SENTIENT Challenge Themes

The overriding bottom line up front (BLUF) message is touting the SENTIENT capability to contribute to national security by helping connect the proverbial informational “dots” out of the vast morass of overwhelming data sources quicker, more automated, with better utilization of computers and algorithms and humans interacting—the concept of human aided machine-to-machine learning. The point of the interaction is to better leverage human and machine collaboration to be more predictive and to ferret out activity or events faster, before they become actual threats and to support decision makers taking action (prevention) before things elevate to a crisis, response or reaction (e.g., act left of the boom.)

From any viewpoint we have myriad information challenges to solve in bringing about this vision of human and machine aided collaboration to improve our security. We have a Large or Big data challenge because of known increases in social media devices which have exponentially exploded the signal to noise ratio between relevant information and background noise. We formerly had a problem trying to find a needle in a haystack and the key was identifying and dealing with a finite number of the former to work on the latter. We structure information and visualization capabilities based on a largely analog model. We now have almost an infinite number of haystacks to go through and we must prioritize the resulting needles through threat matrix techniques in order to deal with the needles. Coupled with transition to the digital age--increases in sensor input caused by proliferation of high definition photos and video, as well as quantitative phenomenology (best example probably medical imaging, but the concept of measurement or signature based information such as [redacted] (b)(1) (b)(3)
as well as radar and [redacted] the average analyst is simply overloaded and overwhelmed by sheer information volume. The technique of burning through the haystacks by running algorithms against templates or models of know behavior to deal with only the needles through machine aided automation or upstream processing—for instance an astronomy pipeline

approach, where machines automatically report on 99% of the available collection, while analysts look at less than 1% consisting of mostly those that have been flagged or marked by the machine computer algorithm screener--has not come to fruition.

This process deals with finding issues and identifying problems and patterns of behavior that we know about and that match templates and/or models of observable activity and signatures that we have decomposed and applied; the so-called known knowns or the known unknowns. The real challenge is to go after unknown unknowns where we have things like enigma problems or cyber related efforts or problems that have simply not been captured in signatures or observables up to this point. The plan is to have machines conducting some type of alternative hypothesis testing or "red teaming" in an attempt to identify behavior that stands out because it defies characterization or simply does not fit with anything else.

We know analysts benefit from using a multi-source or multi-information environment, but the workflow for GEOINT analysts has somewhat calcified because of the need to tightly couple the information request process to the exploitation and subsequent reporting or production process; a tight engineering coupling was necessary to ensure compatibility between sources of imagery (metadata), exploitation and mensuration tools, and storage devices—necessary so the analyst could get the right image with the right information at the right time to meet reporting deadlines and to prevent waste of time and effort. Also, because of the large target data sets we have on a worldwide basis—some 750K discrete reportable entities with associated baselines—the information or data base connection is also tightly coupled and structured in accordance with decades old production standards. The baseline information has not matured and is not structured in a way that accommodates our emerging reporting methodology, for instance, Activity Based Intelligence or Object Based Reporting, from the standpoint that we deal mostly with targets and fixed facilities and nation states and political boundaries—so we don't deal with entities below a certain level, we lack good status and activity flags, movement and terrain models, etc. Also, analysts do not have good and consistent access within their workflow (e.g., when they are exploiting GEOINT and producing intelligence) to all sources of information, for instance, unclassified open source, foreign sources, special access program information, law enforcement sensitive information, activity related to US based citizens, etc. All to say that there are real impediments to connecting the dots between and among disparate data bases when the information is not discoverable or consumable within the analyst's workflow. Finally, all analysts have deadlines to meet similar to commercial publishing battle rhythms' and therefore must by necessity drastically cut back on the amount of information they typically grind through in order to focus on production. Such production is most often in support of a particular customer or consumer and analysts endeavor to answer those specific requirements vice general or generic or standing requirements. In meeting these needs analysts often work around existing legacy workflow data shortfalls and report out via a variety of non-traditional or non-standard means which don't fit well with the heretofore baseline data base structure. Since these products are often one offs or non-standard, we add the additional challenge of these reports looking like unstructured data challenges to the mainline system, resulting in discovery and consumption problems for other analysts and a lack of reflection in the baseline information system used as the basis for the community knowledge on topics of interest or concern. The net result of the above is that only about 55% of analysts actually report into the baseline information system intended to house all reporting and production. Analysts spend nearly 65% of their time wrestling data and information issues. Customers and consumers of information and intelligence find that even if they know what specifically they need or are looking for, they simply can't find it with the current production systems. Separately, we miss observation and collection opportunities because we lack the capability or agility to connect and act on the information we have to pursue alternative hypothesis or to act with the threat execution timeline. There is a need to better integrate what we do with what we know in consideration of what we don't know.

The bottom line down the bottom (BLDB) is the emerging theme of providing better information to the decision makers to allow them to make more timely and relevant decisions; to be more predictive in our reporting, to prevent bad things from happening and to provide greater overall efficacy in the process by leveraging the power and speed of machines to deal with big data issues and to enable analysts to do more analysis and less searching—and to connect the dots!

Some good exemplars;

Outbreak of the Ebola virus in Africa—a massive data problem and a huge signal to noise ratio, big data problem, resulting in a minor correlation that can be acted upon.

Google taking on the challenge of being able to do better tracking of flu outbreaks than the federal health organizations by focusing on search queries and narrowing down the correlation of search terms with flu outbreaks, resulting in a reduction of nearly two weeks in the use of a predictive model.

[redacted] (b)(1)
(b)(3)

Enigmas that remain in our system because there simply is not enough information known about facilities, places or things.

Forensic tradecraft for a variety of activity but particularly for Counter-Improvised Explosive Devices exploitation where machines perform automatic identification, detection and track information to reduce 10s of terabytes to much smaller file sizes to enable analysts to work in "dot space," resulting in a significant reduction in the time it takes to reconstruct events.

Regards, [redacted] (b)(3)