



virus

BULLETIN

Covering the global threat landscape

VBSPAM COMPARATIVE REVIEW SEPTEMBER 2015

INTRODUCTION

A few weeks ago, a journalist contacted me to check whether an email sent to Hillary Clinton's personal email account in 2011 could have been a Russian spear-phishing attempt that used a New York parking ticket as a lure, as had been reported by some media sources¹. It was actually pretty clear that this was no targeted attack against Ms Clinton, but that her email address, like billions of others, had ended up on spammers' lists. (Gary Warner has details of this particular campaign on his blog².)

While spear-phishing is a serious problem for many organizations, the fact that a target as prominent as the US Secretary of State received an email in a run-of-the-mill spam campaign shows that spam remains the background noise of the Internet. And though it is neither the most advanced nor the most exciting kind of cybercrime, its impact remains huge, both for systems administrators and for the millions of users who accidentally open email attachments every day.

Thankfully, as our VBSpam tests have regularly shown, there are many anti-spam solutions that help to mitigate the problem of spam at the network level. In this test, all 13 full solutions reached the required standard to earn a VBSpam award, and five of them achieved a VBSpam+ award.

Moreover, for the first time this month's tests gave some insight into how long products took to filter emails.

THE TEST SET-UP

The VBSpam test methodology can be found at <http://www.virusbtn.com/vbspam/methodology/>. As usual,

¹ <http://www.nytimes.com/2015/10/02/us/politics/malware-on-hillary-clinton-server-prompts-look-at-suspected-russian-hacking.html>.

² <http://garwarner.blogspot.com/2011/08/new-york-city-uniform-traffic-ticket.html>.

emails were sent to the products in parallel and in real time, and products were given the option to block email pre-DATA (that is, based on the SMTP envelope and before the actual email was sent). However, on this occasion no products chose to make use of this option.

For those products running on our equipment, we use *Dell PowerEdge* machines. As different products have different hardware requirements – not to mention those running on their own hardware, or those running in the cloud – there is little point comparing the memory, processing power or hardware the products were provided with; we followed the developers' requirements and note that the amount of email we receive is representative of that received by a small organization.

To compare the products, we calculate a 'final score', which is defined as the spam catch (SC) rate minus five times the weighted false positive (WFP) rate. The WFP rate is defined as the false positive rate of the ham and newsletter corpora taken together, with emails from the latter corpus having a weight of 0.2:

$$\text{WFP rate} = (\# \text{false positives} + 0.2 * \min(\# \text{newsletter false positives}, 0.2 * \# \text{newsletters})) / (\# \text{ham} + 0.2 * \# \text{newsletters})$$

$$\text{Final score} = \text{SC} - (5 * \text{WFP})$$

As of this test, there are some additional criteria for the VBSpam award, based on the speed of delivery of emails in the ham corpus. To qualify for a VBSpam award, not only must the value of the final score be at least 98 but in addition, the speeds of delivery at 10 and 50 per cent must be classified as 'green' or 'yellow', and at 95 per cent they must be classified as 'green', 'yellow' or 'orange'.

To earn a VBSpam+ award, products must combine a spam catch rate of 99.5% or higher with a lack of false positives and no more than 2.5% false positives among the newsletters – in addition, the speeds of delivery at 10 and

50 per cent must be classified as ‘green’, and at 95 and 98 per cent they must be classified as ‘green’ or ‘yellow’. The colour-coded speed classifications are explained later in this review.

THE EMAIL CORPUS

The test started on Saturday 15 August at 12am and finished 16 days later, on Monday 31 August at 12am. There was a small problem which lasted a little over an hour in the early evening of 19 August (emails sent during this period were not included in the corpus), but it didn’t affect the rest of the test.

The test corpus consisted of 139,292 emails. 129,208 of these emails were spam, 64,790 of which were provided by *Project Honey Pot*, with the remaining 64,418 spam emails provided by *spamfeed.me*, a product from *Abusix*. They were all relayed to the products in real time, as were the 9,773 legitimate emails (‘ham’) and 311 newsletters.

Figure 1 shows the catch rate of all full solutions throughout the test. To avoid the average being skewed by poorly performing products, the highest and lowest catch rates have been excluded for each hour. We can see that

the high catch rates that we observed in July continued in this test.

SPEED

SMTP, which defines the delivery of email, is a store-and-forward protocol. Alice can send an email to Bob without Bob being online at that moment. Even when the email travels over the public Internet from Alice’s mail server to that of Bob, it’s not a problem if the latter is temporarily offline: a new delivery attempt is made a little while later.

Still, most people expect email to be delivered immediately. Indeed, a spam filter that decided to delay all emails by an hour – something which could allow it to respond better to new spam campaigns and thus improve its performance – would soon find itself losing customers.

In the VBSpam tests, we care about how email works and about how spam filters interfere in practice, not within the theoretical boundaries of SMTP. Hence we have always kept an eye on whether products artificially improve their performance by holding onto emails for a significant amount of time. While we have never had to ask products to change their behaviour, we thought it important to be open

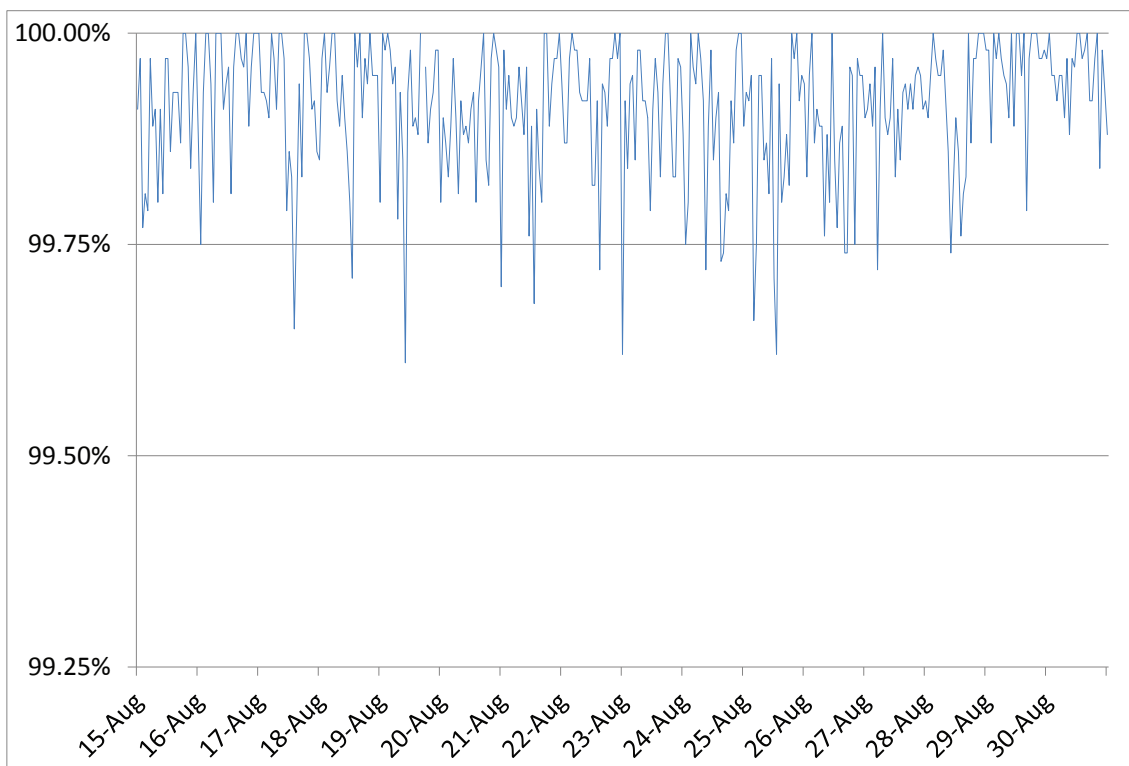


Figure 1: Spam catch rate of all full solutions throughout the test period.

about by how long products delay the delivery of emails. As of this test, we will publish this data.

For this measure, we only look at the emails from the ham corpus. These are the emails that people expect to receive immediately (no one – other than the spammer – cares how long it takes to deliver a spam email, if it gets delivered at all). Moreover, some products use a variant of greylisting, in which the delivery of certain emails is delayed until some time has passed³, hoping to be able to block the latest spam campaigns in this way. If the selection of emails to be delayed is done well, with few legitimate emails, this is a valid method that shouldn't affect how users experience a product.

When looking at speed, we ignored false positives.

For each email from the ham corpus, we measured the time (in seconds) it took for a full email to be delivered back to us after our server had made a connection to the product. However, because a few seconds don't make a noticeable difference⁴ and there are many external factors (including the hardware on which the product runs) that influence the delivery time, we have decided not to publish the *actual* time it takes for products to deliver emails.

Instead, for each product we sort the emails by the time it took for them to be delivered back to us and then look at the emails at the 10, 50, 95 and 98 percentiles. For these emails, we use colour-coded categories to report the length of delay: 0–30 seconds (green); 30 seconds to two minutes (yellow); two minutes to 10 minutes (orange); and more than 10 minutes (red). For user experience, these values could be interpreted as near real-time delivery, a small delay, a significant delay, and a huge delay.

As mentioned earlier in the report, we also set some minimum values for this 'speed' in order for products to qualify for a VBSpam or VBSpam+ award. It was good to see that, based on their speed alone, all products qualified for a VBSpam award, and all but one would have qualified for a VBSpam+ award.

RESULTS

In the text that follows, unless otherwise specified, 'ham' or 'legitimate email' refers to email in the ham corpus – which excludes the newsletters – and a 'false positive' refers to a message in that corpus that has erroneously been marked by a product as spam.

³ Greylisting normally refers to a situation in which a 4xx 'temporary error' is sent to the sending mail server, under the often correct assumption that spammers haven't set up their servers to re-attempt delivery. This kind of greylisting cannot be tested in the VBSpam set-up.

⁴ Email isn't suitable for situations where seconds can make a difference, including high-frequency trading.

Axway MailGate 5.3.1

SC rate: 99.76%

FP rate: 0.03%

Final score: 99.49

Project Honey Pot SC rate: 99.55%

Abusix SC rate: 99.98%

Newsletters FP rate: 3.9%

Speed: 10%: ●; 50%: ●; 95%: ●; 98%: ●

Axway's MailGate virtual appliance was among the majority of products that returned at least 98% of emails within 30 seconds – which its customers and their users will no doubt appreciate. Customers will also be pleased with the catch rate of more than 99.75%, only marginally lower than that recorded in July, and just three false positives. With this performance, *Axway* achieves its tenth VBSpam award.



Bitdefender Security for Mail Servers 3.1.2

SC rate: 99.98%

FP rate: 0.00%

Final score: 99.98

Project Honey Pot SC rate: 99.98%

Abusix SC rate: 99.98%

Newsletters FP rate: 0.0%

Speed: 10%: ●; 50%: ●; 95%: ●; 98%: ●

Bitdefender has achieved a VBSpam award in every one of the last 38 VBSpam tests, and this test proves that nothing funny is happening behind the scenes: at least 98 per cent of emails are returned within half a minute, thus creating a near real-time experience for the product's users.

Even more importantly, and certainly more impressively, the product blocked 99.98% of spam – missing just 26 spam emails – and did so without a single false positive in either corpus of legitimate emails. *Bitdefender's* 39th VBSpam award therefore comes with a big plus attached.



Egedian Mail Security

SC rate: 99.95%

FP rate: 0.02%

Final score: 99.84

Project Honey Pot SC rate: 99.91%

Abusix SC rate: 99.99%

Newsletters FP rate: 0.3%

Speed: 10%: ●; 50%: ●; 95%: ●; 98%: ●

Egedian's users may see a small percentage of email take slightly longer to arrive, but with 98 per cent arriving within



	True negatives	False positives	FP rate	False negatives	True positives	SC rate	Final score
Axway	9770	3	0.03%	309	128899	99.76%	99.49
Bitdefender	9773	0	0.00%	26	129182	99.98%	99.98
Egedian	9771	2	0.02%	65	129143	99.95%	99.84
FortiMail	9772	1	0.01%	15	129193	99.99%	99.93
GFI	9772	1	0.01%	723	128485	99.44%	99.38
IBM	9773	0	0.00%	165	129043	99.87%	99.87
Kaspersky LMS	9771	2	0.02%	88	129120	99.93%	99.83
Libra Esva	9773	0	0.00%	24	129184	99.98%	99.98
McAfee SaaS	9772	1	0.01%	67	129141	99.95%	99.84
OnlyMyEmail	9773	0	0.00%	1	129207	99.999%	99.98
Scrollout	9761	12	0.12%	762	128446	99.41%	98.77
Sophos	9771	2	0.02%	289	128919	99.78%	99.67
SpamTitan	9773	0	0.00%	47	129161	99.96%	99.95
Spamhaus DBL*	9773	0	0.00%	75570	53638	41.51%	41.51
Spamhaus ZEN*	9773	0	0.00%	10172	119036	92.13%	92.13
Spamhaus ZEN+DBL*	9773	0	0.00%	6235	122973	95.17%	95.17

*The Spamhaus products are partial solutions and their performance should not be compared with that of other products. (Please refer to the text for full product names and details.)

a little over half a minute, that’s not going to have a major impact on user experience.

The main focus of these tests remains the products’ performance on the various corpora, and I was pleased to see the catch rate of *Egedian Mail Security* increase to 99.95%. This time, the product missed two legitimate emails, which were both from the same Russian sender, and thus the virtual solution came very close to winning its first VBSpam+ award. A regular VBSpam award was easily within its grasp though.

Fortinet FortiMail

SC rate: 99.99%
FP rate: 0.01%
Final score: 99.93
Project Honey Pot SC rate: 99.98%
Abusix SC rate: 100.00%
Newsletters FP rate: 0.3%
Speed: 10%: ●; 50%: ●; 95%: ●; 98%: ●



A move to a new (virtual) network means that we may soon have to replace *Fortinet’s FortiMail* appliance, which has been in our tests since July 2009 and which has never missed a VBSpam award, with a virtual version of the product. For

now, this old appliance continues to perform very well – in fact, the spam catch rate of 99.99% (a mere 15 spam emails were missed) is *FortiMail’s* highest performance in 38 tests.

Unfortunately, there was a single false positive, which prevented the product from achieving a VBSpam+ award, but with the fifth highest final score, and an excellent speed too, *Fortinet* has many reasons to be pleased.

GFI MailEssentials

SC rate: 99.44%
FP rate: 0.01%
Final score: 99.38
Project Honey Pot SC rate: 99.48%
Abusix SC rate: 99.40%
Newsletters FP rate: 0.3%
Speed: 10%: ●; 50%: ●; 95%: ●; 98%: ●



End-users whose emails are protected by *GFI MailEssentials* don’t have to wait long for their emails to arrive: the product filtered at least 98 per cent of emails in under 30 seconds.

This month, users would have seen a few more spam emails arrive in their inbox though, as the catch rate dropped to just

	Newsletters		Project Honey Pot		Abusix		STDev [†]	Speed			
	False positives	FP rate	False negatives	SC rate	False negatives	SC rate		10%	50%	95%	98%
Axway	12	3.9%	293	99.55%	16	99.98%	0.56	●	●	●	●
Bitdefender	0	0.0%	14	99.98%	12	99.98%	0.12	●	●	●	●
Egedian	1	0.3%	56	99.91%	9	99.99%	0.14	●	●	●	●
FortiMail	1	0.3%	15	99.98%	0	100.00%	0.07	●	●	●	●
GFI	1	0.3%	339	99.48%	384	99.40%	0.48	●	●	●	●
IBM	0	0.0%	150	99.77%	15	99.98%	0.23	●	●	●	●
Kaspersky LMS	0	0.0%	69	99.89%	19	99.97%	0.16	●	●	●	●
Libra Esva	0	0.0%	24	99.96%	0	100.00%	0.10	●	●	●	●
McAfee SaaS	6	1.9%	66	99.90%	1	100.00%	0.15	●	●	●	●
OnlyMyEmail	2	0.6%	1	99.999%	0	100.00%	0.01	●	●	●	●
Scrollout	3	1.0%	456	99.30%	306	99.52%	1.00	●	●	●	●
Sophos	0	0.0%	192	99.70%	97	99.85%	0.34	●	●	●	●
SpamTitan	1	0.3%	44	99.93%	3	100.00%	0.13	●	●	●	●
Spamhaus DBL*	0	0.0%	21933	66.15%	53637	16.74%	10.58	N/A	N/A	N/A	N/A
Spamhaus ZEN*	0	0.0%	8799	86.42%	1373	97.87%	3.31	N/A	N/A	N/A	N/A
Spamhaus ZEN+DBL*	0	0.0%	5001	92.28%	1234	98.08%	2.25	N/A	N/A	N/A	N/A

*The Spamhaus products are partial solutions and their performance should not be compared with that of other products.

†The standard deviation of a product is calculated using the set of its hourly spam catch rates.

● 0–30 seconds; ● 30 seconds to two minutes; ● two minutes to 10 minutes; ● more than 10 minutes.

(Please refer to the text for full product names.)

below 99.5%. That is not a serious problem, and with low false positive rates, GFI easily earns another VBSpam award.

IBM Lotus Protector for Mail Security

SC rate: 99.87%

FP rate: 0.00%

Final score: 99.87

Project Honey Pot SC rate: 99.77%

Abusix SC rate: 99.98%

Newsletters FP rate: 0.0%

Speed: 10%: ●; 50%: ●; 95%: ●; 98%: ●

IBM's Lotus Protector for Mail Security has performed well in the past year, earning three VBSpam+ awards. I am pleased that we are now able to show that this performance wasn't achieved by holding onto the emails for an



unreasonable amount of time: IBM's 'speed colours' were all green.

In this test it also equalled the spam catch rate we recorded in the last test and combined this with a clean sheet in the ham and newsletter corpora. The product's fourth VBSpam+ award is thus well deserved.

Kaspersky Linux Mail Security 8.0

SC rate: 99.93%

FP rate: 0.02%

Final score: 99.83

Project Honey Pot SC rate: 99.89%

Abusix SC rate: 99.97%

Newsletters FP rate: 0.0%

Speed: 10%: ●; 50%: ●; 95%: ●; 98%: ●



Hosted solutions	Anti-malware	IPv6	DKIM	SPF	DMARC	Multiple MX-records	Multiple locations
McAfee SaaS	McAfee	√	√	√		√	√
OnlyMyEmail	Proprietary (optional)		√	√	*	√	√

* OnlyMyEmail verifies DMARC status but doesn't provide feedback at the moment.

(Please refer to the text for full product names.)

Local solutions	Anti-malware	IPv6	DKIM	SPF	DMARC	Interface			
						CLI	GUI	Web GUI	API
Axway MailGate	Kaspersky, McAfee	√	√	√				√	
Bitdefender	Bitdefender	√				√		√	√
Egedian	Bitdefender; ClamAV	√				√		√	√
FortiMail	Fortinet	√	√	√		√		√	
GFI	Five anti-virus engines	√		√				√	
IBM	Sophos; IBM Remote Malware Detection			√		√		√	
Kaspersky LMS	Kaspersky	√		√		√		√	
Libra Esva	ClamAV; others optional		√	√		√		√	
Scrollout	ClamAV			√		√		√	
Sophos	Sophos		√	√				√	
SpamTitan	Kaspersky; ClamAV	√	√	√		√		√	√

(Please refer to the text for full product names.)

For *Kaspersky Lab*, the introduction of our speed ‘traffic lights’ did show that a minority of emails are returned with a noticeable delay. Whether this is something organizations will consider a problem depends on their particular needs. In this test, it would have prevented the company from achieving a VBSpam+ award – had two false positives not already done so.

However, another VBSpam award is still easily earned, especially considering the product’s high spam catch rate of 99.93%.

Libra Esva 3.5.1.0

SC rate: 99.98%
 FP rate: 0.00%
 Final score: 99.98
 Project Honey Pot SC rate: 99.96%
 Abusix SC rate: 100.00%
 Newsletters FP rate: 0.0%
 Speed: 10%: ●; 50%: ●; 95%: ●; 98%: ●



Libra Esva will no doubt be pleased that we can show that the product doesn’t hold onto emails to improve its performance: at least 98% of them were returned within 30 seconds. And when the emails did return, the product was almost always correct in its classification: only 24 spam emails were missed and no legitimate emails were misclassified. This resulted in a clean sheet, the highest final score of all products, and another VBSpam+ award for the product’s developers in Italy.

McAfee SaaS Email Protection

SC rate: 99.95%
 FP rate: 0.01%
 Final score: 99.84
 Project Honey Pot SC rate: 99.90%
 Abusix SC rate: 100.00%
 Newsletters FP rate: 1.9%
 Speed: 10%: ●; 50%: ●; 95%: ●; 98%: ●



It was nice to see *McAfee’s SaaS Email Protection* return

Products ranked by final score	
Libra Esva	99.98
Bitdefender	99.98
OnlyMyEmail	99.98
SpamTitan	99.95
FortiMail	99.93
IBM	99.87
Egedian	99.84
McAfee SaaS	99.84
Kaspersky LMS	99.83
Sophos	99.67
Axway	99.49
GFI	99.38
Scrollout	98.77

(Please refer to the text for full product names.)

the emails very quickly, as using a hosted solution adds an extra hop over the public Internet which could delay delivery.

When it comes to classifying the emails correctly, *McAfee* performed very well too, blocking 99.95% of all spam emails, and with only one false positive – an email in Brazilian Portuguese – it was but a whisker away from earning a VBSpam+ award. Another VBSpam award for the industry giant is well deserved.

OnlyMyEmail's Corporate MX-Defender

SC rate: 99.999%

FP rate: 0.00%

Final score: 99.98

Project Honey Pot SC rate: 99.999%

Abusix SC rate: 100.00%

Newsletters FP rate: 0.6%

Speed: 10%: ●; 50%: ●; 95%: ●; 98%: ●

OnlyMyEmail's performance has regularly stunned readers (and authors!) of the VBSpam reports and this was no exception: the product missed only one spam email in a corpus of almost 130,000 messages and erroneously blocked just two newsletters.



It was good to see that this impressive performance is not achieved by holding onto the emails for a long time – the fact that a small percentage of emails takes (just) over 30 seconds to be filtered is probably a natural consequence of the fact that this is a hosted solution, which requires an extra hop (and whose servers are on the other side of the ocean). With a very high final score, yet another VBSpam+ award goes to the product's developers in Michigan.

Scrollout F1

SC rate: 99.41%

FP rate: 0.12%

Final score: 98.77

Project Honey Pot SC rate: 99.30%

Abusix SC rate: 99.52%

Newsletters FP rate: 1.0%

Speed: 10%: ●; 50%: ●; 95%: ●; 98%: ●

Scrollout F1 had an unfortunate run earlier this year, failing to achieve a VBSpam award three times in a row. As mentioned in the last VBSpam report, we had come to suspect that the product might not have been fully adjusted to the test environment, and indeed that does seem to have been the case.

In this test, the product's false positive rate dropped significantly: at 0.12% it remained higher than that of other products, but its final score was well above the VBSpam threshold. And thus, with four green speed markers, *Scrollout* ends its run of bad luck and earns another VBSpam award.

Sophos Email Appliance

SC rate: 99.78%

FP rate: 0.02%

Final score: 99.67

Project Honey Pot SC rate: 99.70%

Abusix SC rate: 99.85%

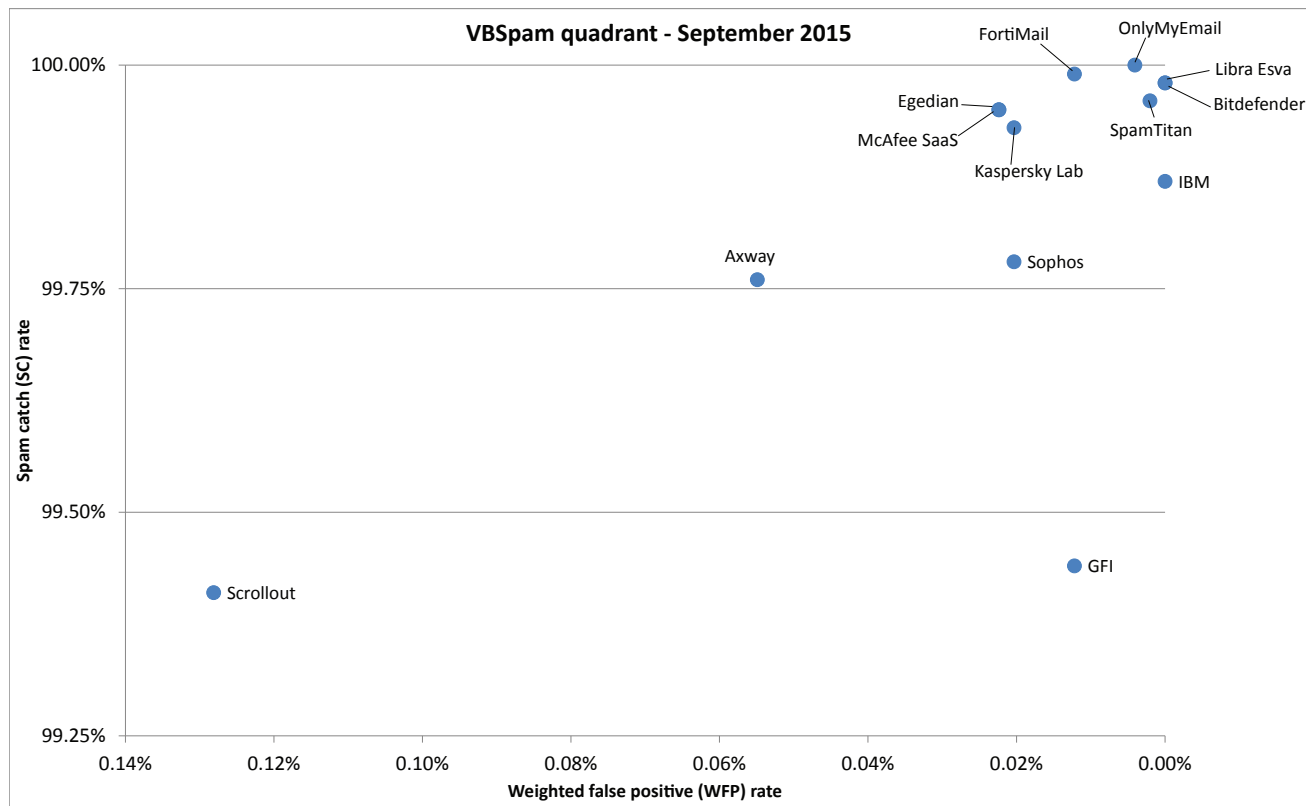
Newsletters FP rate: 0.0%

Speed: 10%: ●; 50%: ●; 95%: ●; 98%: ●

For *Sophos's Email Appliance*, speed isn't an issue: at least 98% of legitimate emails are returned well within half a minute, showing that the appliance – which has been running in our lab for almost six years – hasn't ceased to perform well in all this time.

The same is true for the core metrics, where it blocked 99.78% of spam – roughly the same as in the last test. Due to two false positives, a VBSpam+ couldn't be awarded, but with its 34th VBSpam award in succession, *Sophos* extends its unbroken run.





(Please refer to the text for full product names.)

SpamTitan 6.00

SC rate: 99.96%

FP rate: 0.00%

Final score: 99.95

Project Honey Pot SC rate: 99.93%

Abusix SC rate: 100.00%

Newsletters FP rate: 0.3%

Speed: 10%: ●; 50%: ●; 95%: ●; 98%: ●

The yellow speed marker at the 98% mark means that *SpamTitan* returned a minority of emails with a delay of a little over 30 seconds. This isn't something that many organizations would be likely to mind though, and with no false positives and an impressive 99.96% spam catch rate, the product earns its seventh VBSpam+ award.



Spamhaus DBL

SC rate: 41.51%

FP rate: 0.00%

Final score: 41.51

Project Honey Pot SC rate: 66.15%

Abusix SC rate: 16.74%

Newsletters FP rate: 0.0%

Spamhaus ZEN

SC rate: 92.13%

FP rate: 0.00%

Final score: 92.13

Project Honey Pot SC rate: 86.42%

Abusix SC rate: 97.87%

Newsletters FP rate: 0.0%

Spamhaus ZEN+DBL

SC rate: 95.17%

FP rate: 0.00%

Final score: 95.17

Project Honey Pot SC rate: 92.28%

Abusix SC rate: 98.08%

Newsletters FP rate: 0.0%

As the DNS protocol is designed to give a response almost instantly, it is very popular for blacklists like those

of *Spamhaus*, for whom the added delay under normal circumstances like those in our test is negligible. Hence we don't measure or report on the speed.

In this test, it was good to see the domain-based *DBL* improve its performance further – well over 40% of emails were blocked simply because they contained a blacklisted domain. This suggests that spammers aren't as good at avoiding blacklisted domains as they have been recently.

The catch rate for *Spamhaus*'s IP-address based *ZEN* blacklist remains high at just over 92% of spam, thus showing how IP blacklists in general, and that of *Spamhaus* in particular, can help remove the bulk of the spam with no false positives (as in this test).

CONCLUSION

Although, with 13 participating full solutions, this was the smallest VBSspam test in a long while, it was the first time since July 2014 that all full solutions achieved VBSspam certification. The good news was made all the better by the fact that products rarely delayed emails – or if they did, they managed to avoid doing so for legitimate emails.

The next VBSspam test will run in October–November 2015 (and is about to start at the time of writing this report), with the results scheduled for publication in November/December. Developers interested in submitting products should email martijn.grooten@virusbtn.com.

Editor: Martijn Grooten

Chief of Operations: John Hawes

Security Test Engineers: Scott James, Tony Oliveira, Adrian Luca, Lewis Jones, Ionuț Răileanu

Sales Executive: Allison Sketchley

Editorial Assistant: Helen Martin

Developer: Lian Sebe

Consultant Technical Editor: Dr Morton Swimmer

© 2015 Virus Bulletin Ltd, The Pentagon, Abingdon Science Park, Abingdon, Oxfordshire OX14 3YP, England

Tel: +44 (0)1235 555139 Fax: +44 (0)1865 543153

Email: editorial@virusbtn.com Web: <http://www.virusbtn.com/>