



HERVÉ SCHAUER CONSULTANTS
Cabinet de Consultants en Sécurité Informatique depuis
1989
Spécialisé sur Unix, Windows, TCP/IP et Internet

Lutte contre le spam

Spam-image :

Présentation et moyens de

lutte

Louis Nyffenegger
<Louis.nyffenegger@hsc.fr>

- Problématiques de la lutte antispam
- Qu'est ce qu'un spam-image ?
- Pourquoi il faut combattre le spam-image ?
- Le problème des images
- Les méthodes simples
- La reconnaissance de texte
- Méthode à base de Datamining
- Comparaison des différentes méthodes
- Conclusion

- Le nombre de spams augmente chaque année
- Les spammeurs se « professionnalisent »
- Les faux-positifs ne sont **pas acceptables**
- Le temps d'apprentissage peut être très long, seul le **temps de reconnaissance doit être minimisé.**

- Phénomène relativement nouveau
- Message contenant une image
- Différents format d'images utilisés
- 2006 : l'année du spam-image ?

THE BEST REPLICA WATCHES FOR SALE

We offer only Luxury trademarks!

Christmas Special discount prices!

Get 25% off total price for 2+ watches!

98% Perfectly Accurate Markings!

[Click here \(www.timeopera.com \)!!!](http://www.timeopera.com)

HELLO

CHRISTMAS MEN'S POWER CHARGE!

CIALIS + VIAGRA

10 + 10 = \$129.95

20 + 20 = \$249.95

30 + 30 = \$319.95

20% DISCOUNT!!

and most popular products

Cialis Soft, Viagra professional, Viagra Soft, Cialis, Valium, Generic Viagra, Xanax, Soma, Ambien and more more more!

[CLICK HERE!](#)

HELP TO SAVE THE CHILDREN!

<http://www.savechildren.net>

We are all like this... We do not believe that anything bad might happen to us. It seems to us that we will never find the trouble round our door. Seeing a woman with her handicapped child in the street, we pretend not to notice them with a mixed feeling of shame and sympathy. And, to be honest, we are glad that this misfortune is not ours.

The Childrens Clinical Hospital is the largest federal institution providing medical aid to all handicapped children. We try to accept any child in need of our assistance irrespective of its parents' income.

Please donate at <http://www.savechildren.net>

(the website of the charity fund for handicapped and socially challenged children).

We are extremely grateful to you for your help. It means that you are next to us and contribute to our common case. Thank you for your assistance rendered to our Hospital as it is truly invaluable because it is an expression of the feeling of comradeship and brotherly love. God can miraculously unite us and make brothers and sisters those who did not know each other yesterday.

Head Physician of the hospital, Dr. Nikolai N. Vaganov

Content-Type: text/html;

charset="windows-1251"

Content-Transfer-Encoding: quoted-printable

[...]

**<IMG alt=3D"" =hspace=3D
src=3D"cid:721221ac240a\$09110bba\$fe127bc0@*****" >**

[...]

-----=_NextPart_000_0000_82A6958B.FFC27482

Content-Type: image/gif;

name="ixm.gif"

Content-Transfer-Encoding: base64

Content-ID: <721221ac240a\$09110bba\$fe127bc0@***>**

- Tout le monde n'utilise pas *mutt* ;)
- Utilisation de plus de ressources informatiques :
 - Espace disque
 - Bande passante
- Empoisonnement de filtre Bayesiens
- Peut contenir du contenu explicite ou des malwares
- C'est du spam !!!

- Pas de possibilité simple de mesure d'égalité ou de distance entre 2 images :
 - Le contenu d'une image n'est pas plus grand que celui d'une autre
 - Une image n'est pas égale à une autre
- Ce n'est pas un problème avec 2 solutions possibles : spam et pas spam, une image peut être un paysage, un texte, un dessin, ...
- Il n'est pas possible de reconnaître efficacement une partie d'une image comme un mot dans un texte
- Une image peut changer dans le temps : gif animé

- Sur un serveur de particulier
- Sur 5000 spams (14 $\text{score spamassassin} > 8$) :
 - 2200 contiennent une image :
 - 70 images apparaissent plus d'une fois :
 - La plupart apparaissent 2 fois
 - Une dizaine apparaissent 10 fois
 - 1 apparaît 1000 fois
 - En dimensionnant les images en 40x40 :
 - 670 images apparaissent au moins 2 fois

Méthode simple : toute image est un spam

- Reconnaissance de tous les messages contenant une image :

```
full      __JPG_ATTACH  /imageVjpeg/i
```

```
full      __GIF_ATTACH  /imageVgif/i
```

- Mise en place d'une règle SpamAssassin qui est validée quand un message au format HTML avec une adresse de retour erronée est reçu:

```
describe  IMG_SPAM    fake reply in html with image
```

```
meta      IMG_SPAM    (__JPG_ATTACH || __GIF_ATTACH) && UN_FAKE_REPLY &&  
HTML_MESSAGE
```

```
score     IMG_SPAM    3
```

Tiré de :

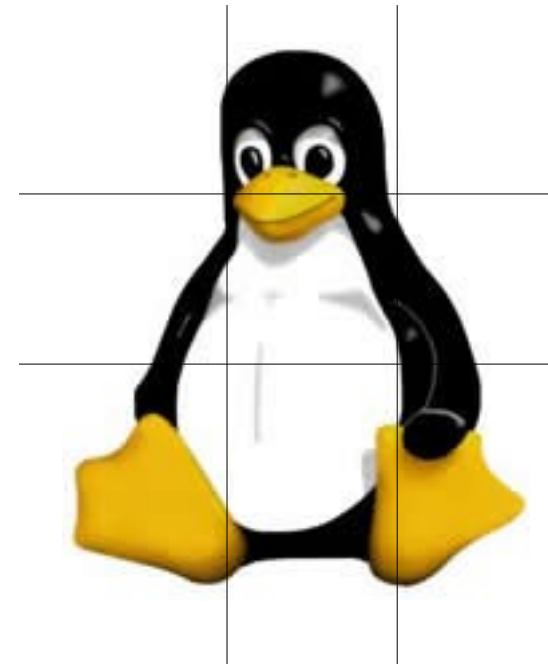
<http://blog.devnull.fr/post/2006/12/16/Catching-image-spam-with-Spamassassin>

- Fonctionnement : réalisation d'une **empreinte** (md5) de chaque image et comparaison à une **whitelist et/ou une blacklist**
- Avantages : **Temps de calculs faibles**
- Limites :
 - Apprentissage manuel ou simpliste
 - Très facilement contournable et déjà contourné
- Améliorations : réduire la taille de l'image avant l'application du md5

- Fonctionnement:
 - Calcul de la signature de l'image
 - Si l'image est référencée comme spam : ajout du score associé
 - Si l'image n'est pas référencée : le score spamassassin permet de classer automatiquement l'image
- Avantages :
 - « apprentissage » simpliste
 - Rapide
- Inconvénient :
 - Facilement contournable

- Fonctionnement : une reconnaissance de texte est réalisé sur l'image, puis l'antispam cherche des mots clés connus
- Avantages :
 - Difficile à contourner
 - Pas d'apprentissage nécessaire
- Limites :
 - Temps de calcul élevé
 - Les spammeurs utilisent des captchas
- Solutions existantes : un plugin SpamAssassin : FuzzyOCR



- Calcul de la couleur moyenne d'une image :
- Découpage de l'image en N parties puis calcul de la valeur moyenne de chaque partie
- Exemples :

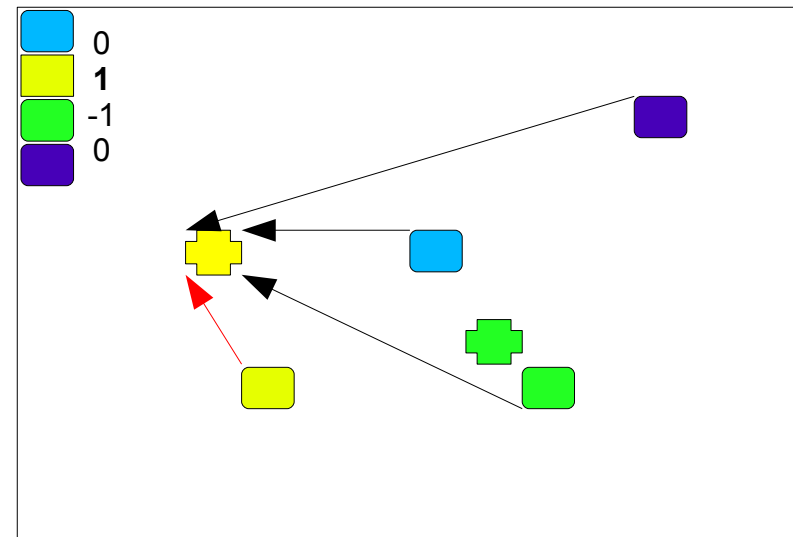
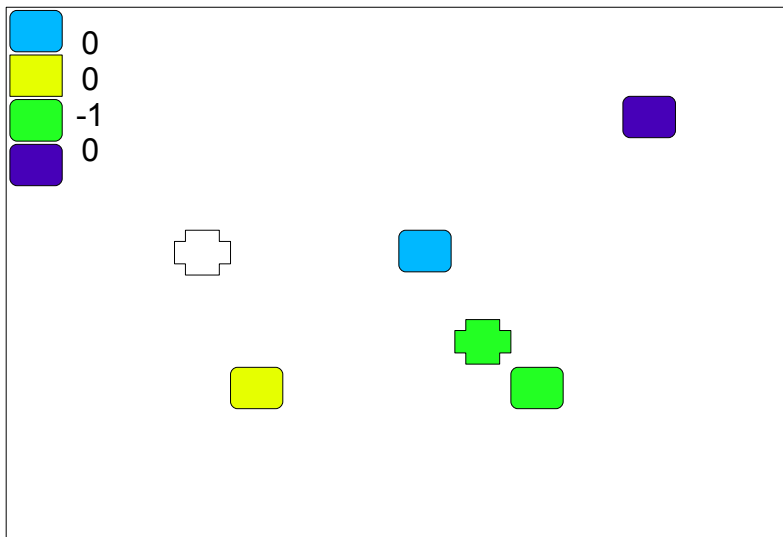
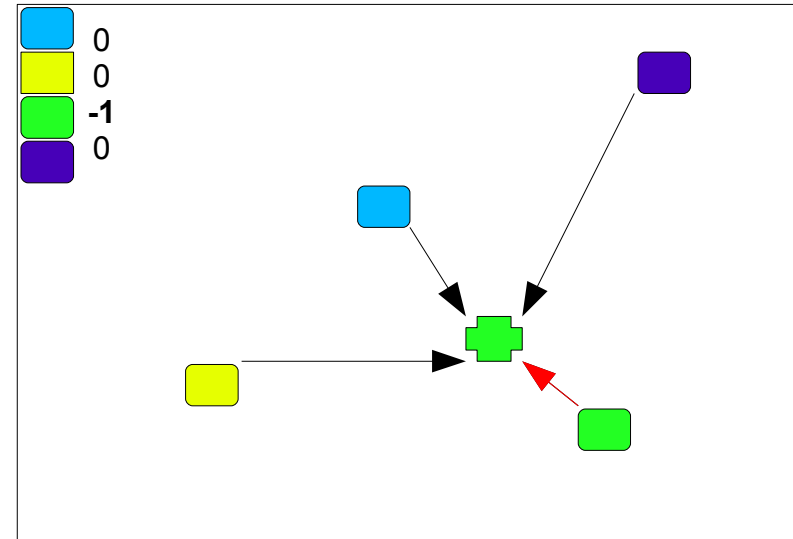
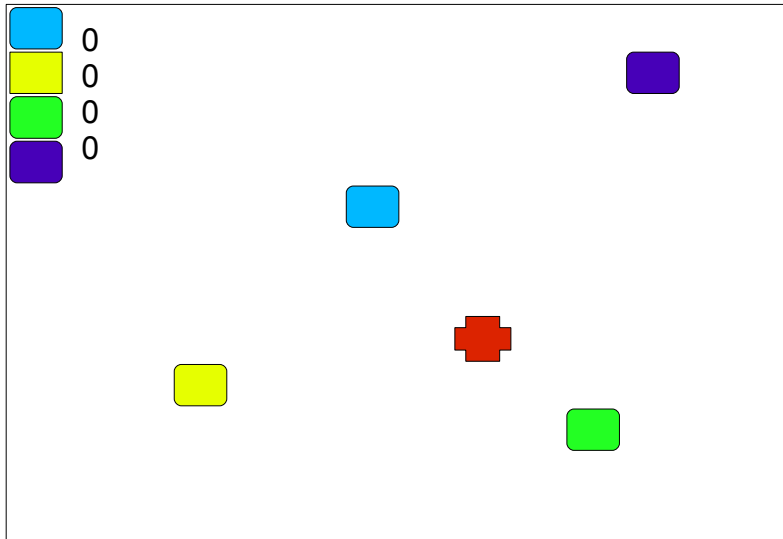


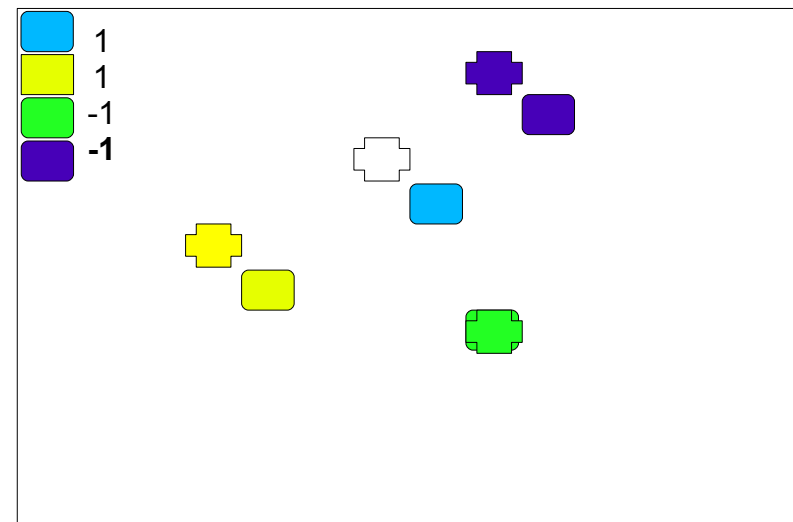
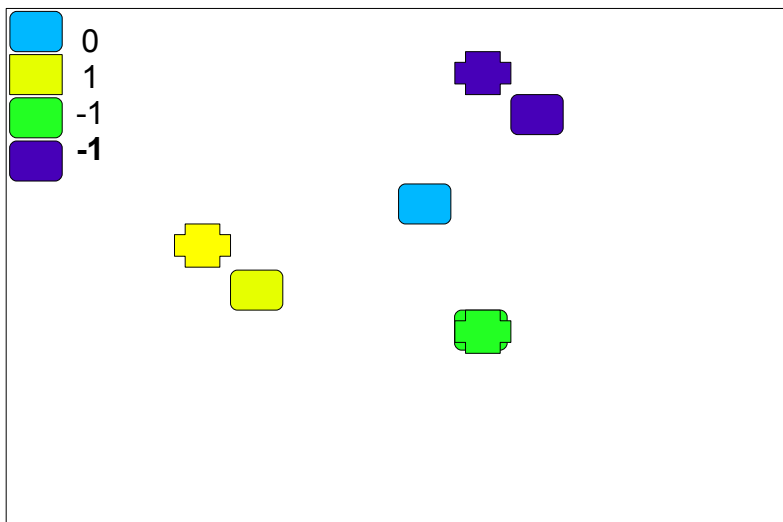
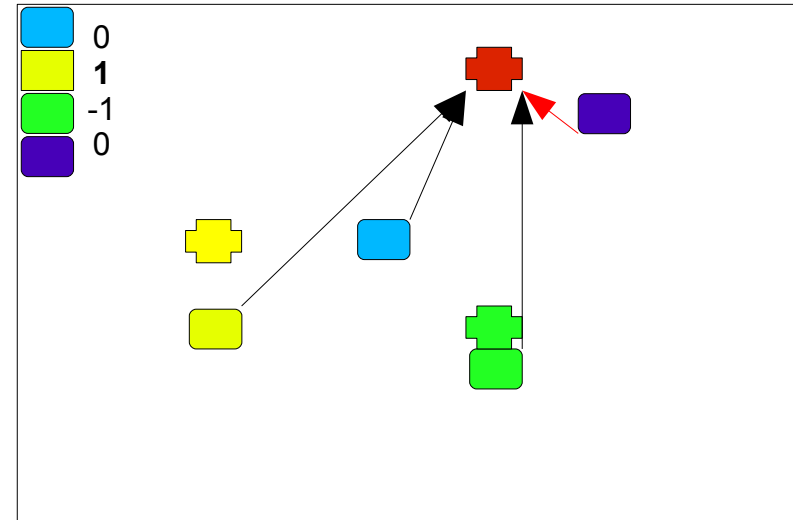
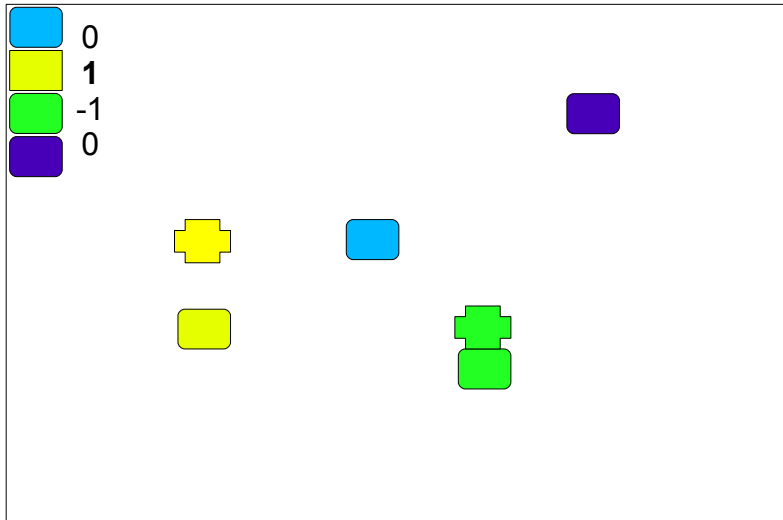
- Répartition RGB de l'image
- Répartition RGB de chaque partie de l'image
- Exemples:

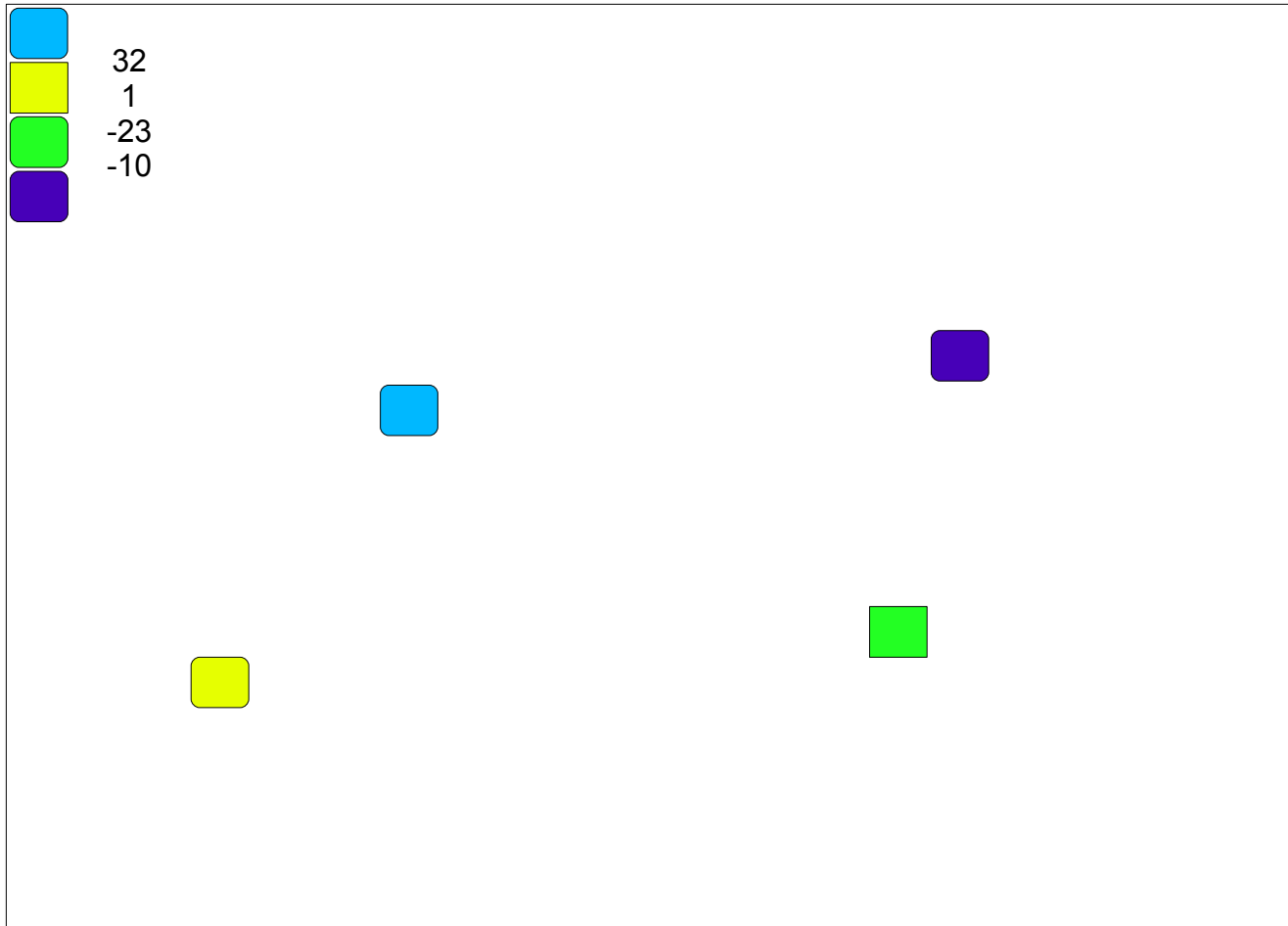


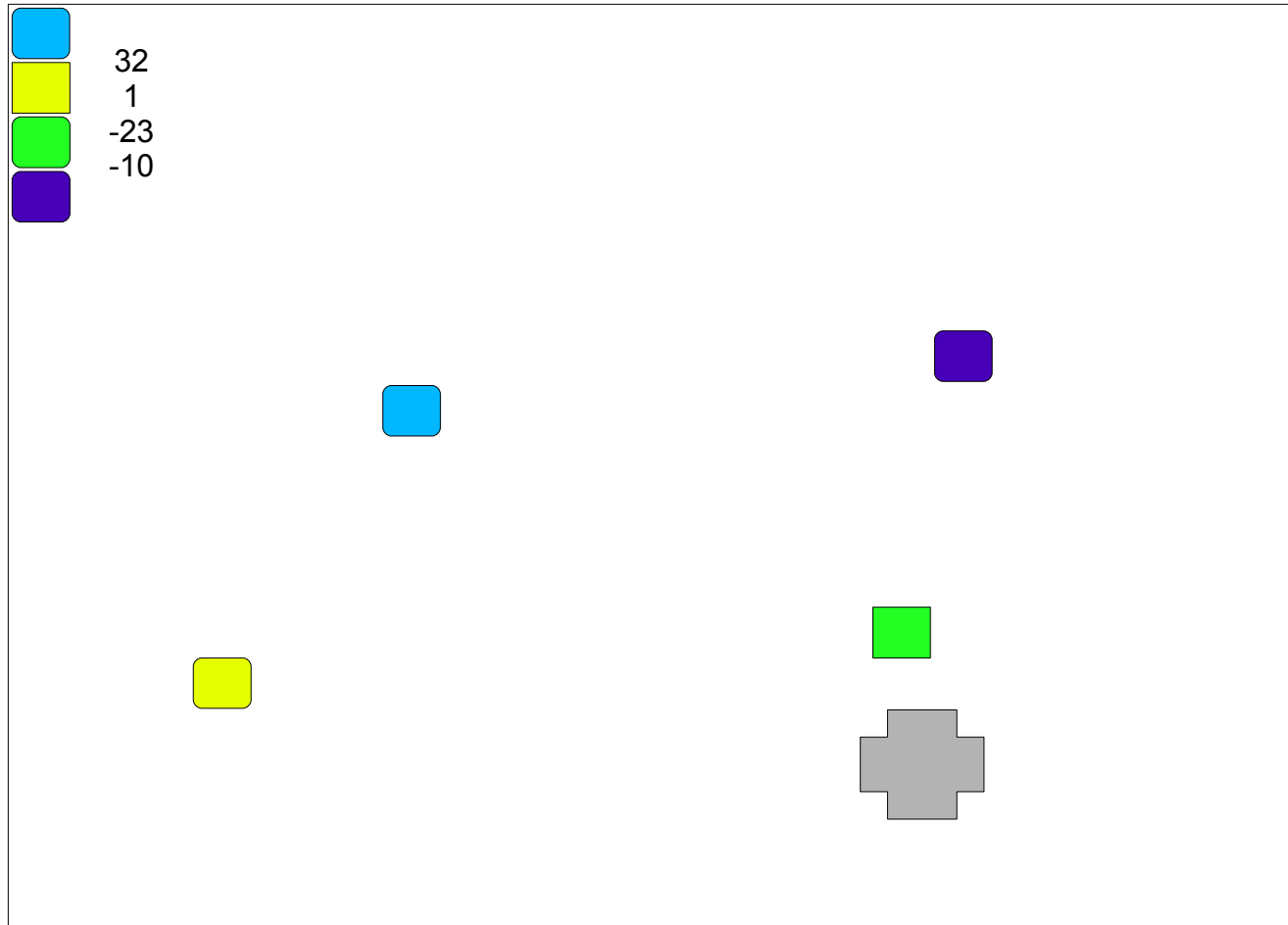
- K plus proches voisins :
 - Recherche des K images les plus proches de l' image et comptabilisation du nombre de spams et de hams
 - Problème : énormément de calcul à réaliser
- K-means : classification automatique de données
 - Classification automatiquement des données en n groupes
 - Temps de calcul faible mais l'utilisation de 2 groupes (spams/hams) ne permet pas d'avoir de bons résultats

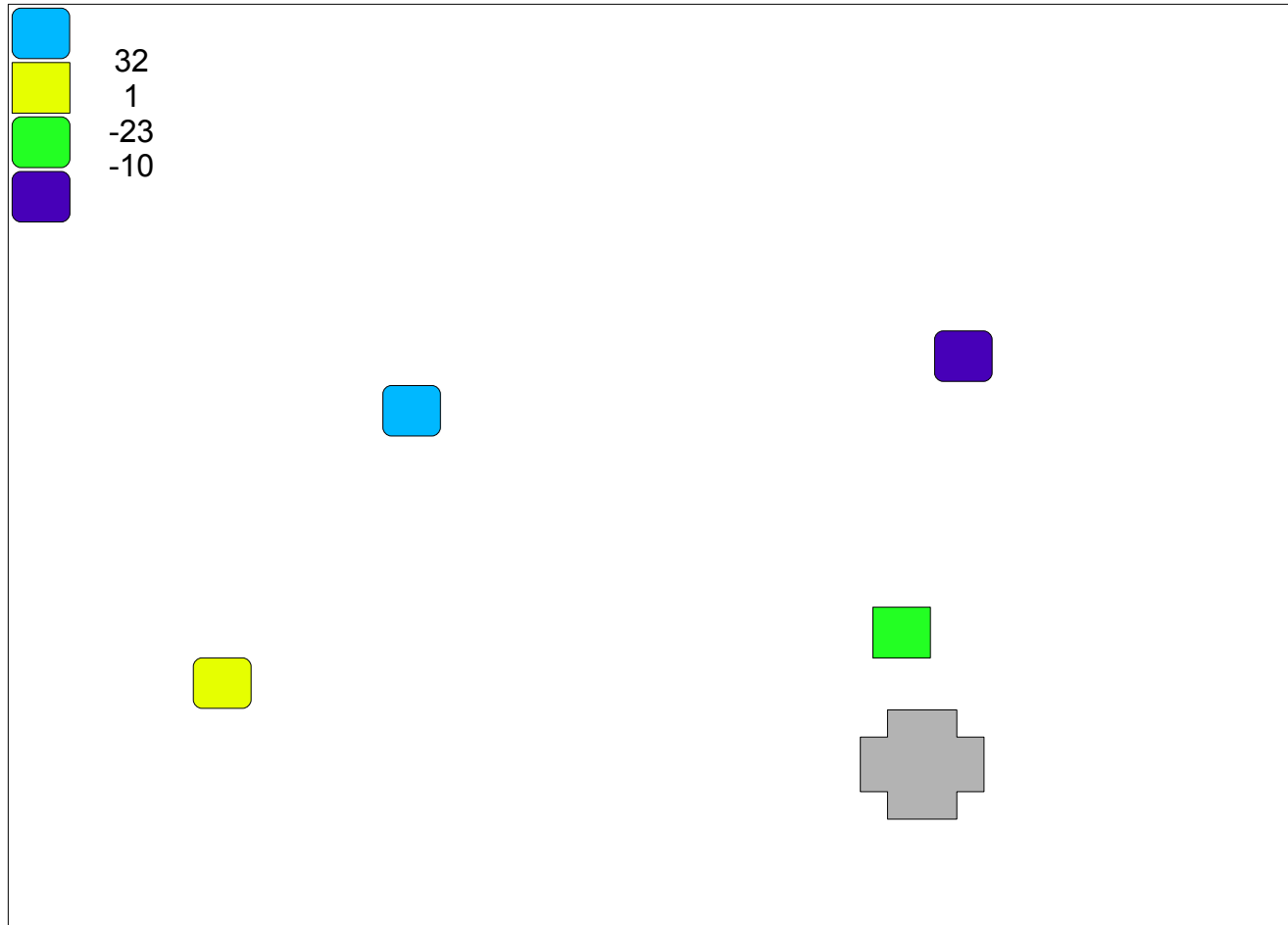
- Étude d'un exemple simple en 2 dimensions
- Base d'apprentissage contenant :
 - Des images légitimes 
 - Des images provenant de spam-image 
- Calcul « d'images moyennes » par l'algorithme à partir de ces images et d'un score associé:
 - Diminution du score d'un point à chaque spam-image
 - Augmentation du score d'un point à chaque image légitime











SPAM !!

- Dimension du problème plus grande
- Plus la quantité d'informations différentes récupérée sur l'image augmente plus le temps de calcul est long.
- Distinguer les informations importantes de l'image de celles qui ne le sont pas (Analyse en Composante Principale)
- Optimisation du nombre de neurones
- Images simples donc détection efficace

- Des méthodes simples déjà dépassées
 - Images bruitées
 - GIF animés
- Les méthodes complexes arrivent à de bons résultats :
 - FuzzyOCR
 - Datamining
- Les méthodes complexes sont gourmandes :
 - Temps de calcul
 - Ressource nécessaire

- La détection du spam-image en est encore à ces débuts, des solutions existent et sont de plus en plus efficaces
- La solution se trouve peut-être dans une augmentation de l'interaction avec l'utilisateur
- Une gestion de la dimension temporelle du spam peut sans doute apporter des améliorations
- Éviter la création d'usine à gaz antispam
- Une lutte sans fin, bientôt la VoIP :(

- **Formation Postfix et antispam :**

- <http://www.hsc.fr/services/formations/>



Paris : 12 juin

Questions ?

Louis.Nyffenegger@hsc.fr
www.hsc.fr