A Mathematical Cyber Ontology (CybOnt) for Cyber Data Fusion and Cyber Data Exchange

SBSI
SILVER BULLET SOLUTIONS, INC.
CUBRC
Edutainiacs

1 (U) Introduction

(U) Silver Bullet Solutions, Inc., with teammates CUBRC, Inc. and Edutainiacs, Inc., is developing a mathematical ontology for cyber events, entities, behaviors, associations, and intentions – cyber Situation Awareness (SA) – and associated cyber Command and Control (C2) – Network Operations (NetOps), Defensive Cyber Operations (DCO), and Offensive Cyber Operations (OCO). This Cyber Ontology (CybOnt) will improve interoperable data exchange between cyber operations nodes and enable data fusion for detection of cyber attacks as they are being planned and before they become incidents.

2 (U) Cyber Data Fusion

(U) The advantage of a data fusion approach to cyber SA is that it is probabilistic, which allows operators and analysts to adjust the probability-of-false-alarm (p_{FA}) to probability-of-detection (p_D) ratio to the level that supports their operational need, e.g., for timeliness, operator workload, completeness. It is also mathematically principled, building upon decades of data fusion and associated probabilistic AI R&D, e.g., [1], 2, 3, 4], [5). It also has an architecture that started from radar target tracking evolved to imagery and all-source fusion and that fits the cyber SA problem well. The architecture in which we had a role is called the Joint Directors of Laboratories (JDL) data fusion levels shown in Figure 1 [6, 7]. For the purposes of cyber I&W and SA, data fusion at Level 0 extracts features from sensor data, identifies and localizes cyber actors and events at Level 1, links actors and events at Level 2, and assesses risks at Level 3. Level 4 is the process of adjusting cyber fusion in response to new and emerging threat TTP and Courses of Action (CoA).

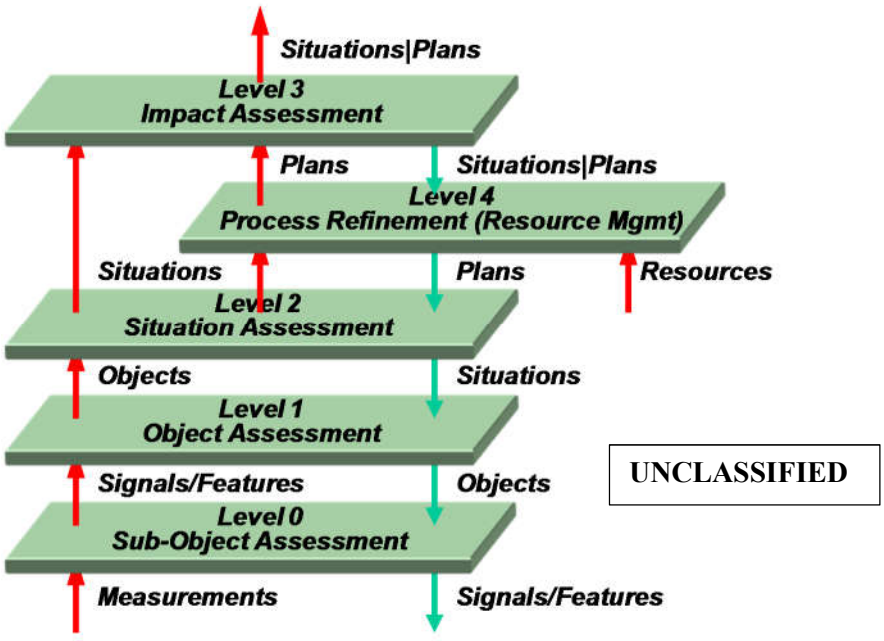


Figure 1. (U) Typical Flow Between Fusion Levels

(U) The C2 analog was pioneered in the 1980's by Silver Bullet engineer Thomas Murphy in his tiers-of-integration and layers-of-C2 architecture that is applicable to the Cyber SA and C2 [8]. Later, our colleague Chis Bowman refined this as a dual data fusion and resource management [9]. Figure 2 shows cyber SA and C2 the fusion levels and C2 layers architecture. On the left are the JDL fusion levels that produce cyber situation awareness. On the right side are command and control layers. The cyber ontology focus is then the structure of data ingested by cyber fusion centers to produce cyber SA and conduct cyber C2 (blue lines).

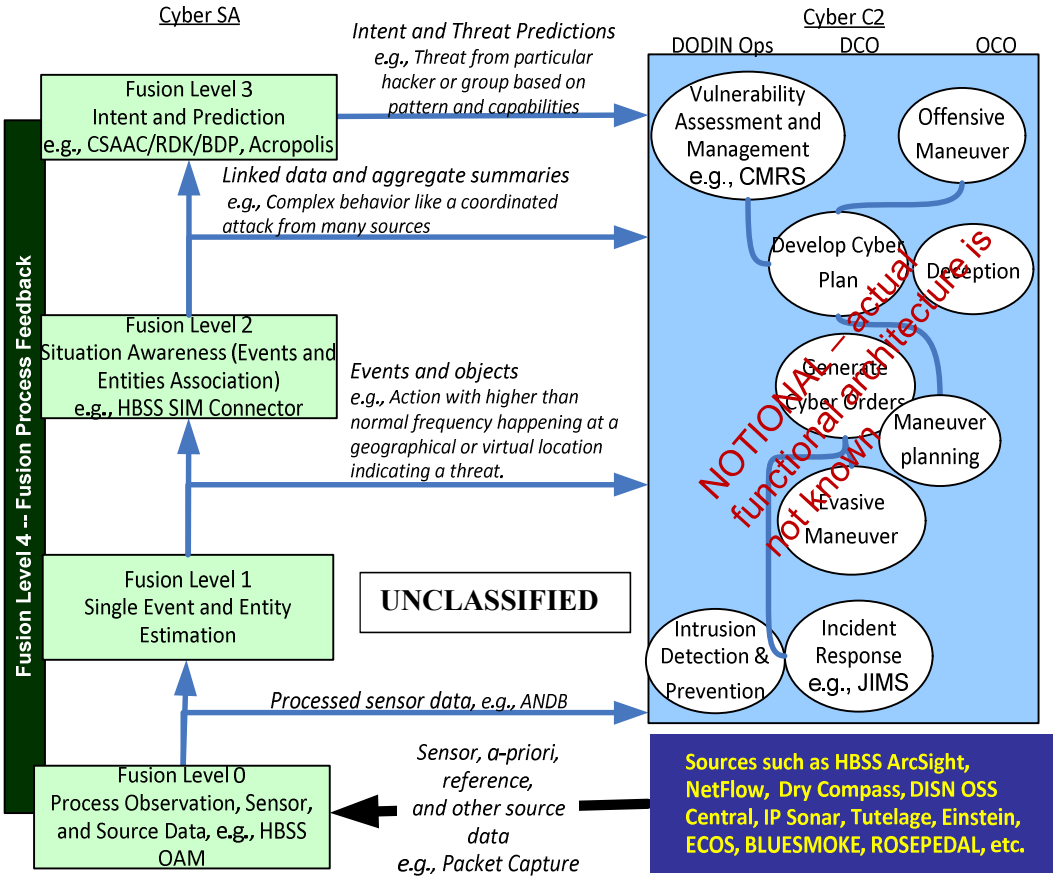


Figure 2. (U) Cyber Fusion Layers and Example C2 Functions

(U) In actual practice, the fusion is distributed over nodes and interoperates more like Figure 3. On the left an individual node's A-Box to T-Box reasoning takes place to determine if a set of sensor data is a typeInstance to T-Box object, event, behavior, TTP, CoA, campaign, and threat actor models. To the right, nodes collaborate at single-INT (outer circle) and then multi-INT levels (inner circle) to refine estimates. Where the cyber ontology (CybOnt) comes in is with the exchange of detailed and unambiguous – mathematically structured – information between the various nodes, National to and from BCT. In the sensor and data fusion world this is called Distributed Data Fusion (DDF) [10, 11] and, for the distributed and diverse algorithms to produce accurate estimates, it is essential that the exchanged data be unambiguous and interoperable.

(U) There has been substantial research on conducting data fusion over ontologically structured data [e.g., 12, 13, 14, 15]. The data fusion community has concluded that distributed data fusion algorithms require an unambiguous ontology so that algorithms and operators can work independently but in coordination via the ontology. Ontologists are looking for more than just structure, what is sought is structure that implies computation, e.g., $a \in A \subset B \Rightarrow a \in B$. These type of assertions are elementary to encode in OWL and entail in a general purpose reasoner but would be difficult to encode in a conventional data structure and would require manually constructed software for the entailment.

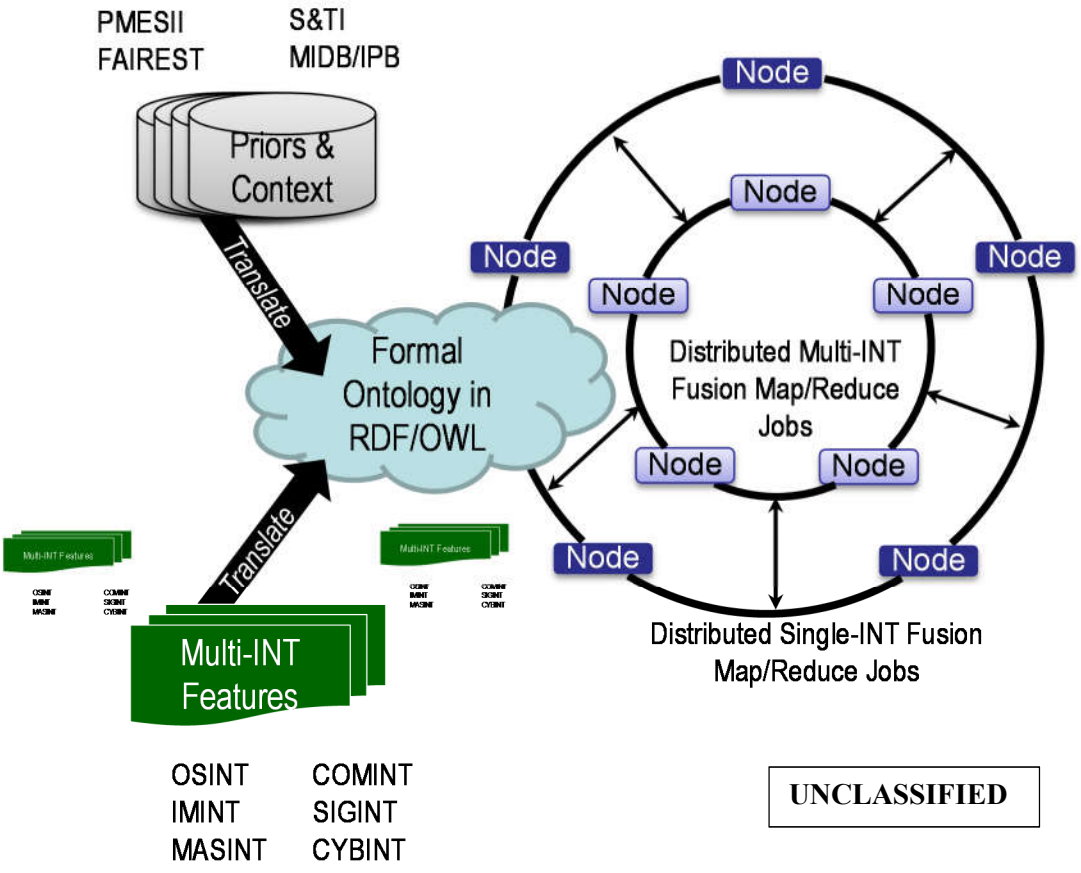


Figure 3. (U) A View of Distributed Cyber Fusion

(U) An analog of cyber fusion for classical Electronic Warfare (EW) is depicted in Figure 4. Reports from the sensors come in on the left. Its values are matched to modes in some database developed by Science and Technical Intelligence (S&TI) organization that then indicate the cause of the signal may be from some radar emitter (properties of its transmitter). Next, if a part causes something, then also does its whole, so we hypothesize about the platforms (or sites) that have that emitter, also knowable from other S&TI databases. Again, we use the same wholePart rule to infer the nation cause of the signal, inferable from operational intelligence databases. In the last step, we use conflicting or common desired effects between us and other nations to infer Friend (common desired effects), Hostile (contrary desired effects), or Neutral (neither). (Cause

in ontologies can be beforeAfter with very high probability (e.g., near 1) accompanied by a notBeforeAfter with very low probability (e.g., near 0)).

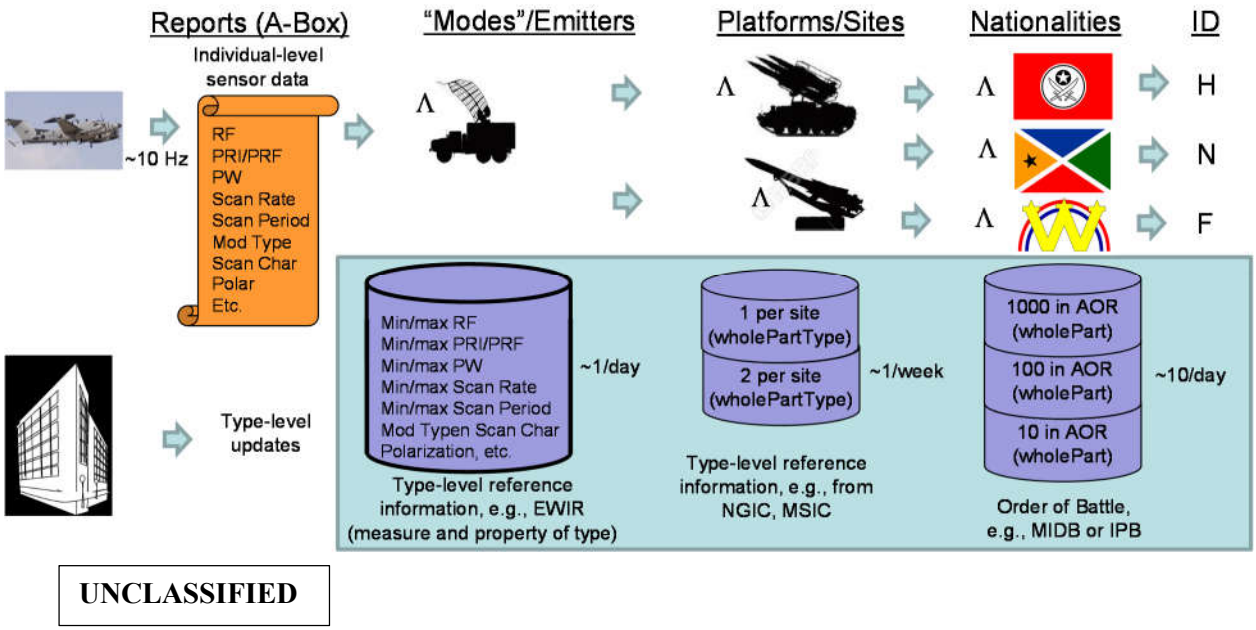


Figure 4. (U) Diagram for an ESM/ELINT Reasoner

(U) In EW this is often computed as a likelihood ratio times priors ratio to compute the likelihood that a detected signal is from an emitter type i as shown in the equation below. In other words, the ratio that the features would be caused by emitter_i to all other causes. The reason the likelihood ratio is so powerful is, 1) the ratio cancels out the P(features) for which there is no rational basis for computation, 2) ratios, and thus hypotheses, can be compared because the null hypothesis normalizes the competing positive hypothesis scores, 3) the scores are mathematically principled, unlike many Artificial Intelligence (AI) rule-based confidence assignment algorithms, so they scale and are adaptable. For ESM/ELINT, the ratio is:

$$\Lambda/\Lambda = \frac{L(\text{emitter}_i \mid \text{signal})}{\sum_{\iota \neq i} L(\text{emitter}_\iota \mid \text{signal})} = \frac{P(\text{signal} \mid \text{emitter}_i)}{P(\text{signal} \mid \neg\text{emitter}_i)} = \frac{\sum_{j \in i} P(\text{signal} \mid \text{mode}_j) P(\text{mode}_j \mid \text{emitter}_i)}{\sum_{\eta \notin i} P(\text{signal} \mid \text{mode}_\eta) P(\text{mode}_\eta \mid \neg\text{emitter}_i)}$$

(U) where the elements can be computed as:

$$\frac{P(\text{signalFeature}_\alpha \mid \text{mode}_{ij})}{P(\text{signalFeature}_\alpha \mid \neg\text{mode}_{ij})} = \frac{\int pdf_{\text{signalFeature}_\alpha} pdf_{\text{mode}_{ij}}}{\sum_{\iota \neq i} \int pdf_{\text{signalFeature}_\alpha} pdf_{\text{mode}_{i\eta}}}$$

(U) Analogously, the likelihood ratio for cyber indicators could be computed from observables and those could be conditionalized to compute likelihood ratios of cyber attack TTP, CoA, campaigns, and threat actors.

(U) We are also experimenting with incorporation of Political, Military, Economics, Social, Infrastructure, and Information (PMESII) [16], Faction-by-faction Actors, Institutions, Resources, Economics, Supra-system, and Timelines (FAIREST), and other contextual influences.

(U) In implementation, the DDF computations will be done as distributed processes, e.g., using Map Reduce distributed data processing. The mapper would put out the job to the Hadoop cluster and those with specified indicators relevant to the activity and actor types would return values that the Reducer would turn into the likelihood ratio. Standard practice is to concatenate relationship pairs in the id field of the key/value pair and have the value be the type of relationship. The process takes CybOnt, turns it into triples and then the triples are put into the NOSQL database as key value pairs. Map Reduce jobs process the database information.

(U) CybOnt supports the computations by having indicators, cyber activity types, actor types, priors, and context in a mathematically related form so the data lends itself to the computations. CybOnt's Individual measures, properties, and patterns (A-Box) are linked (e.g., by typeInstance) to the analogous Type-level measures, properties, and patterns (T-Box). Properties and Measures are common so Individual Properties and Measures and readily compared to Type-level Properties and Measures. The likelihood ratios relate the A-Box and T-Box typeInstance's probabilistically.

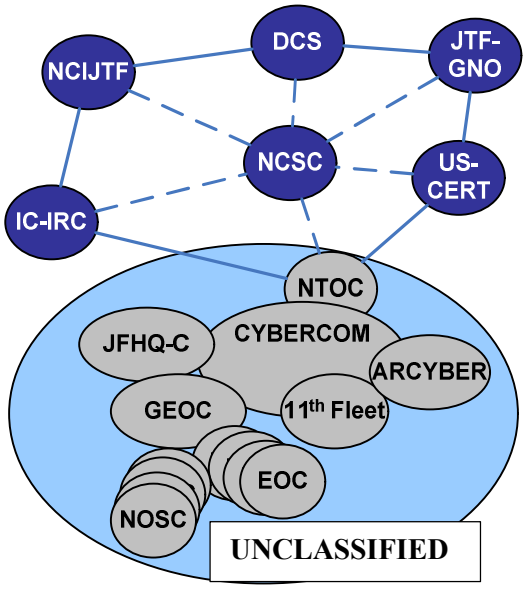
3 (U) Cyber Data Exchange

(U) CybOnt can be a means for data exchange between NetOps, DCO, and OCO operations centers. Because the data has mathematical meaning, distributed analytics should be interoperable and able to contribute to each other's detection, anticipation, response, and planning for cyber events and entities. The machine interpretability of the ontology will enable resilient automation that does not take as long to implement as today's human-readable automation specifications which require extensive testing and trial-and-error to get working right because of the inherent ambiguity in the language-based specifications. This is necessary now to get inside the threat cycle since cyber threats evolve at a faster pace and with greater obscurity than do the kinetic threats. At the Federal level, there will be a benefit from DoD's improved performance as well as DoD's ability to utilize Intelligence Community data (e.g., for sample see the Federal organizations shown in Figure 5.) This is also true at the national level and other mission partner levels. A side benefit of this work will be compliance with the National Defense Authorization Act (NDAA) requirement to harmonize the various cyber event schemas.

(U) There exist today data model and schema elements pertaining to NetOps and DCO such as Malware Attribute Enumeration and Characterization (MAEC) [17], Cyber Observable Expression (CybOX™) [18], Structured Threat Information eXpression (STIX™) [19], Common Weakness Enumeration (CWE™) [20], Common Vulnerabilities and Exposures (CVE) [21, 22], Common Attack Pattern Enumeration and Classification (CAPEC) [23], and others such as cyber device or software formats, heretofore collectively called "Cyber Schemas".

(U) While these are immensely valuable the nature and purposes of the cyber SA and C2 require more than just conventional databases and structures. More than other mission areas, cyber SA and C2 is distributed over many Joint and mission partner nodes. Examples of the required

interaction are shown in Figure 5. As well, the cyber SA process multi-source and multi-INT. In many cases no single detected event or activity is sufficiently suspicious to trigger counter action; that is, the enemy is assumed to be sophisticated enough to be silent, to "stay under the radar". Detection may occur as a result of multiple types of sensor and intelligence data or as the result of associating events and entities, i.e., data fusion. This requires an accumulation of detections and hypotheses for assured detection while minimizing false alarms. In a distributed environment, data exchanges between organizations need to be unambiguous and interoperable. As well, the reaction time for some cyber events necessitates efficiency in the interpretation of shared data and exchanged messages to support automatic and semi-automatic reactions



JFHQ-C	Joint Force HQ - Cyber
NTOC	NSA/CSS Threat Operations Center
GEOC	Global Enterprise Operations Center
EOC	Enterprise Operations Center
NOSC	Network Operations and Security Center
US-CERT/DHS	United States Computer Emergency Readiness Team
JTF-GNO/DoD	Joint Task Force—Global Network Operations Coordination
NCIJTF/FBI	National Cyber Investigative Joint Task Force
IC-IRC/DNI	Intelligence Community—Incident Response Center
DC3/DoD	Defense Cyber Crime Center

Figure 5. (U) Example of Required Cyber Interactions

(U) We use the qualifier "mathematical" in the phrase "mathematical ontology" to mean that in such an ontology, every Type (Class) has a distinct mathematical interpretation to differentiate from many so-called ontologies today differ little from conventional data models or schemas. We leveraged a high-level ontology we developed for the DoD CIO. The highest level of our foundation ontology is consistent with SUMO (Suggested Upper Merged Ontology) and ISO 15926. The ontology foundation is extensional, 4-dimensional, and higher order. Extensionality provides a way to identify and compare things unambiguously using physical existence as the criterion for identity. Four-dimensional [24] mereotopology provides a way to unambiguously and flexibly associate spatio-temporal parts both past and possible futures. Higher-order typing provides a way to employ differing levels of abstraction with hierarchies of patterns. Also signs and representations are separated from referents and relationships.

(U) There is parallelism between Individual and Type patterns, what in OWL is called A-Box (Individual) and T-Box (Type). A type of reasoning in OWL is classifying the A-Box into a T-Box, e.g., recognizing actual cyber attacks from archetype cyber attack models.

(U) In contrast, the Cyber Schemas are constructed as conventional XML Schema Descriptions (XSD) and are not founded on a formal ontology foundation. This means they do not have machine interpretable semantics and require human interpretation. This type of conventional approach will have the same types of interoperability problems that typically plague multi-node and multi-source operations. Human interpretations across that vast a spectrum of operators and systems engineers results in dis-interoperabilities. Algorithms and analytics will not work synergistically to accumulate pieces of information into reliable and actionable intelligence. They will require enormous trial and error testing just to get same types of algorithms to be interoperable.

(U) A mathematically-based ontology offers a better way to achieve cyber SA and C2 than can be supported by today's conventional data schemas / models and rule-based alerting systems. Formal logic (e.g., type, set theory), mereotopology (parts and boundaries), 4-dimensionalism, and other mathematical constructs are embedded in an ontology. So an ontology doesn't just say, column a is in table B (relational) or class C has attribute d (object); it is more precise, saying, $a \subset B$ or $d \in C$. Data schemas do not convey semantics, only syntax, so semantics have to be interpreted by humans looking at the names, descriptions, sample values, etc. For even modestly complex representations, the semantics typically interpreted differently. In contrast, ontologic semantics are computer implementable; indeed it is fairly easy to encode basic set theoretics into a computer. The common language of math can be computer-implemented so that computers may perform inferences allowing for more automation of processes, interoperability of systems, and increased shareability of heterogeneous data and information.

(U) Success in cyber warfare requires faster response and automation. The ontology that satisfies these requirements must be stronger than a data model or conventional schema. And while it is important that the ontology reflect a common perception of terms, for machine interpretation and processing it must, in addition, be supported by something a machine can understand, e.g., an underlying mathematical meaning. Consequently it is necessary to develop a mathematical cyber ontology that is founded on principled mathematics. Principled mathematics will reduce ambiguity and enable common understanding and machine interpretation of cyber indicators.

(U) CybOnt currently has high-level patterns for:

a. (U) Desired Effects
b. (U) Resource and Temporal Flow
c. (U) Information and Data
d. (U) Organizational Structure
e. (U) Capability
f. (U) Criticality and Risk
g. (U) Plans and Campaigns
h. (U) Systems and Services
i. (U) Geopolitical Extent

(U) We have partially developed and are continuing to develop CybOnt A-Box and T-Box patterns for:

a. (U) Cyber Indications Reports. In response to reconnaissance or other suspicious activity, provides details of reconnaissance or suspicious activities. T-Box and A-Box ontologies have been developed for the DoS / DDoS general pattern and more specific layers 2 through 7:

    1) (U) MAC flood indication
    2) (U) ARP spoofing indication
    3) (U) SMURF attack indication
    4) (U) SYN flood indication
    5) (U) Telnet DoS indication
    6) (U) Malformed SSL resource starvation indication
    7) (U) GET and POST resource starvation indication

b. (U) DCO Configuration Order / Advisory. In response to detected enemy capabilities, TTP, vulnerabilities, and activities, provides network, services, and applications configurations for defense and counter-exploit responses. An example was modeled for Perimeter Blocking Order/Advisory.

c. (U) Cyber Attack Order. Provides target(s) and method of attack. In general, these will be reverse STIX and MAEC telling Blue force what type of attack and parameters of attack. The contents of the order will be analogous to those of traditional kinetic engagement orders. .

d. (U) Cyber Attack Request. Provides rationale and requested response including mission criticality (rationale) and desired effects. This models mission criticality and risk in the current ontology since the request will have to provide that information to the Requestee in order for the Requestee to made the decision to accept and, if so, the manner in which the attack will be conducted. This also uses the Desired Effects model so the Requestor can specify the requested effects, not the means, so the Requestee can consider all counter attack options that could achieve the effect.

e. (U) Modus Operandi / TTP Advisory. Based on all-source analysis, enemy cyber modus operandi, tactics, techniques, and procedures. This could include the reverse of a Cyber Attack Order, i.e., what Red force would order. .

f. (U) Capabilities Assessment. Based on all-source analysis, assessment of enemy cyber capabilities. . This borrows heavily from the existing Capabilities and Desired Effects model since Desired Effects are the preferred way to express Capabilities in DoD.

g. (U) Attack Scenarios Advisory. Based on all-source analysis, enemy possible attack scenarios.

h. (U) Vulnerabilities Report. As a result of probing and analysis, friendly, enemy, and host nation vulnerabilities at all network layers and for services and applications. Like the Cyber Attack Request, this uses the mission criticality and risk model in the current

i. (U) Cyber Incident Report. After an attack has occurred, provides details of attack and attacker. Note this also uses Disjoint to represent that the Performers allowed access to the controlled Resource are disjoint from those not allowed.

j. (U) Communications Status Report. This report povides details of QoS such as packet loss, effective baud rate, SNR, Signal strength. For this we expect heavy harvesting from Cybox and maybe some existing tactical messages.

k. (U) NAVWAR Report. Information on location of attacker, techniques employed, signal power, CDMA SNR. For a jammer this may look a lot like a traditional jamming report (structured ontologically, of course.) The estimated Resource and mission target(s) of the attack might also be important. For deception, the technique being employed is probably important.

l. (U) Information Warfare Order. This report provides target(s) parameters and attack method(s) to employ such as jamming frequency, modulation, beamwidth, ERP, and duration.

4 CybOnt and Cyber Fusion CONOPS

(U) The CONOPS for the employment of CybOnt and related DDF for Cyber SA and C2 is shown in Figure 6. Streaming sensor data, becoming existant at many times per millisec, is extracted, translated, and loaded (made cloud available) into CybOnt's Individual level or what in OWL is called "A-Box". This is all in the cloud, i.e., source data does not move. The streams are real-time and high data rate and trigger algorithm execution and SA updates. At a much lower data rate CybOnt's Type level (in OWL, "T-Box") is updated, on the order of a few a day, based on new threat type updates. The Threat Analyst is aided by a Semantic Distance Algorithm (SDA) developed by Silver Bullet that computes the closeness of the unstructured threat reports to the existing CybOnt patterns. The distance is used to rank the CybOnt T-Boxes so the Threat Analyst can extend, specialize, modify, and reuse them for the new threat report. This automation assistance helps the Threat Analyst develop T-Box patterns more quickly and more accurately, i.e., without creating redundant and possibly inconsistent patterns. Natural Language Processing (NLP) is part of SDA but it is much more powerful because it uses a massive term frequency database as priors to compute likelihood ratios of matches in a mathematically principled way. SDA was developed by Silver Bullet in part for a national agency. T-Box updates also trigger batch-type algorithm SA updates for all the changed T-Box references. In the background (not shown), engineers continue using the CybOnt development methodology developed under this project to produce new weekly releases of CybOnt and associated extractors, translators, and algorithms.

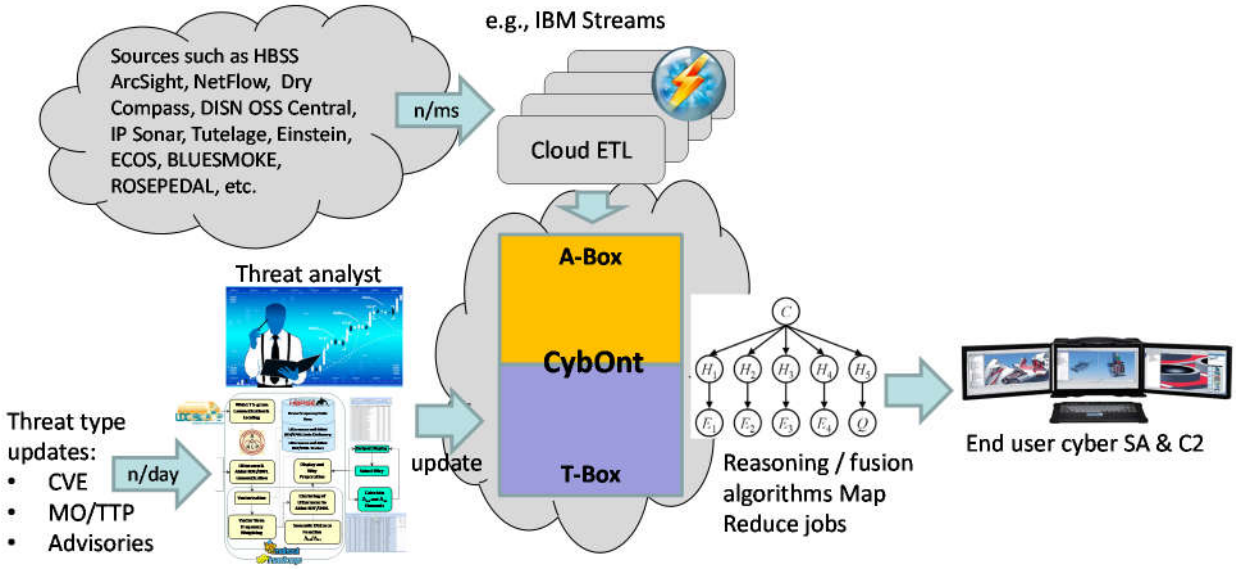


Figure 6. (U) High-level CONOPS for CybOnt Employment

5 (U) References

[1] (U) David L. Hall, Martin Liggins II (Editor), James Llinas (Editor), Handbook of Multisensor Data Fusion: Theory and Practice, Second Edition, CRC Press, (2012)

[2] (U) Yaakov Bar-Shalom, X. Rong Li, Thiagalingam Kirubarajan, Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software, Wiley, (2001)

[3] (U) Pearl, Judea; Probabilistic Reasoning in Intelligent Systems: Patterns of Plausible Inference; 1988

[4] (U) Sindhu Raghavana, Parag Singlab and Raymond J. Mooney, "Plan Recognition Using Statistical– Relational Models", in Plan, Activity, and Intent Recognition: Theory and Practice, Gita Sukthankar, Christopher Geib , Hung Hai Bui, David Pynadath, Robert P. Goldman (eds.), Elsevier Science, 2014.

[5] (U) Lawrence A. Klein, Sensor and Data Fusion Concepts and Applications, SPIE Press, 1999

[6] (U) Office of Naval Technology, "Functional Description of the Data Fusion Process", Data Fusion Development Strategy, Office of Naval Technology, November, (1991)

[7] (U) Steinberg, A. N., Bowman, C. L., White, F. E., "Revisions to the JDL Data Fusion Model", http://www.dtic.mil/dtic/tr/fulltext/u2/a391479.pdf

[8] (U) Murphy, T.H., Kraft, T., "A Conceptual Control Model for Discussion Combat Direction System Architectural Issues", in Proceedings of the Fourth MIT/ONR workshop on Command, Control, and Communications Systems, 1981

[9] (U) Christopher Bowman, "The Dual Node Network (DNN) Data Fusion & Resource Management (DF&RM) Architecture", AIAA 1st Intelligent Systems Technical Conference, 2004
Read More: http://arc.aiaa.org/doi/abs/10.2514/6.2004-6288

[10] (U) Chee Yee Chong, D. Hall, M. Liggins and J. Llinas (editors), Distributed Data Fusion for Network-Centric Operations, CRC Press, Nov 2012

[11] (U) Pramod K. Varshney, Distributed Detection and Data Fusion, Springer, 1997

[12] (U) Waltz, Edward; Ontologies and Data Fusion; Second Annual CMIF Workshop on Critical Issues in Information Fusion; Beaver Hollow, NY; October 2003

[13] (U) Boury-Brisset, Anne-Claire; Ontology-based Approach for Information Fusion; Proceedings of the International Sensor and Information Fusion conference; ISIF; 1993.

[14] (U) McDaniel, D.M., Regian, J.W., and Schaefer, G., "Ontology Based Fusion for E-2D", in *Proceedings of the National Symposium on Sensor and Data Fusion*, Military Sensing Information Analysis Center (SENSAIC), 2005
[15] (U) Mieczyslaw M. Kokar, Christopher J. Matheus, Kenneth Baclawski, "Ontology-based situation awareness", Information Fusion, Vol 10, Elsevier, 2009
[16] (U) R. Hillson, "The DIME/PMESII Model Suite Requirements Project", 2009 NRL Review, NRL, (2009) Information Technology Division
[17] (U) https://maecproject.github.io/
[18] (U) http://cyboxproject.github.io/
[19] (U) http://stixproject.github.io/
[20] (U) https://cwe.mitre.org/index.html
[21] (U) https://cve.mitre.org/
[22] (U) https://nvd.nist.gov/
[23] (U) https://capec.mitre.org/
[24] (U) Sider, Theodore, Four-Dimensionalism: An Ontology of Persistence and Time, Oxford University Press, (2003)