

**(S//SI) Dealing With a 'Tsunami' of Intercept**

FROM: [REDACTED]

Human Language Technology (S23)

Run Date: 08/29/2006

(S//SI) Everyone knows that analysts have been drowning in a tsunami of intercept whose volume, velocity and variety can be overwhelming. But the Human Language Technology Program Management Office (HLT PMO) can predict that in the very near future **the speed and volume of SIGINT will increase even more, almost beyond imagination.** And we are working on ways to help analysts deal with it all.

(S//SI) Of the HLT PMO's five Strategic Thrusts, the one that addresses this problem is **High Speed/ High Volume.** It must deal with today's collection and must plan for tomorrow's. The current collection environment is characterized by huge amounts of data, coupled with severely limited capability to send material forward, and extremely limited number of queries that exactly describe messages of value. That means we are capable of finding huge amounts of data, much of which is not what we really want, and that we cannot send it all back for analyst processing.

(TS//SI) To plan for tomorrow, High Speed/ High Volume is in line with changes in the overall NSA/CSS systems, particularly TURBULENCE and TURMOIL because when they become a reality in the near future, we can expect collection capabilities to increase significantly. TURBULENCE is an umbrella cover term describing the next generation mission environment that will create a unified system. TURMOIL is a passive filtering and collection effort on high-speed networks. This is designed to be flexible and can be modified quickly to deliver data in analyst-ready form.

(S//SI) One of High Speed/ High Volume's first efforts is in developing and implementing ways to **push HLT capabilities very close to the collection points of the SIGINT system.** In particular, HLT is about to demonstrate an operational prototype of language identification for Special Source Operations (SSO) Counterterrorism text targets running at line speeds (STM-16) at the packet-level. Resources permitting, HLT analytic processors will automatically generate content-based events for TURMOIL based on language.

(S//SI) HLT processors will demonstrate the ability to characterize very high speed channels based on content, thus enabling analysts to task the SIGINT system to send back messages based on information found in message content, not just on externals. (Externals can be Signal Related Information (SRI) that comes with each message, such as channel, Time Up/Time Down, etc.) Using HLT services, analysts will be able to build more precise descriptions of the data they want. In addition, content-based metadata will allow SIGDEV analysts to run more detailed surveys. HLT services that work on data content at the collection point can also provide indications or warnings that the SIGINT system must adapt its collection strategy.

(S//SI/REL) Resources permitting, High Speed/ High Volume will deploy **capabilities for voice, text, and image data,** and will take advantage of research being done by a number of organizations including the Research Directorate's Coping With Information Overload Office (R6), Disruptive Technologies Office (DTO), and SID/ Analysis and Production's Advanced Analysis Laboratory (AAL). HLT research and transfer of its technology into operations means the development of algorithms that can incorporate HLT capabilities for the processing of elements such as email attachments and VOIP.

(S//SI/REL) The research and technology transfer also may provide "stealthy," low-profile in-target implants for Tailored Access Operations (TAO) or technologies to enable high speed processing in

very low size, weight and power applications for other CLANSIG customers. And, to help address the "unknown unknown" target analysis problem, HLT is investigating techniques and technologies for high volume voice processing so that all voice data can be scanned for key words before it is selected based on phone numbers.

(S//SI) Ultimately, HLT's High Speed/ High Volume will give the analyst greater ability to influence collection and processing much farther forward in the SIGINT system, as well as help the SIGINT system achieve greater overall filtering and selection effectiveness. That means **more analysts wil be getting better SIGINT** at a time when volume and velocity are maximum.