VirusBuster
www.virusbuster.hu

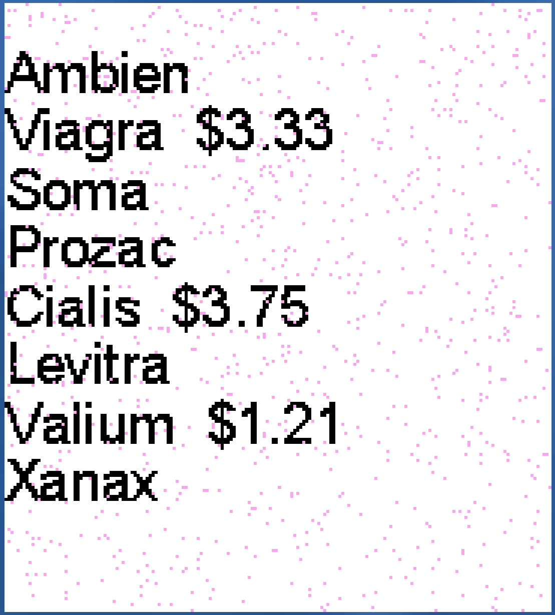
Detecting SPAM pictures using statistical features

Sandor Antal
santal@virusbuster.hu

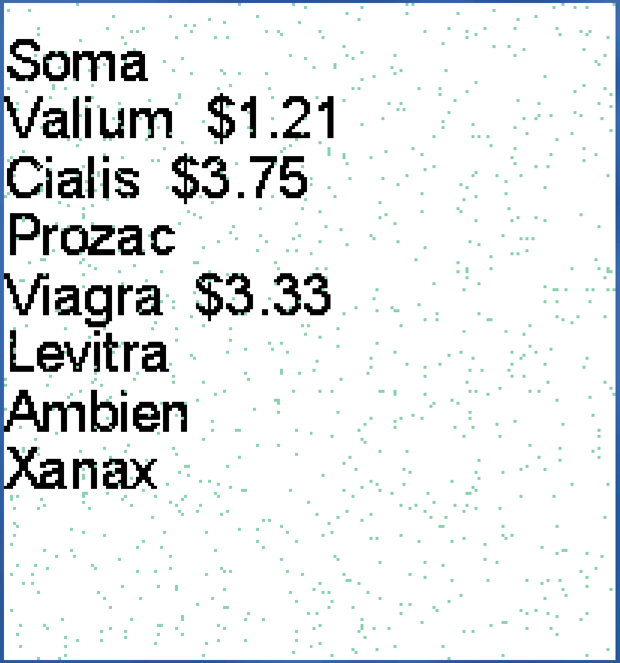
What is *Polimorfic Image Spam*?

- Unsolicited bulk email (*spam*)
- Essential information is in an attached image (*image* spam)
- Image is usually varied randomly to deceive checksum-based methods (*polimorfic* image spam)

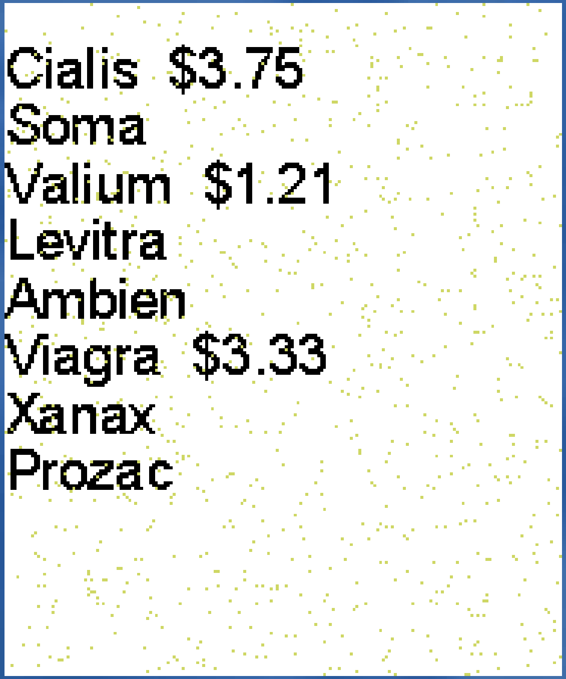
Polymorphic Spam Images Sample



Ambien
Viagra $3.33
Soma
Prozac
Cialis $3.75
Levitra
Valium $1.21
Xanax



Soma
Valium $1.21
Cialis $3.75
Prozac
Viagra $3.33
Levitra
Ambien
Xanax



Cialis $3.75
Soma
Valium $1.21
Levitra
Ambien
Viagra $3.33
Xanax
Prozac

add **noise**
(easily ignored
by humans)

reorder text

change image **size**

Our Goal

… is to develop an image filter method, which performs

- high detection rate in varied *spam* images,
- low false positive rate in *ham* images, and
- acceptable performance.

Image Filtering Methods

- Using checksum-based hash (Accurate Hash Method, AHM)
- Using OCR to get the text of image
- Getting and evaluating file & image attributes by
  - Similarity Hash Method (SHM)
  - Decision Tree Method (DTM)

Accurate Hash Method

- Calculates a checksum of the image as a hash key
- Compares it to the keys of trained spam and ham pictures
- If hits, image considered the same as the trained image

Accurate Hash Method (continued)

- The image doesn’t have to be rendered
  - It is a quite fast method.
- If two images differ, their hash keys will, too.
  - Cannot detect varied instances
- Database tokens exist for every trained image

Optical Character Recognition Method

Recognizes the characters and renders to text which is processed as a normal text part of the mail (*plain / text*).

- Can recognize spam instances of new family
- Can't detect images without text
- Can be deceived by noise
- Very slow and has needs a lot of resources.

Attribute-based Decision Methods

- Attributes
  - File attributes (w/o rendering)
  - Image attributes (rendering)
- Evaluation methods
  - Similarity hash method
  - Decision tree building

File Attributes

Available without rendering

- File format (e. g. JPEG, GIF, BMP etc)
- File length
- Average byte value
- Variance of bytes
- Image dimensions (in most cases)
- etc

Image Attributes

- Brightness
- Contrast
- Number of colors
- etc

For getting them needs to render, process and sometimes transform the image.

Image Transformations

You can get attributes both from original and transformed image

- Filters: Blur, Median etc.
- Gradient image generation
- Transforming into grayscale
- Thresholding (binary image)
- Resizing image
- Fourier transformation, etc...

Similarity Hash Method (SHM)

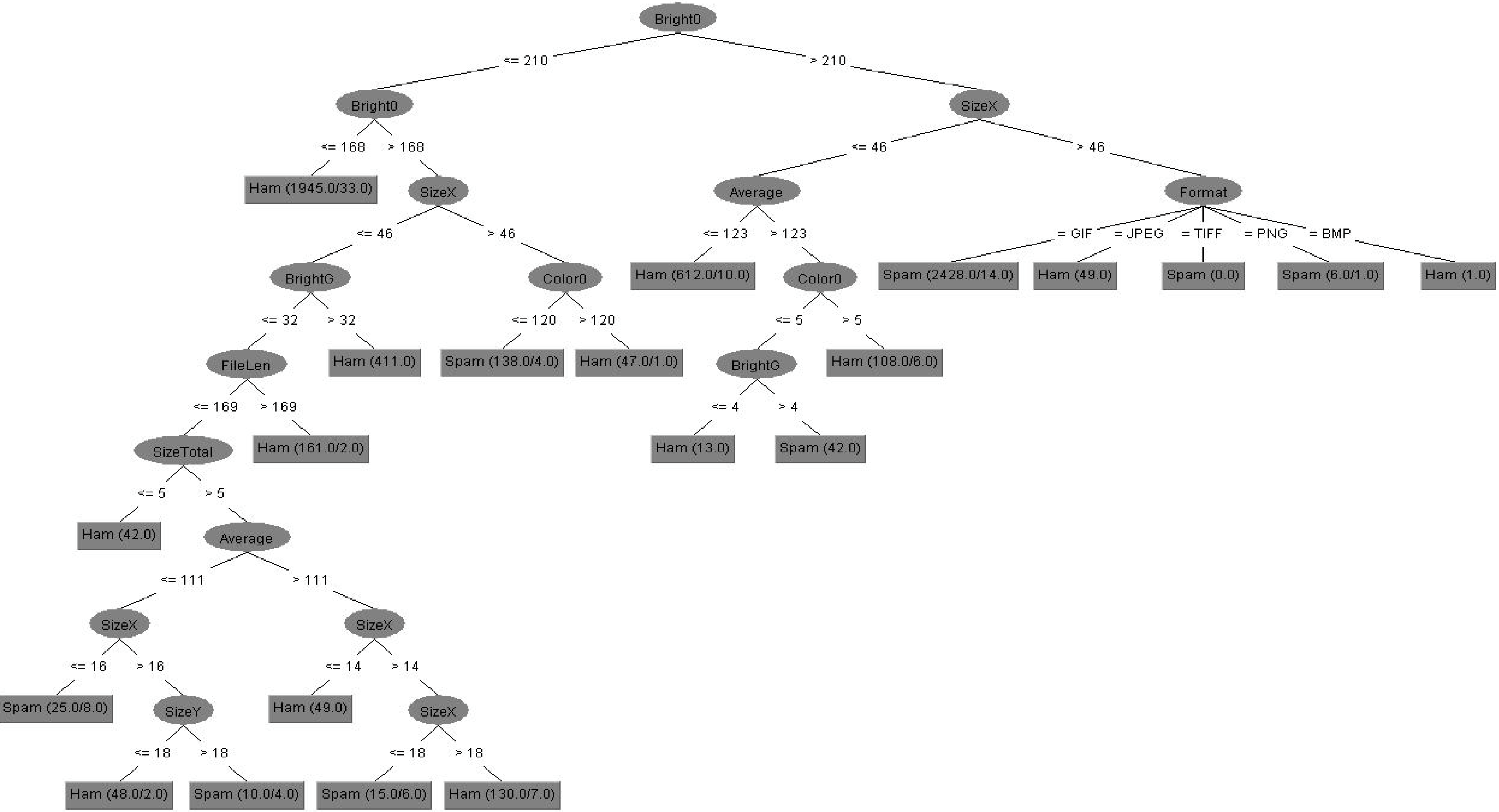
The *similarity hash key* is calculated from the attributes.

If the attributes of two images are close together, the hash keys are similar or equal.

⇓

It may recognize polimorphic spam

Decision Tree



Attribute Evaluation by Decision Trees

A node is considered to be a leaf if

- number of images is less than the size threshold, or
- Number of minorities are less than precision threshold.

Features:

- Every image belongs to exactly one leaf of the tree.
- The database contains only a few token, one for each of tree leaves

Decision Tree Method

Training phase

- Attributes are calculated for each images in the sample training collection.
- Builds a decision tree from this data.

Filter phase

- The attributes of examined image are also calculated
- The decision tree is used to decide whether it is a spam image or not

Case Studies

Test methods

- **Accurate hash method: MD5 (AHM)**
- **Similarity hash method from file attributes only (SHM)**
- **Decision tree method using 9 (both file- and image-) attributes (DTM)**
- **OCR and find spam-like words**

Case Study No. 1: Mixed Images

- Aim: to compare the capabilities of each image processing method itself
- Processed only images
  - Ham images (35,923)
  - Stable spam images (2,847)
  - Polymorphic spam images (24,035)
- Discard the text parts, headers etc.
- Bayesian spam database is not used

Case Study No. 2: Polymorphic Spams

- Aim: to compare spam filters use these methods in the most problematic spam type
- Test sample: families of polymorphic spams (16,578 mails)
- The whole mails (including non-image parts, either) were processed
- Full spam filter products were used
- Bayesian spam database is also used

Case Study No. 3: Wild Test

- Aim: to compare methods in a real set of emails
- Test sample: one hour traffic from a pay-free public e-mail server
- The whole mails (including non-image parts, either) were processed
- Full spam filter products were used
- Bayesian spam database is also used

Results

	Hams False positive (%)		Spams Detected (%)		
Case Filter	#1 Mixed	#3 Wild	#1 Mixed	#2 Poly	#3 Wild
AHM	0.00	0.24	10.03	31.98	94.06
SHM	3.03	0.25	82.45	50.77	95,11
DTM	2.96	0.26	91.25	97.56	99.02
OCR	8.03	8.91	97.11	99.20	97.28

Evaluation of AHM

- Advantages
  - ✓ Very low false positive rate (0 or close)
  - ✓ The highest speed
- Disatvantages
  - – Its detection rate is very low, especially on polymorphic spams
  - – Very big database

VirusBuster
www.virusbuster.hu

Evaluation of OCR

- Advantages
  - ✓ Usually very good detection rate
  - ✓ Can detect spam from an unknown family
  - ✓ Image database not needed at all.
- Disadvantages
  - – Very slow
  - – Cannot process images without text ⇒ Worse wild detection rate than DTM
  - – Very high false positive rate
  - – Easy to disturb

Evaluation of DTM

- Advantages
  - ✓ Usually very good detection rate
  - ✓ Low false positive rate (but higher than MD5 or SHM)
  - ✓ Acceptable performance
  - ✓ Uses only a few database tokens
- Disatvantages
  - – Detection of new spam familiy is not quite good (but better than SHM)

Conclusion

- The DTM can satisfy our original aims:
  - Very good detection rate
  - Quite low false positive rate
  - Performs acceptable running speed
- The AHM can help to avoid some false positive detection
  - White list of common ham images (e.g. smileys, trade logos etc.)

Questions?

Sandor Antal
santal@virusbuster.hu