Phishing signatures creation HOWTO

Török Edwin

September 16, 2006

1.0.1 Example

The following line:

*R http://www\.googl*e\.*(com|ro|it) www\.googl*e\.*com*

Means: R - this is a regex.

1.3 Regular expressions

POSIX regular expressions are supported, and you can consider that internally it is wrapped by ^, and $. In other words, this means that the regular expression has to

1.4 Flags

Flags are a binary OR of the following numbers:

HOST_SUFFICIENT 1

DOMAIN_SUFFICIENT 2

DO_REVERSE_LOOKUP 4

CHECK_REDIR 8

CHECK_SSL 16

CHECK_CLOAKING 32

CLEANUP_URL 64

CHECK_DOMAIN_REVERSE 128

CHECK_IMG_URL 256

DOMAINLIST_REQUIRED 512

The names of the constants are self-explanatory.

These constants are defined in libclamav/phishcheck.h, you can check there for the latest flags.

There is a default set of flags that are enabled, these are currently: (CLEANUP_URL|DOMAIN_SUFFICIENT|CHE
ssl checking is performed only for a tags currently.

You must decide for each line in the domainlist if you want to filter any flags (that is you don't want certain checks to be done), and then calculate the binary OR of those constants, and then convert it into a 3-digit hexnumber. For example you devide that domain_sufficient shouldn't be used for ebay.com, and you don't want to check images either, so you come up with this flag number: 2

2.1 Special characters

[the opening square bracket - it marks the beginning of a character class, see section 2.2

**** the backslash - escapes special characters, see section 2.3

^ the caret - matches the beginning of a line (not needed in clamav regexes, this is implied)

$ the dollar sign - matches the end of a line (not needed in clamav regexes, this is implied)

. the period or dot - matches *any* character

| the vertical bar or pipe symbol - matches either of the token on its left and right side, see section 2.4

? the question mark - matches optionally the left-side token, see section 2.5

***** the asterisk or star - matches 0 or more occurences of the left-side token, see section 2.5

+ the plus sign - matches 1 or more occurences of the left-side token, see section 2.5

(

3.3.2 Undetected phish mails

Using why.py doesn't help here unfortunately (it will say: clean), so all you can do is:

```
$clamscan/clamscan –phish-scan-alldomains undetected.eml
```

And see if the mail is detected, if yes, then you need to add an appropiate line to daily.pdb (see section 3.2 on the previous page).

If the mail is not detected, then try using:

```
$clamscan/clamscan –debug undetected.eml|less
```

Then see what urls are being checked, see if any of them is in a whitelist, see if all urls are detected, etc.

4 Hints and recomandations

5 Examples