

Understanding the Network -Level Behavior of Spammers Anirudh Ramachandran and Nick Feamster

Report by Gessiou Eleni

This paper provides an insight into the techniques spammers use to transmit spam, viewed from the network point of view.

The authors look at spam from the point of view of the IP addresses used to relay spam messages, and they analyze from three main points of view:

- the distribution of IP addresses of spam relays as compared to the distribution non-spam mail senders.
- the behavior of individual bots in a live botnet used to send spam.
- the prevalence of briefly announcing new (hijacked) BGP routes to send spam.

To perform these analyses the authors have used four datasets - a large dataset of spam (10million in number) from a single domain, the mail logs of non-spam email relays (approximately 700K) at a large email provider, a log of all the IP addresses of the bots in a botnet used by spammers, and a trace of all BGP route changes seen by the domain receiving the spam.

The results of the study show that the vast majority of received spam arrives from a few concentrated portions of IP address space. In addition, it is showed that the ASes containing hosts responsible for sending large quantities of spam differ from those sending large quantities of legitimate email and 40% of spam emails are received from the United States. Even if the IP addresses are transient, it seems that the blacklists are quite effective, since nearly 80% of all spam received from mail relays, appear in at least one of eight blacklists tested.

The analysis of botnet behavior show that most bots send spam to a domain only once, and each bot rarely sends many messages to that domain. Presumably each bot sends to many domains though, which indicates that collaborative anti-spam techniques are likely to be needed rather than looking at the behavior of individual bots. Also, using passive OS fingerprinting, they identify that most received spam is sent from Windows hosts. The results indicate that current IP address based blacklisting is of limited effectiveness, in cases that botnets are used for sending spam emails.

Finally the paper looks at the use of "BGP Spectrum Agility", whereby a hijacked IP address range is advertised via BGP, used to send spam and the route for this IP address space is withdrawn, shortly after the spam is sent. This is not the most prevalent technique observed - it is assumed that less than 10% of spam is sent using BGP Spectrum Agility -, but it is perhaps the most sophisticated. Large /8 prefixes are seen to be hijacked, and IP addresses widely distributed within the prefix are used to originate spam. This has clear implications for the need to secure the routing infrastructure.

In total, network-level information may be used in combination with the current techniques, to help mitigate spam. Network-level properties have two important

properties that could potentially lead to more robust filtering: 1. are less malleable than those based on an email's contents and 2. may be observable in the middle of the network, or closer to the source of the spam, which may allow spam to be quarantined or disposed of before it ever reaches a destination mail server.