Short Paper Review of:
Characterizing Botnets From Email Spam Records

Citation:

Likes:

• I have to say right off the bat I liked seeing a paper on spam from a company like microsoft that manages a huge email service like hotmail. This unique vantage point for viewing traffic can't be had by most researchers, and it's hard to beat a real world view.

• I suspected this but it was interesting to note that the results showed that their a correlation between the size of the botnet versus the amount of spam each bot sends out. (the larger botnets send less spam per bot).

• It was interesting that the distribution of spam botnet localizations were distributed all over the world, no segmentation based on IP range. That, on the surface, blows away my idea of tracing propagation of infections for the purpose of reputation based on IP+Rage.

• I was glad that unlike the previous paper “On the Spam Campaign Trail,” that the authors of this paper point out that their method only detects spam related bots.

• I liked the explanation about trust vrs. untrust of IP addresses in a relay list for a forwarded message. I think they missed the boat on spoofing (see dislikes) though.

• There was a good bit of math that I was initially unfamiliar with but the authors did a good break down on what each portion of the equations stood for and how it fit together. I would feel confident using this same approach/math to reproduce similar calculations.

Dislikes:

• Their was no real discussion of the specific reasoning behind the 9 day collection cycle other then spam building up very quickly. It would seem that more data over a longer time would be better for trending analysis.

• While I agree that the method is easier than analyzing IRC/DNS/other logs, it's not an effective way of getting the same complete picture for other email hosting providers.
• I disagree that you can get an idea of the size of a botnet based on one provider's view of spamming bot IP addresses. To do so makes the assumption that that one provider (microsoft) has the greatest number of IP addresses and is targeted evenly in the distribution lists across all infected systems in the botnet.

• I didn't like that the collection sets of IP’s versus messages didn't overlap. If the data was to be used for correlation and notification the datasets would have to overlap.

• The IP dynamics discussion / and estimation of botnet size don't incorporate any prior work on spoofed IP detection / estimation. There's a lot of work from MIT, AFRL, Cornell on calculating how many routers, dns servers, etc that are susceptible to spoofed IP addresses being passed through and the potential number of currently spoofed traffic such as spam. Some work, discussion dealing with this or a direct statement of “we don't take into account the potential that infected spam-bot ip addresses may be spoofed,” is desirable.

• The claim that this work was a first step in the study of botnets from an economical motivation perspective I think was a little far fetched. In fact in the whole of the paper, economics / monetary related items were hardly mentioned.

Repeat / Add to Work?

• I probably would not repeat this research, it's interesting but changes based on the trends of botnet and spam evolutions. It would also most likely take a large amount of time to get the relevant data from a provider like microsoft without having an inside associate. If I was to repeat the work I would want more data over a longer period of time with both IP and spam message sets having direct time overlap at a bare minimum.

Synopsis:

• It was a good paper, and I enjoyed seeing the spam problem from the perspective of an email service provider. I would have done a few things differently, most notably the data-set related items. Overall I had a lot less of a problem with the approach compared to the previous paper “On the Spam Campaign Trail” primarily because this paper explained the approach and datasets more clearly. I would not repeat the research for a number of reasons but I would use this paper as a reference in other related endeavors.