
Subject: Learning spam

Posted by [InforMed Direct](#) on Sun, 25 Sep 2005 13:10:23 GMT

[View Forum Message](#) <> [Reply to Message](#)

We've just upgraded to the latest version having been running the older version for some time. Some nice new features in there but am I right in saying that there still isn't any way for users of Outlook/Exchange to automate the learn process for spam that gets through?

With the old version we had a real fudge of a mechanism whereby users moved the email into a shared mailbox. IT admin then periodically used Outlook Express to collect all the messages from this mailbox. A 3rd party tool (DBXTRACT) was then used to write the contents of the Outlook Express mailbox into EML files which could then be run through the SA-LEARN program. As I said, a real fudge.

I was kind of hoping that this had been improved in the new version.

BTW - has anyone else found a tool that works with Outlook that can export a folder in Outlook into single EML files? It's probably not that hard to write an Outlook macro to do something similar to DBXTRACT.

Cheers, Rob.

Subject: Re: Learning spam

Posted by [support](#) on Mon, 26 Sep 2005 13:27:44 GMT

[View Forum Message](#) <> [Reply to Message](#)

InforMed Direct wrote:

> We've just upgraded to the latest version having been running
> the older version for some time. Some nice new features in
> there but am I right in saying that there still isn't any way
> for users of Outlook/Exchange to automate the learn process for
> spam that gets through?

Did you already have a look at the IMAP2mbox tool in the NoSpamToday! contribution area ?

Please see

<http://www.byteplant.com/support/nospamtoday/contrib.html>

Subject: Re: Learning spam

Posted by [eastwood](#) on Thu, 29 Sep 2005 08:42:17 GMT

[View Forum Message](#) <> [Reply to Message](#)

if you use mozilla thunderbird instead of outlook express to download the shared mailbox, that stors the mail in mbox format.

It what I do anf then each month i simple copy the mbox folder to the server and run SA-LEARN on that

works really well and not too hard to do

Cheers

Subject: Re: Learning spam

Posted by [InforMed Direct](#) on Thu, 29 Sep 2005 15:25:20 GMT

[View Forum Message](#) <> [Reply to Message](#)

> if you use mozilla thunderbird instead of outlook express to
> download the shared mailbox, that stors the mail in mbox
> format.

Thanks - will check that out as well as it'll probably easier than the DBEXTRACT route.

Cheers, Rob.

Subject: Re: Learning spam

Posted by [Heidner](#) on Sun, 02 Oct 2005 17:15:53 GMT

[View Forum Message](#) <> [Reply to Message](#)

Imap2Mbox was written do do exactly what you've asked without using DBEXTRACT or mozilla.

Your admins setup a simple batch job that runs Imap2mbox at prescribed times. Imap2mbox connects to the exchange (or any other IMAP2 email server) and copies the content of a spam mail box to a spam assassin ready mbox file. The script then runs spamassassin in the learn mode.

You can configure imap2mbox to automatically empty the spam wastebasket when done. You can also create a "HAM" folder such that examples of good e-mails are sent to it and Imap2mbox exports them again so spamassassin can learn them as good e-mails.

If you are using exchange there is a link in the documentation for imap2mbox on how to create global public e-mail folders that anyone can drag their spam into.

For example:

```
rem learns an mbox mail folder as spam
rem
rem execute this batch in installation directory
```

```
rem
rem imap2mbox has been previously configured to read from the mail server and
rem export the emails into the mbox container called "spam.mbox"
rem
rem
rem A logfile called spam.log is created so we can monitor the activities of
rem  imap2mbox and spam-learn. Spam.log should be deleted periodically
rem
del spam.mbox
echo *===* >> spam.log
imap2mbox.exe --config=spam.cfg >> spam.log
if exist spam.mbox sa\salsa-learn -C sa\ruleset --showdots --spam --mbox spam.mbox >> spam.log
echo Begin to expire old tokens >> spam.log
sa\salsa-learn.exe -C sa\ruleset --showdots --force-expire >> spam.log
echo Begin syntax check of SA token database >> spam.log
sa\spamassassin.exe -x -C sa\ruleset --lint >> spam.log
echo IMAP-LEARN-SPAM complete >> spam.log
echo                >> spam.log
```

Will result in....

```
*===*
```

```
The server is: thecount.
Sun Sep 11 04:25:03 2005
2 E-mails to process / 2 - messages selected
Extracting mail from: Public Folders/SPAM WASTEBASKET is complete.
Mail written to mbox: spam.mbox
Messages were deleted after writing to mbox.
Learned from 2 message(s) (2 message(s) examined).
Begin to expire old tokens
Begin syntax check of SA token database
IMAP-LEARN-SPAM complete
ECHO is on.
```

Subject: Re: Learning spam
Posted by [kfoutts](#) on Mon, 28 Nov 2005 19:14:38 GMT
[View Forum Message](#) <> [Reply to Message](#)

Okay, got the import to mbox done, I've offered a couple hundred emails to the lean spam db but I'm still getting the same emails through, any ideas?

Post Edited (11-29-05 22:18)

Subject: Re: Learning spam

Posted by [support](#) on Fri, 02 Dec 2005 10:51:23 GMT

[View Forum Message](#) <> [Reply to Message](#)

Check the X-Spam headers of the mails that get through. You can see the Bayes spam probability in % by looking at the BAYES_XXX test result.

Subject: Re: Learning spam

Posted by [MX1](#) on Wed, 08 Feb 2006 19:38:58 GMT

[View Forum Message](#) <> [Reply to Message](#)

I tried to setup this solution, but I get a problem in using imap2mbox.

The log-file tells the following:

The server is: mailserver.

Wed Feb 8 20:19:59 2006

E-mails to process / 1 - messages selected

Extracting mail from: Public Folders/z_HAM is complete.

Mail written to mbox: Y:\(... full path ...)\ham.mbox

But: in the public exchange folder in my EX55 there a 2 ham mails, but only one was extracted.

When opening ham.mbox with text editor it only show the folowing content:

From

The public folders have such a letter:

=ffentliche Ordner

Second thing is:

There's a structure like

Favoriten

z_ham

z_spam

How to address folder z_ham or z_spam ?

Third prob:

when using parameter --delete=1 then I get a message: "Option delete does not take an

argument"; doesn't matter if I use 0 or 1

Thank you in advance for your help.

By the way: NT4 Server, EX5.5SP4, NST2.3.3.4

Subject: Re: Learning spam

Posted by [Heidner](#) on Thu, 09 Feb 2006 17:06:12 GMT

[View Forum Message](#) <> [Reply to Message](#)

Could you paste a copy of your IMAP2mbox config file into the thread so I could look at it.

But before sending remember to blank out the line that the "magic" number on it. For example my config file for spam looks like:

[Server]

Name=the-server

Port=143

[User]

Name=heidner_wa/dennis/heidner

Magic= ---- blanked out ----

[Public Box]

Path=Public Folders/

Folder=SPAM WASTEBASKET

[Output]

Mbox=spam.mbox

[Options]

Verbose=1

Delete=1

Timestamp=WedMar21553222005

[Timestamp]

heidner_wa is the MS NT domain name, dennis is the e-mail box and "heidner" is the alias of the e-mail box dennis. The path to the public folders is "Public Folders/" (the / is important) and the actual spam mailbox is "SPAM WASTEBASKET".

My config file for HAM looks like:

[Server]
Name=the-count
Port=143

[User]
Name=heidner_wa/dennis/heidner
Magic= --- blank out this line ----

[Public Box]
Path=Public Folders/
Folder=HAM

[Output]
Mbox=ham.mbox

[Options]
Verbose=1
Delete=1
Timestamp=WedMar21553222005

[Timestamp]

I use the following script to process the spam

```
rem learns an mbox mail folder as spam
rem
rem execute this batch in installation directory
rem
rem imap2mbox has been previously configured to read from the mail server and
rem export the emails into the mbox container called "spam.mbox"
rem
rem
rem A logfile called spam.log is created so we can monitor the activities of
rem  imap2mbox and spam-learn. Spam.log should be deleted periodically
rem
del spam.mbox

echo *===* >> spam.log

imap2mbox.exe --config=spam.cfg >> spam.log

if exist spam.mbox sa\salsa-learn -C sa\ruleset --showdots --spam --mbox spam.mbox >> spam.log

echo Begin to expire old tokens >> spam.log
```

```
sa\sa-learn.exe -C sa\ruleset --showdots --force-expire >> spam.log  
echo Begin syntax check of SA token database >> spam.log  
sa\spamassassin.exe -x -C sa\ruleset --lint >> spam.log  
echo IMAP-LEARN-SPAM complete >> spam.log  
echo >> spam.log
```

It generates the log entries that look like:

```
*===*  
The server is: the-count.  
Fri Jan 20 04:25:06 2006  
1 E-mails to process / 1 - messages selected  
Extracting mail from: Public Folders/SPAM WASTEBASKET is complete.  
Mail written to mbox: spam.mbox  
Messages were deleted after writing to mbox.  
Learned tokens from 1 message(s) (1 message(s) examined)  
Begin to expire old tokens  
expired old bayes database entries in 108 seconds  
184254 entries kept, 1541 deleted  
token frequency: 1-occurrence tokens: 64.87%  
token frequency: less than 8 occurrences: 20.96%  
Begin syntax check of SA token database  
IMAP-LEARN-SPAM complete  
ECHO is on.  
*===*
```

NOTE NOTE!!!!

Notice when viewing the path of the spam wastebasket in outlook -- it would appear that you must configure IMAP2MBOX with the path of "Path=Public Folders/All Public Folders/" .. this is incorrect.. the actual path that exchange wants is "Path=Public Folders/". Microsoft for some reason has confused the issue by adding in the "All Public Folders" subcategory when viewing from outlook.

Before posting your config file... make sure you blank out the "magic" line.

It would help if you can include copies of an output from learning and the script you might be using to run IMAP2mbox (as I've done above). But remember to delete or clear out the magic line...

Subject: Re: Learning spam
Posted by [Heidner](#) on Thu, 09 Feb 2006 17:11:45 GMT
[View Forum Message](#) <> [Reply to Message](#)

Your path setting should look something like:

The ham folder would be:

Folder=z_ham

and Spam would be

Folder=z_spam

Subject: Re: Learning spam
Posted by [MX1](#) on Thu, 09 Feb 2006 22:48:18 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thanks for the support. I double checked everything, but I have no idea what is running wrong.
Pls find the files as follows:

1. creating the spam.cfg

I call the following command to first create a config-file:

=====

```
--server=192.168.100.100 --username="GUELICH/Administrator/admin" --mbox="spam.mbox"  
--pass=*** >> spam.log
```

=====

Then I get a config-file as follows:

=====

```
[Server]  
Name=192.168.100.100  
Port=143
```

```
[User]  
Name=GUELICH/Administrator/admin  
Magic=***
```

```
[Public Box]
```

```
Folder=z_SPAM
```

```
[Output]  
Mbox=spam.mbox
```


[Options]
Verbose=1
Delete=0
Timestamp=

=====
Then I start the following file learn_spam.bat (similar to yours)
=====

```
del spam.mbox
echo *===* >> spam.log
imap2mbox.exe --config=spam.cfg >> spam.log
if exist spam.mbox sa\salsa-learn -C sa\ruleset --showdots --spam --mbox spam.mbox >> spam.log
```

```
echo Begin to expire old tokens >> spam.log
sa\salsa-learn.exe -C sa\ruleset --showdots --force-expire >> spam.log
```

```
echo Begin syntax check of SA token database >> spam.log
sa\spamassassin.exe -x -C sa\ruleset --lint >> spam.log
```

```
echo IMAP-LEARN-SPAM complete >> spam.log
echo. >> spam.log
```

=====
This creates a logfile like that:
=====

```
*===*
The server is: 192.168.100.100.
Thu Feb 9 23:29:37 2006
E-mails to process / 1 - messages selected
```

```
Mail written to mbox: spam.mbox
Learned tokens from 1 message(s) (1 message(s) examined)
Begin to expire old tokens
Begin syntax check of SA token database
IMAP-LEARN-SPAM complete
=====
```

But there are not only 1 email in the folder z_spam but 48, just for testing purpose.

In my opinion the mistake is in the german translation of Public Folders. In my config file, it is
is displayed as

Are there other people having experience with a german EX55?

My second problem seems to be solved after havin read your above message. I did not make test

about the "delete" parameter, this will follow later on when the most important things are running.

Hoping this posting will help you.

Thanks for your help.

Subject: Re: Learning spam

Posted by [Heidner](#) on Fri, 10 Feb 2006 04:38:17 GMT

[View Forum Message](#) <> [Reply to Message](#)

IMAP2MBOX is a perl application. I used ActiveStates perl SDK for the program. I will check to see if their libraries have a problem with international languages. And if so what steps I must take to fix the problem.

Meanwhile could you run IMAP2mbox with the "--debug" and "--verbose" switches, then copy the output -- removing any sensitive information (like passwords) and paste into thread or e-mail me...

my e-mail is dennis "at" heidners.net replacing "at" with @

Subject: Re: Learning spam

Posted by [Heidner](#) on Fri, 10 Feb 2006 04:47:04 GMT

[View Forum Message](#) <> [Reply to Message](#)

One other possibility... can you try some other user beside "Administrator" to open and read the public mailbox?

The reason being is that with NT4.0 and Exchange 5.5, strange things happen when you try to access the administrators mailbox.

(Hint - don't try adding Outlook onto the exchange server... it may break!).

The reason is that MS creates some special e-mail accounts that are used to manage the information store, routing, etc... and you may be getting caught by that. Several years ago I had a rather nasty experience as a result of trying to change the e-mail profile away from the Exchange defined account to another one... on the exchange server. lots of stuff broke until I returned the settings...

Subject: Re: Learning spam

Posted by [Frostie](#) on Fri, 10 Feb 2006 09:15:10 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hi MX1

I had the same problem (Ex 2000 German), but was able to solve it.

I configured Imap2Mbox so that I can access my normal Inbox.
After that I started Imap2Mbox with the parameter --debug, which gave me, beside other information, a folder list.

In that list the public folders look like:

```
"&ANY-ffentliche Ordner/"
&ANY-ffentliche Ordner/QS-Kundenkontakte
&ANY-ffentliche Ordner/HWM_intern
&ANY-ffentliche Ordner/Kunden
&ANY-ffentliche Ordner/Lieferanten
&ANY-ffentliche Ordner/Vertriebsbesprechungsraum
&ANY-ffentliche Ordner/Termine Gesch&AOQ-fts&APw-hrung
&ANY-ffentliche Ordner/Kaufm&AOQ-nnische Leitung Aufgaben
&ANY-ffentliche Ordner/Kaufm&AOQ-nnische Leitung Kalender
&ANY-ffentliche Ordner/Spam-Mails
&ANY-ffentliche Ordner/Internet Newsgroups
```

I simply copied the name for the public folders from that list in my config file and...voila...it works.

Hope it works for you

Marc

Subject: Re: Learning spam
Posted by [MX1](#) on Fri, 10 Feb 2006 15:15:48 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello Dennis,
hello Marc,

problem is solved, it is running perfectly now.

It was the problem of the spelling. After using Marc's trick it was running like it never has done anything else before. Great software. Thanks (for software as well as for the support).

Michael

Subject: Re: Learning spam
Posted by [Heidner](#) on Fri, 10 Feb 2006 16:43:06 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thanks, Marc and Michael

I will add the suggestion of using debug to look at the current folder names to the document file.

I am pleased to hear it is working for both of you.

Subject: Re: Learning spam
Posted by [MX1](#) on Fri, 10 Feb 2006 18:24:19 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dennis, one other question: is it possible to start 2 tasks of imap2mbox at the same time on the same machine, eg. 4:00 a.m. learningham.bat and learningspam.bat or should I wait, e.g. 1/2 hour until I start the next process?

Subject: Re: Learning spam
Posted by [Heidner](#) on Sat, 11 Feb 2006 06:33:25 GMT
[View Forum Message](#) <> [Reply to Message](#)

You should be able to run multiple copies of IMAP2MBOX at the same time. Especially since the config files would be different and the resulting ham/spam mbox would be unique.

You should change the scripts so that you export the ham, export the spam. Then in a new batch file learn spam, then learn ham and force the expire.

It might also work just to merge the scripts into one batch file. export ham and spam then immediately learn both of them. The process should go pretty quick.

Subject: Re: Learning spam
Posted by [MX1](#) on Sat, 11 Feb 2006 08:02:48 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you, I did not think about the easiest way to bring both batch files together.

Subject: Re: Learning spam
Posted by [Anonymous](#) on Thu, 18 Jan 2007 20:27:06 GMT
[View Forum Message](#) <> [Reply to Message](#)

When I run imap2mbox I get an error "Invalid string supplied for user name" when trying to logon to the server. I've tried every combination of username I can think of. I am putting in domain\username\alias (which are the same by the way) tusers9 and tusers9 for both username and alias. Any suggestions to get this to work?
