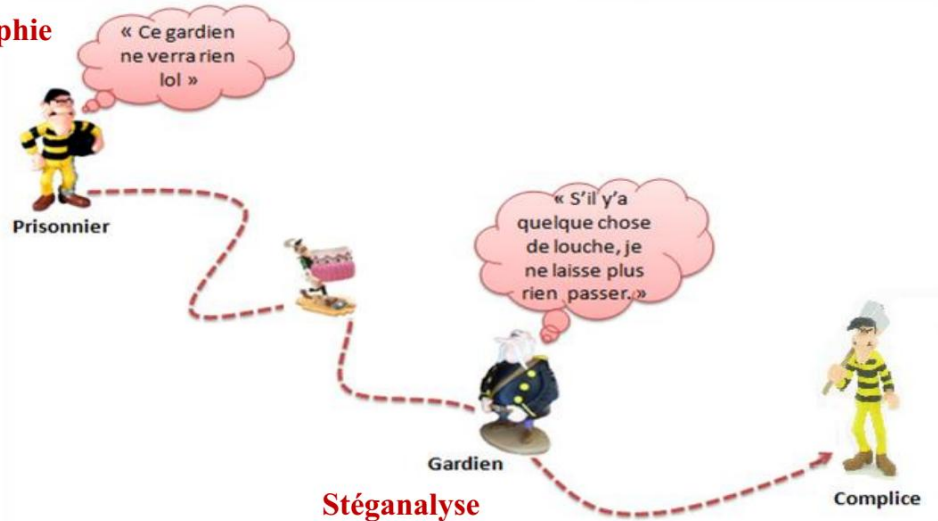


## La stéganalyse

La stéganalyse a pour objectif de détecter la présence de données dissimulées à l'aide d'un algorithme stéganographique. Elle est la discipline duale de la stéganographie.

### Stéganographie



## TYPES DE STEGANALYSE

La stéganalyse peut être appliquée par deux types de personnes. L'attaquant actif, qui connaît la présence de l'information et tente de la modifier ou de l'extraire et l'attaquant passif, c'est-à-dire la personne qui arrive à déceler la présence du message et qui ne fait que constater sa présence.

### 1. Attaque active:

dans ce type d'attaque on souhaite non seulement détecter le message caché mais, en plus, on va chercher à extraire, modifier ou supprimer ces données.

Cette destruction aura souvent lieu par l'intermédiaire de modification de support (e.g: Compression, changement de format, recadrage, symétrie,...).

### 2. Attaque passive:

il s'agit simplement de détecter la présence de messages dissimulés. Ce type d'attaque peut prendre plusieurs formes:

- La lecture ou l'écoute de fichier,
- La comparaison avec le fichier original (s'il est disponible),
- Certaines attaques statistiques (attaque sur le LSB),
- La détection des signatures des logiciels utilisés (étude du code hexadécimal).

## LES METHODES DE LA STEGANALYSE

Selon le type des mesures effectuées pour la distinction entre les images de couverture et les stégosimages, nous distinguons deux types de stéganalyse :

- **La stéganalyse universelle**
- **la stéganalyse spécifique.**

## **1 . STEGANALYSE UNIVERSELLE**

Si les mesures utilisées pour la détection sont indépendantes des algorithmes que nous essayons de détecter, la stéganalyse est dite universelle. La stéganalyse universelle permet alors de répondre à la question « le médium est-il Stéganographié ? ».

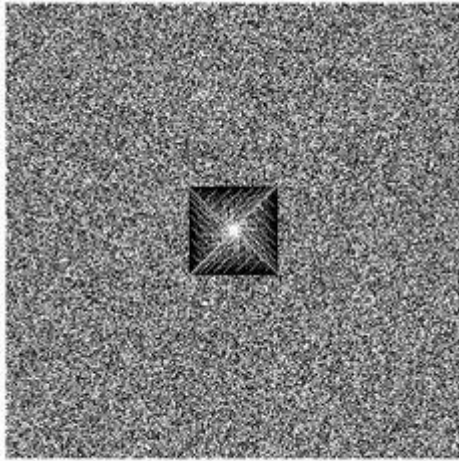
### **1.1. ATTAQUE VISUELLE**

L'insertion d'un message dans le dernier plan de bit peut se faire de façon aléatoire sur l'ensemble des pixels de l'image ou de façon séquentielle à partir de début de l'image.

L'idée de cette attaque est basée sur le fait que une image peu texturée, le plan LSB est corrélé avec l'image d'origine. L'insertion du message perturbe le plan LSB en proportion de la taille de message. Les attaques visuelles appliquent des filtres sur l'image originale et l'image stéganographiée, supprimant les composantes les plus visibles (bits de poids forts) et renforçant les autres (bits de poids faible),

- Si une dissimulation est détectée grâce à des anomalies de couleur l'algorithme de stéganographie a été attaqué avec succès.
- Si un recouvrement n'a pas été détecté par l'observateur, les plans de bits de l'image sont ensuite examinés, en commençant par le plan le moins significatif.

Une dissimulation est typiquement mis en évidence par une partie localisée du bruit dans ce plan (voir Fig. 3) .Si ce bruit est facilement détectée, il est possible d'extraire les données à partir de ce plan. La figure 4 propose un exemple simple d'une attaque visuelle.



**Figure 3:** Une dissimulation est visible au milieu du plan de bits.



**Figure 4:** Affichage du LSB plan de cet objet stego révèle son message caché.

## 1.2. ATTAQUES STATISTIQUES

Attaques statistiques sur LSB dissimulation sont beaucoup plus efficaces qu'une attaque visuelle. Les attaques statistiques font usage de la relation entre les plans de bits dans une image ou la relation entre les pixels dans un plan binaire pour déterminer si un message est inséré dans une image.

- Attaques statistiques sont généralement accordés à travailler contre un algorithme d'intégration particulier, puisque les différentes stratégies de dissimulation affectent la perturbatrice de valeurs de pixels d'une manière unique. Par exemple, un type particulier d'attaque statistique peut détecter tous les algorithmes de dissimulation qui intègrent dans le DCT d'une image JPEG comme Jsteg. Cette même attaque ne sera probablement pas efficace contre la forme la plus simple de LSB dissimulation.
- Le **RS stéganalyse** développé par Fridrich et. al sert une attaque statistique exemplaire. RS stéganalyse commence par définir un ensemble de groupes discrets  $G$  de pixels dans l'image. Les pixels de chaque groupe sont basculés selon un masque  $M$ .

## ▪ STEGANALYSE BASEE SUR DES PAIRES DE VALEURS DE L'IMAGE

Différentes sont les méthodes de stéganalyse basées sur l'analyse statistique des paires de valeurs de pixels. Le principe de ces méthodes se base sur le choix des sous ensembles des paires de pixels ou bien le choix des groupes de pixels vérifiant des hypothèses proposées pour la stéganalyse (eg: égalité des sous ensembles).

La détection se base sur le fait que l'insertion d'un message dans les bits de poids faibles peut modifier ou ne vérifier pas une des hypothèses proposées. Dans cette partie, nous présentons

- l'analyse statistique à base de  $\chi^2$ ,
- le schéma de Memon basée sur les paires de pixels
- le schéma proposée par Fridrich basée sur les groupes de pixels

### Analyse statistique à base de $\chi^2$

La stéganalyse  $\chi^2$  est basée sur le principe que, les fréquences d'apparition des nuances d'une paire de valeurs tendent à l'égalité sous l'action d'insertion LSB (voir les histogrammes en figure 5).

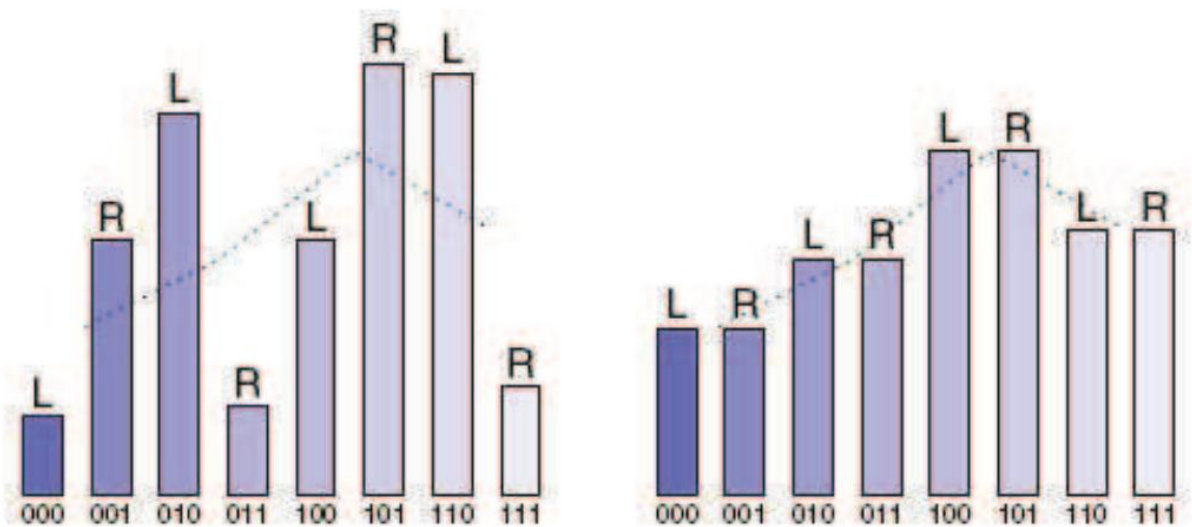


Figure 5. : Partie d'un histogramme d'une image avant et après insertion

Sur la figure précédente, la ligne en pointiez correspond à la moyenne des deux valeurs composant une paire de valeur de pixel. Comme on peut le remarquer, cette moyenne n'est pas affectée par l'insertion.

Cette moyenne sera donc utilisée afin de créer un modèle théorique à partir d'un fichier dont on ne sait pas si il a été modifié. La moyenne théorique sera donc calculée à partir de la formule suivante

$$y_i^* = \frac{n_{2i} + n_{2i+1}}{2} \quad (3.1)$$

La fréquence mesurée sur l'image analysée est  $y_i = n_{2i+1}$ .

Alors la valeur du  $\chi^2$  mesurant la différence des distributions est :

$$\chi_{k-1}^2 = \sum_{i=1}^k \frac{(y_i - y_i^*)^2}{y_i^*} \quad (3.2)$$

Où  $k-1$  représente le degré de liberté (nombre des paires de valeurs de pixels).

La probabilité que les deux distributions soient identiques est donnée par l'expression suivante

$$p = 1 - \frac{1}{2^{\frac{k-1}{2}} \Gamma\left(\frac{k-1}{2}\right)} \int_0^{\chi_{k-1}^2} e^{-\frac{x}{2}} x^{\frac{k-1}{2}-1} dx \quad (3.3)$$

Avec  $\Gamma$  représente la fonction gamma d'Euler qui s'écrit :

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt \quad (3.4)$$

Pour mieux visualiser l'application de cette méthode à une image, on applique cette dernière à l'exemple suivant.

Soient l'image (C) de 4×4 pixels et le message à insérer (1001101000111001).

$$c = \begin{bmatrix} 00 & 00 & 10 & 10 \\ 01 & 11 & 10 & 00 \\ 00 & 11 & 10 & 00 \\ 00 & 00 & 00 & 00 \end{bmatrix} + 1001101000111001 = \begin{bmatrix} 01 & 00 & 10 & 11 \\ 01 & 10 & 11 & 00 \\ 00 & 10 & 11 & 01 \\ 01 & 00 & 00 & 01 \end{bmatrix}$$

Les paires des valeurs de pixels sont [00,01] et [10,11]. La première étape consiste à calculer le nombre d'occurrences de chaque élément, ensuite chaque élément sera classé à l'aide d'un indice suivant le tableau suivant.

Indice	Elément
0	00
1	01
2	10
3	11

En classe ensuite les éléments des deux images (image originale et image stéganographiée) dans les divers indices, cela donne le tableau d'occurrences suivant.

Indice	0	1	2	3
Image originale	9	1	4	2
Image stéganographiée	5	5	3	3

### Tab.3.2 Occurrences de chaque élément

Ensuite, à l'aide de la formule (3.1), la distribution des occurrences théorique peut être calculée.

Dans cet exemple formé de deux paires de valeurs, elle sera formée de deux v valeurs.

$$y_0^* = \frac{n_0 + n_1}{2} = 5$$

$$y_1^* = \frac{n_2 + n_3}{2} = 3$$

Ensuite, avec la formule (3.2), il est possible d'appliquer le test du  $\chi^2$  à notre image stéganographiée, ainsi qu'à notre image originale

$$\chi^2 = \sum_{k \text{ catégories}} \frac{(\text{fréquence observée} - \text{fréquence théorique})^2}{\text{fréquence théorique}}$$

$$\chi^2_{k-1 \text{ Steganographiée}} = \chi^2_{k-1} = \sum_{i=1}^k \frac{(y_i - y_i^*)^2}{y_i^*} = \frac{(5-5)^2}{5} + \frac{(3-3)^2}{3} = 0$$

$$\chi^2_{k-1 \text{ Steganographiée}} = \chi^2_{k-1} = \sum_{i=1}^k \frac{(y_i - y_i^*)^2}{y_i^*} = \frac{(1-5)^2}{5} + \frac{(2-3)^2}{3} = \frac{53}{15}$$

Il est ensuite possible de calculer la probabilité d'avoir un message dissimulé via l'équation (3.3). Le résultat du  $\chi^2$  intervenant dans les bornes d'intégration de la fonction de densité, en ce qui concerne l'image stéganographiée, le résultat est aisément identifiable. Le résultat de l'intégration étant nul, le membre de droite de la soustraction s'annule. Cela donne donc une probabilité de 1.

On appelle loi du  $\chi^2$  à r degré de liberté la répartition des carrés de r variables aléatoires indépendantes, dont chacune est répartie suivant un loi normale d'espérance mathématique nulle et de variance unité. La densité de cette répartition est

$$k_r(u) = \begin{cases} \frac{1}{2^{\frac{r}{2}} \Gamma(\frac{r}{2})} u^{\frac{r}{2}-1} e^{-\frac{u}{2}} & \text{pour } u > 0 \\ 0 & \text{pour } u < 0 \end{cases}$$

avec

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$$

la fonction  $\Gamma$  d'Euler.

$$p = 1 - \frac{1}{2^{\frac{k-1}{2}} \Gamma(\frac{k-1}{2})} \int_0^0 e^{-\frac{x}{2}} x^{\frac{k-1}{2}-1} dx = 1 - 0 = 1$$

En ce qui concerne le même calcul pour l'image originale, le même raccourci n'est pas possible. Il faut donc calculer l'entier de l'équation.

$$p = 1 - \frac{1}{2^{\frac{1}{2}} * 1.772} \int_0^{\frac{53}{15}} e^{-\frac{x}{2}} x^{\frac{k-1}{2}-1} dx = 0.144$$

Il y a donc une probabilité de 14.4 % que l'image originale dissimule des données, ce qui est bien sûr erroné.

En conclusion



- cette méthode donne de bon résultat, mais elle permet uniquement la détection de données cachées de manière séquentielle au niveau des LSB. De plus, elle ne permet qu'une estimation de la taille des données approximative.
- cette méthode a été adaptée avec succès aux images JPEG. Ceci a été rendu possible en appliquant la même logique que ci-dessus, mais cette fois-ci sur les coefficients DCT des images JPEG

## --- RS stéganalyse

### Principe

Cette méthode, développée par Fridrich, Miroslav Goljan et Rui Du, est basée sur la classification de groupes de pixels en catégories distinctes.

En se basant sur une image de  $L \times H$  pixels, dont chaque pixel a une valeur comprise dans l'ensemble  $P$ . Pour une image en 256 niveaux de gris (8 bits),  $P = \{0, \dots, 255\}$ .

La première opération à effectuer, est de diviser l'image en groupes disjoints de  $n$  pixels adjacents  $G = (x_1, \dots, x_n)$ . Ce nombre  $n$  est défini par les auteurs à 4. Une fonction de discrimination  $f$  permettant d'évaluer la rugosité de chacun des groupes  $G$  sera définie.

Elle attribuera une nombre réel à chacun des groupes, plus le groupe sera bruité, plus cette valeurs sera grande. Un exemple de cette fonction de discrimination peut être donné par :

$$f(G) = f(x_1, \dots, x_n) = \text{somme}_{i=1, n-1} |x_{i+1} - x_i| \quad (5)$$

Cette fonction peut être déterminée par rapport au propriétés statistiques de l'image. Pour la suite, la fonction (5) sera utilisée comme fonction de discrimination.

Ensuite, il est défini des fonction réversible sur l'ensemble  $P$ . Ces fonctions consiste essentiellement en des permutations de valeurs. Ces fonctions sont au nombre de trois :

$F_1 : 0 \leftrightarrow 1, 2 \leftrightarrow 3, \dots, 254 \leftrightarrow 255$

$F_{-1} : -1 \leftrightarrow 0, 1 \leftrightarrow 2, \dots, 255 \leftrightarrow 256$

$F_0$  : identite

Afin d'appliquer ces fonctions de permutation sur les groupes  $G$  de pixels définis plus haut, Il faut définir une matrice de permutation  $M(\text{masque})$ , ainsi que la matrice inverse  $-M$  (Le masque  $-M$  est défini à partir de  $M = (x_j)$  par  $-M = (-x_j)$  pour tout  $j$ ). Ces matrices auront une taille de  $1 \times n$ , avec des valeurs comprises dans  $\{-1, 0, 1\}$ . Cela



définira la fonction de permutation à appliquer à chacun des membre du groupes de tel sorte de donner le groupe permuté  $F(G) = (FM(1)(x_1), FM(2)(x_2), \dots, FM(n)(x_n))$ .

Ceci étant défini, il est possible de classier chaque groupe de pixels dans une des trois catégories distinctes. Ces catégories sont définies de la manière suivante :

Groupe Régulier :  $G \in R \Leftrightarrow f(F(G)) > f(G)$

Groupe Singulier :  $G \in S \Leftrightarrow f(F(G)) < f(G)$

Groupe inutilisable :  $G \in U \Leftrightarrow f(F(G)) = f(G)$

Les fonctions de permutation des valeurs des pixels simulent l'ajout de bruits à l'image source. Dans le cas d'une image, l'ajout de bruit aura comme influence une augmentation de la fonction de discrimination. Au niveau de la classification des groupes, cela représentera une augmentation de ceux dans le groupe R.

En définissant RM comme le pourcentage de groupes réguliers après l'application des transformations de la matrice M, SM en étant le pendant pour les groupes singuliers. De la même manière, il faut définir R-M et S-M pour la matrice -M. Il est possible d'en déduire les relations suivantes :

$$RM + SM \leq 1 \text{ et } R-M + S-M \leq 1$$

En se basant sur l'équation (6), il est possible de faire l'hypothèse que l'application de la matrice M ou de sa matrice inverse, ne va pas changer de manière significative la distribution des groupes. Donc sur une image, les relations  $RM \sim R-M$  et  $SM \sim S-M$  devrait être vérifiées.

$$F^{-1}(x) = F(x + 1) - 1 \quad (6)$$

D'après les expériences menées par les auteurs, cette relation est vérifiée, et devient un bon indicateur pour les images issues d'appareil photo numériques ou scanners dans des formats avec ou sans perte d'informations.

- On note RM le pourcentage de groupes de type R calculé à l'aide du masque M, et SM le pourcentage de groupes de type S. L'hypothèse sur laquelle se base l'analyse est le fait que sur des images naturelles RM et R-M sont égaux, ainsi que SM et S-M. Le bruitage du plan LSB par la stéganographie LSB modifie cette égalité et fait tendre la différence R - S vers 0 à mesure que la taille du message inséré augmente. Après permutation de 50% du dernier plan de bits (la taille du message est alors égal à la taille du plan de bit), on a  $RM = SM$ .
- L'objet de la stéganalyse RS est l'extrapolation des courbes RM, R-M, SM et S-M en fonction de la taille du message p afin de calculer leurs intersections et d'en déduire la taille p. Les courbes R-M et S-M sont modélisées par des droites, alors que les courbes RM et SM le sont par des équations du second degré. Soit donc à analyser une image contenant un message de longueur p, en

pourcentage du nombre total de pixels. L'algorithme de stéganalyse consiste donc d'abord en le calcul du nombre de groupes  $RM(p/2)$ ,  $SM(p/2)$ ,  $R-M(p/2)$  et  $S-M(p/2)$  de l'image analysée.

- Puis, en permutant par LSB tous les pixels de l'image (i.e. application de  $F_1$  à tous les pixels) on peut calculer les cardinaux et les pourcentages des groupes  $RM(1 - p/2)$ ,  $SM(1-p/2)$ ,  $R-M(1-p/2)$  et  $S-M(1-p/2)$ . En supposant que les courbes  $RM$  et  $R-M$  se coupent en la même abscisse que les courbes  $SM$  et  $S-M$ , et que  $RM(1/2) = SM(1/2)$ , on déduit les équations des paraboles et leur intersection.

Un petit exemple permettant de visualiser l'impact sur la classification de l'adjonction d'un message à une image peut être donné. Pour ce faire, en se basant sur l'image suivante :

$$I_{base} = \begin{bmatrix} 139 & 144 & 149 & 153 & 155 & 155 & 155 & 155 \\ 144 & 151 & 153 & 156 & 159 & 156 & 156 & 156 \\ 150 & 155 & 160 & 163 & 158 & 156 & 156 & 156 \\ 159 & 161 & 162 & 160 & 160 & 159 & 159 & 159 \\ 159 & 160 & 161 & 162 & 162 & 155 & 155 & 155 \\ 161 & 161 & 161 & 161 & 160 & 157 & 157 & 157 \\ 162 & 162 & 161 & 163 & 162 & 157 & 157 & 157 \\ 162 & 162 & 161 & 161 & 163 & 158 & 158 & 158 \end{bmatrix} \quad (9)$$

Cette image de 8x8 pixels sera utilisé comme image de base, donc non stéganographiée, pour le reste de la démonstration. Sur cette base, des clusters de  $n$  pixels adjacents doivent être défini. Ici,  $n$  sera égal à 4, ce qui donne, pour la première ligne de l'image, les groupements suivant.

$$G_1 = (139, 144, 149, 153) \text{ et } G_2 = (155, 155, 155, 155)$$

Sur l'ensemble des groupes  $G$  définis, on applique la fonction de discrimination (5). Le tableau 21 donne les résultats obtenu sur l'image originale. Ces résultats seront nécessaires pour déterminer la classification de chaque groupes après avoir appliqué la matrice de permutation.

G	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
f(G)	14	0	11	3	13	2	5	1	3	7	0	3	3	5	1	5

TAB. 21 – Résultat de la fonction de discrimination sur l'image de base

Après avoir obtenu ces résultats, il faut maintenant perturber notre image d'origine à l'aide des fonctions réversible, ainsi que de la matrice de permutation  $M = \begin{bmatrix} 0 & 1 & 1 & 0 \end{bmatrix}$ . L'image résultante aura l'allure suivante :

$$I_{base_{perm}} = \begin{bmatrix} 139 & 145 & 148 & 153 & 155 & 154 & 154 & 155 \\ 144 & 150 & 152 & 156 & 159 & 157 & 157 & 156 \\ 150 & 154 & 161 & 163 & 158 & 157 & 157 & 156 \\ 159 & 160 & 163 & 160 & 160 & 158 & 158 & 159 \\ 159 & 161 & 160 & 162 & 162 & 154 & 154 & 155 \\ 161 & 160 & 160 & 161 & 160 & 156 & 156 & 157 \\ 162 & 163 & 160 & 163 & 162 & 156 & 156 & 157 \\ 162 & 163 & 160 & 161 & 163 & 159 & 159 & 158 \end{bmatrix} \quad (10)$$

Comparativement à (9), seul les éléments centraux sont modifiés en utilisant la fonction  $F_1$ . Le tableau 22 compare les valeurs obtenues pour chaque groups, ainsi que leur classification.

G	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$f(G)$	14	0	11	3	13	2	5	1	3	7	0	3	3	5	1	5
$f(F(G))$	14	2	12	3	13	2	7	3	4	9	2	5	7	7	5	5
Classification	U	R	R	U	U	U	R	R	R	R	R	R	R	R	R	U

TAB. 22 – Résultat de la fonction de discrimination et classification

La théorie énoncée plus haut, disant que la perturbation de l'image provoque une augmentation de la fonction de discrimination, est vérifiée. En effet, la majorité des groupes sont défini comme étant réguliers. En dissimulant le message de taille maximum (11), il est possible d'observer l'impact sur une image contenant un message.

$$m = 0000\ 0001\ 0010\ 0011\ 0100\ 0101\ 0110\ 0111\ 1000\ 1001\ 1010\ 1011\ 1100\ 1101\ 1110\ 11(11)$$

L'image résultant de la dissimulation du message (11), de manière séquentielle par surécriture du LSB au sein du conteneur est la suivante :

$$I_{Stego} = \begin{bmatrix} 138 & 144 & 148 & 152 & 154 & 154 & 154 & 155 \\ 144 & 150 & 153 & 156 & 158 & 156 & 157 & 157 \\ 150 & 155 & 160 & 162 & 158 & 157 & 156 & 157 \\ 158 & 161 & 163 & 160 & 160 & 159 & 159 & 159 \\ 159 & 160 & 160 & 162 & 163 & 154 & 154 & 155 \\ 161 & 160 & 161 & 160 & 161 & 156 & 157 & 157 \\ 163 & 163 & 160 & 162 & 163 & 157 & 156 & 157 \\ 163 & 163 & 161 & 160 & 163 & 159 & 159 & 159 \end{bmatrix} \quad (12)$$

Comme sur l'image d'origine, il faut appliquer la fonction de discrimination (5), afin de calculer les valeurs de  $f(G)$ . Ensuite, à l'aide de la même matrice  $M$  que précédemment, il faut perturber cette image pour obtenir nos valeurs de  $F(f(G))$ . Le tableau 23 est une récapitulation des valeurs obtenues au cours de cet exemple.

G	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$f(G_{base})$	14	0	11	3	13	2	5	1	3	7	0	3	3	5	1	5
$f(F(G_{base}))$	14	2	12	3	13	2	7	3	4	9	2	5	7	7	5	5
$Classification_{base}$	U	R	R	U	U	U	R	R	R	R	R	R	R	R	R	U
$f(G_{stego})$	12	1	12	3	12	3	8	1	3	10	3	6	5	8	3	4
$f(F(G_{stego}))$	14	1	12	3	12	3	6	3	3	8	1	6	3	8	3	6
$Classification_{stego}$	R	U	U	U	U	U	S	R	U	S	S	U	S	U	U	R

TAB. 23 – Récapitulation des résultats obtenus

Contrairement à l'image de base, celle contenant un message stéganographié possède plus de groupes de pixel classé comme singulier que ceux étant défini comme régulier. L'image d'exemple étant très petite, la différence n'est que minime. Cette méthode nécessite une plus grande surface pour permettre d'exploiter son potentiel. Les phases suivantes ne seront donc pas traitées avec cet exemple, car le résultat en serait certainement décevant.

Cependant, d'après les expériences menées par les auteurs, l'analyse RS se révèle très efficace pour la détection de contenu stéganographié à l'aide de la méthode de surécriture des LSB.

Les faiblesses de la méthode sont, d'une part qu'elle est inutilisable sur des images fortement bruitées. Se basant sur l'ajout de bruit, elle ne permet pas un seuil de détection suffisant sur des images déjà bruitées. D'autre part, cette méthode est plus efficace sur des images modifiées de manière aléatoire. Si les modifications sont ciblées sur une région précise, les résultats de l'analyse risquent de ne pas laisser apparaître ces modifications. Cependant, en complément à la méthode du  $\chi^2$  (qui arrive à détecter les modifications séquentielles), une large palette d'algorithmes procédant par surécriture du LSB peuvent être détectés.

## DÉTECTION DES AUTRES TECHNIQUES DE STÉGANOGRAPHIE

-- LA SPIRALE EMBEDDING