
Subject: Advanced Statistics

Posted by [Jon](#) on Tue, 01 Nov 2005 11:51:29 GMT

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

NST! is doing a fantastic job of filtering and killing our (previously) awful spam problem, and I'd like to monitor it's performance closely if possible.

What I would like is a computer-readable version of the Spam Filtering Report that is mailed daily, so that I can write a script to import those totals and produce a graph and various other stats.

Is this possible?

I have tried the SpamLogs utility, but I cannot get it to produce anything.

Thanks,
Jon

Subject: Re: Advanced Statistics

Posted by [support](#) on Wed, 02 Nov 2005 10:47:53 GMT

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

You should find all the info you want (and more) in the nospamtoday_Statistics.csv file.

You can import this file into MS Excel, or OpenOffice Calc, or you can view it with a text editor.

Subject: Re: Advanced Statistics

Posted by [Jon](#) on Wed, 02 Nov 2005 16:34:13 GMT

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

Ok, great - that seems to be perfect.

What is the "behaviour" of this file? If it is deleted, does it get recreated the next time it is written (I'm guessing it does). How does the .csv~ file factor into this?

What I'm thinking is, each night I'll have a script that runs at midnight. If that script were to read the .csv file and do "stuff" with the data for the previous day; then delete the file, how would that impact the running of NST?

I don't need to delete the file, but it would save some processing if there file were not full of x months worth of data, when I only ever want to read the last entry.

Thanks,
Jon

Subject: Re: Advanced Statistics
Posted by [support](#) on Thu, 03 Nov 2005 10:50:03 GMT
[View Forum Message](#) <> [Reply to Message](#)

> What is the "behaviour" of this file? If it is deleted, does
> it get recreated the next time it is written (I'm guessing it
> does). How does the .csv~ file factor into this?

The .csv file is written on shutdown, and once at midnight. A new record is added for the day, a new line at the end of the file.

The .csv~ file is always the backup of yesterday's file.

>
> What I'm thinking is, each night I'll have a script that runs
> at midnight. If that script were to read the .csv file and do
> "stuff" with the data for the previous day; then delete the
> file, how would that impact the running of NST?

Only the statistics display in the admin client would no longer show anything.

> I don't need to delete the file, but it would save some
> processing if there file were not full of x months worth of
> data, when I only ever want to read the last entry.

Subject: Re: Advanced Statistics
Posted by [Jon](#) on Thu, 03 Nov 2005 10:56:50 GMT
[View Forum Message](#) <> [Reply to Message](#)

Ok, so I will not delete the file, so the Statistics in the Admin Client are unaffected... I'll simply read the whole file and find the entry I need (i.e. the last). Having RTFM after posting the queries above, I found the description of the usage of the file, and I can see it will not grow beyond a month of stats, so it'll never get huge.

Yesterday, I did rename the .csv file to see what would happen, and one confusing difference was the number of columns that were recorded in the newly-created file... it had far, far fewer columns than the saved one. The old one had lots of "SpamAssassin_XYZ" columns, while the new one has only one: "SpamAssassin_AWL". This isn't a huge problem, as I am not interested in those really 'granular' stats, but I did wonder how I should handle changes in the data?

Thanks for all your help so far.

Jon

Subject: Re: Advanced Statistics

Posted by [support](#) on Tue, 08 Nov 2005 10:39:45 GMT

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

> Yesterday, I did rename the .csv file to see what would happen,
> and one confusing difference was the number of columns that
> were recorded in the newly-created file... it had far, far
> fewer cols than the saved one. The old one had lots of
> "SpamAssassin_XYZ" columns, while the new one has only one:
> "SpamAssassin_AWL". This isn't a huge problem, as I am not
> interested in those really 'granular' stats, but I did wonder
> how I should handle changes in the data?

As SpamAssassin rules come and go, the number of statistics file columns change. But most of the columns' names, such as the mail counts, won't change (unless you change port or filter names in your configuration).

So if you reference the columns by name instead of by position, you should be fine.

Subject: Re: Advanced Statistics

Posted by [eastwood](#) on Mon, 14 Nov 2005 13:28:58 GMT

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

looking at the csv file the first column says its the date but the actual data in my csv file looks like this:

```
38313;  
38314;  
38315;  
38316;  
38317;  
38318;  
38319;  
38320;
```

how is that the date?

Sorry if I am being thick?

Subject: Re: Advanced Statistics

Posted by [Jon](#) on Mon, 14 Nov 2005 13:40:57 GMT

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

eastwood,

It is a date stored as "Number of days since 01/01/1900", and it's caused me a few headaches. To be clear though, this is a pretty common way to store dates. You will likely find that anything you are trying to analyse the data with will handle this type of date storage - Excel is quite happy with it, fyi... you just need to format the column as a date instead of a number.

The problem I've had is getting PHP to handle it, as it wants to work with "...days since 01/01/1970" - again, a very common C format for storing dates.

I've not fully debugged this yet, but I'm currently subtracting 25569 from the NST date to try and convert it into the more popular "...1970" format. That number might be wrong. $70 * 365 = 25550$; $70 * 365.25 = 25567.5$. I manually tweaked that "magic number" to come up with 25569, which seems to work (currently). I'm not sure why it is subtly different, but I suspect leap years come into it somehow.

HTH,
Jon

Subject: Re: Advanced Statistics
Posted by [eastwood](#) on Wed, 16 Nov 2005 09:28:05 GMT
[View Forum Message](#) <> [Reply to Message](#)

sounds like fun...

When I get a spare minute I am going to try to use ASP and mysql to get it in to a db to make life a little easier and like you provide stats the customers and display it on my web site.

Subject: Re: Advanced Statistics
Posted by [Heidner](#) on Sat, 19 Nov 2005 18:11:25 GMT
[View Forum Message](#) <> [Reply to Message](#)

Jon, did you ever get spamlogs working?

It wasn't intended to provide statistics - but instead fold the many lines from the extended logging into one line that you could easily be viewed or sent to users -- so they would know what mail was deleted or rejected.

It pretty much requires that you enable all the extended logging options with NST.

Subject: Re: Advanced Statistics
Posted by [Jon](#) on Mon, 21 Nov 2005 19:54:55 GMT
[View Forum Message](#) <> [Reply to Message](#)

I've done a simple PHP script that runs every daily at midnight. This reads the entire .csv log and builds a graph based on the data, using OWTChart. The graph contains as much data as there is

in the CSV (i.e. the past month).

Here's an example, covering just a couple of weeks (since I cleared the log - see post above)

http://southdown.co.uk/nst_stats.jpg

There's no legend on the image, but basically the green areas are mail that made it to the mail server (i.e. passed the NST tests, but including those marked as spam in their subject); red is the flat-out-denied spam (that exceeds the threshold) and pink are the mails denied by AV scan, illegal attachments and MIME failures.

Jon

Subject: Re: Advanced Statistics
Posted by [Heidner](#) on Tue, 22 Nov 2005 04:52:00 GMT
[View Forum Message](#) <> [Reply to Message](#)

Had you given any thought of contributing the php script to the NST user library?

Subject: Re: Advanced Statistics
Posted by [eastwood](#) on Sun, 18 Dec 2005 11:08:51 GMT
[View Forum Message](#) <> [Reply to Message](#)

Jon.

Managed to now get my stats working with ASP and pie charts. I am working on reproducing the details from the email for each day further.

Sample picture can be seen at: <http://www.spiralsites.com/images/spam.gif>

I will keep you posted as I do more.
