(U) Hold the Spam, Please

FROM: ████████████

Target Access, Collection, and Techniques (S2G23)
Run Date: 02/03/2005

Separating the wheat from the chaff when collecting e-mail traffic (S//SI)

(S//SI) Spam, or junk e-mail, makes up over 60% of the world's e-mail traffic. This is an annoyance (at the very least) for the recipient. And when NSA intercepts target e-mails heavily laden with this same spam, it becomes the Agency's problem as well. Spam affects NSA by impeding our collection, processing and storage of DNI* traffic. Unfortunately, filtering out spam has proven to be an extremely difficult and cumbersome task. It demands the use of hundreds of constantly changing filter terms and runs the risk of false hits (i.e. eliminating valid, intelligence-bearing e-mails instead of the unwanted spam).

(S//SI) Analysts in Proliferation and Arms Control SIGDEV (S2G23) and CES Data and Metadata Services (S31541) have tackled the problem from a different angle. They have developed a way to filter spam in-house through a metadata-tagging process that uses existing dictionary and SCISSORS processing. This content- and volume-oriented approach is not a panacea, but it is the only process actually doing something about spam. Presently, they are tagging an average of 150,000 spam sessions a day in about two-thirds of data flows containing common e-mail. Some analysts are reporting as much as a 40-percent reduction of spam in their daily searches, but results vary by target.

(S//SI) Furthermore, these metadata tags should provide a more accurate means of measuring and studying the overall effect of spam, and ultimately should help us adjust our front-end collection. If you would like to know more about this anti-spam effort, contact ████████████ at ████████ (s).

*(U) Notes:
DNI = Digital Network Intelligence

"(U//FOUO) SIDtoday articles may not be republished or reposted outside NSANet without the consent of S0121 ([DL sid_comms](#))."