The Notion of Global Data Fusion and its Application to Cyber Security

Mario LA MANNA
Evoelectronics
Rome, Italy

ABSTRACT

Global Data Fusion is one of the main technologies used in complex systems. While the application of data fusion has already been proposed for the implementation of specific tools, its extension to the overall design process of a complex system is far from a desired target. The development of advanced architectures based on an interdisciplinary design approach makes this extension possible, especially at the higher levels of the architecture, involving situation assessment, impact assessment and process refinement. This paper analyses one of the advanced cyber security architectures and explores the capability of this architecture to include data fusion tools at the top level of the architecture. The effects of the generalisation of data fusion techniques are then analysed and the consequent improvements in the network security of critical infrastructures are described and quantified.

Keywords: Data fusion, cyber security, situational awareness, environment monitoring, critical infrastructures, network security.

1. INTRODUCTION

An advanced security system for critical infrastructures (Fig. 1) is composed of the following subsystems: a) Sensor subsystem, which collects data coming from the external environment; b) Data and Information Fusion subsystem, which merges data collected by sensors and data and information coming from external intelligence sources; c) Human Agent in the Loop, which performs operations by means of a human operator in the decision loop; d) Core Processor, which combines all the information produced by the previous subsystems with internal data extracted by statistical learning, on the basis of the historical database.
The system is based on an interdisciplinary (human/machine) approach, having to cope with the heterogeneity of the data produced by the system itself, those collected by the monitoring tools and secure information coming from different sources. Monitoring data are derived from different types of sensing units, while secure information comes from intelligent external sources, human in the loop agents and internal system intelligence. Specific intelligence data and information travel through Internet, by analysing messages exchanged through the social networks. For the above reasons, the application of Data Fusion to cyber security systems for critical infrastructures needs a global approach and its extension to the complete design process.

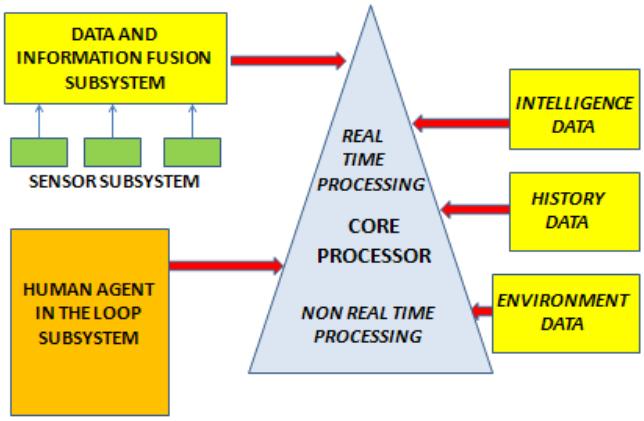

Fig. 1: Advanced Cyber Security Architecture for Critical Infrastructures.

According to the JDL definition of Data Fusion [7] (Fig. 2), Data Fusion is a multilevel process, with five different levels, ranging from Level 0 (lowest level) up to Level 4 (highest level).

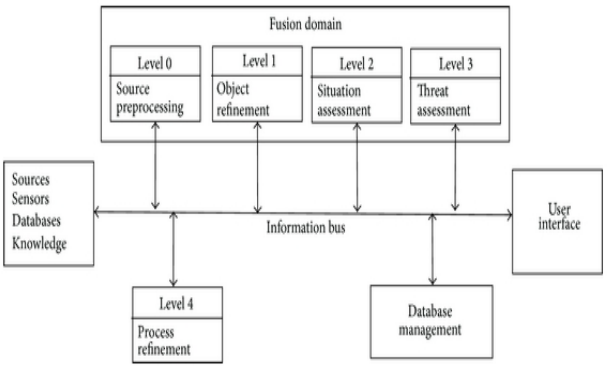

Fig. 2: The JDL Data Fusion Framework.

Level 0, Source Preprocessing: includes the information extraction process. This level reduces the amount of data and maintains useful information for the high-level processes.

Level 1, Object Refinement: employs the processed data from the previous level. The output results of this stage are the object discrimination (classification and identification) and object tracking (state of the object and orientation). This stage transforms the input information into consistent data structures.
Level 2, Situation Assessment: establishes relationships between the objects (i.e., proximity, communication), to determine the significance of the entities or objects in a specific environment. The aim of this level includes performing high level inferences and identifying significant activities and events (patterns in general). The output is a set of high-level inferences.
Level 3, Impact Assessment: evaluates the impact of the detected activities in order to perform a future projection and identify possible risks, vulnerabilities, and operational opportunities. This level includes an evaluation of the risk or threat and a prediction of the logical outcome.
Level 4, Process Refinement: provides resource and sensor management. The aim is to achieve efficient resource management while accounting for task priorities, scheduling, and the control of available resources.
The JDL model allows Data Fusion techniques to be extensively employed on multisensor environments with the aim of merging and aggregating data from different sensors. Data Fusion in multisensor environments makes it possible to obtain better detection probability and higher reliability by using data from multiple distributed sources. The application of Data Fusion techniques to the highest levels of the architecture (Global Data Fusion) is the winning methodology to design robust and secure systems, by overcoming the drawbacks still existing in traditional cyber security applications.

In the next sections, we first review the application of Data Fusion to Level 0 and Level 1 (already implemented in the present state of the art systems). Then we discuss the application of Data Fusion techniques to the higher levels, namely Level 2, Level 3 and Level 4 (Global Data Fusion concept), taking into account the above described architecture. The graphics reported in the rest of the paper are the result of merging and simplifying the data collected in a series of experiments on Global Data Fusion applied to critical infrastructures. The main scope of this discussion is to assess the improvement achievable in a cyber security system by adopting the Global Data Fusion concept.

2. LEVEL 0 AND 1 : SOURCE PREPROCESSING AND OBJECT REFINEMENT

The main goals of the low JDL levels (Level 0 and Level 1) are to identify, detect and characterize the system environment, which includes cyber entities (e.g. computers, networks, data flow, etc.), their relative information (e.g. operating systems, hardware, patches, etc.) and intrusion detection data (e.g. security logs, adversary presence data, etc.). The system environment data would aid either a cyber defense automated system or a human analyst in being able to manage a limited amount of detected attacks. In order to be more effective, Data Fusion has to be extended at the higher levels, as explained in the next sections.

3. LEVEL 2: SITUATION ASSESSMENT

The Level 2 process combines the multiple features of a multisensor network into a comprehensive picture of the current situation. In particular, this process provides an overall understanding of the current state of the system. The system data, such as operating system patches, installed antivirus software definitions, list of running processes and other security related data give an estimation of the system robustness and its ability to counteract a known set of attacks. The representation of all these data for the whole network gives a precise estimation of the level of awareness in the current system state.

In the general case, system awareness is a combination of two different factors, namely system health assessment and knowledge of attacker capabilities. The algorithms applied in Level 2 Data Fusion are mainly pattern matching and machine learning. The assessment about the system health (e.g. patch level, antivirus definition, etc.) can be analysed with respect to a known desired security state. Suitable actions have to be carried out either when the desired state is not matched or if the system exhibits undesired behaviour like high CPU load, low available drive space, high network traffic, etc. As a result of these actions, the final assessment recognizes a suitable level of the network defensive posture.

Moreover, there must be a precise understanding of the capabilities of the potential attacker, including the estimation of the kind of attacks that are more likely to happen, by learning from historical data.

Level two data fusion represents an advance beyond the creation of raw sensor data, as occurs at the first level, and supports the synthesis of more meaningful information for guiding human decision-making. Bayesian decision theory is one of the most common techniques employed in level two data fusion. It is used to generate a probabilistic model of uncertain system states by consolidating and interpreting overlapping data provided by several sensors. It also determines conditional probabilities from a priori evidence. On this level is used one of two most popular techniques which are: Bayesian Decision Theory and Dempster-Shafer Evidential Reasoning.

The Bayesian Networks Bayesian networks are useful for both inferential exploration of previously undetermined relationships among variables as well as descriptions of these relationships upon discovery.

The Dempster-Shafer method has several advantages over Bayesian decision theory. Most importantly, hypotheses do not have to be mutually exclusive, and the probabilities involved can be either empirical or subjective. As Dempster-Shafer sensor data can be reported at varying levels of abstraction, a priori knowledge can be presented in different formats.

4. SYSTEM AWARENESS

The combination of the known state of the defensive posture and the capability set of the potential attackers results in the definition of the awareness of the current system security level. System awareness involves three critical areas, namely computing and network components, threat information and mission dependencies. Achieving a high level of system awareness requires to focus on data collection, data management and environment analysis, in order to get a real time picture of the scenario, in particular about computer systems, networks and users. System awareness includes three sub topics, namely network awareness, threat awareness and mission awareness. Network awareness: disciplined asset and configuration management, routine vulnerability auditing, patch management and compliance reporting, recognize and share incident awareness across the organization. Threat awareness: identify and track internal incidents and suspicious behavior, incorporate knowledge of external threats, participate in cross-industry or cross-government threat-sharing communities on possible indicators and warnings. Mission awareness: develop a comprehensive picture of the critical dependencies to operate in cyberspace, understand these critical dependencies to support mission impact in forensic analysis (after a situation); triage and real time crisis action response (during a situation); risk readiness assessment prior to task execution (anticipating and avoiding situations) and informed defense planning (preparing to mitigate the impact of a future situation).

A general pattern of the system awareness as a function of the number of environment parameters under control is reported in Figure 3, where the system awareness ranges from 0 (no awareness) to 1 (complete awareness) and the environment parameters include both system health parameters and external threat parameters.

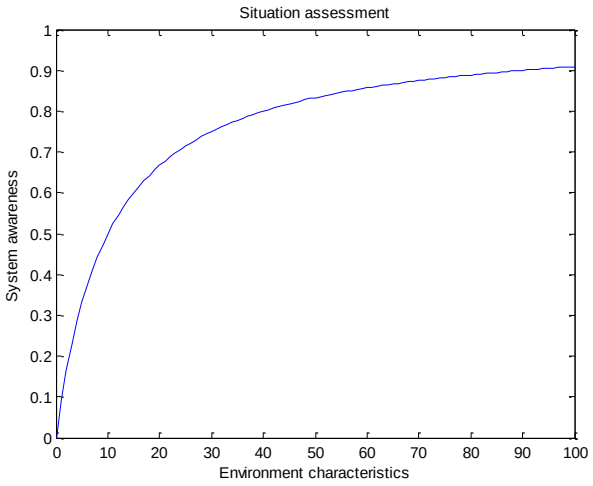

Fig. 3: System awareness as a function of the multiplicity of environment parameters under control.

5. LEVEL 3: IMPACT ASSESSMENT

The Level 3 process has the goal to coordinate the defensive action of the network after a suitable evaluation of the possible attacker's options. This task relies on the system awareness gained from Level 2 and exploits a further knowledge about common vulnerabilities and exposures of the network, together with a deep understanding of the attacker's strategies. An automated fusion process combines the knowledge about the capabilities of the attack deriving from previously learned strategies with the estimation of the current health of the network components. As a further step, the security system, with the possible intervention of the human analyst (through a human in the loop sub-process) provides additional actions to recover the system and neutralize the effects of the attack. In order to carry out the defense action, it is paramount to understand as much as possible the possible attacker's strategies and the multiplicity of defense capabilities of the system.

The Level 3 process extends the current situation into the future to draw inferences about threats and opportunities for operations. The most used techniques in this level are: Expert Systems, Blackboard Architecture and Fuzzy Logic. An expert system is regarded as the personification within a computer of a knowledge based component from an expert skill in such a form that the system can offer intelligent advice or take an intelligent decision about processing function. A blackboard system consists of three major components: the software specialist modules, providing specific expertise needed by the application, the blackboard, namely a shared repository of problems, partial solutions, suggestions, and contributed information and the control shell, which controls the flow of problem solving activity in the system. Fuzzy Logic is a mathematical technique for dealing with imprecise data and problems that have many solutions rather than one. Fuzzy logic is derived from fuzzy set theory dealing with reasoning that is approximate rather than precisely deduced from classical predicate logic.

6. PROBABILITY OF RECOVERY

The impact assessment/ threat refinement analysis contains a number of threat perspective models, in order to derive an estimation about the probability of recovery, in case of attack. These models form a central repository of threat intent inference information. Each model is concerned with reasoning about the adversary's strategy from a single analysis perspective. Each threat perspective model is represented by a hierarchical graph structure. The graphs provide a structure for collecting factlets under a specific perspective and then reasoning as to the likelihood of one or more adversary strategy. The computational mechanisms to perform the fusion of factlet evidence within each graph are provided by the Data Fusion at this level.

A typical situation, expressed in terms of probability of recovery, as a function of the threat capabilities and the multiplicity of the network defense resources, is represented

in Figure 4, where the probability of recovery, ranging from 0 (no recovery) to 1 (full recovery), is a function of both the threat capability and the multiplicity of the network defense capabilities.

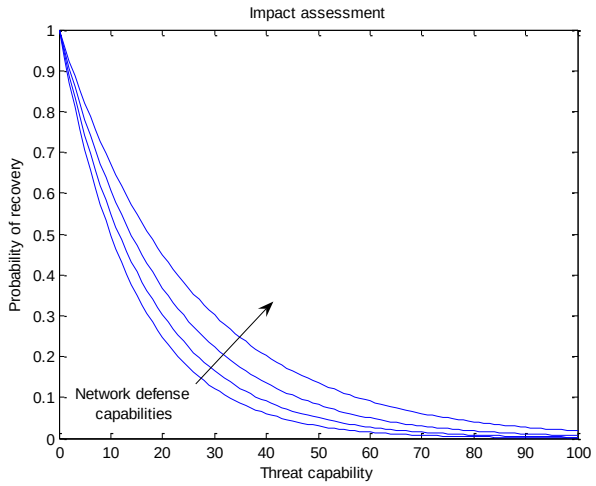

Fig. 4: Probability of recovery as a function of threat capability and network defense capabilities.

7. LEVEL 4: PROCESS REFINEMENT

The Level 4 process is mainly a decision process, regarding the observation of the overall data fusion system and the selection of the capabilities to detect new attack methods. This decision process includes monitoring for specific types of traffic in the network, such as connections to unknown hosts or servers in foreign countries, or high number of connections through the same ports, etc. As a consequence of this investigating activity, which can also be carried out in a human in the loop mode, the Level 4 process can deploy specific monitoring tools to make the same process more efficient.

According to Hall & McMullen (2004) human computer interaction (HCI) research in the fusion domain has mainly considered interaction between the user and a geographical information display (based on a geographical information system) through menus and dialogs. However, the current research interest in this area is growing, and techniques such as gesture recognition and natural language interaction are currently of interest.

8. PROBABILITY OF BLOCKING THREATS

Process refinement can then be defined as a meta-process or as a decision making task, taking viewpoint from decision theory, determining the most appropriate sensor action to be taken in order to achieve maximum utility. The application of this level to the whole Data Fusion process serves to increase the probability of blocking threats, by knowing the possible threat options and customizing the active defense tools in order to deploy the most effective monitoring tools.

A general pattern of the probability of blocking threats as a function of the number of options owned by the threats under control is reported in Figure 5, where the threat defeating probability, ranges from 0 to 1, is a function of both the multiplicity of threat options and the multiplicity of monitoring tools deployed by the system.

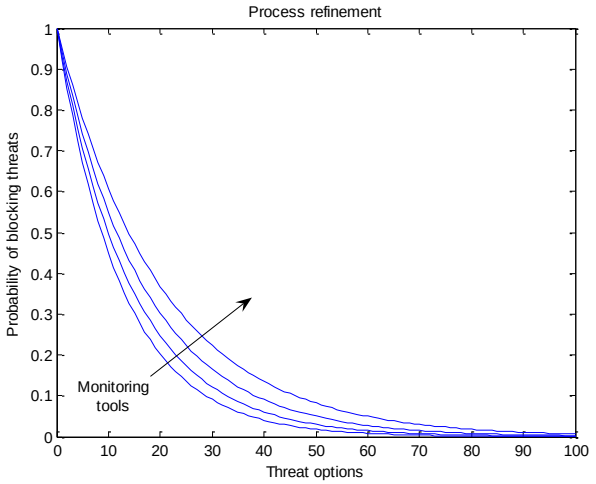

Fig. 5: Probability of blocking threats as a function of threat options and monitoring tools.

9. PERFORMANCE EVALUATION

In the last sections, we have explored the capability of a multisensor architecture based on an interdisciplinary design approach to exploit the Global Data Fusion concept. The effects of this new type of data fusion techniques have been analyzed step by step, in order to gather some measure of the improvement in the network security of critical infrastructures, as a consequence of the application of these techniques. This section has the goal to refine this estimation, relying on the previous findings about system awareness, probability of recovery and probability of blocking threats.

In order to obtain the final results, we will briefly resume the achievements contained in the previous four sections (Section 2, Section 3, Section 4 and Section 5).

In Section 2, we reviewed the tasks performed at the lowest Data Fusion levels. In Section 3, we showed that system awareness grows up with the multiplicity of environment parameters under control. System awareness includes both knowledge of the threat capabilities and knowledge and control of defense capabilities. As much as the network is aware of threat capabilities, as more efficiently the network can deploy its defense capabilities. In Section 4, we showed that, when controlling both threat capability and defense capability, the network can reach a high level of probability of recovery from external attacks. A fast recovery from external attacks corresponds to leave the attacker with less options

available. In Section 5 we discussed about the network capacity of blocking the maximum number of threats that can attack the system. Leaving the threat with less options and using as much as possible suitable monitoring tools contribute to achieve more efficiency in blocking threats and, as a consequence, to reach a higher level of security.

After merging the results of Section 2, Section 3, Section 4 and Section 5, we produce a final estimation, which is graphically represented in Fig. 6. In particular, we evaluate the increased network security as a function of the data fusion capabilities of the network and the number of threats that can potentially attack the system. The final estimation shows that the network security can increase sensibly as much as the system can rely on high number and quality of data fusion capabilities and as much as the system can exploit a high level of threat detection capability.

10. INCREASED NETWORK SECURITY

The effect of the application of the highest levels of the Data Fusion techniques have been analyzed step by step, in order to find the improvements in the network security of critical infrastructures as a consequence of this improved process.

A general pattern of the increased network security as a function of the Data Fusion capabilities and the threat detection capability is reported in Figure 6.

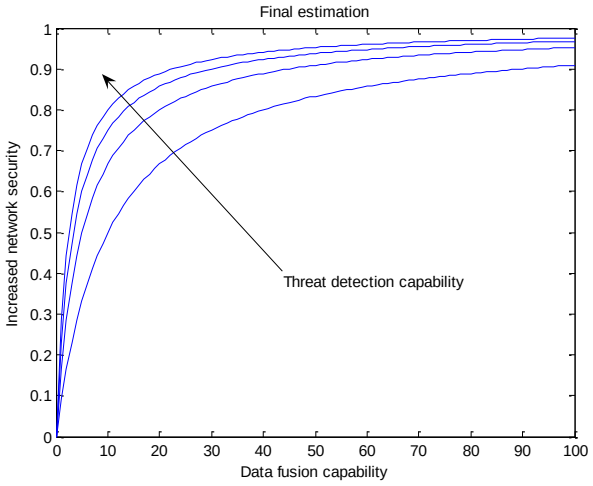

Fig. 6: Increased network security as a function of data fusion capabilities and threat detection capability.

We observe that, in a general case, where the data fusion techniques described in this paper are not applied, the network security is estimated as having performance figures belonging to the left bottom section of the graphic in Figure 6. When applying our fusion strategies, the network security grows up and reaches the top sector of the same graphic. An additional consideration, that was already introduced in the last sections, regards the cooperation between human and machine, which is necessary to achieve the estimated results. Even if the human contribution cannot be easily embedded in our measurements, the cooperation between human and machine is decisive to achieve the maximum level of security.

11. CONCLUSIONS

The implementation of the full range of data fusion capabilities for cyber security is still an open issue. Data Fusion algorithms need to be developed at all levels of the JDL model. The recent development in advanced cyber security architectures, which makes use of a novel interdisciplinary approach, allows to extend Data Fusion to the higher levels of the JDL architecture, namely situation assessment, impact assessment and process refinement. This paper focuses on a proposed cyber security architecture and explores the effectiveness of this extension at the top level of the architecture.

With regard to situation assessment (Level 2 in the JDL model), a general pattern of the system awareness as a function of the number of environment parameters under control is derived, where system awareness is found as strongly dependant from the accurate knowledge of the environment characteristics.

Impact assessment (Level 3 in the JDL model) corresponds to the Data Fusion capability which coordinates the defensive action of the network by means of the knowledge of the attacker techniques. Our analysis shows that the probability of recovery can be strongly increased by a suitable number of defense measures and consistently decreased by the threat capabilities.

Regarding process refinement (Level 4 in the JDL model), the goal is to analyse how much the probability of blocking threats can depend on the threat capabilities and on the defense capabilities.

At the end of the overall analysis, we show that the network security can be consistently increased by applying more data fusion capabilities and by incrementing the network capability of threat detection through a suitable combined process based on the Global Data Fusion concept.

12. REFERENCES

[1] M. LaManna ”Urban Environment Monitoring: System and Technology Issues”, IMCIC 2012, 25-28 March 2012, Orlando, FL.

[2] M. LaManna “Data Fusion of Heterogeneous Sensors in Urban Environment Monitoring” IMCIC 2013, 9-12 July 2013, Orlando, FL.

[3] M. LaManna “Cost-Benefit Analysis of Automated Systems for the Control of Urban Critical Infrastructures” WMSCI 2015, 12-15 July 2015, Orlando, FL.

[4] M. LaManna “Future Trends for Cyber Security for Critical Infrastructures” WMSCI 2016, 5-8 July 2016, Orlando, FL.

[5] M. LaManna “Technology Intercepts for Cyber Security applied to Critical Infrastructures” WMSCI 2017, 8-11 July 2017, Orlando, FL.

[6] JDL, Data Fusion Lexicon. Technical Panel For C3, F.E. White, San Diego, Calif, USA, Code 4^{20}, 1991.

[7] S. Schreiber-Ehle, W. Koch “The JDL model of data fusion applied to cyber-defence - A review paper”, Workshop

on Sensor Data Fusion: Trends, Solutions, Applications, Bonn, 4-6 September 2012.

[8] T. Armending: “Cybercrime: much more organized”, CSO Report, 23 June 2015.

[9] I. Hashem et al. “The rise of Big Data on Cloud Computing: review and open research issues”, Information Systems, Vol. 47, January 2015.

[10] B. Berkowitz “Intelligence for the Homeland.” SAIS Review of International Affairs 24, no. 1, 2004.

[11] W.Karwowski “International encyclopedia of ergonomics and human factors”, CRC Press, 2006.

[12] Kalyan Veeramachaneni, CSAIL,MIT et al. “AI2: Training a big data machine to defend.” AI2 IEEE International Conference on Big Data Security, April 2016, New York.

[13] I. Arel, D. C. Rose, and Thomas P. Karnowski “Deep Machine Learning. A New Frontier in Artificial Intelligence Research” IEEE Computational Intelligence Magazine, 2013.

[14] G. Dahl, W. Stokes, Li Deng, Dong Yu, “Large-Scale Malware Classification using Random Projections and Neural Networks”, IEEE Conference on Acoustics, Speech, and Signal Processing, 2013.

[15] E. Blasch, et al. “High-Level Information Fusion Management and System Design”. Norwood, MA: Artech House Publishers, 2012.

[16] P. Shinkman: “Reported Russian Cyber Attack Shuts Down Pentagon Network”, US News, 6 August 2015.

[17] E. David, “Deep Learning for Automatic Malware Signatture Generation and Classification”, IEEE Intl. Conference on Neural Networks, Killarney, Ireland, July 2015.

[18] Andler, S. F. Information Fusion from Databases, Sensors and Simulations, Annual Report 2005, June 2006.

[19] Hall, D. & Llinas, J. Handbook of multisensor data fusion. CRC Press.

[20] Hughes, T.J. “Sensor Fusion in a Military Avionics Environment.” Measurement and Control. Sept. 1989.

[21] Hall, D. & McMullen, S.A.H. (2004) Mathematical techniques in multisensor data fusion. Artech House.

[22] Hughes, T.J. “Sensor Fusion in a Military Avionics Environment.” Measurement and Control. Sept. 1989

[23] Ramsvik, H. AIS as a tool for Safety of Navigation and Security - Improvement or not?

[24] Svensson, P. Technical survey and forecast for information fusion. In: RTO IST. Symposium on Military Data and Information Fusion. 20-22 October, 2003.

[25] Wald L., 1999, Some Terms of Reference in Data Fusion, IEEE Transactions on Geoscience and Remote Sensing Vol.37 No.3 May 1999.