Measuring Performance of Cyber Situation Awareness Systems

George P. Tadda
Air Force Research Laboratory
Rome Research Site
Rome, New York, U.S.A.
george.tadda@rl.af.mil

Abstract – Enabling situation awareness necessitates working with processes capable of identifying domain specific activities. This paper addresses metrics developed to assess research level systems and to measure their performance in providing those processes. The metrics fall into four dimensions; confidence, purity, cost utility, and timeliness. The bulk of the discussion will provide an overview of each of these dimensions, the anticipated usefulness of measures within each dimension, and a discussion of observations resulting from the use of each measure in building Cyber Network Defense capabilities. The paper concludes with a brief mention of ongoing activities and identifies ideas for future research.

Keywords: Data Fusion, Cyber Fusion, JDL, Situation Awareness, Situation Assessment, Impact Assessment, Metrics, Measures of Performance

1 Introduction

For the last several years, novel work has concentrated on developing and applying ideas for implementing Higher Level Fusion. Higher Level Fusion is a term that describes level 2 (Situation Assessment) and level 3 (Impact Assessment) which are defined in [2]. Techniques for implementing levels 0, 1, and 4 have been ongoing since (and before) they were clearly defined by [2] but levels 2 and 3 have been difficult and eluding.

It was quickly apparent that Situation and Impact Assessment would require something more then the fairly direct data-driven approach used for the other data fusion levels. Context, point of view, and greater understanding of the environment were all missing both from the definitions of fusion levels and from the technical approaches. What meaning does a situation have? How can the situation be understood and leveraged for decision making? What is a situation? To help address the missing pieces, literature searches quickly identified work in dynamic human decision making and situation awareness. This work is characterized by the research performed by Dr. Mica Endsley and described in [3]. Recognizing the complementary components of the Joint Directors of Laboratories (JDL) Data Fusion Model and the Situation Awareness (SA) Model defined by Dr. Endsley, a merged model was developed and presented in [4]. This merged model captures both data-driven and knowledge-driven aspects for enabling situation awareness. Any implementations based on the model only enable situation awareness rather then provide or achieve situation awareness because awareness, as Dr. Endsley states in [3], is a human "knowledge state" and only something a computer can assist in providing. To develop and test the SA Reference Model introduced in [4], research began to apply it to the cyber domain. The SA Reference Model has since been updated and more completely defined in [6] as shown in Figure 1.

The difference between situation awareness and situation assessment is described in [3]. Dr. Endsley indicates that it's important to "distinguish the term *situation awareness*, as a state of knowledge, from the processes used to achieve that state. These processes, which may vary widely among individuals and contexts, will be referred to as *situation assessment* as the process of achieving, acquiring, or maintaining SA." Situation Assessment is further defined in [6] as "the understanding of the current situation and its impact/threat and the projection of the current situation into the future and its future impact/threat." The goal then is to measure a system's ability: 1) to identify and present the current situation; 2) to assess that situation's impact/threat; and 3) to project the situation and assess the projection's impact/threat.

Cyber Situation Awareness (Cyber SA) is still a relatively new term with specific research in the area only a few years old. Thinking in terms of maintaining awareness ("a state of knowledge") of a cyber environment to facilitate decision making for defensive or responsive actions is a growing area of interest. Current work has focused on enabling awareness of computer networks for cyber defense, specifically with attack identification. In [7] and [8], the application of the SA Reference Model to the defensive cyber environment is shown as well as a description of some initial efforts to measure the performance of software systems that implemented the SA Reference Model. The assessment was based on the metrics defined in [5].

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUL 2008		**00-00-2008 to 00-00-2008**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Measuring Performance of Cyber Situation Awareness Systems	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Air Force Research Laboratory,Rome Research Site,Rome,NY,13441**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
11th International Conference on Information Fusion, June 30 ? July 3, 2008, Cologne, Germany.

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**	

The remainder of this paper will briefly describe related work, the metrics or measures of performance, lessons learned from the development and assessment of Cyber SA systems, and a brief overview of the process used to assess Cyber SA systems and their results concluded by briefly describing current and future work in this area.

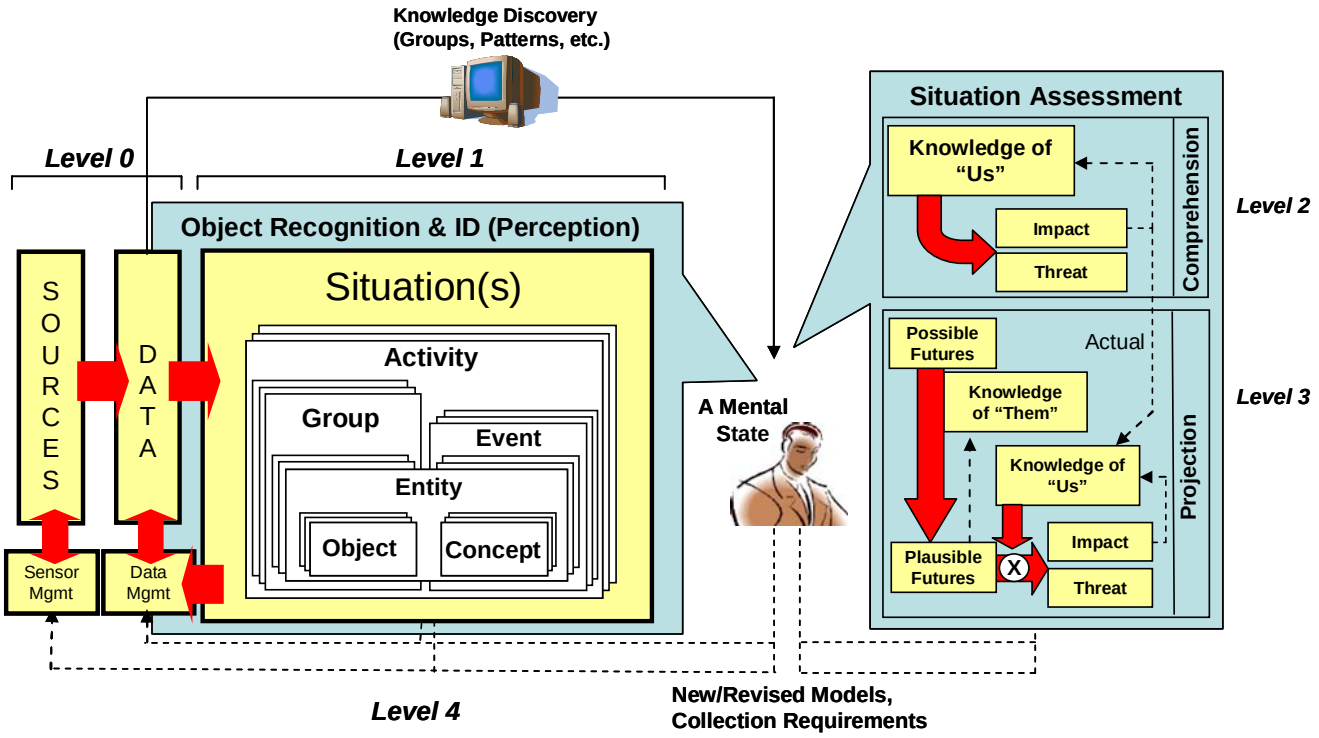

Figure 1 - Proposed Revision of Reference Model (from [6])

2 Related Work

Detecting and identifying computer attacks have leveraged many different technologies since Dorothy Denning first proposed using expert systems for finding computer attacks in 1987, [9]. Since then, as described in [10], a myriad of techniques have been used including anomaly detection, pattern matching, agent-based systems, and numerous others many of which are in use today. While these approaches have met with varying levels of success, it's important to note that the majority, from a fusion perspective, would fall under the definition of level 0 or early level 1 data fusion. This can be exemplified when you consider that a Transmission Control Protocol (TCP)/Internet Protocol (IP) packet could be considered "signal data". Then an Intrusion Detection System (IDS) that employs a technology described above is a sensor that "senses" a pattern or anomaly in the signal data and produces an event related to that signal (or more commonly referred to as an alert). The result of work in IDSs have been valuable to better detect possible network intrusions but because of the false alarm rates and the fairly low information level of the events, users are being inundated with data and have to rely on high-levels of expertise to understand the alerts generated. The high false alarm rate led to work in alert correlation that attempts to tie similar alerts together based on some feature or a priori information of the alert data. Alert correlation systems have been successful in reducing the numbers of false alarms but they typically stop after a few common features are correlated.

From the perspective of measuring these types of systems, they've almost exclusively been measured in terms of false alarm rates, false positives, and false negatives. Additionally, the events are considered to be not just indicators of an attack but an attack themselves regardless of the level of information they provide (e.g., a simple port scan or a more complex buffer overflow pattern are both considered an attack). A primary focus of alert correlation systems has been to reduce the number of false alarms produced by the first-level IDS.

When considering the need to provide situation awareness in the cyber domain, very little is described in the overall body of research. This also means that very little work is being done to define measures of performance for systems enabling Cyber SA.

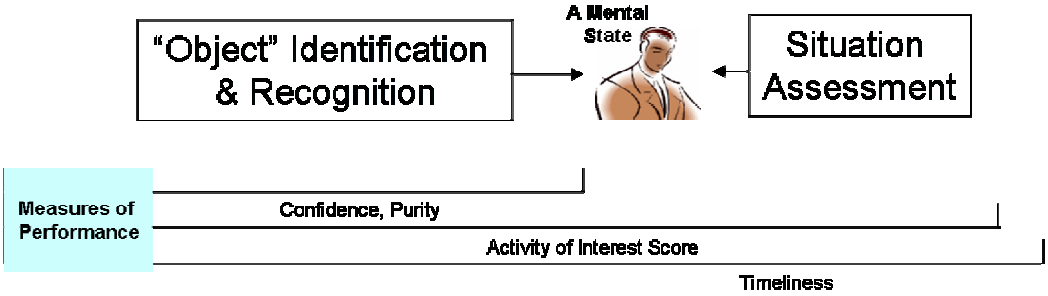

Figure 2 - Metrics Mapped to Reference Model (from [6])

3 Metrics

In [6], there are very detailed descriptions and examples of metrics for generic SA systems. [8] provides an initial attempt to define the same measures specifically for Cyber SA. Rather then repeat details of those two papers, only a brief summary of the metrics as applied to the cyber domain is included below. Note that the metrics were initially developed in [5] independent of domain with the intent that they could be applied to any system that enables SA and that [6] updates that work based on information gained when the metrics were applied to the cyber domain as captured in [8]. The current set of measures of performance have four dimensions; confidence, purity, cost utility, and timeliness.

Before describing the metrics, a few definitions are needed. The "raw data" or input streams for a Cyber SA system are the events generated by network sensors. Network sensors include output data streams from such tools as intrusion detection systems, firewall logs, system logs, network flow or connection data, etc. These events are considered the *evidence* of a Cyber SA system. When the evidence from multiple data streams is fused together in such a way as to identify a potential attack, we call this collection of evidence an *attack track*. An attack track should contain all of the evidence associated with a complete attack regardless of the attacks complexity. A *situation* is the set of all attack tracks at a given point in time.

These definitions are built from the ideas captured in Figure 1. In previous work, [4], [7], [8], this model was also referred to as the SA Reference Model and is used to provide a guide to developing SA systems. The definitions of the previous paragraph implement the various sections of the model with input streams representing "sources"; the output from network sensors and other evidence representing "data"; and attack tracks corresponding to activities and situations. Figure 2 shows at a high level how each of the metrics that will be discussed map onto the reference model.

The metrics currently measure a Cyber SA system's ability to correctly fuse evidence, produce attack tracks, and prioritize the attack tracks into a meaningful order for a user. One of the metrics, *attack score*, has applicability to systems capable of assessing impact/threat and should improve in value (get closer to 1.0) as a system can more completely analyze the attack track in the context of a network or in importance to effects on a mission. Attack score is in the process of being updated to measure more information then just attacks. As described in [6] the updated measure is being called an "activity of interest score."

One last comment on metrics before getting into the bulk of the discussion concerns data reduction. Early work, [5], spoke at length about a Data-Information Ratio (DIR) as shown in equation (1). The DIR was intended to measure the overall reduction in the amount of "stuff" that was presented to a user. When data is presented, a user tends to be: 1) overwhelmed by volume and lack of context; 2) has to rely on individual expertise for understanding; and, 3) has to mentally process (fuse and assess) the data.

$$\text{DIR} = \frac{\text{Number of Observations}}{\text{Number of Complex Entities}} \quad (1)$$

The initial thought was that by automatically analyzing data into more useful information the user would have less to deal with and could more productively and more effectively maintain awareness of the environment. The DIR has proven to be very informative but at a fairly high level. It tends to indicate the capability of a class of work rather then the capability of individual systems. For instance, in the cyber domain, we've observed an on average data reduction of two orders of magnitude when processing "alerts" (observations or events) into attack tracks (complex entities or activities). The advantage to the user then is that instead of tens of thousands of individual pieces of data to consider they now only have to consider a few hundred possible attack tracks. An attack track only reduces the information initially presented while maintaining the ability to "drill down" into the more detailed data that makes up the track. When combined with a mechanism to prioritize importance of the possible attacks, the power of this general class of analysis begins to become apparent. However, beyond the general data reduction, the DIR

doesn't provide a lot of insight into the performance of particular technologies or implementations.

The rest of this section will discuss each of the metrics in more detail. Also, some general observations about the metrics based on their use to assess research efforts are included in each subsection. The metrics, or measures of performance, are discussed according to the four dimensions; confidence, purity, cost utility, and timeliness. As shown by the mapping in Figure 2, the confidence and purity measures address levels 0 and 1, cost utility adds levels 2 and 3, and timeliness covers the entire model.

3.1 Confidence

As discussed in [8], for the cyber domain, **confidence** is a measure of how well the system detects the true attack tracks. The confidence dimension consists of four metrics; (1) recall, (2) precision, (3) fragmentation, and (4) misassociation. Consider the diagram shown in Figure 3; it represents the space of attack tracks identifiable by a Cyber SA system. The attack tracks can be classified into three categories; (1) known tracks, (2) detected tracks, and (3) correctly detected tracks.

Known tracks are the attack tracks given by ground truth and contain all the evidence for a particular attack. *Detected Tracks* are the attack tracks hypothesized by a Cyber SA system under evaluation and contain all the fused evidence. Finally, *Correctly Detected Tracks* are known attack tracks detected by the Cyber SA system. Known tracks not detected would be false negatives and detected tracks not known would be false positives.

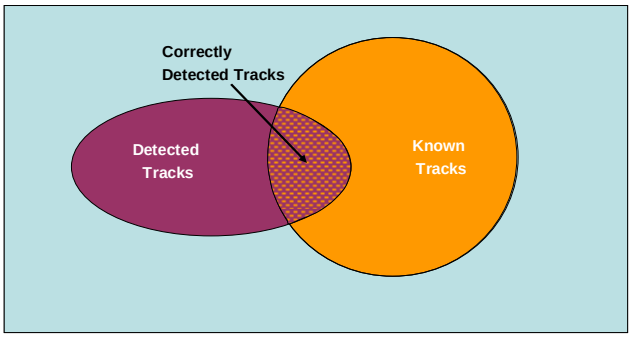

Figure 3 - Space of Attack Tracks

Given this description, the confidence metrics are as described by the equations below:

$$\text{Recall} = \frac{\text{Correctly Detected Tracks}}{\text{Known Tracks}} \quad (2)$$

$$\text{Precision} = \frac{\text{Correctly Detected Tracks}}{\text{Detected Tracks}} \quad (3)$$

$$\text{Fragmentation} = \frac{\#\text{of Fragments}}{\text{Detected Tracks}} \quad (4)$$

$$\text{Misassociation} = \frac{\#\text{IncorrectNonFragmentTracks}}{\text{Detected Tracks}} \quad (5)$$

Two additional definitions are needed to complete the discussion of confidence metrics. A *fragment* is said to be an attack track that should have been included within another track (a more rigorous definition is in [8]). For example, in an island-hopping attack, a targeted computer is compromised and then used to originate attacks on other computers. To correctly detect this attack, any time a target becomes an attacker, all the evidence of this attack should be included in same track as that from the original attacker through the island all the way to the subsequent targets. Often a fusion engine will not correctly associate the subsequent evidence with the original attack reporting them as two or more attack tracks. In other words, a single attack is reported as two or more attacks only one of which would be counted as the correctly detected track. A fragment can give the appearance of a false positive when, in reality, it usually indicates a portion of a more complex attack. *Misassociation* is a little simpler and captures all "other" detected tracks. A misassociated track is one which is neither a fragment nor correctly detected. Summed together, the values for precision, fragmentation, and misassociation should sum to 1 or cover 100% of the attack tracks produced by the Cyber SA system.

Assessing Cyber SA systems under research and development showed the confidence metrics to be the most useful when determining the overall capability of a system thus far. The traditional tension between recall and precision was observed in that high recall usually meant low precision – detected lots of attacks but couldn't clearly identify them. While, in contrast, high precision (knew the exact attack type and details) would often result in low recall or missed attacks. The fragmentation metric was interesting from the perspective of complex attacks and the level of data reduction. High fragmentation led to more attack tracks being presented which lowered data reduction but still typically maintained the two orders of magnitude reduction. The greatest value in the fragmentation metric was in indentifying that it would be better to keep more complex attacks in a single attack track and also in generating discussion concerning attribution of an attack. For example, if there was evidence of two attack tracks from different original attackers attacking the same target, *A*. Then, there was subsequent evidence that *A* was attacking *B*. Which of the original two attackers would be attributed to the follow-on attack on B? How would the attribution be made? Could you be sure that one of those attackers continued on and that it wasn't a new attack originating with *A*? These remain open questions and interesting areas of research.

3.2 Purity

As defined in [8], **purity** characterizes the quality of the correctly detected tracks. Purity metrics also "look into" a track at the evidence and provide indications as to how well the evidence is being correlated and aggregated into an attack track. There are two metrics used to measure purity; (1) Misassignment rate, and (2) Evidence Recall. The equations for these metrics are given below:

$$\text{Misassignment Rate} = \frac{\#\text{ of Incorrect Evidence}}{\text{Total Evidence Detected}} \quad (6)$$

$$\text{Evidence Recall} = \frac{\#\text{ of Correct Evidence}}{\text{Total Evidence in GT}} \quad (7)$$

GT in equation 7 is an abbreviation for ground truth.

By looking at the quality of the correctly detected tracks, the thought was that the metric could indicate how well the Cyber SA system was using the available evidence. Misassignment rate could answer the question about whether the system was assigning evidence to a track that wasn't relevant or if it only considered directly useful evidence. While evidence recall was intended to tell us how much of the evidence available was truly being used?

When applying the purity metrics to the cyber domain, neither proved to be particularly useful. Misassignment rate was probably the stronger of the two in that when the rate was very high it would indicate an incorrect correlation or association of the underlying data. This would essentially indicate a "bug" or flaw in the system's fusion engine. However, it rarely indicated anything about the quality of the detected attacks. Extraneous evidence didn't necessarily have any correlation to lower or higher detection rates. Evidence Recall was even less useful. The thought was that as more evidence was used, the attack detection would be more accurate (higher recall and precision). However, empirically we found almost no relationship between the amount of evidence used and the quality of the detections. In fact, it almost appeared that the less evidence used the better the detections. This almost indicates that there are only a few truly relevant network events that indicate attacks. An open area of research or question is whether there's a minimally complete set of data or events that could indicate the presence of an attack.

3.3 Cost Utility

Referring to [8] again, we see that **cost utility** is defined as the ability of a system to identify the "important or key" attack tracks with respect to the concept of cost. In [8], two cost utility metrics were described. Since that paper was written, the *weighted cost* metric as applied to the cyber domain is no longer in use. The intent of the metric was to capture or gauge the usefulness of the system by considering the types of attacks detected with a positive weight and penalizing the system for false positives with a negative weight. Different weights were also assigned to different categories of attacks. Weighted Cost is then a simple sum of the values assigned to the types of attacks detected including false positives divided by the sum of the values of the attack tracks in ground truth. Observations when using the weighted cost showed that it didn't add any value in measuring the performance of a Cyber SA system.

The metric that appears to have great value in measuring the performance of a Cyber SA system is the attack score. The attack score is an earlier version of the "activities of interest (AOI) score" described in [6]. Attack score is calculated as shown in equation 8.

$$\text{Attack Score} = \frac{\text{NAGT} \times \text{NTGT} - \sum_{i=1}^{\text{NAD}} P_i}{\text{NAGT} \times \text{NTGT} - \sum_{i=1}^{\text{NAGT}} i} \quad (8)$$

Where:

NAGT is the number of attacks in ground truth
NTGT is the number of tracks in the ground truth
NAD is the number of attacks detected
P_i is the position of the i^{th} attack in the results

Then, as described in [6], "If any (or all) of the AOIs *[attack tracks]* are not part of the results list or if their position is greater than the total number of activities in the ground truth, we set the position value (P_i) for those AOI(s) equal to the Total Number of Activities in the Ground Truth. By adding this condition, if there are no AOIs included/identified in the results list, the AOI score will equal 0. Whereas, if there is only a subset of the actual AOIs identified, the system will get credit for only those."

Attack score tries to measure the presentation of a prioritized list of hypothesized attacks by counting how many actual attacks occur and how close their priority is to the top of the list. In effect, the attack score measures the ability of a system to perform situation assessment as defined in Figure 1. The attack score is considered a cost utility measure because the lower in a prioritized list that the actual attack appears implies that more work is performed before the actual attack would be considered or could be acted on. For example, if the actual attack was at the top of the prioritized list, a user's attention would be drawn to it first and action taken. If the actual attack appeared at position 25, then the user would have to consider or "look into" 24 other attack tracks before getting to the actual attack track that was important or required action. Analyzing the 25 tracks indicate a cost in time before being able to take an appropriate action. Thus, an ideal attack score, with all actual attacks (true positives) at the top of the list, would be 1.0. Anything

less then an attack score of 1.0 means that some level of effort is expended considering incomplete tracks or false positives. Another way of looking at the attack score is that it measures a system's capability to assess the situation. How the attacks are listed (prioritized) is determined algorithmically and could be influenced by the desire of the user to include; most critical, most damaging, most likely, greatest mission impact, etc. The different prioritization algorithms could influence the ordering of the list which in turn affects the attack score which could also provide differing assessments of the situation. How well this measure can assess the situation and what other metrics may be needed is a potential area of future research.

The attack score has tremendous opportunity to be an indicator of improved analysis as more sensor types are considered, as better models are used in the fusion process, or as we improve our capability to add additional methods for situation analysis as described in [6]. With improving analysis, the attack score would approach the ideal of 1.0.

3.4 Timeliness

The final dimension, timeliness, was defined in [8], as the ability of the system to respond within the time requirements of a particular domain. More specifically, timeliness would need to measure the time elapsed before a decision could be made or action taken. Because we're interested in awareness, it's not simply the time it takes to present or detect an attack or activity but also includes the time it takes to enable awareness of the activity so that a user could make a decision. Timeliness touches the border between a measure of performance and a measure of effectiveness. This is an area of future research.

To conclude this section on metrics, it's important to note that a single metric is probably inadequate to characterize the performance of a system. Rather, a set of metrics measuring the various dimensions of the problem are needed to fully characterize and provide insight into the performance of the systems.

4 Assessment Process Overview

The metrics described in Section 3 and the corresponding observations were the result of two research assessments performed during 2005 and 2006. The 2005 assessment was a manual process using primarily the DIR metric but also tried to apply all of the metrics described in Section 3. Because a manual process was used, the complexity of the data set, and the volume of data we quickly identified the need for an automated metric calculation tool. The final results of the 2005 assessment were inconclusive because of having to rely on a manual process.

For the 2006 assessment, the process shown in Figure 4 was implemented in an automated tool. The tool was implemented using JAVA and it provided an automatic mechanism to both generate a ground truth file and assess a system's results file.

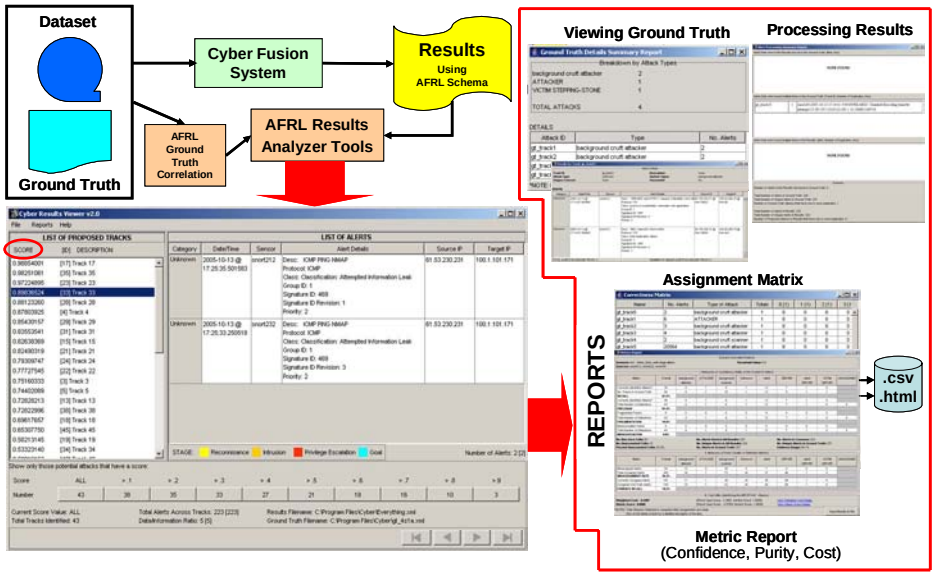

Figure 4 - Assessment Process

Note the phrase, "generate a ground truth file." An additional observation is that data, as currently available, isn't naturally organized into a form useful for directly determining a system's ability to do situation assessment. We still need to abstract the data to a higher level and organize it into something appropriate for measuring situation assessment. To do this, we took the raw data in combination with ground truth (known attack steps, known roles of network nodes, known background traffic, and known network topology) and organized the

data into attack track based ground truth. The ground truth was then stored in an XML file. Any system under evaluation would then run their system against the raw data and produce a "results file" in a similar XML format. The automated assessment tool would then compare the results file to the ground truth file and produce the metrics as well as information (right side of Figure 4) for empirical analysis and checks for correctness.

A total of 75 runs for several different systems and the technology they implement were performed. Because of the uniqueness of this work and no prior results to compare our findings against, we instead compared the results to an average across the 75 runs. Then counted the number of times a particular system exceeded the average. For recall, precision, and evidence recall, a "good" count is if it was higher then the average. While for fragmentation, misassociation, and misassignment rate a "good" count is if it was below the average. Table 1 shows anonymous scores for each metric for two of the research systems assessed.

Table 1 - Metric Summary

Metric	Percent Runs Exceeding Goal	Percent Runs Exceeding Goal
Recall	100.0%	23.3%
Precision	20.0%	93.3%
Fragmentation	20.0%	90.0%
Misassociation	36.7%	90.0%
Misassignment Rate	36.7%	60.0%
Evidence Recall	50.0%	40.0%
Number of Attacks Found	96.7%	46.7%
Attack Score (AOI)	0.5156	0.1580

While not exceptionally exciting, the results do show the traditional tension between recall and precision and the attack score shows that an additional effort needs to be applied to improving or reducing the amount of work needed by an analyst. Specific observations about each metric are embedded in the discussion in Section 3 as each metric is discussed.

5 Current Research

Current research in Cyber SA has concentrated on producing attack tracks and in various methods for visualizing this information. Referring back to Figure 1, the current work has implemented the left side of the figure. Recently started research efforts are beginning to explore additional analysis techniques to understand the impacts of the identified situations – comprehension in the upper right section of Figure 1. To do the further analysis requires additional data sources, such as; vulnerabilities, mission supported by an asset, applications running on an asset, services supported by an asset, and network flows and paths each of which could be relevant. A critical piece of the newer research efforts are to identify and capture the data listed above, perform the analysis, and then to extract and fuse the information needed. The output of new analysis then becomes an information source for the Cyber SA system. The current thought is that the metrics defined above will be able to also measure the newer work without significant modification. The one exception is the attack score's transformation into an activity of interest score. As more data and analysis is performed, the goal is to enable awareness of more then just attacks on a network. The activity of interest score will have to evolve to better measure this new information.

6 Future Research

The largest gap in current research is how to perform situation projection or anticipation – projection in the lower right section of Figure 1. The work is currently able to identify the situation as it unfolds but projection would require projecting the situation one or more time steps forward. Then, once projected, the new situation could be analyzed using techniques similar to those used to analyze the current situation. Part of the analysis would include a paring down of the projections from possible futures into a more manageable set of plausible futures based on ideas such as the most critical situation, most damaging situation, most likely situation, etc. Techniques for managing large state spaces resulting from the possible futures or specifically detecting coordinated attacks are also research areas of interest not currently being addressed.

Additional research would need to identify the measures that would be used in order to determine the success or capability of the systems performing the projections. At this time, these are undefined and it has yet to be determined if the current set of metrics would be useful in measuring projections. Finally, once an SA system is operational or nearly operational, what are the measures of effectiveness that would be used to indicate the system has enabled situation awareness for the cyber operator? These measures are unique to a domain and are distinct from the measures of performance discussed in this paper because they would address measures of the cognition achieved by an operator … or not.

Acknowledgements

The author thanks Dr. John Salerno, AFRL/RIEA, and Mr. Mike Hinman, AFRL/RIEA for their reviews, comments, and valuable insights.

References

[1] U.S. Department of Defense, Data Fusion Subpanel for the Joint directors of Laboratories, Technical Panel for C3, "Data Fusion Lexicon," 1991.

[2] Alan N. Steinberg, Christopher L. Bowman, and Franklin E. White. Revisions to the JDL Data Fusion Model, presented at the Joint NATO/IRIS Conference, Quebec. October 1998.

[3] Mica R. Endsley, *Toward a Theory of Situation Awareness in Dynamic Systems*. Human Factors Journal, Volume 37(1), pages 32-64, March 1995.

[4] John J. Salerno, Michael Hinman, and Douglas Boulware, *A Situation Awareness Model Applied to Multiple Domains*. Proceedings of the Defense and Security Conference, Orlando FL, March 2005.

[5] John J. Salerno, Michael Hinman, and Douglas Boulware, *Evaluating Algorithmic Techniques in Supporting Situation Awareness*, Proceedings of the Defense and Security Conference, Orlando FL, March 2005.

[6] John J. Salerno, *Measuring Situation Assessment Performance through the Activities of Interest Score*. To be published at Fusion 2008, Cologne GE, 2008.

[7] John J. Salerno, et al., *Achieving Situation Awareness in a Cyber Environment*. Proceedings of the Situation Management Workshop of MILCOM 2005, Atlantic City NJ, October 2005.

[8] George Tadda, et al., *Realizing Situation Awareness within a Cyber Environment*. In Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2006, edited by Belur V. Dasarathy, Proceedings of SPIE Vol. 6242 (SPIE, Bellingham, WA, 2006) 624204, Kissimmee FL, April 2006.

[9] Dorothy E. Denning, *An Intrusion Detection Model*, IEEE Trans. Software Engineering, Vol. 13, No. 2. pp 222-232, Feb. 1987.

[10] Stefan Axelsson, *Intrusion Detection Systems: A Survey and Taxonomy*, Chalmers University of Technology, Dept. of Computer Engineering, Goteborg Sweden, Technical Report 99-15.

[11] Moises Sudit, Adam Stotz, Michael Holender, William Tagliaferri, and Kathie Canarelli, *Measuring situational awareness and resolving inherent high-level fusion obstacles,"* In Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2006, edited by Belur V. Dasarathy, Proceedings of SPIE Vol. 6242 (SPIE, Bellingham, WA, 2006) 624205, Kissimmee FL, April 2006.

[12] Moises Sudit, Adam Stotz, and Michael Holender, *Situational awareness of a coordinated cyber attack,"* Proceedings of SPIE Vol. 5812, 114, 2005.

[13] Jared Holsopple, Shanchieh Jay Yang, and Moises Sudit, *TANDI: threat assessment of network data and information,"* In Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2006, edited by Belur V. Dasarathy, Proceedings of SPIE Vol. 6242 (SPIE, Bellingham, WA, 2006) 62420O, Kissimmee FL, April 2006.