ieşirile filtrelor trece-sus. Acest lucru

conduce atît la benzi de frecvenţă înguste cât şi la captarea detaliilor imaginii. În plus permite precizie şi flexibilitate mai ridicate în selecţia benzilor care vor fi folosite în procesul de inserţie.

Procesul de insertie

Paşii principali ai procesului de inserţie sunt prezentaţi în figura următoare:

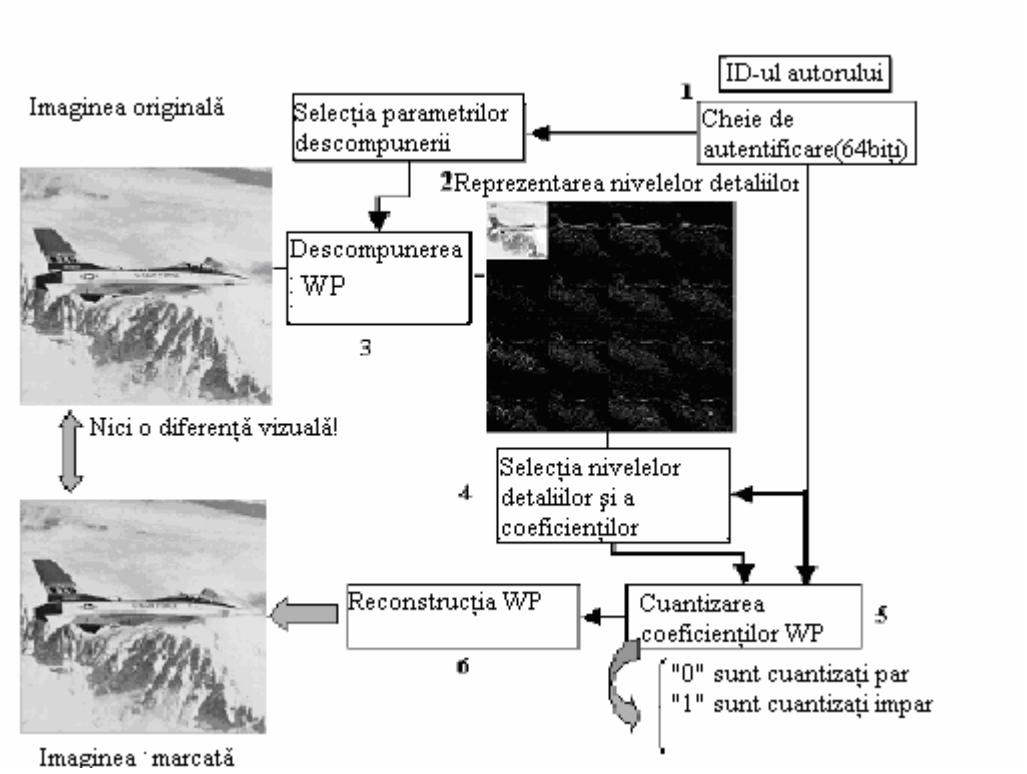

Figura 3.7. Schema de inserţie a marcajului [PaqWar02].

1. Se alege cheia de autentificare a autorului pe 64 biţi. Aceşti 64 de biţi sunt suficienţi pentru a garanta unicitatea cheii; de altfel reprezintă şi standardul actual folosit în schemele de identificare.

2. Cheia e folosită pentru a selecta decompunerea aplicată imaginii (funcţia wavelet şi numărul de niveluri). Unul din avantajele transformarii wavelet este flexibilitatea oferită de multitudinea tipurilor wavelet disponibile. În cazul de faţă, acest număr mare de funcţii creşte securitatea schemei, fiind practic imposibil pentru un potenţial pirat să ştie care domeniu wavelet a fost utilizat în procesul de inserţie. Numerele selectate sunt alese astfel încât să ofere o bună rezoluţie în frecvenţă şi un număr destul de mare de coeficienţi per bandă pentru inserţie, în timp ce păstrează complexitatea în limite rezonabile. Autorii au selectat undisoare mama (wavelet mother) specifice care ofera suport compact si finit şi rezoluţie în frecvenţă îmbunătăţită. Un pas important este alegerea parametrilor astfel încât să obţină rezoluţie spaţială optimă.

3. Descompunerea WP ale imaginii e efectuată în conformitate cu parametrii indicați de cheie (conform pasului 2).

4. Cheia unică a autorului e folosită încă o dată la selecția nivelelor de detalii şi a coeficienților în care marcajul va fi inserat. Sistemul identifică în primul rând M niveluri de detalii principale în care cele N regiuni de coeficienți wavelet packets (WPC) sunt definite. Benzile sunt distribuite în scopul de a acoperi întreg spectrul de frecvențe în timp ce regiunile spațiale ale WPC sunt translatate de la o bandă la alta pentru a acoperi întreaga imagine. Apoi, se definesc K benzi secundare, în relație cu fiecare banda principală şi sunt definiți grupuri WPC corespunzători. În scopul autentificării imaginii e necesar ca întreg spectrul de frecvențe şi întreaga imagine să fie acoperite.

5. Cheia secretă e folosită pentru ultima dată pentru inserare. Zero-urile logice sunt inserate prin cuantizare pară a mediei regiunilor selectate ale WPC; unu logic e inserat prin cuantizare impară. Pasul optim Δ e folosit pentru a cuantiza media regiunilor principale selectate iar $\Delta/2$ e utilizat pentru regiunile secundare. pasul de cuantizare trebuie măsurat ca funcție a distribuției benzii de inserție. Pentru aceasta se estimează energia conținută în fiecare bandă. Sistemul consideră mai importante marcajele inserate în regiuni cu mai multe detalii (reprezentate de coeficienți cu amplitudine mai mare) pentru care HVS e mai puțin sensibil.

6. În cele din urmă coeficienții (unii modificați iar alții nu) sunt utilizați pentru a forma imaginea marcată prin reconstrucția WP.

Cuantizare optimă

Alegerea pasului optim de cuantizare Δ este deosebit de importantă pentru a maximiza capacitatea cât şi de a minimiza distorsiunile introduse. Acest lucru înseamnă că vrem să minimizăm media erorii pătratice de cuantizare:

$$\sigma_q^2 = 2\sum_{i=1}^{M/2-1}\int_{(i-1)\Delta}^{i\Delta}(x-(\frac{2i-1}{2})\Delta)^2 f_x(x)dx+$$
$$2\int_{(M/2-1)\Delta}^{\infty}(x-(\frac{M-1}{2})\Delta)^2 f_x(x)dx$$

În relația de mai sus $fx(x)$ este densitatea de probabilitate a variabilei x; μ este media ei şi M este numărul de nivele de cuantizare folosite. Minimizarea erorii ca funcție de Δ este:

$$\frac{\delta\sigma_q^2}{\delta\Delta} = -\sum_{i=1}^{M/2-1}(2i-1)\int_{(i-1)\Delta}^{i\Delta}(x-(\frac{2i-1}{2})\Delta)f_x(x)dx-$$
$$(M-1)\int_{(M/2-1)\Delta}^{\infty}(x-(\frac{M-1}{2})\Delta)f_x(x)dx = 0$$

Coeficienţii WP au distribuţie Laplaciană. În figura următoare se prezintă distribuţia asociată cu banda de frecvenţe înalte a unei descompuneri WP cu 4 niveluri a unei imagini. Problema minimizării a fost deja rezolvată pentru acest tip de distribuţie, ceea ce înseamnă că paşii de cuantizare optimi sunt disponibili ca o funcţie a numărului de nivele de cuantizare. Astfel pasul utilizat reflectă gradul de protecţie obţinută.

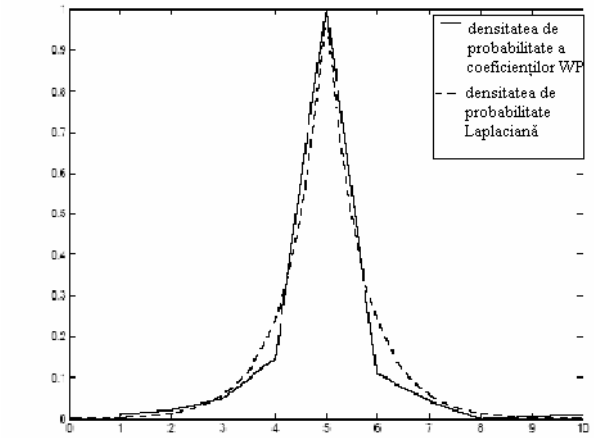

Figura 3.8. Densitatea de probabilitate a coeficienţilor WP.

Procesul de detecţie

Primii 4 paşi ai procesului de detecţie sunt identici cu cei de inserţie. Cheia unică a autorului e folosită pentru a extrage marcajul fără a utiliza imaginea originală.

5. Regiunile principale de cuantizare sunt scanate şi se verifică corelaţia cu cheia de autentificare. Regiunile secundare sunt şi ele examinate pentru a verifica autenticitatea fiecăreia.

6. Comparaţia în interiorul benzi (Intraband): se fac asocieri între regiunile WPC aparţinând aceluiaşi nivel de detalii pentru a decide dacă imaginea a suferit de pe urma modificărilor intenţionate de frecvenţă.

7. Comparaţia între benzi (Interband): se face verificarea între regiunile WPC asociate aceleiaşi regiuni spaţiale pentru a vedea dacă imaginea a fost modificată sau nu.

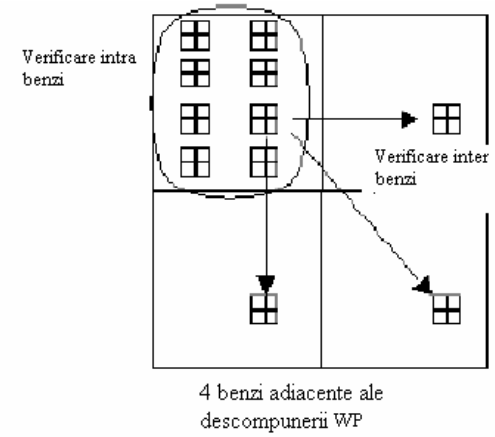

Figura 3.9. Schema de verificare intra/inter benzi [PaqWar02].

8. Pe baza rezultatelor de la 5,6 şi 7 se ia decizia în privința autenticității imaginii. Dacă se decide că a fost atacată se identifică frecvența sau regiunile spațiale modificate.

Rezultatele experimentale din [PaqWar02] confirmă că probabilitatea de alarmă falsă este extrem de redusă. Folosirea transformarii WP pentru inserția watermark-ului creşte robustețea marcajului, făcând ca sistemul să se comporte bine în prezența compresiei JPEG. Performanțele schemei pot fi sporite în continuare prin folosirea imaginilor color deoarece capacitatea de a accepta un marcaj invizibil este mai mare decât în cazul imaginilor alb-negru.

Bibliografie

[BajBor01] Titu I. Bajenescu, Monica E. Borda, "Securitatea in Informatica si Telecomunicatii", Editura Dacia, Cluj – Napoca, 2001

[BarBartCapPiv98] M Barni, F. Bartolini, V. Cappelini and A. Piva, "ADCT Domain System for Robust Image Watermarking", Signal Processing (Special Issue on Watermarking), vol. 66, no.3, 1998

[CoxKilLeiSha96] I.J. Cox, J. Killian, T. Leighton and T. Shamoon, Secure Spread Spectrum Watermarking for Images, Audio and Video, *Proc. of ICIP*, vol. 3, pp.243-246, 1996.

[DifHel76] W. Diffie and M. E. Hellman, "New Directions in Cryptography", IEEE Transactions on Information Theory, vol. IT-22, Nov. 1976, pp: 644-654.

[HarGir98] Frank Hartung, Bernd Girod, "Watermarking of Uncompressed and Compressed Video", Signal Processing 66 (1998) 283 – 301

[InoMiyKat99] H Inoue, A Miyazaki, T Katsura, An image Watermarking Method Based on the Wavelet-Transform, *Proc. of ICIP*, vol. 3, pp.296-300, 1999.

[Isa02] Alexandru Isar, "Securitatea transmiterii informației pe INTERNET", 2002

[IsaNaf98] Alexandru Isar, Ioan Nafornita , "Reprezentari Timp – Frecventa", Editura Politehnica , Timisoara, 1998

[KunHat98] D. Kundur, D. Hatzinakos, "Digital Watermarking using Multiresolution Wavelet Decomposition", Proc. IEEE Int. Conf. On Acoustics, Speech and Signal Processing, Seattle, Washington, Vol. 5, pp. 2969-2972, May 1998.

[Naf04] Corina Nafornita, "Balizarea Imaginilor Statice Folosind Transformata Wavelet Discreta", Raport de Cercetare din Cadrul Grantului CNCSIS, de tip D, cod 47, numar 33385/29.06.04

[Naf05] Corina Nafornita, "Digital Watermarking in the Wavelet Domain", Editura Politehnica, 2005

[Naf05b] Corina Nafornita, "Studiul comportarii la atacuri a imaginilor marcate transparent a imaginilor", Research report for CNCSIS grant, type TD, code 47 no. 34702/24.06.05:

[NafIsa04] Corina Nafornita, Alexandru Isar, "A Wavelet – Based Watermarking for still Images", Buletinul Stiinific al Universitatii Politehnica din Timisoara, Seria Electonica si Telecomunicatii, Tom 49(63), Fascicola 2, 2004

[PaqWar02] A. H. Paquet, R. K. Ward, "Wavelet-Based Digital Watermarking for Image Autenthication", Accepted for Publication at the IEEE Canadian Conference on Acoustics Speech and Signal Processing (ICASSP) 2002

[PerHer99] Fernando Perez – Gonzales, Juan R. Hernandez, "A Tutorial on Digital Watermarking" In Proc. of the 33rd IEEE Annual Carnahan Conference on Security Technology, Madrid, Spain, October 1999.

[PetAndKuh99] Fabien A. P. Peticolas, Ross J. Anderson and Markus G. Kuhn, "Information Hiding –A Survey", Proceedings of the IEEE, special issue on protection of multimedia content, 87(7): 1062-1078, July 1999

[SchTirOsb94] R. G. van Schyndel, A. Z. Tirkel, and C. F. Osborne, "A Digital Watermark", *Proc. of the IEEE Int. Conf. on Image Processing*, vol. 2, pp. 86–90, Austin, Texas, Nov 1994.

[VoyPit99] G. Voyatzis and I. Pitas, "Problems and Challenges in Multimedia Networking and Content Protection", TICSP 30 Series, 1999

[Wal95] S. Walton, "Information Authentication for a Slippery New Age", *Dr. Dobbs Journal*, vol. 20, no. 4, pp. 18–26, Apr 1995.

[WolDel96] R. B. Wolfgang and E. J. Delp, "A Watermark for Digital Images", *Proc. IEEE Int. Conf. on Image Processing*, vol. 3, pp. 219–222, 1996.

[Won98] P.W.Wong, "A watermark for image integrity and ownership verification", in Proc IS&T PIC, Portland, Oregon, 1998.

[WuLiu98] M. Wu, B. Liu: "Watermarking for Image Authentication", ICIP, 1998

[XiaBonArc97] Xiang-Gen Xia, C.G. Boncelet and G.R. Arce, A Multiresolution Watermark for digital images, *Proc. of International Conference on Image Processing*, vol. 3, pp.48-51, 1997.

[XieArc98] L. Xie, G. Arce, Joint Wavelet Compression and Authentication Watermarking,. In IEEE International Conference on Image Processing, Chicago, IL, Oct 1998

[YeuMin97] Yeung, M., Mintzer, F.: An invisible watermarking technique for image verification. In: Proc. IEEE Int. Conf. Image Processing. Volume 2. (1997) 680--683

[YuLuLiaShe00] G. Yu and C. Lu and H. Liao and J. Sheu, "Mean Quantization Blind Watermarking for Image Authentication", Proc. IEEE Int. Conf. on Image Processing, Vancouver, Canada, Vol. III, pp. 706-709, 2000.