UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

Tesi di dottorato di ricerca in

Ingegneria Informatica ed Automatica

Ciclo XXVI

Marzo 2014

An approach for joint estimation of physical and logical security by semantic modelling

---

Mariana Esposito

Supervisors:

Prof. Nicola Mazzocca

Ing. Concetta Pragliola

Coordinator :

Prof. Francesco Garofalo

thesis submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy

in the

Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione

UNIVERSITY OF NAPLES “FEDERICO II"

Abstract

Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione

Doctor of Philosophy

An approach for joint estimation of physical and logical security by semantic modelling

by Mariana Esposito

Key activities in critical systems are the monitoring, observation and comprehension of different phenomena, aimed at providing an updated and meaningful description of the monitored scenario, as well as its possible evolutions, to enable proper decisions and countermeasures for the protection and safety of people and things. The threats coming from many different sources, internally and externally. The diffusion of new technologies have made more accessible the assets of a system.

In this thesis we demonstrate that the use of a semantic model for the information management it is suitable in order to meet these issues. In particular, thesis proposes and implement a methodology and approach for the early situation awareness recognizing a threat situation on time, for decision support to automatically activate recovery strategies. The threat on which the thesis focus on are regarded the logical and physical security. In particular for the logical security estimation will be presented a an approach guided by metrics. Then will be presented some results and example of real application.

Marzo 2014

Preface

Some of the research and results described in this Ph.D. thesis has undergone peer review and has been published in, or at the date of this printing is being considered for publication in, academic journals and conferences. In the following list all the papers developed during my research work as Ph.D. student:

- Mariana Esposito, Inaki Eguia, Francesco Flammini, Alfio Pappalardo and Erkuden Rios, "Formalizing SPD metrics for Embedded Systems Multilayer approach", *Second Eastern European and Mediterranean Software Process Improvement Conference (EuroMed SPI II),* Zamudio, Spain, October 6-7, 2011.
- Valentina Casola, Mariana Esposito, Francesco Flammini, and Nicola Mazzocca, "Monitoring railway infrastructures, a case-study for the pShield project" *The 2nd International Workshop on Mobile Commerce, Cloud Computing, Network and Communication Security 2012 at IMIS 2012."*
- Casola V., De Benedicitis A., Drago A., Esposito M., Flammini F., Mazzocca N. "Securing Freight Trains for Hazardous Material Transportation: a WSNBased Monitoring System*". Proceedings of International Defense and Homeland Security Simulation Workshop (DHSS 2012),* Vienna (Austria), September 2012 (I3M Multiconference)
- Amato F., Casola V., Esposito M., Mazzocca N., Mazzeo A. "A Smart Monitoring System Based on a Fast Classifier and a Semantic Post Reasoner". *Proceedings of International Defense and Homeland Security Simulation Workshop (DHSS 2012),* Vienna (Austria), September 2012 (I3M Multiconference)
- Casola, V., Esposito, M., Flammini, F., Mazzocca, N., & Pragliola, C. (2013, January). Performance Evaluation of Video Analytics for Surveillance On-Board Trains. *In Advanced Concepts for Intelligent Vision Systems (pp. 414-425). Springer International Publishing*.
- Buemi, F., Esposito, M., Flammini, F., Mazzocca, N., Pragliola, C., & Spirito, M. (2013). Empty Vehicle Detection with Video Analytics*. In Image Analysis and Processing–ICIAP 2013* (pp. 731-739). Springer Berlin Heidelberg.

- Amato, F., Casola, V., Esposito, M., Mazzeo, A., & Mazzocca, N. (2013). A smart decision support systems based on a fast classifier and a semantic post reasoner. *International Journal of System of Systems Engineering, 4(3), 317-336.*
- M. Esposito, F. Flammini, A. Fiaschetti: "The New SHIELD Architectural Framework ". *In: ERCIM News No. 93, April'13, Special Issue on Mobile Computing: pp. 53* (ERCIM EEIG, Sophia Antipolis Cedex, France, ISSN: 0926-4981).
- G. Garibotto, P. Murrieri, A. Capra, S. De Muro, U. Petillo, F. Flammini, M. Esposito, C. Pragliola, G. Di Leo, R. Lengu, N. Mazzino, A. Paolillo, M. D'Urso, R. Vertucci, F. Narducci, S. Ricciardi, A. Casanova, G. Fenu, M. De Mizio, M. Savastano, M. Di Capua, A. Ferone: *White Paper on Industrial Applications of Computer Vision and Pattern Recognition. ICIAP (2) 2013*: 721-730

Index

Preface ... **III**
Index ... **V**
Index of tables ... **VII**
Index of figures ... **VIII**
Introduction ... **9**
Chapter 1 ... **12**
1.1 The problem of CIP ... 12
1.2 Application Domain ... 13
1.3 Critical Infrastructure the Event Cycle ... 18
1.4 Critical Infrastructure: Threats and countermeasures ... 20
1.5 Security of Critical Infrastructure ... 25
1.6 Thesis Contributions ... 26
Chapter 2 ... **28**
2.1 Semantic interoperability: XML and RDF ... 28
2.2 Semantic Formal Language ... 31
2.2.1 The ontology ... 33
2.3 Semantic Technologies ... 35
2.4 Ontology representations ... 37
2.4.1 Semantic Web Ontology languages ... 40
Chapter 3 ... **43**
3.1 Physical and Logical security convergence ... 43
3.2 Complex Event Processing ... 46
3.3 Decision Support System ... 49
3.3.1 Semantic and Ontological models in DSS and data integration ... 50
Chapter 4 ... **54**
4.1 Why semantic model? ... 54
4.2 A model for monitoring system ... 55
4.3 The system architecture ... 58
4.4 Data model and processing ... 63
4.5 A typical case study in Railways Security Domain ... 65
Chapter 5 ... **72**
5.1 The New SHIELD Architectural framework ... 72
5.2 nSHIELD Attack Surface Metric ... 74
5.3 SPD level ... 75
5.3.1 Porosity ... 75
5.3.2 Controls ... 76
5.3.3 Limitations ... 78
5.4 Mathematical model of SPD Level ... 79
5.5 The nShield attack surface metrics ontology ... 82

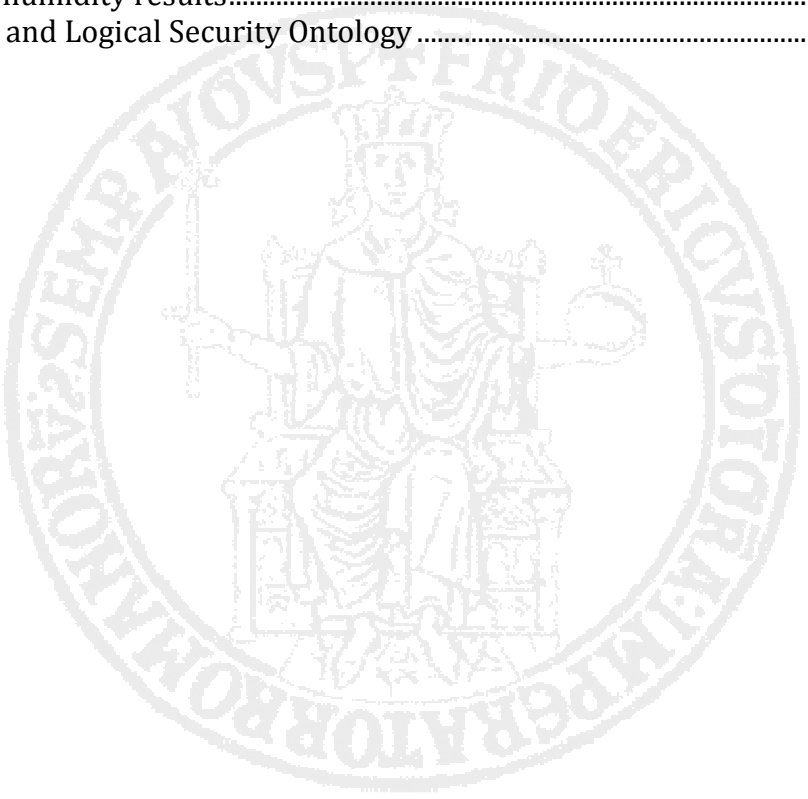
Chapter 6 ..... **85**
6.1 Physical Security a WSN Application: Post Reasoner application 85
6.2 Joint estimation of physical and logical security 88
6.2.1 Integration as event correlation 89
6.2.2 Integration with SPD metrics elaboration 90
Conclusions **93**
References ..... Errore. Il segnalibro non è definito.

Index of tables

Table 1 Controls and Limitations 79
Table 2 SecLimsum variable calculating 81

Index of figures

Figure 1 Classification of main critical infrastructure sectors 14
Figure 2 Event Life Cycle 19
Figure 3 Assets of transportation systems 21
Figure 4 Semantic web layers 35
Figure 5 System Model 58
Figure 6 System architecture 61
Figure 7 Example of rule:tree branch 62
Figure 8 Example of Ontology for Physical Security 65
Figure 9 SPARQL Result of Events raising Alarm 70
Figure 10 SPARQL results Composed Events 71
Figure 11 nSHIELD Framework 74
Figure 12 Porosity Ontology 82
Figure 13 Control Ontology 83
Figure 14 Limitation Ontology 83
Figure 15 SPD attribute integration 84
Figure 16 Temperature, Voltage results 87
Figure 17 Voltage, humidity results 88
Figure 18 Physical and Logical Security Ontology 92



Introduction

The recent interests on the adoption of cyber infrastructures for developing smart surveillance system and for homeland protection has led the international scientific community in an effort aimed at the definition of new security strategies and tools. Key activities in security systems are the monitoring, observation and comprehension of different phenomena, aimed at providing an updated and meaningful description of the monitored scenario, as well as its possible evolutions, to enable proper decisions and countermeasures for the protection and safety of people and things.

The threats coming from many different sources, internally and externally. The diffusion of new technologies have made more accessible the assets of a system. The asset and information can also be accessed physical by the insider, but using, for instance the network, they are accessible by the outsiders who can explore and penetrate the network and the logic perimeter of a system. In addition, outsider access can be enhanced by the recruitment of insiders to furnish important information on the protection in place or key applications.

In these scenarios, not only smart surveillance and alert systems are needed but enriched decision support systems (DSS) are desirable. Such systems rely on heterogeneous data acquisition tools and on data elaboration to prune non-significant information; nevertheless, this is not enough as there is the need to interpret what data really represents to reduce false alarms and detect even weak risk conditions.

The availability of advanced monitoring techniques and heterogeneous information sources has increased the accuracy in observing, measuring and describing the nature of phenomena: the current level of technology in this field represents an opportunity to improve the understanding about observed phenomena but, at the same time, it introduces a high degree of complexity in the data elaboration and fusion. However, many automatic and intelligent detection systems generate unnecessary warnings (false

alarms); this problem, unfortunately, dramatically limits the use of these systems to enable automatic or partially automatic countermeasures.

Smart DSSs are needed to enable, when possible, the automatic adoption of countermeasures in case of alarms or to support end-users during decision making activities (when a too large number of sensors, devices, or cameras placed inside the site to be protected produce a wide amount of data to be processed).

In recent years, scientific world's attention has been devoted to both the information management with data and decision fusion approaches. On the other hand, to improve the situation assessment, it is possible to adopt different types of models for the description of the knowledge-base, for event correlation and for the definition of the situation and threat identification. Very promising approaches are based on semantic and ontological models

A semantic model can be used for understanding observed phenomena. All sensors should use the same data model and the same interpretation of data, a shared data model provides syntactic interoperability mechanisms but this is not enough, a more complex system model is needed to even provide semantic interoperability. In the literature, some approaches based on semantic inference rules for phenomena comprehension are available. Nevertheless, due to the introduced overhead, the knowledge base is usually inferred in offline mode.

The use of this kind of system is often associated to the management of physical security of an infrastructure, but during last years it has widespread the need of evaluation of logical security. The two kind of security are not separated, they came together and are strictly connected.

In this thesis we demonstrate that the use of a semantic model for the information management it is suitable in order to meet these issues. In particular, thesis proposes a methodology and approach for the early situation awareness recognizing a threat situation on time, for decision support to automatically activate recovery strategies. The threat on which the thesis focus on are regarded the logical and physical security. In particular for the logical security estimation will be presented a an approach guided by metrics, developed during the collaboration in European Project (nSHIELD). The summary of activities are:

- Starting from state of art, providing a innovative methodology for security information (physical and logical) fusion and correlation by semantic model in order to ensure a correct and shared information interpretation into events and situation assessment before raising an alarm.

- Developing a smart Decision Support System based on semantic and ontological approach. The semantic enrichment process is automatically performed to build a knowledge base, which will be inferred on-line by a light smart classifier that will raise an alarm in case of risk detection. The decision approach will be based on two steps: (1) a smart in-line classifier based on the semantic model to raise an alarm, in case of threat event detection, (2) a post reasoner offline inference engine, in order to further comprehend the event and its causes.

This thesis is structured as follows. The chapter 1 presents a panoramic view of the problem of critical infrastructures protection. It introduces the main sector in which is required a protection by malicious and non malicious threats. Then it underlines the problem of security of an infrastructure, introducing the issues not only on physical security protection but even at logical security level. Finally, the chapter presents the contribution and aim of this thesis work.

The chapter 2 introduces the problem of integration of information by a semantic model perspective. The chapter aims to introduce the readers in this world proposing the basic concepts and main solution presented in the scientific literature. In this way the readers have a awareness of the problem and on the solution proposed in this thesis.

The chapter 3 gives a view of a state of art on different topic discussed in this thesis, in particular on convergence of physical and logical security for the infrastructure protection. Then it presents the state of art on complex event processing, decision support system and then on semantic modelling of decision support systems.

The last 3 chapter are the description of the contribution on this thesis. In the chapter 4 will be presented the approach for information integration based on semantic model. The process for the semantic enrichment of the information from different king of source, the population of ontology domain and the inference methodology for the detection. Then it w ill described the process of the post resoner.

The chapter 5 introduces the metric for the estimation of logical security. The metrics derived from a collaboration on an European project called nSHIELD. The metrics have a proper semantic model, so the are suitable for an elaboration by the semantic classifier and realized a first approach to the convergence of logical an security convergence.

Finally in the chapter 6 present some application results and some example of application of security convergence.

Chapter 1

The Critical Infrastructure Protection

In the recent year, Critical Infrastructure Protection has became an important issues, for the protection against terrorism and any other form of criminality. It requires the development of innovative approaches in order to match the challenges of Homeland Security. This section provides the description on the current background with its related issues and involved technologies countermeasures; finally it illustrates the motivations and the main contributions of this thesis.

1.1 The problem of CIP

A Critical Infrastructure (CI) represents for an infrastructure or asset the incapacity or destruction of which would have a debilitating impact on the national security and the economic and social welfare of a nation. Such infrastructures could be damaged by non malicious threats as well as by malicious attacker. In recent years the architectural of CIs has changed because of several economic, technological and social reasons related to privatization and globalization processes. This aspect offers capabilities to the global economy and allowed to improve the quality of services provided, but on the other one it introduced new vulnerabilities. Each infrastructures (e.g. telecommunication, network, etc..) now depend on services provided by other infrastructures (e.g. information system, transportation system, air-traffic control, etc…). These means the creation of extensive troubles because a cascade effects and an interdependencies of malfunctions. All components of the infrastructures are linked and are part of the same chain to be protected. For this reason is important to identify innovative approach to mitigate and manage the threats.

In response to these challenges one of the fundamental aspects consists in territorial monitoring, aimed at providing an updated and meaningful description of the monitored scenario, as well as its possible evolutions, allowing interventions and countermeasures, when required for the protection of people and things.

Hence descends the crucial role of precocious alert systems, of decision support systems, and of intelligent surveillance systems. All such solutions rely on data acquisition by means of heterogeneous sensing tools, data elaboration aimed at pruning non significant information, and on the interpretation of the high-level meaning of what data represent, with the aim of reducing false positives as well as detecting weak alarm conditions. The availability of advanced monitoring techniques and heterogeneous information sources has increased the accuracy in observing, measuring and describing the nature of phenomena: the current development level of technology in this field represent an opportunity to improve the understanding about observed phenomena, but at the same time introduced a high degree of complexity in the data elaboration and fusion. The international scientific community has reached significant results in those fields. Despite that, the development of methodologies and tools allowing to implement automatic situation understanding and decision support tools which comply with security standards required in common is still an open research issue. Furthermore, information fusion and decision support systems are complementary research areas, yet typically separately dealt with, and few methodological approaches try to combine benefits from both areas. Furthermore, current decision support systems suffer from numerous limitations, partly due to such separation of research fields, since they are typically designed and developed for specific domains and hardly adaptable, both to domain changes, both to novel threat scenarios.

1.2 Application Domain

Critical infrastructures include material and ITC assets, networks, services, and installations. These can be applied in different sectors (Figure 1).

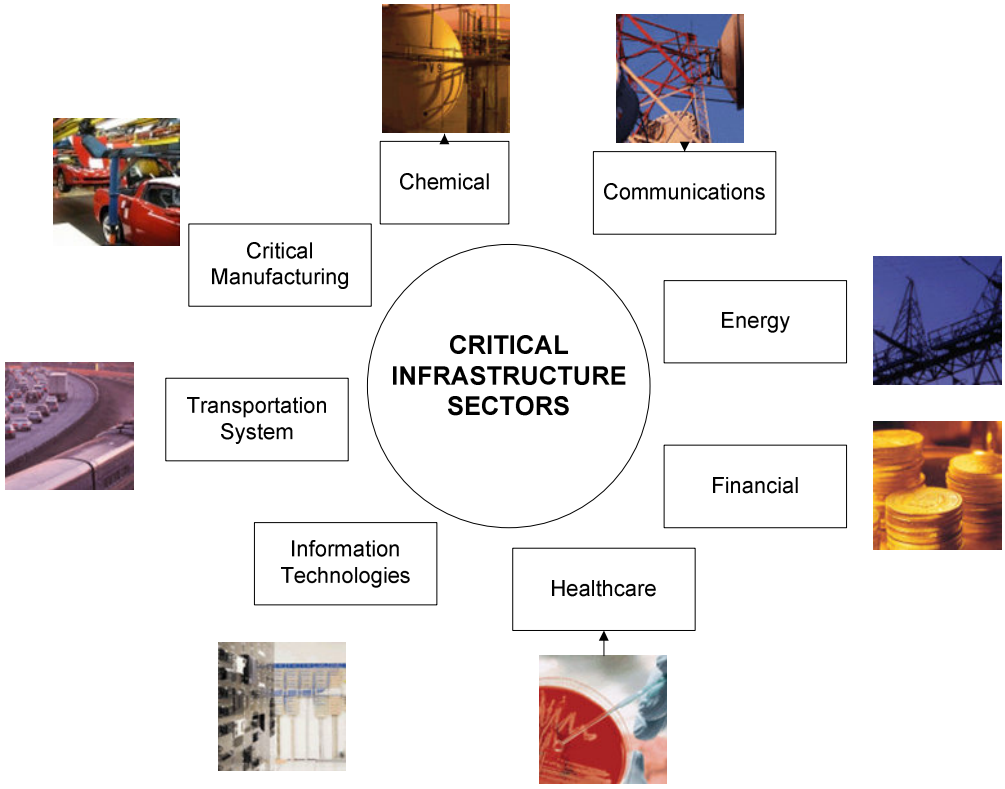


Figure 1 Classification of main critical infrastructure sectors

As specified in [1]. A classification of critical infrastructure sector is shown in figure 1:

- **Chemical Sector:** The sector can be divided into five main segments, based on the end product produced:
  - Basic chemicals
  - Specialty chemicals
  - Agricultural chemicals
  - Pharmaceuticals
  - Consumer products

  Each of these segments has distinct characteristics, growth dynamics, markets, new developments, and issues. The majority of Chemical Sector facilities are privately owned, requiring to work closely with the private sector and its industry associations to set goals and objectives, identify assets, assess risks, prioritize needs, and implement protective programs.

  The Chemical Sector is dependent on and depended on by a wide range of other sectors, including: Communications, Critical Manufacturing, Emergency Services, Energy, Food and Agriculture, Healthcare and Public Health, Information Technology, Transportation Systems, and Water and Wastewater Systems.

- **The Communications Sector** is an integral component of the economy, underlying the operations of all businesses, public safety organizations, and government. The sector has evolved from predominantly a provider of voice services into a diverse, competitive, and interconnected industry using terrestrial, satellite, and wireless transmission systems. The transmission of these services has become interconnected; satellite, wireless, and wired providers depend on each other to carry and terminate their traffic and companies routinely share facilities and technology to ensure interoperability.

  The private sector, as owners and operators of the majority of communications infrastructure, is the primary entity responsible for protecting sector infrastructure and assets. Working with the federal government, the private sector is able to predict, anticipate, and respond to sector outages and understand how they might affect the ability of the national leadership to communicate during times of crisis, impact the operations of other sectors, and affect response and recovery efforts.
- **Energy Sector:** The reliance of virtually all industries on electric power and fuels means that all sectors have some dependence on the Energy Sector. The Energy Sector is well aware of its vulnerabilities and is leading a significant voluntary effort to increase its planning and preparedness. Cooperation through industry groups has resulted in substantial information sharing of best practices across the sector. Many sector owners and operators have extensive experience abroad with infrastructure protection and have more recently focused their attention on cyber security.
- **The Financial Sector** represents a vital component of nation's critical infrastructure. Large-scale power outages, recent natural disasters, and an increase in the number and sophistication of cyber attacks demonstrate the wide range of potential risks facing the sector. Financial institutions provide a broad array of products from the largest institutions with assets greater than one trillion dollars to the smallest community banks and credit unions. Whether an individual savings account, financial derivatives, credit extended to a large organization, or investments made to a foreign country, these products allow customers to:
  - Deposit funds and make payments to other parties;
  - Provide credit and liquidity to customers;
  - Invest funds for both long and short periods; and
  - Transfer financial risks between customers.

- **The Healthcare and Public Health Sector** protects all sectors of the economy from hazards such as terrorism, infectious disease outbreaks, and natural disasters. Because the vast majority of the sector's assets are privately owned and operated, collaboration and information sharing between the public and private sectors is essential to increasing resilience of the nation's Healthcare and Public Health critical infrastructure. The sector plays a significant role in response and recovery across all other sectors in the event of a natural or manmade disaster. While healthcare tends to be delivered and managed locally, the public health component of the sector, focused primarily on population health, is managed across all levels of government: national, state, regional, local, tribal, and territorial.
- **The Information Technology Sector** is central to the nation's security, economy, and public health and safety. Businesses, governments, academia, and private citizens are increasingly dependent upon Information Technology Sector functions. These virtual and distributed functions produce and provide hardware, software, and information technology systems and services, and - in collaboration with the Communications Sector - the Internet. The sector's complex and dynamic environment makes identifying threats and assessing vulnerabilities difficult and requires that these tasks be addressed in a collaborative and creative fashion.

  Information Technology Sector functions are operated by a combination of entities - often owners and operators and their respective associations - that maintain and reconstitute the network, including the Internet. Although information technology infrastructure has a certain level of inherent resilience, its interdependent and interconnected structure presents challenges as well as opportunities for coordinating public and private sector preparedness and protection activities.
- **The transportation system** quickly, safely, and securely moves people and goods through the country and overseas. The Transportation Systems Sector consists of seven key subsectors, or modes:
  - **Aviation** includes aircraft, air traffic control systems, and approximately 450 commercial airports and 19,000 additional airports, heliports, and landing strips. This mode includes civil and joint use military airports, heliports, short takeoff and landing ports, and seaplane bases.
  - **Highway Infrastructure and Motor Carrier** encompasses nearly 4 million miles of roadway, almost 600,000 bridges, and some 400 tunnels in 35 states.

Vehicles include automobiles, motorcycles, trucks carrying hazardous materials, other commercial freight vehicles, motor coaches, and school buses.

- **Maritime Transportation System** consists of about 95,000 miles of coastline, 361 ports, 25,000 miles of waterways, 3.4 million square miles of Exclusive Economic Zone, and intermodal landside connections, which allow the various modes of transportation to move people and goods to, from, and on the water.
- **Mass Transit and Passenger Rail** includes service by buses, rail transit (commuter rail, heavy rail--also known as subways or metros--and light rail, including trolleys and streetcars), long-distance rail--namely Amtrak and Alaska Railroad--and other, less common types of service (cable cars, inclined planes, funiculars, and automated guide way systems).
- **Pipeline Systems** consist of vast networks of pipeline that traverse hundreds of thousands of miles throughout the country, carrying nearly all of the nation's natural gas and about 65 percent of hazardous liquids, as well as various chemicals. These include approximately 2.2 million miles of natural gas distribution pipelines, about 168,900 miles of hazardous liquid pipelines, and more than 109 liquefied natural gas processing and storage facilities.
- **Freight Rail** consists of seven major carriers, hundreds of smaller railroads, over 140,000 miles of active railroad, over 1.3 million freight cars, and roughly 20,000 locomotives. Further, over 12,000 trains operate daily. The Department of Defense has designated 30,000 miles of track and structure as critical to mobilization and resupply of U.S. forces.
- **Postal and Shipping** moves over 574 million messages, products, and financial transactions each day. Postal and shipping activity is differentiated from general cargo operations by its focus on letter or flat mail, publications, or small- and medium-size packages and by service from millions of senders to nearly 152 million destinations.

- **The Critical Manufacturing Sector** is crucial to the economic prosperity and continuity of the United States. A direct attack on or disruption of certain elements of the manufacturing industry could disrupt essential functions at the national level and across multiple critical infrastructure sectors. The Critical Manufacturing Sector identified the following industries to serve as the core of the sector:

- Primary Metal Manufacturing: Iron and Steel Mills and Ferro Alloy Manufacturing; Alumina and Aluminum Production and Processing; and Nonferrous Metal (except Aluminum) Production and Processing
- Machinery Manufacturing: Engine, Turbine, and Power Transmission Equipment Manufacturing
- Electrical Equipment, Appliance, and Component Manufacturing: Electrical Equipment Manufacturing
- Transportation Equipment Manufacturing: Vehicle Manufacturing, Aviation and Aerospace Product and Parts Manufacturing; and Railroad Rolling Stock Manufacturing

Products made by these manufacturing industries are essential in varying capacities to many other critical infrastructure sectors. The Critical Manufacturing Sector focuses on the identification, assessment, prioritization, and protection of nationally significant manufacturing industries within the sector that may be susceptible to manmade and natural disasters.

1.3 Critical Infrastructure the Event Cycle

For each system can be developed an assurance plan in order to:

- assess its vulnerabilities to both physical and cyber attacks;
- plan to reduce vulnerabilities;
- develop systems to identify and, if possible, prevent attempted attacks;
- contain attacks and trigger adequate countermeasures.

The infrastructure assurance plan includes several tasks and follows the six-step U.S. DoD *(Department of Defense)* life cycle (see ref. [2]) shown in Figure 2. the six steps describes activities that occur for and efficient event management and comprehend the effects.

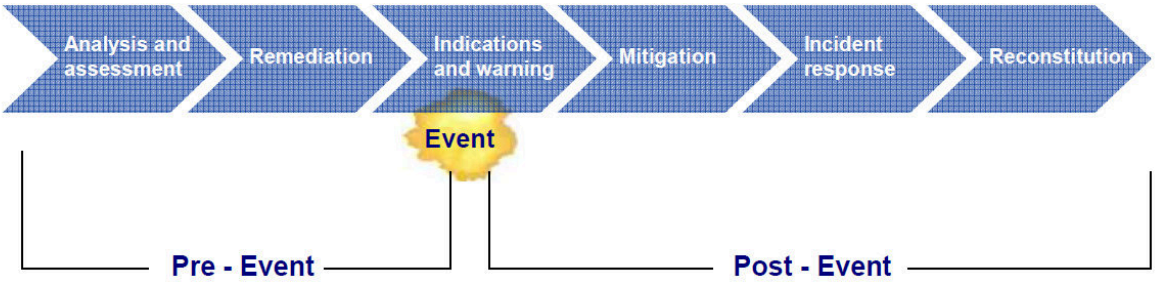


Figure 2 Event Life Cycle

1. **Analysis and assessment.** This activity is the foundation and the most important element of the life cycle phases. It aims to indentify critical assets and determine their risks and vulnerabilities, as well as their interdependencies, configurations and characteristics. The assessment also evaluates the operational impact of infrastructure loss or degradation.
2. **Remediation.** This phase involves precautionary measures and actions taken before an event occurs to fix the known cyber and physical vulnerabilities. It aims at developing and installing short/long range strategies to reduce risks. Remediation actions may include the choice and installation of protection mechanisms, but also education and awareness, procedural changes or system configuration and component changes. Obviously remediation actions have a cost based on the nature of vulnerabilities and a cost/benefit analysis could be useful to support the design of a security system to protect a CI.
3. **Indications and warning.** This phase involves the continuous monitoring to assess the assurance capabilities of critical infrastructure assets and to determine if there are event indications to report. Indications are preparatory actions or preliminary infrastructure conditions that signify that an incident is likely, is planned, or is under way. Warning is the process of notifying asset owners of a possible threat or hazard. Integration platforms for heterogeneous input data sources could come to support this phase as well.
4. **Mitigation.** It comprises actions taken before or during an event in response to warnings or incidents. The aim is to minimize the operational impact of the loss or debilitation of a critical asset by taking the opportune countermeasures.
5. **Incident response.** This phase comprises the plans and activities taken to eliminate the cause or source of an infrastructure event.

6. **Reconstitution.** The last phase of the CIP life cycle, involves actions taken to rebuild or restore a critical asset capability after it has been damaged or destroyed. This phase is the most challenging and least developed process.

This steps assure the coordination of protection and coordination of activities, ensuring a high security degree. The bad coordination and execution of one of these steps can be expose the system to potential risks of failure and compromise the services of CI. For these reasons the most important activities is the first phase (Analysis and Assessment), it refers to risk assessment focuses on asset and their threats. The risk assessment process measures the expected risk of and infrastructure basis on a careful analysis of threats and relative vulnerability. Furthermore the risk mitigation process is able to choose the countermeasures and forecast their impact on risk. The overall iterative process is named risk management.

1.4 Critical Infrastructure: Threats and countermeasures

As described in [7], a CI, such as Mass Transportation System have the variety of settings of a and asset (Figure 3).

Transit system are constantly faced with the challenge of managing risks to their assets. Each asset has its own level of risk based on its attractiveness as a target, vulnerabilities, accessibility, and criticality to the system. The process of evaluating risk and implementing countermeasures requires a high effort.

Since funding for security efforts is limited, security measures for each asset must be commensurate with the threats and vulnerabilities of that particular asset and the potential consequences of an attack or other disaster.

The diversity of assets of a mass transit system leads to a range of possible threats and countermeasures. Some assets might be targets for a terrorist attack intended to inflict civilian injuries; others might be means for providing misinformation to the public, others for obtaining sensitive information about the system. Transit systems or their components could also be affected indirectly by an attack elsewhere, which may compromise communications, operations, or maintenance capabilities. Results of attacks or incidents might include:

- Loss of life or physical damage to passengers, staff
- Physical damage to infrastructure, and possibly to the surrounding environment
- Loss of power through direct attack or by external event

- Non authorized access to the information of CI
- Alteration or disruption of information

Several threats can be identify in a mass transit system, some of them are discussed in [7].

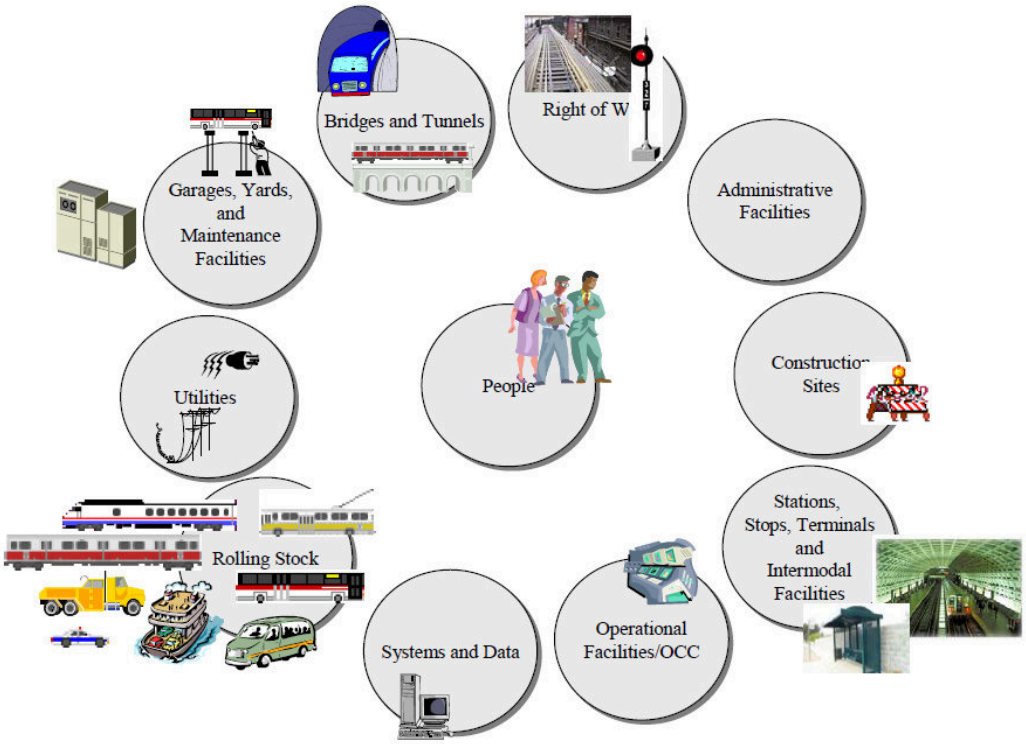


Figure 3 Assets of transportation systems

Arson

The hazards of arson, an intentionally set fire, in a transit facility include the destruction of assets within the facility, structural damage to the facility itself, and injuries or fatalities due to direct exposure to fire or to smoke and fumes.

Burning fuel, oil, plastics, and some paints can cause dense smoke and toxic fumes. Toxic fumes present a serious health threat and may cause death by asphyxiation. In addition, smoke can reduce visibility, obscuring exit pathways and making escape more difficult for victims. Since fires may occur accidentally as well as intentionally, there is crossover between protection against accidental fires and protection from arson. Arson and explosion-related fires, however, may cause more severe damage because they tend to target or cluster around critical systems and equipment.

Explosives

The hazards of an explosive blast include the destruction of assets within a facility, structural damage to the facility itself, and injuries or fatalities. In addition, explosions may start a fire, which may inflict additional material damage, injuries, or fatalities due to direct exposure or to heat, smoke, and fumes. An explosion is an instantaneous or almost instantaneous chemical reaction resulting in a rapid release of energy. The energy is usually released as rapidly expanding gases and heat, which may be in the form of a fireball. The expanding gases compress the surrounding air creating a shock wave or pressure wave. The pressure wave can cause structural damage to the structure while the fireball may ignite other building materials leading to a larger fire. The strength of a blast depends on the type and amount of explosive material used. A bomb that a person can carry is capable of a smaller blast than an explosive-laden truck.

Weapons of Mass Destruction

Weapons of mass destruction (WMD) typically refer to nuclear, radiological, chemical, and biological weapons capable of inflicting mass casualties. WMD can also refer to radioactive materials and other contaminants intended to quickly harm large numbers of people, such as any powders, liquids, gases, and dirty bombs; most of these come in a liquid, vapor, gas, or powder form, and are spread through air movement.

The hazards of WMD include fatalities or deleterious health effects, as well as potentially permanent contamination of a facility that may render it unusable. Many agents have little or no plainly discernable characteristics, so symptoms may be the first sign that an attack has occurred. While some chemical agents induce immediate symptoms, other agents will not produce symptoms for hours after the attack. Some biological agents may have an incubation period of up to a few days before symptoms appear.

Tampering

Tampering with transit facilities' assets may be a means to achieve any of the above events, such as starting a fire or spreading an airborne chemical agent, or it may be a stand-alone act, such as tampering with track to induce derailment. It can also include the intentional ramming of a facility, with a truck, boat, or airplane, in order to cause structural damage to a facility or injury to its users.

The ramming vehicle may be laden with explosives. Depending on the situation, tampering may lead to asset damage, structural damage, contamination, injuries, and/or fatalities.

Network Failure/Cyber Attack

Transit systems rely on computerized networks to facilitate operations and enhance efficient service delivery, which makes them vulnerable to network failure and cyber attacks. While this document does not offer specific considerations on how to protect computer networks, it is crucial to understand their importance to operating and communicating among agency staff as well as with partner organizations and the public-at-large. Network failure may be caused by faulty or damaged internal components, direct cyber attack to the agency's network, direct attack to a peripheral system or network, or even a blanket computer virus. The result may be loss of communications or operations capabilities as well as misinformation by hacking into a Web site or server.

The principal strategies to counter attacks can be: deter attackers from attempting an attack, detect potential threats promptly, minimize the impact from an attack and respond and recover (or resume critical operations as quickly as possible). Applying these concepts to the physical design of infrastructure leads to several general strategies that are applicable to transit assets.

Surveillance and monitoring of Infrastructure is based on sensor system deployed along assets in order to monitor several different parameters. The protection has the aim to reduce the vulnerabilities but also to be a deterrent (to reduce the rate of threats occurence). This make easy the management of emergency procedures due to a more precise monitoring.

A typical PSIM system, for mass transit security, is composed by several kind of sensors [6], described follow.

Anti-intrusion and Access control systems.

Devices has the ability to know when someone has entered a secured area. Access control is the ability to determine who can or cannot enter specific fields, areas or access particular assets. Traditional intrusion detection systems, made up by different devices such as:

- volumetric sensors for motion detection;
- magnetic contacts to detect illicit doors opening;
- glass break detectors;

- microphones cables for fence/grill vibration detection;
- active infrared barriers for detecting intrusions inside the tunnels.

To distinguish between authorized and unauthorized accesses, it is necessary to employ access control systems to identify authorized staff by using:

- possess (keys, badges, etc.)
- knowledge (PIN, password, etc.)
- individual biometric features (e.g. fingerprint reading)

Intelligent video surveillance.

Intelligent video surveillance is composed by

- advanced cameras with special features;
- digital video processing and recording, using efficient data compression protocols;
- video-analytics of the scenes, using computer vision algorithms;

The presence of surveillance works as deterrent not only because an area is being monitored remotely, but also because activities can be detected in automatic mode.

Wireless Sensor Networks (WSN). This kind of sensor are more widespread due to their potential for providing diverse new capabilities to a wide variety of applications (environmental monitoring, health, object tracking and military applications).

Other (smart) sensors. Other kind of sensors most used in the mass transit protection system are:

- smoke and heat detectors for fire protection;
- vibration and movement detectors
- CBRNe (chemical, biological, radiological, nuclear, explosive) sensors aimed at detecting harmful substances within the infrastructure assets.

This heterogeneity of sub-systems technologies mentioned above and their measured data make difficult the interoperability among different systems, introduce the problem of integration, phenomena comprehension and manage the problem of information security.

1.5 Security of Critical Infrastructure

The first aim in order to guarantee the correct working of a CI is to protected by malicious and non malicious threats. The foremost concepts is named Security the latter is safety. In the CI domain the term security can be separated in two different means, physical and logical. Physical security describes security measures that are designed to deny unauthorized access to facilities, equipment and resources, and to protect personnel and property from damage or harm (such as espionage, theft, or terrorist attacks). Physical security involves the use of multiple layers of interdependent systems. Meanwhile logical security consist in protection of information and guarantee their integrity, availability and confidentiality.

The introduction of security measures at the design phase is more effective than adopting a retroactive solution. An other issue is the quantity and the heterogeneity of assets located in several areas and with different exposition to threats. A Physical Security Information Management (PSIM) System should be able to protect different king of assets, having different risks, different vulnerabilities and criticalities. For these reasons the adoption of security measures is different for each assets and is a consequences of relative threats, as well as with the evaluation of potential attacks or malfunctioning of the asset.

During last year ([3],[4],[5]) is born an other important issue, the logical security in correlation with physical. The security means physical (e.g. access control) or logical (e.g. virus detection, unauthorized network access). Today, entities the manage these two types of security are separate and, often, do not collaborate. These have a dramatic impact on the protection of CI, as well as on security and safety.

In general physical security system for an infrastructure is composed by:

- Physical access control, such as card readers, biometric devices.
- Power supply system: Uninterrupted Power Supplies (UPS), generators, electric distribution system.
- Physical blocking mechanism: electromagnetic lock devices
- Fire control system,
- Life support system: heating, cooling, ventilation monitoring of temperature, humidity, condensation and etc.
- Video and audio surveillance systems.

These sub-system work in conjunction with others services, such as network, information technology infrastructures. For instance the door reader is connect to fire protection that is interfaced with CCTV (Closed Circuit television system) sub-system. Physical security focuses of protection of asset, people and infrastructure against threats. Add to this the need to protect the privacy and security of information by attacker and identity theft for instance is important. So the logical and physical security became a key issue.

The access facilities can compromise the logical security of asset, the network security, the confidentiality of information, the authorized access to the information and the manipulation are a key issues. A combination of two kind of security gives a more comprehension of attack across the physical and logical environments. The availability of sensor based on TCP/IP communication protocol are just some of the ways where logical security is tied to the physical access system. In the following sub-section it will be illustrated a panoramic of most widespread threats of a critical infrastructure and their countermeasures.

1.6 Thesis Contributions

As discussed in the previous sections, CI are highly complex and composed by different sensor for monitoring, and then is requested high level of information security. For this reasons the protection of Critical Infrastructure has became and important activity which requires a multidisciplinary and innovative approach in order to identify the threats, detect the malicious event for the physical security requirements, assure the information security and then provide automatic countermeasures.

Today, PSIM system needs to have reliable and secure information on time and integrate and merge them in a common vision in order to comprehend and detect event and then determinate the opportune countermeasures. So is required the use of appropriate technologies and methodologies.

In order to achieve a correct methodology to overcomes the aforementioned issues the most important steps are:

2. evaluate the risk of infrastructure to be protected.
3. develop a methodology for determinate and prevent the threats, that correlate and fuses different information for several sources.
4. determinate the activities for an automatic countermeasures.

This thesis proposes a methodology and approach for the early situation awareness recognizing a threat situation on time, for decision support to automatically activate recovery strategies. The detailed activities are:

- Starting from state of art, providing a innovative methodology for security information (physical and logical) fusion and correlation by semantic model in order to ensure a correct and shared information interpretation into events and situation assessment before raising an alarm.
- Developing a smart Decision Support System based on semantic and ontological approach. The semantic enrichment process is automatically performed to build a knowledge base, which will be inferred on-line by a light smart classifier that will raise an alarm in case of risk detection. The decision approach will be based on two steps: (1) a smart in-line classifier based on the semantic model to raise an alarm, in case of threat event detection, (2) a post reasoner offline inference engine, in order to further comprehend the event and its causes.

Chapter 2

Integration and interoperability by Semantic modeling

2.1 Semantic interoperability: XML and RDF

In the research topic of interoperability, the use of data model provide to the whole system and its components a time consuming and the introduction of artificial software layers than can impact on control and performance. On the other hand this kind of techniques provides many advantages such as:

- Availability of complete information,
- Improvement of *existing analysis* and application of the *new analysis*.
- Cost reduction resulting from the *multiple use of existing information sources*.
- Avoidance of redundant data and conflicts.

In literature exist various integration levels, that it is possible to investigate in order to provide complete integrated access to information:

- **Syntactic Integration:** Many standards have evolved that can be used to integrate different information sources. Such as database interfaces (e.g. ODBC, HTML and XML).
- **Structural Integration:** The first problem that passes a purely syntactic level is the integration of heterogeneous structures. This problem is solved by mediator systems defining mapping rules between different information structures.
- **Semantic Integration:** is the resolution of semantic conflicts, that make a one to one mapping between concepts or terms impossible.

The contribution of this theses is to provide an general approach to the problem of information integration, taking into account the integration and combination several

technologies, including standard mark-up languages and ontologies. In order to overcome the aforementioned limitations it the issues can be solved considering the three level of integration.

In particular the work is focused on semantic heterogeneity which is based on different semantics. In order to manage semantic heterogeneities, a formal representation is needed. Due to the diffusion of Web, some standard languages are developed, such as XML (eXtensible Markup Language) and RDF (Resource Description Framework) by the W3C community for this purpose [34], [35]. In order to overcame the semantic heterogeneity of these standard we will described the peculiarities of them and them we will proposes an integration/interoperation process based on Ontologies and XSLT (eXtensible Stylesheet Language Transformations) Transformation [36] and [34].

XML and RDF have been developed for the semantic description of information sources. As described in [37] and [38], XML was proposed as an extensible language allowing the user to define new kind of tags in order to indicate the type of content. Therefore, the best benefit of XML is in the opportunity to exchange data in a structured way. A data object is said to be XML document if it follows the guidelines for well-formed XML documents provided by the W3C community. The specification provide a formal grammar used in well-formed documents. In addition is possible to extend the general grammar with the introduction, by users, of new grammatical constraints on the structure of a document using a document type definition (DTD). A XML document is valid if it has an associated type definition and complies to the grammatical constraints of that definition. A DTD specifies elements that can be used in an XML document. In the document, the elements are delimited by a start and an end tag. It has a type and may have a set of attribute specifications consisting of a name and a value. The additional constraints in a DTD refer to the logical structure of the document, this especially includes the structure of tags allowed and/or required. Further restrictions that can be expressed in a DTD concern the type of the attributes and default values to be used when no attribute value is provided.

An XML schema itself is, an XML document defining the valid structure of an XML document in the spirit of a DTD. The elements used in a schema definition are of the type 'element' and have attributes that are defining the restrictions already mentioned above. The information in such an element is a list of further element definitions that have to be nested inside the defined element. Furthermore, XML schema has some several characteristics, such as: support of basic data types, Constraints on attributes, type definition, name-space mechanism. This features enable to map data-models of applications from whose

information we want to share with others on an XML schema This procedure has a big potential in the actual exchanging of data. However, the user must to commit to our data-model in order to make use of the information. Therefore, it lacks an important advantage of meta-information. XML is designed to provide an interchange format for weakly structured data by defining the underlying data-model in a schema and by using annotations, from the schema, in order to clarify the role of single statements. Two things are important in this claim from the information sharing point: XML has a syntactic/structural model and describes data on the object level, without semantic.

In the research community was developed a new language able to fill these gaps, the RDF standard proposed as a data model for representing meta-data about web pages and their content using an XML syntax. In RDF, a resource is expressed in terms of a triple (resource, property, value) [39]. The property is a two-placed relation that connects a resource to a certain value of that property. This value can be a simple data-type or a resource. Additionally, the value can be replaced by a variable representing a resource that is further described by nested triples making assertions about the properties of the resource that is represented by the variable. Furthermore, RDF allows multiple values for a single property. For this purpose, the model contains three *built-in data types* called collections, namely an unordered lists (*bag*), ordered lists (*seq*), and sets of alternatives (*alt*) providing some kind of an aggregation mechanism. A further requirement arising from the nature of the web is the need to avoid name-clashes that might occur when referring to different web-sites that use different RDF-models to annotate meta-data. RDF defines name-spaces for this purpose. Name-spaces are defined by referring to an URL that provides the names and connecting it to a *source id* that is then used to annotate each name in an RDF specification defining the origin of that particular name.

However, if people want to share this information, there has to be an agreement on a standard core of vocabulary in terms of modeling primitives that should be used to describe meta-data. RDF schemes (RDF/S) attempt to provide such a standard vocabulary. RDF/S provides a notion of concepts (*class*), slots (*property*), inheritance (*SubclassOf, SubslotOf*) and range restrictions (*Constraint Property*). Unfortunately, no well-defined semantics exist for these modeling primitives in the current state. Further, parts such as the re-identification mechanism are not well defined even on an informal level. Lastly, there is no reasoning support available, not even for property inheritance.

Before we have seen that XML defines structures as well, except there are no sophisticated mechanism for mapping different structures. RDF is designed to provide some information

on the semantic level, by enabling us to include meta-information in the description of a web-page. But RDF doesn't provide semantic descriptions. Rather it provides a common syntax and a basic vocabulary that can be used when describing this meta-data. Fortunately, the designers of RDF are aware that there is a strong need for an additional 'logical level' which defines a clear semantics for RDF-expressions and provides a basis for integration mechanisms.

XML and especially XML schemata are suitable for exchange data with a well defined syntax and structure. Simple RDF provides a uniform syntax for exchanging meta-information in a machine-readable format. However, in their current state neither XML nor RDF provides sufficient support for the integration of heterogeneous structures or different meanings of terms. There is a need for semantic modeling and reasoning about structure and meaning.

2.2 Semantic Formal Language

The term 'Ontology' [40] has been used in many ways and across different communities. In general, each person has an individual view on the world and the things he/she has to deal with every day. However, there is a common basis of understanding in terms of the language we use to communicate with each other. Terms from natural language can therefore, be assumed to be a shared vocabulary relying on a (mostly) common understanding of certain concepts with very little variety. These conceptualizations provide a terminology that can be used for communication between people. The example of our natural language demonstrates, that a conceptualization cannot be universally valid, but rather a limited number of persons committed to that particular conceptualization. This fact is reflected in the existence of different languages which differ even more (English and Japanese) or much less (German and Dutch). Confusion can become worse when we are considering terminologies developed for a special scientific or economic areas. In these cases, we often find situations where one term refers to different phenomena. The consequence of this confusion is, a separation into different groups, that share terminology and its conceptualization. These groups are then called information communities. The main problem with the use of a shared terminology according to a specific conceptualization of the world is that much information remains implicit. When a mathematician talks about a binomial normal he is referring to a wider scope than just the

formula itself. Possibly, he will also consider its interpretation (the number of subsets of a certain size) and its potential uses (e. g. estimating the chance of winning in a lottery). Ontologies set out to overcome this problem of implicit and hidden knowledge by making the conceptualization of a domain (e. g. mathematics) explicit. This corresponds to one of the definitions of the term ontology most popular in computer science [41]: An ontology is an explicit specification of a conceptualization. An ontology is used to make assumptions about the meaning of a term available. It can also be viewed an explication, of the context a term, it is normally used in [42] for example, describes context in terms of twelve independent dimensions that have to be know in order to understand a piece of knowledge completely. There are many different ways in which an ontology may explicate a conceptualization and the corresponding context knowledge. Jasper and Uschold [43] distinguish two ways in which the mechanisms for the specification of context knowledge by an ontology can be compared:

- **Level of Formality:** The specification of a conceptualization and its implicit context knowledge, can be done at different levels of formality. As already mentioned above, a glossary of terms can also be seen as an ontology, despite its purely informal character. A first step to gain more formality, is to describe a structure to be used for the description. A good example of this approach is the standard web annotation language XML (see section). The DTD is an ontology describing the terminology of a web page on a low level of formality. Unfortunately, the rather informal character of XML encourages its misuse. While the hierarchy of an XML specification was originally designed to describe a layout, it can also be exploited to represent sub-type hierarchies, [44] which may lead to confusion. Fortunately, this problem can be solved by assigning formal semantics to the structures used for the description of the ontology. However, a formalization is only available for the structural part of a specification. Assertions about terms and the description of dynamic knowledge is not formalized which offers total freedom for a description.
- **Extend of Explication:** The other comparison criterion is, the extend of explication that is reached by the ontology. This criterion is strongly connected with the expressive power of the specification language used. We already mentioned DTD's which are mainly a simple hierarchy of terms. Furthermore, we can generalize this by saying that, the least expressive specification of an ontology consists of an organization of terms in a network using two-placed relations. The idea of this goes back to the use of semantic networks in the seventies. Many extensions of the basic

idea examined have been proposed. One of the most influential ones was, the use of roles that could be filled out by entities showing a certain type [94]. This kind of value restriction can still be found in recent approaches. RDF schema descriptions [95] which might become a new standard for the semantic descriptions of web-pages, are an example of this. An RDF schema contains class definitions with associated properties that can be restricted by so called constraint-properties. However, default values and value range descriptions are not expressive enough to cover all possible conceptualizations. A more expressive power can be provided by allowing classes to be specified by logical formulas. These formulas can be restricted to a decidable subset of first order logic. This is the approach of description logics [96].

2.2.1 The ontology

Ontologies are useful for many different applications, which has different requirements on the level of formality and the extend of explication provided by the ontology. Information communities are useful because they ease communication and cooperation among members with the use of shared terminology with well defined meaning. On the other hand, the formalization of information communities makes communication between members from different information communities very difficult. Generally, because they do not agree on a common conceptualization. This situation demands for an explication and explanation of the use of terminology. Informal ontologies with a large extend of explication are a good choice to overcome these problems. While definitions have always played an important role in scientific literature, conceptual models of certain domains are rather new. Nowadays systems analysis and related fields like software engineering, rely on conceptual modeling to communicate structure and details of a problem domain as well as the proposed solution between domain experts and engineers. Prominent examples of ontologies used for communication are Entity-Relationship diagrams and Object-oriented Modeling languages such as UML [97]. ER-diagrams as well as UML are not only used for communication, they also serve as building plans for data and systems guiding the process of building (engineering) the system. The use of ontologies for the description of information and systems has many benefits. The ontology can be used to identify requirements as well as inconsistencies in a chosen design. Further, it can help to acquire

or search for available information. Once a systems component has been implemented, its specification can be used for maintenance and extension purposes. Another very challenging application of ontology-based specification is the reuse of existing software. In this case, the specifying ontology serves as a basis to decide if an existing component matches the requirements of a given task. Depending on the purpose of the specification, ontologies of different formal strength and expressiveness are to be utilized. While the process of communication design decisions and the acquisition of additional information normally benefit from rather informal and expressive ontology representations (often graphical), the directed search for information needs a rather strict specification with a limited vocabulary to limit the computational effort. At the moment, the support of semiautomatic software reuse seems to be one of the most challenging applications of ontologies, because it requires expressive ontologies with a high level of formal strength.

The previously discussed considerations might provoke the impression that the benefits of ontologies are limited to systems analysis and design. However, an important application area of ontologies is the integration of existing systems. The ability to exchange information at run time, also known as interoperability, is an valid and important topic. The attempt to provide interoperability suffers from problems similar to those associated with the communication amongst different information communities. The important difference being the actors are not people able to perform abstraction and common sense reasoning about the meaning of terms, but machines. In order to enable machines to understand each other, we also have to explicate the context of each system on a much higher level of formality.

Each system that wants to inter-operate with other has to transfer its data information into this common framework. Interoperability is achieved by explicitly considering contextual knowledge in the translation process.

As reported in [29] some ontologies are diffused in scientific literature. A common domain ontology describes the semantics of the domain in the SIMS mediator [45], in which all terms of a domain are arranged in a complex structure. Each information source is related to the terms of the global ontology. However, the scalability of such a fixed and static common domain model is low [46], because the kind of information sources which can be integrated in the future is limited. In OBSERVER [47] and SKC [46] it is assumed, that a predefined ontology for each information source exists. Consequently, new information sources can easily be added and removed. But the ontologies use their own vocabulary. In MESA [35] the third hybrid approach is used. Each source is related to its source ontology.

In order to make the source ontologies comparable, a common global vocabulary is used, organized in a common domain ontology. This hybrid approach provides the biggest flexibility because new sources can easily be integrated and, in contrast to the decentralized approach, the source ontologies remain comparable.

2.3 Semantic Technologies

The World Wide Web represents a huge repository of information which can be retrieved and used. Unfortunately, information is represented with no meaning associated, since the meaning of retrieved information can be (re-)established only in the process of interpreting the information by humans. As a result, information scattered throughout the current (and traditional) version of the web is almost totally useless for software, non-human users (machine agents).

In attempt to respond to this situation, the term "Semantic Web" was coined by Tim Berners-Lee [49] referring to a "web for machines" as opposed to a web to be read by humans. In their understanding, "The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation" (Figure 4)

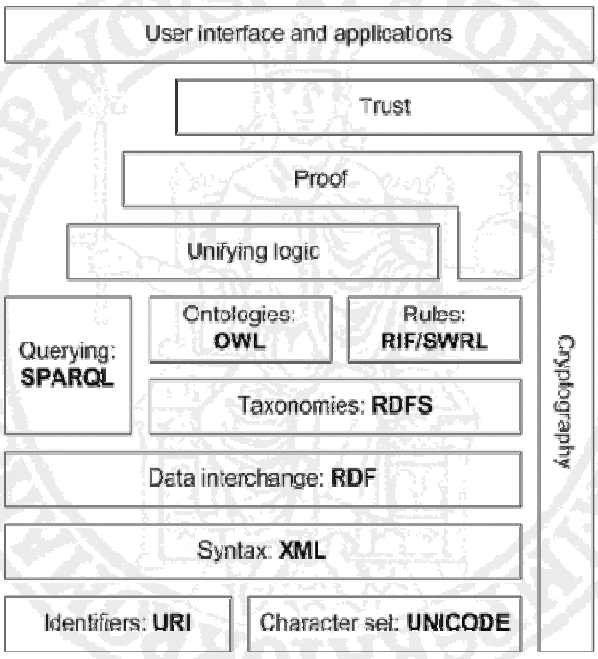


Figure 4 Semantic web layers

The Semantic Web is the opportunity for providing, finding and processing information via the Internet with the help of (machine) agents which are capable of dealing with the semantics of the information. The idea is to transform information into something meaningful to actors who seek to enhance their knowledge in order to satisfy a specific concern or accomplish a specific task related to their particular context.

The vision of the Semantic Web is based on the employment of semantic technologies that allow the meaning of information and the meaning of associations between information to be known and processed at execution time. To fulfil the promises and enable semantic technologies to work, there must be a knowledge model (of some part) of the world that is used to provide meaning to information to be processed within an application. The knowledge model has the form of a semantic model which differs from other kind of models [50]:

- Connections: The meaning of terms, or concepts, in the model is established by the way they connect to each other.
- Multiple views: a semantic model expresses multiple viewpoints and several interconnected models could be used to represent different aspects.
- Sharing: semantic models represent knowledge about the world in which systems operate and are shared across applications.
- Reasoning capability: use of a model is referred to as "reasoning over the model". The reasoning can range from graph search to intricate inferencing.

Although the role of a semantic model can be played by a simple taxonomy, nowadays use of semantically richer ontologies (ontological models) dominates.

New knowledge can be derived by examining the connections between concepts. Simple ontologies are just connections, richer ontologies include rules and constraints governing these connections. The semantic web is not so much a technology as an infrastructure, enabling the creation of meaning through standards, mark-up languages, and related processing tools. To represent ontologies in a formal way, several languages can be used. The Semantic Web principles are implemented in the layers of Web technologies and standards. The most common ontology languages are briefly described follow.

At the beginning, the idea of the semantic web tried just to enhance the current version of the web. It started out with a document oriented approach. The basic idea was to make web pages identifiable by computers as information resources carrying not only information (readable only by humans) but the meaning of this information as well.

Ontologies in this thesis define a shared conceptualization of the application domain at hand and provide the basis for defining metadata, that have a precisely defined semantics, and that are therefore machine-processable.
Advanced applications can use ontologies to relate the information to a semantic model of a given domain. In this way semantic technologies offer a new way to integrate different applications. Nowadays, the field of semantic interoperability is the most addressed problem connected with the idea of the semantic web.

2.4 Ontology representations

Most of languages can be used to represent ontologies; many of them evolved from creation of ontology construction methodologies. The Open Knowledge Base Connectivity (OKBC) [51] model and languages like "Knowledge Interchange Format (KIF)" [52] are examples that have become the bases of other ontology languages.
Several languages use frame logic which is basically an object-oriented approach defining frames and attributes (classes and properties). There are also several languages based on description logic, e.g. Loom [53], DAML+OIL [71], or later evolved Web Ontology Language (OWL) [55] standard.
Representation languages can be divided in terms of different abstraction levels used to structure the representation itself:

- **Extensional level:** the model is formulated by specifying every object from the domain.
- **Intentional level:** objects are defined by means of (necessary and sufficient) conditions for belonging to the domain.
- **Meta-level:** concepts from intentional level are abstracted, higher level concepts are specified, and previous concepts are seen as instances of new meta-concepts.

Some issues emerge from analysis of ontology representation, concerning the scope and modality of context expression: these criteria consider the basic formal nature of languages and that various languages deal with the representation of incomplete information in different way:

- *Class and relations:* languages aiming at representing objects, classes and relations.

- *Actions and processes:* languages that provide specialized representation structures for describing dynamic characteristics of the domain, such as actions, processes, and workflows (they usually can represent static aspects of domain too, but only in elementary level).
- *Everything*: languages that may be used for any kind of contexts and applications.

The context can be expressed in the following ways:

- **Programming languages:** allow representation and manipulation of data in several ways and according to various paradigms, leading to a cleaner separation between data structures and algorithms that handle them. Object oriented paradigm is preferred in recent years. This approach is generally associated with a number of concepts, such as complex objects, object identity, methods, encapsulation, typing and inheritance. Example can be language F-logic [56], logical formalism that tries to capture the features of object-oriented approaches to computation and data representation. F-Logic forms the core of systems such as Ontobroker [57].
- **Conceptual and semantic database models:** semantic (or conceptual) models were introduced as schema design tools. Examples of proposed semantic data models are ER and Extended ER data model, FDM (Functional data model), SDM (Semantic Data Model). Semantic models provide more powerful abstractions for the specification of databases.
- **Information system/software formalisms:** here belong different formalisms for information system design, especially in object-oriented design. Most widely used formalism is Unified Modelling Language (UML). UML was designed for human-to-human communication of models for building systems in object-oriented programming languages. Over the years its use has been extended to a variety of different aims, including the design of databases schemas, XML document schemas, and knowledge models.
- **Logic-based:** very important class of languages is based on logic. Such languages express a domain-ontology in terms of the classes of objects that are of interest in the domain, as well as the relevant relationships holding among such classes. These languages have a formal well-defined semantics. Three different types of logic-based languages exist – languages based on first-order predicate logic (e.g. KIF

[52]), languages based on description logics (e.g. OWL [55]), and process-action specification languages (e.g. PSL [58]).

- **Frame-based:** frame is a data structure that provides a representation of an object or a class of objects or a general concept or predicate. Some systems define only a single type of frame, other have two or more types, such as class frames and instance frames. The slots of a frame describe attributes of represented concept. They may also have other components in addition to the slot name, value and value restrictions, for instance the name of a procedure than can be used to compute the value of the slot – facets. Frames are usually organized into taxonomies. Through taxonomic relations, classes may be described as specializations of more generic classes with inheritance capability. Frame-based ontology languages were often used in many knowledge-based applications, like Ontolingua [59], OCML [60], OKBC [61] or XOL [62].
- **Graph-based:** formalisms based on various kinds of graph based or graph-oriented notations. Semantic networks [63] and conceptual graphs [64] originated from the Artificial Intelligence community. OML/CKML (Conceptual Knowledge Mark-up Language) [65] is a framework and mark-up language for knowledge and ontology representation based on conceptual graphs. Topic Maps [66] are recent proposal originated from the XML community.
- XML-related formalisms: XML [67] is a tag-based language for describing tree structures with a linear syntax and it is a standard language for exchange of information in the Web. Given the popularity of XML in exchange of information, XML-related languages have been considered as suitable for ontology representation. Important languages are based on Resource Description Framework (RDF) [68]. These provide a foundation for processing metadata about documents.

The expression can be interpreted in single model and several model. In the first case, ontology should be interpreted in such a way that only one model of the corresponding logical theory is a good interpretation of the formal description. In the other one, ontology should be interpreted as specifying what we know about the domain with the reservation that the amount of knowledge we have about the domain can be limited (e.g. first-order logic based languages).

2.4.1 Semantic Web Ontology languages

Description Logics (DLs) are a family of logic-based knowledge representation formalisms designed to represent and reason about the knowledge of an application domain in a structured and well understood way. The basic notions in DLs are concepts and roles, which denote sets of objects and binary relations, respectively. Most of today's semantic web ontology languages are DL-based. Also many of them are XML-related, or they possible XML notation. Several ontology languages have been designed for use in the web. Among them, the most important are OIL [69], DAML-ONT [70] and DAML+OIL [71]. More recently, a new language, OWL [55], is being developed by the World Wide Web Consortium (W3C) Web Ontology Working Group, which had to maintain as much compatibility as possible with pre-existing languages and is intended to be proposed as the standard Semantic Web ontology language. The idea of the semantic Web is to annotate web pages with machine-interpretable description of their content. In such a context, ontologies are expected to help automated processes to access information, providing structured vocabularies that explicate the relationships between different terms.

Extensible Markup Language (XML) [67] - was widely accepted and used as a convenient information representation and exchange format. XML itself don't carry semantics, but is serves as the base syntax for the leading ontology languages that we shall survey. Later additions like XML-DTD (Document Type Definition) and XML-Schema, added some syntactic rules like enumerations, cardinality constrains, and data types, but still lacked even simple semantics like inheritance. The purpose of XML Schema is therefore to declare a set of constraints that an XML document has to satisfy in order to be validated. With respect to DTD, however, XML Schema provides a considerable improvement, as the possibility to define much more elaborated constraints on how different part of an XML document fit together, more sophisticated nesting rules, data-typing. Moreover, XML-Schema expresses shared vocabularies and allows machines to carry out rules made by people. Among a large number of other rather complicated features.

Resource Description Framework (RDF) [68] is a standard way for defining of simple descriptions. RDF is for semantics - a clear set of rules for providing simple descriptive information. RDF enforces a strict notation for the representation of information, based on

resources and relations between them. As referred to in its name, RDF strength is in its descriptive capabilities, but is still lacks some important features required in an ontology language such as inferences for example. However, ontology languages built on top of RDF as a representation and description format. The RDF data model provides three object types: resources, properties, and statements. Resource may be either entire Web page, a part of it, a whole collection of pages or an object that is not directly accessible via the Web, property is a specific aspect, characteristic attribute, or relation used to describe a resource, statement is a triple consisting of two nodes and a connecting edge. These basic elements are all kinds of RDF resources. According to the latter description, a subject is a resource that can be described by some property. The predicate defines the type of property that is being attributed. Finally, the object is the value of the property associated with the subject.
RDF Schema (RDFS) [62] enriches the basic RDF model, by providing a vocabulary for RDF, which is assumed to have certain semantics. Predefined properties can be used to model instance of and subclass of relationships as well as domain restrictions and range restrictions of attributes. Indeed, the RDF schema provides modelling primitives that can be used to capture basic semantics in a domain neutral way. That is, RDFS specifies metadata that is applicable to the entities and their properties in all domains. The metadata then serves as a standard model by which RDF tools can operate on specific domain models, since the RDFS meta-model elements will have a fixed semantics in all domain models. RDFS provides simple but powerful modelling primitives for structuring domain knowledge into classes and sub classes, properties and sub properties, and can impose restrictions on the domain and range of properties, and defines the semantics of containers.

Web Ontology Language (OWL) The next layer in the Semantic Web architecture is Web Ontology Language (OWL) [55], a language for Web ontologies definition and instantiation. OWL enhances RDF vocabulary for describing properties and classes: relations between classes (e.g. subclasses), cardinality, equality, richer typing of properties, characteristics of properties (e.g. symmetry) and instances. OWL is the W3C recommendation for ontology definition, but other standards also support similar characteristics (DAML+OIL). Several tools support modelling with OWL and DAML+OIL. The OWL language also provides three increasingly expressive sublanguages: OWL Lite, OWL DL, and OWL Full, each offers a different level of expressiveness at the trade-off for simplicity, thus offering a suitable sub language parts available for use according to needs. There also exists OWL based enhancement to web services oriented languages, aiming to handle semantic descriptions of

such services. OWL-S [73] is framework for containing and sharing ontological description of the capabilities and characteristics of a Web service.

An OWL-S specification includes three sub-ontologies that define essential types of knowledge about a service – service profile describes the outlining interface and characteristics of the service, a process profile defines the control flow of the service and the service grounding provides mapping with communication-level protocols. OWL-S has similar characteristics with a number of related protocols. The popularity of OWL-S in the Semantic Web community, as Web services description language, adds to the attractiveness of the language.

Chapter 3

The State of Art and motivation

In this section, he motivation of this thesis will be described according to the state of the art, limitations and research open issues.

3.1 Physical and Logical security convergence

Security is a word that generates negative sensations in most people's minds. It describes the uncertainty of the safety of their property or themselves. The evolution of technologies and infrastructure has added more complexity in the world of security. Now the threats come from different sources, internally and externally (crime). As described in Section 1.5 most system infrastructures include different kind of physical sensors and that provides an strong issues for information and integration management. The scientific community has addressed more effort for the introduction of new solutions. Different sensors provides different alarms or events each of them is managed separately without an shared information sharing methodology. For example, in a surveillance system the object detection and motion detection of a single cameras can be improved by other event and measures coming from adjacent cameras or sensors (e.g. audio surveillance system). In video analytics applications, the object detection and tracking performed by the single camera can be improved by means of additional information from other cameras or sensors ([17],[18],[19]). For example in case of occlusion or evacuation. The use of sensors correlation has three important objectives:

- Reduce the False Alarm rate and improve the Probability of Detection;
- Extend the capabilities of a monitoring system in order to cover the user requirements
- Manage the limitation of detection of a single sensors.

In literature are present most architecture that proposes a solution for physical security information integration, but some of them are for a specific kind of devices [20], some of them proposes also a technique for the reasoning [21], based on attack scenarios but they not are so scalable and extensible. The most popular are PSIM (Physical Security Information Management) system [22]. As described in [23] the PSIM are composed by:

- **Data Gathering:** Device level information gathered from a broad range of disparate security systems (video surveillance, intrusion detection, access control, environment sensors, etc…) that includes devices from different vendors
- **Data Evaluation:** The PSIM software should have the ability to evaluate the information that is gathered and based upon analytical algorithms, identify and prioritise, real incidents or situations
- **Confirmation**: The monitored 'situation' should be presented to a system operator in a clear, concise, yet comprehensive format, enabling an accurate and speedy response to a 'confirmed' security incident.
- **Resolution**: The PSIM system software should facilitate the presentation of logically displayed and clearly communicated actions that the Security Operators should carry out when managing a real time incident or situation.
- **Reporting**: All activity should be monitored and 'recorded', including all Operator actions, to aid compliance management, provide training scenarios and as an auditable record of activity subsequent to a security incident

But there is an other problem and issues in this system that has widespread in the last year, the logical security [24].

All asset and information about them are accessible from the inside by approved network access, without proper control mechanism they can be also accessed physically by insider. In the literature there are many works about the consideration of physical and logical security. As described in [25] often the areas of physical security and information technology security are worlds apart. Because these functions are in place and because they at least in part achieve their goals, management tends to perceive that major risks they try to mitigate are being addressed. Physical security systems and devices, process control systems, and IT infrastructures are being integrated without sufficient consideration of the security risks that the increasing intermingling of these systems and infrastructures introduces. For Instance, security system are distributed and composed by sensors that are physically separated from

other components such as central processors. Networks are used to connect these physically separated components. At the same time, information technology (IT) are composed by large numbers of workstations, servers, network devices, and networks that not only connect internal hosts and devices to each other, but also provide intranet and extranet connectivity. IT infrastructures are dynamic and their functionalities are out of control. But today, few paper s faces this problem. One first is written by the National Research Council during 2002, which described the failure of regional transmission grid. It could be happen by the damage of critical component failed in cascade manner due to their interconnection pointed [26]. In [27] some case of convergence are presented and it gives some recommendations concerning how to do so, but it did not specify the issue of vulnerabilities resulting from the convergence of physical security and their systems with IT infrastructures.

In [25] Schultz analyses the problems to give information security professionals a high level view of the physical convergence problem. It declares that research funding agencies would be well-advised to start soliciting research proposals in this area and to provide funding to researchers who appear capable of delivering promising research results related to ways of effectively identifying and mitigating physical convergence- related security risks. As the National Research Council asserted, special systems as well as other types of systems connected to the same networks provide a target rich environment for would-be evildoers. Research can and will provide answers to many of the issues that must be addressed.

The main problem of physical and logical security starts from the enterprises department separation. Physical security systems are in the scope of a physical security department that assesses and mitigates risks in large part resulting from the necessity of allowing physical access to employees, contractors, and visitors. In the same time, physical security systems have evolved considerably in terms of sophisticated computing systems connected to networks, physical security staff members are not likely to have much training and knowledge in computing and networking, let alone information security.

The IT department is responsible to ensure that the infrastructure and components are in place and operating efficiently. IT staff have considerable knowledge concerning computing, networking, and programming but not about physical security and physical security systems. Physical security and IT security are typically also very disparate functions.

Security-related risks associated with deploying systems and devices used to physical security and to support process control are increasing because progressively more they are connected to networks. In the past it was not a problem, the physical system was isolated, simple, and

protected by physical security measures. The connection and the openness to networks provides different security risks that costly and disruptive security- related incidents could easily result. An attacker can access locally or remotely to target of the systems and to devices. The potential for unauthorized remote access now is an high problem exists due to network connection.

According to IMS Research, which estimates the Internet-capable equipment, devices connected to an Internet Network can pass from the 5 billion milestone in 2010, to 22 billion by 2020. This assumption reflects the spread of personal devices such as Smartphone and tablet computers, and also includes all the sensors, cameras, and devices used in security that are now IP-enabled because of the convergence of the IP network. This can have a bad impact on network performance and security. This situation reveals new security challenges, the convergence of Internet-connected devices, voice, video, and data also provides ways to integrate logical and physical security.

In [3] this problem is introduced by CISCO enterprise. It declares that the lack of integration creates the following challenges:

- No single system to identify a person's identity because each functional security department controls its own identity database
- Increased potential for theft
- Lack of IT management and application of best practices applied to physical security devices, or a lack of best practices applied consistently across departments or organizations
- Lack of physical monitoring of logical security devices that can detect tampering; that is, unauthorized access to a logical security device console

Obviously, a convergence methodologies requires different efforts for network connection because increases the data traffic, for installation device for the detection of malicious attacks and etc..

3.2 Complex Event Processing

A complex monitoring system for the infrastructure protection managing different information has the objective to detect events that can be occurred and determinate the

situations corresponding to these events trough opportune procedures. This kind of application are called event-driven detection system.

One of the main research field involved in these kind of application if the Complex Event Processing (CEP) [8]. CEP addresses two important concepts in order to develop scalable and dynamic systems. The CEP have a double function provide information and process into events, but also it detect the relationship between events. Such as temporal correlation, by the definition of rules of correlation, called Event Pattern. Through the aggregation of single events can be generated more abstract events. CEP provides a separation coupling between basic events with a strong relationship to the semantics of the underlying technology and complex events closer to the semantics of the application. Critical applications are interconnected and imposes event load to be processed by CEP system. CEP will be a tool to derive understandable information on the basis of a large number of events.

In this context it is important to manage the event correlation in the presence of highly dynamic systems and support mechanisms for self-organization. Guaranteeing non-functional properties, such as, reliability, availability, performance, and security pose major challenges on the technical infrastructure. According to the application, specific complex event patterns should be developed despite of occurred events during the working time of the system.

Regarding the physical security for infrastructure protection, the event correlation is useful for interpretation of complex event composed by simple event (event form single sub-system/sensor). Correlation is the contemporary of event (temporal correlation), sequence of specific events.

In literature ([9], [10], [12]) there are several approach and techniques for event correlation, the main are described following :

- Rule-Based

  One of the approach to event correlation is the rule-based analysis. It is based on the combination of events, the rule engine analyzes data until it reaches the final state.

  In rule based event correlation, the system uses a set of predefined assumption to evaluate incoming observations until a conclusion is reached. The correlation ability depends on the depth and capability of the rule set. In a rule based event correlation engine, information is represented in three levels [9]:

  - **Knowledge level:** Domain-specific expert information is available in a knowledge base, which is the rule repository.
  - **Data level:** Information about the problem at hand is allocated in the working memory. At this level, facts are stored.

- **Control level:** The response about how to apply the rules from the knowledge base to solve a given problem is located at the inference engine.

Typically, rule-based systems have a offline execution mode.
For very accurate results it is needed of expert knowledge in order to respect the criteria and update the rules in case of changes.

- State Transition Based

Finite State Machine [11] are most used in computer science for their power and easiness. It allows to describe the behaviour and states of a system. the sequence of events it is defines as a state transition, a state of system is reached after a sequence of events. It allow efficient and effective computation. But due to the complexity of new systems make difficult to model them in details and some are allows abstractions. Many solutions consider the use of artificial intelligence, graph theory, neural networks, information theory, and automata theory. For instance, Finite State Machines represent a type of models which can be applied to model-based event correlation approaches. Deep knowledge of the system may describe its structure (static knowledge) and function (dynamic knowledge). Finite State Machine have been considered in some work on diagnosis of Discrete Event Systems (DES) [12]. Machines called diagnosers are automatically synthesised and are able, under some conditions, to detect the occurrence of unobservable events at runtime. The goal of the diagnoser is to infer the presence of alarms or interesting situation from the sequence of observed events. The diagnoser is a Finite State Machine that is built from the model of the system, which is also assumed to be a Finite State Machine. At runtime, the diagnoser observes the behaviour of the system and estimates the state which it has reached. The diagnoser possesses information about the possible failures in the different states of the system and can detect them in a finite time. The advantage of FSM based event correlation is the formally definition and automation. On the other hand it requires the availability of system behaviours of the which is unfortunately rarely in practice.

- Model Based

Model Based Correlation uses a model of the physical world to represent the structure and behaviour of the system under observation, as an inference method. It is not an approach or technique, but a paradigm instead. It has a connection with rule based systems, it is not

as a rule based system because it specifies a system model, with events as consequences of certain model states and transitions, while rule based system specifies event patterns as conditions for certain actions.

- Classification Based

It classifies an event in a predefined class. The most popular technique is the Support Vector machine for classification [14]. The input is a vector contains a low level features, from that is able to detected high level concepts. An other approach is the Bayesian network [15]. It is an acyclic graph which models probabilistic relations, e.g. between the threats and the primitive events detected by sensors. It requires well defined a-priory and conditional probabilities of the. An other most popular approach is Artificial Neural Network [16]. It models in artificial mode the work of human brain. The node of network performs operations on weighted inputs to get an output. The processing can be performed by different techniques (mathematical, temporal, etc.). it is able to solve complex problems, but the selection of a suitable network architecture can be difficult and the training can take lot of time.

3.3 Decision Support System

Decision support systems (DSS) are interactive, computer-based systems that help to choose activities countermeasures. They provide traditional information access (data storage) and retrieval functions with support for model building and model-based reasoning. Typical application areas of DSSs are management and planning in business, health care, the military, and any area in which management will encounter complex decision situations.

As described in [28] a typical DSS id composed by:

- Database management system (DBMS). A DBMS serves as a data bank for the DSS. It stores large quantities of data that are relevant to the class of problems for which the DSS has been designed and provides logical data structures (as opposed to the physical data structures) with which the users interact. A DBMS separates the users from the physical aspects of the database structure and processing. It should also be capable of informing the user of the types of data that are available and how to gain access to them.

- Model-base management system (MBMS). The role of MBMS is analogous to that of a DBMS. Its primary function is providing independence between specific models that are used in a DSS from the applications that use them. The purpose of an MBMS is to transform data from the DBMS into information that is useful in decision making. Since many problems that the user of a DSS will cope with may be unstructured, the MBMS should also be capable of assisting the user in model building.
- Dialog generation and management system (DGMS). The main product of an interaction with a DSS is insight. As their users are often managers who are not computer-trained, DSSs need to be equipped with intuitive and easy-to-use interfaces. These interfaces aid in model building, but also in interaction with the model, such as gaining insight and recommendations from it. The primary responsibility of a DGMS is to enhance the ability of the system user to utilize and benefit from the DSS. In the remainder of this article, we will use the broader term user interface rather than DGMS.

A DSS in order to provide a decision-making process shall have an high-qualitiy and unambiguous information integration. The semantic web assure the integration of information coming from different source and the transformation of DSS in a SematicWeb-DSS. But these systems do not have reasoning and integration of incomplete information capabilities [106], [107]. The scope of this thesis is to improve the capabilities of DSS in order to cover the limitations.

3.3.1 Semantic and Ontological models in DSS and data integration

New open standards are widespread in order to define techniques for cooperative environments. They are based on the adoption of common data models to formally define and represent data knowledge. For instance, the Open Geospatial Consortium provides as data model an XML schema (the Sensor Model Language) for defining geometric, dynamic and observational characteristics of sensors and other standards as Observation and Measurement to describe observed phenomenon [29]. This standard provide some general information on sensors for data discovery, processes and analyzes measurement of sensors, localizes the sensors, provides information about performance (e.g. threshold,

accuracy). But, as aforementioned before, these standards focus on syntactic aspect for the interoperability and not on semantic form.

The semantic enriched data model and ontological model instead grant interoperability among multi-technology systems, providing a formal model for the integration of data gathered by different and heterogeneous sources. A formal data model specifies data relations, terminology and meanings, it is implemented through a set of ontologies formally described in OWL (Web Ontology Language) . Furthermore, the use of ontological models provides advantages in the use of web services architecture, the most popular and used in critical infrastructure applications.
The integration and interoperability among heterogeneous data can be done for:

- Overcame syntactic technical heterogeneity;
- Overcame semantic ambiguities and interpretation;
- Deliver re-usable application and components for different domain
- Evaluate the meaning of phenomena and observation from temporal, spatial and thematic perspective.

In literature are available some models for interoperability, which provide syntactic integration but not semantic relations. One of this is the Sensor Web Enablement (SWE) [30] , a suite of specifications to model sensor characteristics and services. In the suite are included a Sensor Model Language(Sensor ML), an Observation & Measurement and Sensor Observation Service. They allow to model sensor and sensor observations, data retrieval mechanism. Furthermore it is possible to specify information as coordinates and timestamps, but they provide static representation of data without give details on data meaning of sensor observations, the evaluation of the phenomena situation awareness [31].
In [30] is proposed the SSW framework, a fisrt definition of sensor Web, in which there is a first attempt of semantic annotation to existing standard sensor languages of the SWE, in order to increase interoperability and provide contextual information for situation awareness.
Several attempts have been done in the data modelling field in terms of different ontologies. For instance in [32] a data combination and relation as an ontology. In [33] a sensor network ontology is presented with semantic representation of information. In

order to assure the interoperability among monitoring system and integration of data source, different methodologies have been developed.

Nowadays solution for interoperability and integration exist, but a proposal architecture comprising phenomena observation, interoperability, integration of different data, detection of dangerous events, does not exist yet.

The use of ontology model is needed in order to give a shared, semantic model for the information process.

Many works have demonstrated that increment of precision in service discovery in presence of semantic representation.

In this work it is defined a common data model to describe data information and observation.

The proposed model is in RDF language and it is compliant with the Semantic Sensor Web [21][22] approach, too.

The adoption of an ontology helps to fuse data and meaning, so, for the model description, we adopted open standards for interoperability, enriched with semantic information. In fact, a semantic based technology allows us to deal with raw data and manage them as a global Knowledge Base, we obtained an explicit representation of the meaning of data and services that is useful for extracting relevant information and for integrating them. The knowledge can be managed as a database that can be queried in a structured way enabling advanced operation as logic reasoning. Furthermore, it is possible to enforce consistency verification on modelled data to verify the compliance to the model and the acceptability of sensed value

In the literature, some semantic approaches to manage heterogeneous data from sensors are available and some semantic decision models are beginning to be used ([98], [99]). In [100] an architecture for sensor information description and processing, named semantic web architecture for sensor network (SWASN), is proposed. The architecture is based on four layers: the first is the physical level composed by different sensor networks. Each sensor networks manage its own data format. The data are processed in an ontology layer, in which each network has a local ontology. A global ontology is built upon a common vocabulary and it is processed in the semantic layer for the knowledge extraction, through inference and semantic reasoning. Finally, at user level, it is possible to query the ontology in order to process and elaborate data. In this case, the architecture proposed is responsible to process data semantically and then a client can request them for post

elaboration, there is no in-line event detection. Similar architectures are presented in [101], [102], [103] and [104]. In particular,[101] and [103], an automatic process for transformation of XML data into RDF is proposed, the transformation process is driven by semantic reasoning and mapping rules. The transformation is in real-time but not any detection system is proposed. In [104], a middleware architecture to manage event detection in real-time is presented, this is an architecture for automated, real-time, unsupervised annotation of low-level context features and corresponding mapping to high-level semantics. It enables the composition of simple rules through specific interfaces, which may launch a context aware system that will annotate content without the need for user technical expertise. The middleware has a semantic model only for the event management. There are no common models for the data acquired by sensor. The presented approaches use in-line techniques for semantic enrichment data model and in some case propose a model for in-line event detection. In this paper, we have proposed an improvement of these techniques in order to provide a semantic common data model for heterogeneous data interoperability, for real-time event detection with semantic model and a base for a development of a semantic DSS for post elaboration.

Chapter 4

A smart decision support system based on fast classifier and semantic post reasoner

In modern decision support systems there is the need to improve the performance in terms of detection, reliability and real time capabilities. These features are usually in inverse proportion. In this section we propose an innovative approach for a smart event detection and enriched phenomena comprehension. In particular, the proposed approach is based on a two steps process that tries to quickly identify an alarm and then elaborate the acquired knowledge base with a post reasoner to refine the final decision and give operators more feelings about the situation assessment and raised alarms.

4.1 Why semantic model?

In these complex scenario, not only smart surveillance and alert systems are needed but enriched decision support systems (DSS) are desirable. Such Systems, as described in section 3, rely on heterogeneous data acquisition tools (sensors, video, historical and simulated data, ...) and on data elaboration to prune non significant information; nevertheless, this is not enough as there is the need to interpret what data really represents to reduce false positives and detect even weak alarm conditions. In recent years, scientific world's attention has been devoted to both the information management with information and decision fusion approaches, and to the quantitative security estimation of these systems. On the other hand, to improve the situation assessment, it is possible to adopt different types of models for description of knowledge-base, event correlation and for the definition of the situation and threat identification. Very promising approaches are based on semantic and ontological models.

The semantic model can be used for understanding observed phenomena. In particular all sensors must share the same data model and the same interpretation of data. The data model must provide a syntactic interoperability mechanisms and procedures for semantic enrichment to build models in order to:

- ensure a correct and shared information interpretation,
- aggregate raw data into events (simple and composed), that will be used for the situation assessment before a final decision.

In the literature some approaches for event detection and decision support based on semantic inference rules for phenomena comprehension are available.
Nevertheless, due to the introduced overhead, the knowledge base is just inferred in offline mode.
In this work we propose an innovative approach for smart event detection and enriched phenomena comprehension: the knowledge base will be inferred in real time, for the event detection, and a light smart classifier will raise an alarm.

4.2 A model for monitoring system

A monitoring system is, in general, a very complex architecture ([72], [76]) as it involves different components and subsystems that manage data and information in very different ways.
In this section, we propose a data and a monitoring system models in order to cope with the complexity of heterogeneous data sources and to support the development of a reliable DSS. At a very course grain, it is composed of two main layers, namely the sensor network and the monitoring system, these are, in turn, modelled in deeper details to cope with such complexity.
The proposed models enable to define simple data and atomic events coming from different sensors, it allows to model the complexity of the sensor network and the correlation among different events to define composed events to improve the knowledge about the system and locate critical scenarios. The proposed monitoring system models the detection system as a two steps process

- a real-time reaction by means of a fast classifier

- an offline activity through a semantic post reasoner.

The former aiming at providing proper alarms when dangerous events occur, the latter aiming at providing a complete and detailed picture of the situation, useful for operators both in understanding the situation and for decision supporting. In next sections, we will illustrate the details about the system and data models and we will present an innovative architecture to implement the two steps detection strategy. In Figure 5 the model layers are reported. This structure helps to better represent the different features of the underlying sensor network and the monitoring and decisional approaches.

Sensor network model

This block models the physical sensor network responsible of the monitoring activity. The sensing elements can be various: wireless sensor networks, cameras, microphones, infrared sensors, etc. To better identify the network features and the heterogeneity of different sensing elements, this block is divided into three levels:

1. *Sensor physical features*: this level describes the physical characteristics of individual sensors and the type of parameters they measure as: on-board sensors, identifier, measured parameters and communication channels (wired or wireless).
2. *Measurement typology:* this level models the different types of measurements made by the sensors and the associated proprieties.
3. *Topology:* this level models network topology characteristics as sensor localisation, deployment features, sensors communication topologies and distances.

Monitoring system model

This block is responsible for the reasoning on measurements taken by the sensors, for event classification and for triggering reliable alarms. It is composed by three levels that correspond to three different operating phases; the first two levels operate in real-time, they are designed for the early detection of dangerous situations launching alarms, the post reasoner level aims at providing information on the state of the system, and it usually works at a later time in order to refine/re-parameterise the on-line configuration.

In particular, the operating phases can be described as follows:

Real-time acquired knowledge

This is the phase where the system collects all the measurements from the individual heterogeneous sensors. It provides a level of integration for the measurements made by the sensor networks. It models the typology, the structure and the values of raw and structured data acquired and transmitted by sensors, enriched with semantic information about them [77]. At this level, the data is modelled and processed by both the fast classifier and the post reasoner. The encoding language, used to process and transmit information is the RDF standard.

Real-time classified knowledge

This is the phase where the classification of events is performed on-time. The in-line classified knowledge layer models the knowledge derived by the application, on the sensed data. This level already works on semantically enriched data. The overhead for linking data with information about them is necessary at this level because there is a multitude of events that can be detected only by combining information from collections of sensors, which are heterogeneous both for typology of measurement carried out and for the data format in which they are sent to centraliser nodes. Furthermore, at this level, not all the details on the current situation are taken in consideration, in order to allow the classifier to perform efficient decision tasks, even if it is not able to derive the full knowledge about the monitored environment. The cut information is then re-considered into the abstract model of the post reasoner knowledge, which works without real-time constraints.

Post reasoner knowledge

This level traces all system events and applies rules of inference on collected data, in order to analyse retrospectively the relationship between the sensor measurements, the classification of events and associated alarms notified. The level is designed to simplify the

evaluation of the causes of critical events, allowing operators to refine the classification criteria based on the information provided by the tracking of events that led to false alarms (false positive or false negative). This layer is focused to derive useful knowledge to have a detailed view of the situation, finalised to:

- help in situation awareness
- support in the decision process.

The system implementing the semantic reasoner is much more computational expensive than the classifiers used for the real-time decision. At this level, in fact, the outputted inferred data is designed to give support to users with offline reasoning and data mining features, which can be exploited to get a complete knowledge of the situations, even at a later time.

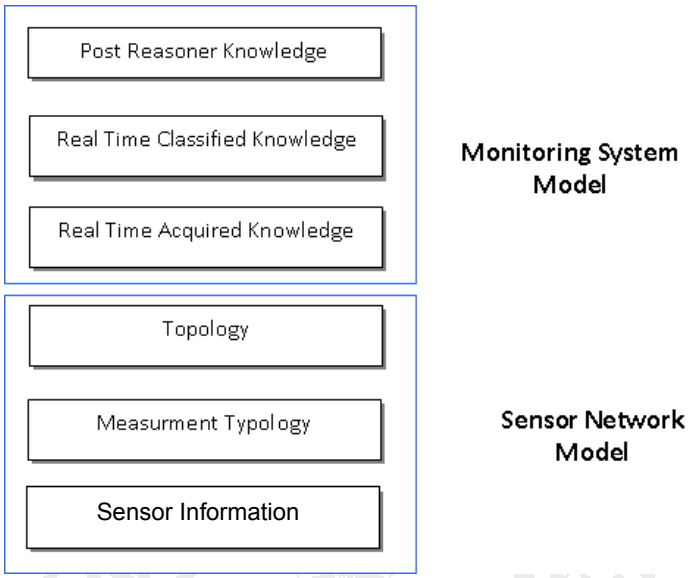


Figure 5 System Model

4.3 The system architecture

The core of the monitoring system is made of:

- a smart event classifier (implementing the real-time acquired and classified knowledge layers)
- a post reasoner (implementing the post reasoner knowledge layer).

We implemented a fast classifier in order to detect potential dangerous condition and then, if necessary, raise an alarm. The event detection is carried out by correlating data coming from different sensors. In fact, in real situations the potential hazard cannot be

detected by using data coming from a single device. According to the above considerations, we designed and developed different modules to:

- semantically enrich the data
- implement the smart real-time classifier
- implement the post-reasoner.

The proposed architecture is illustrated in Figure 6. We suppose that the sensor data source consists of heterogeneous sensors or groups of sensor (WSN, camera, intrusion device, etc.) measuring different parameters. Data are gathered from sensor nodes and can be accessed through a specific sensor gateway. Gateway sensors code the data in XML.
The system is made of three modules:

- **Transformation and integration module:** It gathers and integrates data coming from heterogeneous sensors, in order to make them suitable for the population of the ontological model implemented into the 'semantic classifier' module. Data are semantically labelled, to add information and description about measures and sensors ([78],[79]). This information is then automatically encoded in RDF triples by using a XSLT transformation engine implemented by the transformation module. The RDF files, produced by any network, contain unique references to descriptions of the measured variables and parameters. They are then integrated (integrator module) into a single RDF file, containing instances of the measured values that will then populate the ontology.

- **Semantic classifier module:** this component implements the fast classifier. The events detection is carried out by correlating data coming from different sensors. The implemented classifier is rule-based: the combination of measurements allows classifying events, if there is a particular event, classified as critical; the system raises the corresponding alarm. The smart classifier operates on semantically enriched data, so, a rule can be expressed by combining atomic events from heterogeneous sources. Information about events are managed in the RDF DB (triple store) that contains all the event instances codified in RDF; events combination is codified trough SWRL rules [80]. The chosen classifier has a standard structure to build a predictive model, based on a learner and a predictor component. We adopted a decision tree classifier [81]. In decision tree

mechanisms, the set of decision rules is modelled as a tree in which leaves represent class associated to atomic events to be detected and branches represent conjunctions of features, i.e., condition on the sensed data that lead to those dangerous composed event classes. In order to define the branch rules of the decision tree, a domain expert manually defines a training set made of already classified data; the learner module uses these data in order to set the predictor parameters, which tune the automatic detection of an alert condition. The predictor is responsible to classify the data to decide if alarm conditions occur. To increment the system performance, the rules are pruned recurring to a manual refinement made by domain experts. In Listing 1, a small example of rule is reported; it has been codified as a tree branch (Figure 7). The rule states that if two different sensors (S1 and S2) in a given position detect a pressure value above a given threshold, then an alarm must be arisen. Note that the pressure is expressed with different measure units but this problem is completely overcome thanks to the semantic enrichment that is not reported in this example.

```
( S1.location=41°53’24” N, 12° 29’ 32” E,
  S1.Pressure >101.325 kPa,
  S2.Pressure >30inHg,
  S2.location=41° 53’ 37” N, 12° 29’ 11” E)
  ➔ Alarm
```

Listing 1 Example of rule

The predictor is a parametric system, and parameters are used to tune the reliability of the classifier output. Periodically, or even when misclassified events occur, the learner can recalculate the parameters of the predictor on the basis of a new training set, properly built to refine the behaviour of the classifier. Furthermore, the classifier may be synthesised with a reconfigurable hardware as an FPGA to boost up the performance and meet real-time constraints [82]:

- **Post reasoner module:** This module provides functionality to query the data correlated to events using the query language [83] and enables the ability to perform further analysis with other methods of classification. This allows to analyse the causes of the critical events that have taken place (basic function for an operator) and to refine the procedures and rules of classification if a false alarm

rate is not acceptable. Reasoning operations are performed on the data in order to extract inferred knowledge from them, recurring at Pellet [84] reasoner.

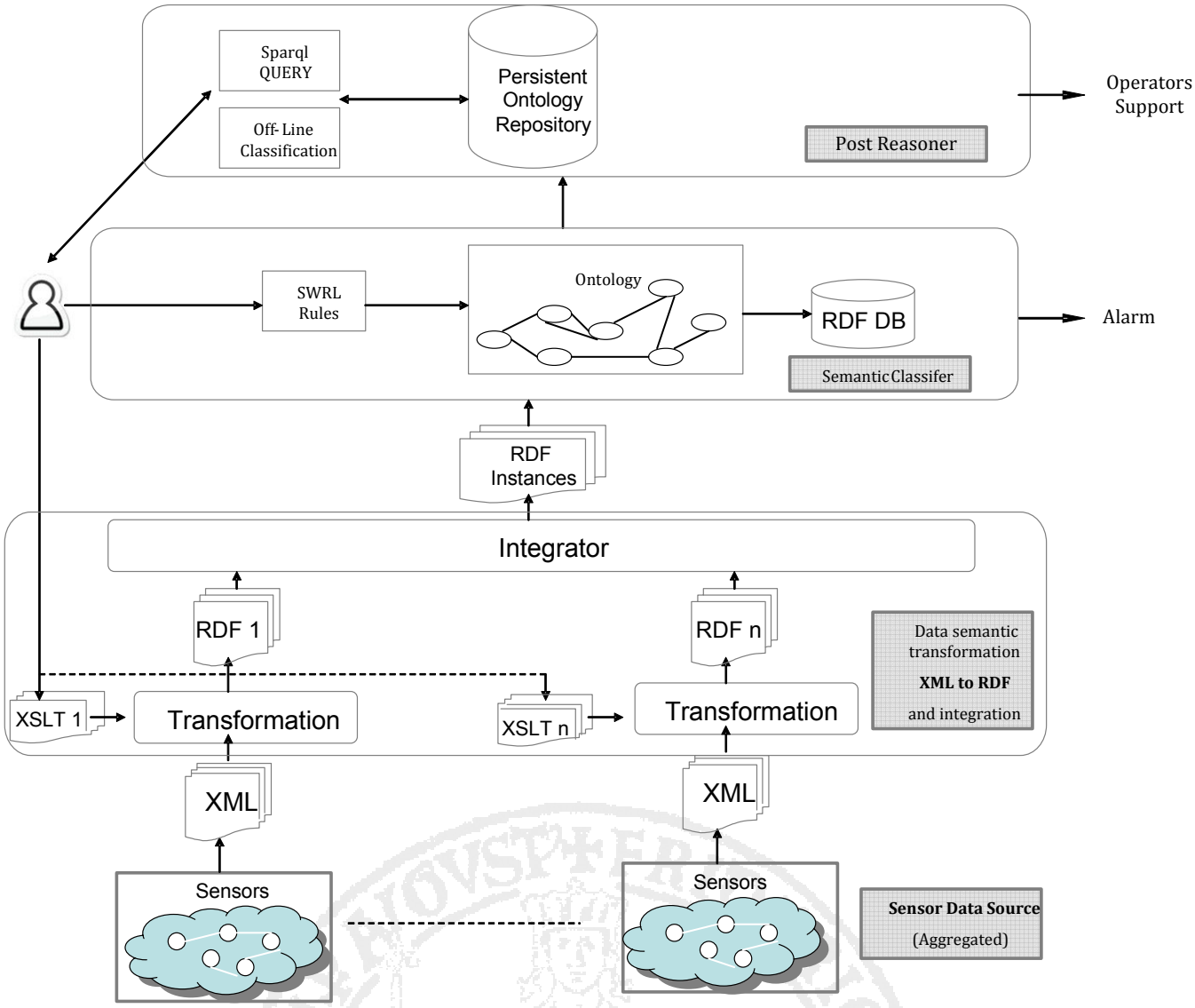


Figure 6 System architecture

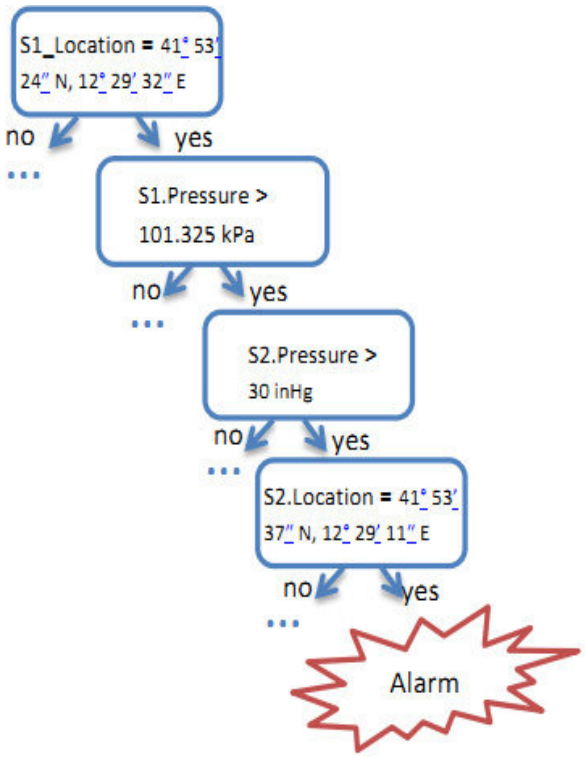


Figure 7 Example of rule:tree branch

The classification actions performed in real-time feed the knowledge base for the post reasoner component. The just built ontology is stored in a triple store repository [85] and can be used offline through the adoption of semantic query languages as SPARQL [86]. Through queries, the post reasoner is able to understand and explain to end users the meaning of the alarms and their causes.

The knowledge base can be seen as an information repository about a particular domain of interest. Typical knowledge bases consist of concepts, properties and instances. We encoded the knowledge base by using an ontology, i.e., a set of classes, properties and instances defined as follows. The classes define the domain concepts; the properties define the relation between classes (domain to range) or attributes (a property of a class). The ontology depends in part on the environment and it will be illustrated in next sections with a case study.

During reasoning, inferences are made, classifying instances of the ontology and associating new properties to instances while maintaining logical consistency.

The reasoner, based on Pellet [87], is able to infer logical consequences from a set of asserted facts about the monitoring system defined by user experts. In particular, it is composed of two components, one implementing the general inference rules and one the specialist rules, defined by domain experts in order to capture the relevant knowledge

about the environment to be monitored. The system uses first-order predicate logic to perform reasoning. The inferences proceed both by forward chaining and by backward chaining.

4.4 Data model and processing

In this section, we explain the data processing workflow, detailing how the data is gathered, integrated and analysed. The relevant domain knowledge is encoded with the help of domain experts using a proper ontology, which models the elements of interest in terms of concepts and relationships relating to the phenomena to be monitored, the events and the associated actions to be performed.

The proposed system manipulates sensed data for raising alarms, performing the following steps:

1. Sensor networks gather data and format them in XML files [104], they encode both sensor properties and measured values.
2. Using XLST engine, XML files are semantically enriched and transformed in RDF files. This file is compliant with the domain ontology and it is suitable to perform semantic reasoning.
3. Composition rules are defined and applied to simple events in order to build composed events, they are coded in SWRL [80] and used to automatically populate the ontological model.
4. By exploiting JENA primitives, the ontological model is automatically populated with the RDF triple instances.
5. The classifier module is able to detect critical events. The classification is performed on the basis of a training set containing properly labelled composed events.

Classified data populate the domain ontology. The data can be queried for the offline analyses and rules refinement for preventing critical events misclassification. To detail the information processing steps, we will adopt a running example on a real use case. As already said, a critical scenario consists of a sequence of simple and composed events. To model such scenario we have built an ontology. An example of ontology for physical security in railway domain system, reported in Figure 8 aims at representing the domain of interest (including measures, events and alarms), i.e., a subway station monitoring system.

The ontology is implemented through JENA (http://jena.apache.org/) library. According to the system model, the ontology has been structured in three parts: the class describing the Sensors features, the class describing the Measurement and the class describing the *Detected_Events.* The Sensor class includes the sensors description and has subclasses for each device sensor type (e.g., *Infrared_Barrier, Intelligent_Camera*…). The Sensor class is associated to different classes (*Sensor_Type, Location, Measurement, Detect_event*). The *Sensor_Type* class specifies for each sensor the measured parameters and their properties. Location class specifies the sensor location, while the Measurement class provides description of the measures performed by each sensor (e.g., *Train_Passing, Line_cross*, …). For modelling the events, we have defined the *Detected_Event* and *Alarm* classes describing the properties of several events that can be detected by the system from the measurements made by the sensors. The *Detected_Event* class is specialised into simple events (e.g., average temperature and high humidity) and composed events that defined by composing different events and measurements from multiple sensors. Simple events are detected by the measure of a single sensor and can be correlated in composed event. *Composed_Event*, classified as critical, are associated to proper alarms described by the *Alarm_Class*. As already said, event composition is performed by using rules written in SWRL, with the help of domain experts.

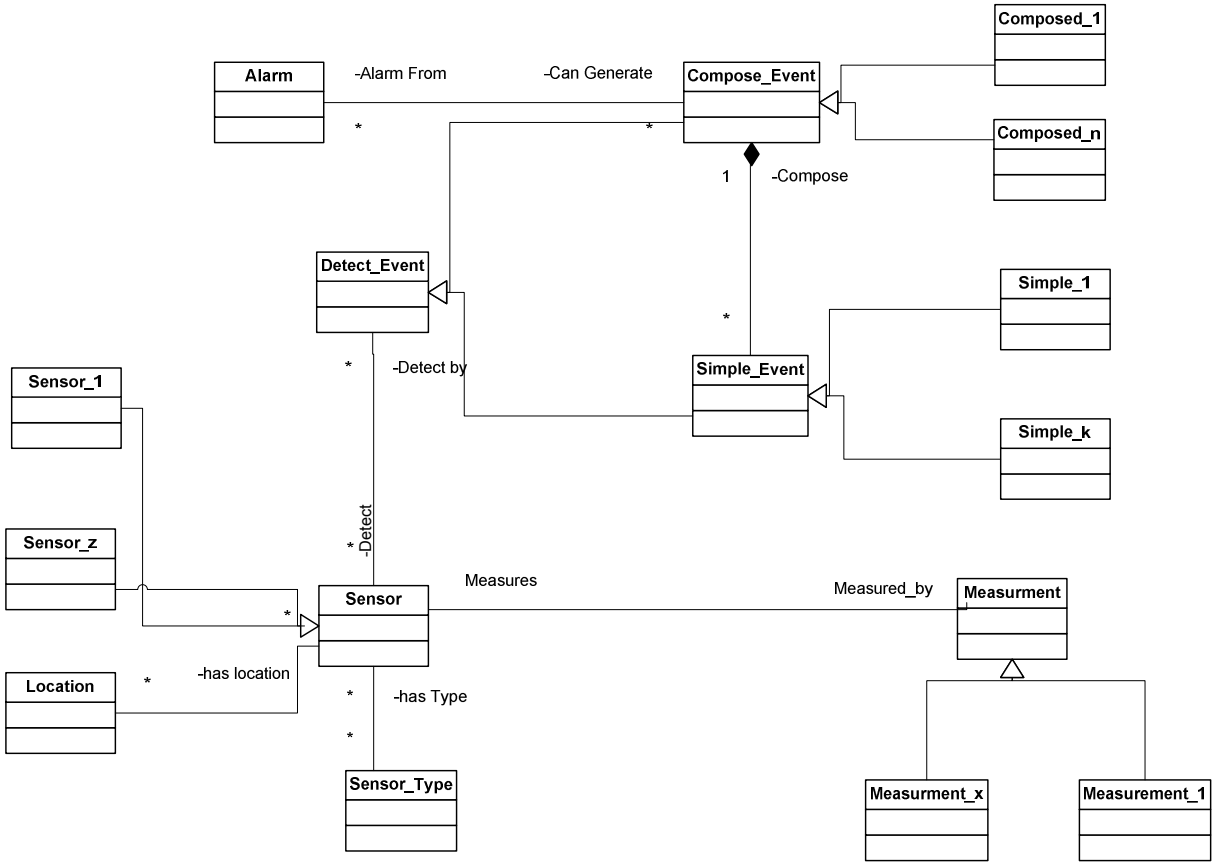


Figure 8 Example of Ontology for Physical Security

4.5 A typical case study in Railways Security Domain

This section illustrates the application of our system in a typical case study of a subway station [88]. The station is supervised through different sensor technologies (smart-cameras, infrared sensors, etc…). The correlation of the different measures, gathered by the sensors, allows to detect some events (e.g., physical intrusions, explosions, …) and, if necessary, raise a proper alarm to the operator.

The station is equipped with a security system including intelligent cameras (S1), active infrared barriers (S2) and explosive sniffers chemical biological radiological and nuclear explosive (CBRNe) (S3) for tunnel portal protection.

The critical scenarios are expressed through the composition of simple events detected by these sensors. For example, we show the detection of the *Drop_Explosive_Tunnel* event, regarding the release of explosives in an underground tunnel.

Let us suppose that the dynamic of the scenario follows the steps reported below:

1. the attacker stays on the platform for the time needed to prepare the attack, missing one or more trains
2. the attacker goes down the tracks by crossing the limit of the platform and moves inside the tunnel portal
3. the attacker drops the bag containing the explosive device inside the tunnel and leaves the station.

These events are detected by sensors Si and can be specified as follows:

- E1. Loitering presence on the platform (E1 by S1)
- E2. train passing (E2 by S1)
- E3. platform line crossing (E3 by S1)
- E4. tunnel intrusion (E4 by S2)
- E5. explosive detection (E5 by S3)
- C1. Dangerous_Presence C1 <-(E1, E2)
- C2 Possible_Explosive C2 <-(E4, E5).

Where E1, E2, E3 and E4 are simple events, C1 and C2 are composed events. The combined event *Drop_Explosive_Tunnel* occurs if one of the two composed events takes place:

1. if (E1, E2) and then (E4, E5)
2. if E3 and then (E4, E5).

In the remainder of this section, we illustrate in details how the system detects the first case. Listing 2 reports the sensed data codified in XML. The listing contains basic information about the CBRNE sensor, detecting the presence of an explosive (value = true). In particular for a sensor is reported the node ID, the measured values and temporal information.

```
<?xml version='1.0' encoding='UTF-8'?>
   <result>
    <nodeid value='1'/>
    <location value='Station1'>
    <name value='Chemical_Presence'/>
    <data value='true'/>
<timestampvalue='2012-05-13T09:00:03+01:00'/>
      </result>
```

Listing 2 CBRNE Sensor Output in XML format

The output information is then semantically enriched by exploiting the proper domain ontologies. Using XLST engine, the XML file is transformed in RDF, as reported in Listing 3. The RDF file, containing the sensed data and their semantic descriptions, is automatically created.

```
<!-- http://www.owl ontologies.com/Ontology1.owl#Chem2 -->
 <owl:NamedIndividual rdf:about="http://www.owl-
ontologies.com/Ontology1.owl#Chem2">
   <rdf:type rdf:resource="http://www.owl-
ontologies.com/Ontology1.owl#Chemical_presence"/>
   <atTimeDate rdf:datatype="&xsd;dateTimeStamp">2012-05-
13T09:00:03+01:00</atTimeDate>
   <hasChem rdf:datatype="&xsd;boolean">true</hasChem>
 </owl:NamedIndividual>
```

Listing 3 Semantic enriched information of Simple Event in RDF

Listing 4 reports the composition, performed by SWRL defined rules, of two events already integrated and coded in RDF. The listing describes the sensors CBRN1, instance of CBRNe class, able to observe chemical presence, and CHEM2, instance of Chemical_Presence Class, positioned in the station 1 (S1).

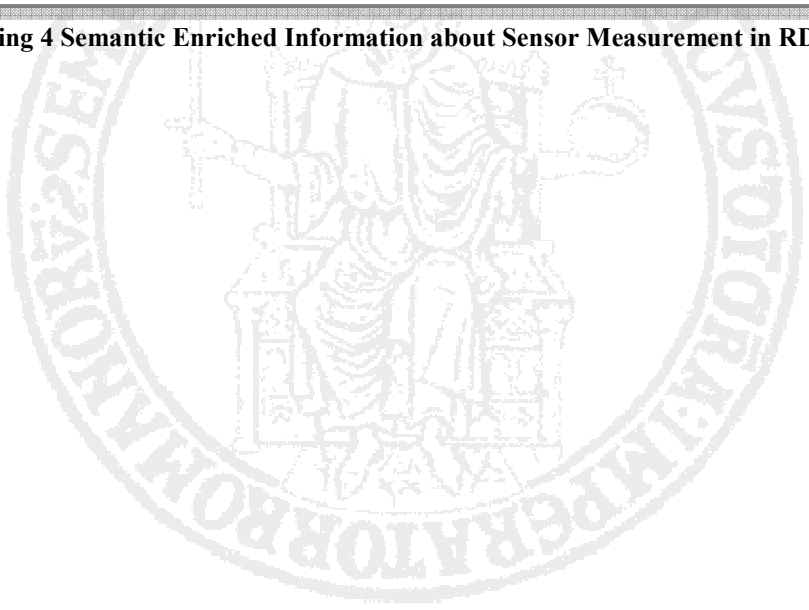
The Boolean CHEM2 allows the smart classifier to infer the presence of explosive event, in fact, it firstly detects simple events, compose them and raise the alarm condition. The composition of these simple events produces the following composed events C1 (*Dangerous_Presence*) and C2 (*Possible_Explosive*).

C1 states that if both events E1 and E2 occur in the same time range, then '*Dangerous_presence'* event is triggered, the second one states if E4 and E5 events occur, the composite event '*Possible_Explosive'* is detected. The combination, with temporal constraints, of '*Dangerous_Presence*' and '*Possible_Explosive'* events triggers the '*Drop_Explosive_Tunnel'* event, launching the corresponding alarm.

Listing 5 represents the activation event E5 '*Detect_Explosive'* is triggered by condition on *'is_Explosive_Detection'*. In the second part of the listing, the event *'Drop_Explosive_Tunnel'* is composed as composition of *'Dangerous_Presence'* and *'Possible_Explosive'* that occur in temporal succession.

```
<!-- http://www.owl-ontologies.com/Ontology1335263048.owl#CBRN1 -->


<owl:Thing rdf:about="http://www.owl-ontologies.com/&Ontology1.owl#CBRN1">
<rdf:type rdf:resource="http://www.owl-ontologies.com/&Ontology1.owl #CBRNe"/>
<rdf:type rdf:resource="http://www.owl-ontologies.com/&Ontology1.owl #Sensor"/>
<rdf:type rdf:resource="http://www.w3.org/2002/07/owl#NamedIndividual"/>
<&Ontology1:Measure rdf:resource="http://www.owl-
ontologies.com/&Ontology1.owl#Chem1"/>
<&Ontology1:Measure rdf:resource="http://www.owl-
ontologies.com/&Ontology1.owl#Chem2"/>
<&Ontology1:hasType rdf:resource="http://www.owl-ontologies.com/&Ontology1.owl#
Chemical"/>
<&Ontology1:Detect rdf:resource="http://www.owl-
ontologies.com/&Ontology1.owl#Detect_Explosive"/>
<&Ontology1:hasLocation rdf:resource="http://www.owl-
ontologies.com/&Ontology1.owl#Station1"/>
</owl:Thing>


<!-- http://www.owl-ontologies.com/&Ontology1.owl#Chem2 -->

<owl:Thing rdf:about="http://www.owl-ontologies.com/&Ontology1.owl#Chem2">
<rdf:type rdf:resource="http://www.owl-
ontologies.com/&Ontology1.owl#Chemical_presence"/>
<rdf:type rdf:resource="http://www.owl-
ontologies.com/&Ontology1.owl#Detect_Event"/>
<rdf:type rdf:resource="http://www.owl-
ontologies.com/&Ontology1.owl#Measurement"/>
<rdf:type rdf:resource="http://www.owl-ontologies.com/&Ontology1.owl#Sensor"/>
<rdf:type rdf:resource="http://www.w3.org/2002/07/owl#NamedIndividual"/>
<&Ontology1:atTimeDate
rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTimeStamp">
2012-05-13T09:00:03+01:00 </&Ontology1:atTimeDate>
<&Ontology1:hasChem rdf:datatype="http://www.w3.org/2001/XMLSchema#boolean">
true </&Ontology1:hasChem>
<&Ontology1:MeasureFrom rdf:resource="http://www.owl-
ontologies.com/&Ontology1.owl
#CBRN1"/>
<&Ontology1:Measure
rdf:resource="http://www.owlontologies.com/&Ontology1.owl#Detect_Explosive"/>
</owl:Thing>
```

Listing 4 Semantic Enriched Information about Sensor Measurement in RDF



```
<!-- http://www.owl-ontologies.com/Ontology1.owl#Detect_Dangerous_Presence -->

<owl:Thing rdf:about="#Detect_Dangerous_Presence">
  <rdf:type rdf:resource="&Ontology1#Composed_Event"/>
  <rdf:type rdf:resource=" #Dangerous_Presence"/>
  <rdf:type rdf:resource="&Ontology1#Detect_Event"/>
  <rdf:type rdf:resource="&Ontology1#Simple_Event"/>
  <rdf:type rdf:resource="&Ontology1#NamedIndividual"/>
  <Ontology1:Detect_Time rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">
    2012-05-13T09:00:05+01:00
  </Ontology1:Detect_Time>
  <Ontology1:is_Dangerous_Presence
rdf:datatype="http://www.w3.org/2001/XMLSchema#boolean">
   true
  </Ontology1:is_Dangerous_Presence>
 <Ontology1:Composed_From rdf:resource="&Ontology1#Detect_Long_Presence" />
<Ontology1:Composed_From rdf:resource="&Ontology1#Detect_Train"/>
</owl:Thing>

<!-- http://www.owl-ontologies.com/&Ontology1.owl#Detect_Possible_Explosive -->

<owl:Thing rdf:about="http://www.owl-
ontologies.com/&Ontology1.owl#Detect_Possible_Explosive">
<rdf:type rdf:resource="http://www.owl-ontologies.com/&Ontology1.owl#Composed_Event"/>
<rdf:type rdf:resource="http://www.owl-ontologies.com/&Ontology1.owl#Detect_Event"/>
<rdf:type rdf:resource="http://www.owl-
ontologies.com/&Ontology1.owl#Possible_Explosive"/>
<rdf:type rdf:resource="http://www.owl-ontologies.com/&Ontology1.owl#Simple_Event"/>
<rdf:type rdf:resource="http://www.w3.org/2002/07/owl#NamedIndividual"/>
<&Ontology1:Detect_Time rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">
2012-05-13T09:00:07+01:00 </&Ontology1:Detect_Time>
<&Ontology1:is_Possible_Explosive
rdf:datatype="http://www.w3.org/2001/XMLSchema#boolean"> true
</&Ontology1:is_Possible_Explosive>
<&Ontology1:Composed_From rdf:resource="http://www.owl-
ontologies.com/&Ontology1.owl#Detect_Explosive" />
<&Ontology1:Composed_From rdf:resource="http://www.owl-
ontologies.com/&Ontology1.owl#Detect_Intrusion"/>
</owl:Thing>

<!-- http://www.owl-ontologies.com/Ontology1.owl#Detect_Drop_Explosive -->

<owl:Thing rdf:about="&Ontology1#Detect_Drop_Explosive">
<rdf:type rdf:resource="&Ontology1#Composed_Event"/>
<rdf:type rdf:resource="&Ontology1#Detect_Event"/>
<rdf:type rdf:resource="&Ontology1#Drop_Explosive_Tunnel"/>
<rdf:type rdf:resource="http://www.w3.org/2002/07/owl#NamedIndividual"/>
<Ontology1:is_Drop_Explosive_Tunnel
rdf:datatype="http://www.w3.org/2001/XMLSchema#boolean">true</Ontology1:is_Drop_Explosiv
e_Tunnel><Ontology1:CanGenerate rdf:resource="&Ontology1#Allarme"/>
<Ontology1:Composed_From rdf:resource="&Ontology1#Detect_Dangerous_Presence"/>
<Ontology1:Composed_From rdf:resource="&Ontology1#Detect_Possible_Explosive"/>
</owl:Thing>
```

Listing 5 Semantic Enriched Information about Composed Event in RDF format

Listing 6 shows the activation of the alarm raised by the '*Detect_Drop_Explosive*' event. The conditions used to manage and understand the cause of the alarms may be queried offline, through a user friendly interface that exploits SPARQL language for querying the semantic

enriched data about the situation, like the alarms that have been triggered and the events detected.

In the Post Reasoner level it is possible, for an operator, to analyse the alarm activation. In our example, the operator can query the system in order to obtain all instances involved in the relation *'Alarmfrom'* that associated every alarms to the *Composed_Events* generating it. In this way, he can know all composed events which have generated the alarm (Figure 9, Listing 7).

```
    <!-- http://www.owl-ontologies.com/Ontology1.owl#Alarm -->
     <owl:Thing rdf:about="&Ontology1;Alarm">
       <rdf:type rdf:resource="&Ontology1;Alarm"/>
       <rdf:type rdf:resource="&owl;NamedIndividual"/>
       <Ontology1:message rdf:datatype="&xsd;string"></Ontology1:message>
             <Ontology1:message          rdf:datatype="&xsd;string">Attention
          Explosive Presence </Ontology1:message>
            <Ontology1335263048:Alarmfrom
rdf:resource="&Ontology1;Detect_Drop_Explosive"/>
        </owl:Thing>
```

Listing 6 Semantic Enriched Information about Raising Alarm Event

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ontology: <http://www.owl-ontologies.com/Ontology1.owl#>


SELECT ?x ?y
WHERE { ?x ontology:Alarmfrom ?y}
```

Listing 7 SPARQL query of Events raising alarms

x	y
Allarme	Detect Drop Explosive

Figure 9 SPARQL Result of Events raising Alarm

In our example the system output the result showed in Figure 10. The operator can know the simple events composing a given *Composed_Events.* Performing the query reported in the Listing 8, the system shows all instances of *Simple_Events* involved in the relation

'Composed_From' with *Composed_Events*. In this way, the operator will be aware of all events composing each composed events, Figure 10.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX Ontology: <http://www.owl-
ontologies.com/Ontology1335263048.owl#>


SELECT ?x ?y
WHERE { ?x Ontology:Composed_From ?y}
```

Listing 8 SPARQL query of Simple Events associated to a Composed Event.

x	
Detect Dangerous Presence	Detect Long Presence
Detect Drop Explosive	Detect Possible Explosive
Detect Drop Explosive	Detect Dangerous Presence
Detect Dangerous Presence	Detect Train
Detect Possible Explosive	Detect Explosive
Detect Possible Explosive	Detect Intrusion

Figure 10 SPARQL results Composed Events

Chapter 5

Measurement of logical security by nSHIELD metrics methodology

As introduced in the first chapter, the logical security, nowadays has widespread as a vital need in order to protect a critical infrastructure by not only physical attack but also by attacks to the information system. In this section we provide a methodology for measurement logical security, developed in the European project named nSHIELD (new embedded Systems arcHItecturE for multi-Layer Dependable solutions).

5.1 The New SHIELD Architectural framework

New SHIELD (nSHIELD) is a European research project co-funded by the Artemis Joint Undertaking (Subprogramme SP6) focused on the research of innovative solutions for security, privacy, dependability (SPD) in the context of embedded systems (ES) [89].
The nSHIELD project aims at addressing SPD issues as "built in" rather than as "add-on" functionalities, by adopting an innovative holistic approach. We perceive this strategy as being the first step towards SPD certification for future ES. The leading ideas at the basis of this research are:

- To enrich the state-of-the-art with new SPD solutions
- To enable the composability of these (new or already existing) solutions

This will be achieved in two steps. First, starting from current SPD solutions, the project will develop new technologies and consolidate those already explored in pSHIELD (a SHIELD pilot project) in a solid basement that will become the reference milestone for a new generation of

"SPD-ready" ES [90]. Second, these technologies will be then enhanced with the "composability" functionality that is being studied and formalized.

In a nutshell, composability is the possibility of dynamically activating one or more SPD functionalities in order to achieve a desired SPD level. This is possible with the implementation of the following enabling mechanisms and technologies:

- Semantic description of security domain and system components, in order to have a machine-understandable language to drive the automatic composition.
- SPD Metrics, in order to quantify the security needs and the achieved security level over heterogeneous environments
- Security Agent, the engine is in charge of continuously monitoring the environment to look for new components or new security needs
- Policies and control algorithms to provide a solution for the "composition problem", ie how to put together the available SPD technologies in order to achieve the security target.

nSHIELD will approach SPD at 4 different levels: node, network, middleware and overlay (see Figure 11). For each level, the state of the art in SPD of individual technologies and solutions (ranging from hardware and communication technologies to cryptography, middleware, smart SPD applications, etc.) is expected to be significantly improved and integrated into the so-called SHIELD architectural framework, which will represent the breakthrough result of the project.

The main objective of the project is to conceive and design an innovative, modular, composable, expandable and high dependable architectural framework. nSHIELD will achieve the desired SPD level in the context of integrated and interoperating heterogeneous services, applications, systems and devices, and will develop concrete solutions capable of achieving this objective in specific application scenarios with minimum engineering effort. Four scenarios have been carefully selected in order to cover a wide and significant range of expected industrial needs.

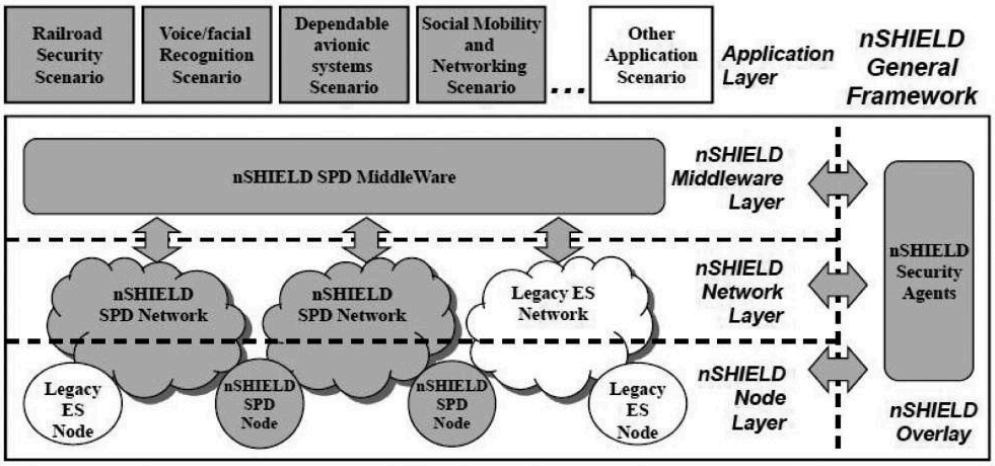


Figure 11 nSHIELD Framework

One of these scenarios addresses dependable surveillance systems for rail-based transit security, but the aim is to extend applicability also to safety-critical (the so-called "vital") subsystems in railway signalling, control and supervision.

In these contexts, the composability of the SHIELD architectural framework will have great impact on the system design costs and time to market of new products and solutions.

At the same time, the integrated use of SPD metrics in the framework will impact on the development cycles of SPD in ES because the qualification, (re-)certification and (re-) validation process of a SHIELD framework instance will be faster, easier and widely accepted.

5.2 nSHIELD Attack Surface Metric

This section introduces the detail of "nSHIELD's attack surface" and present the methodology . The approach is an integration of three different theory of scientific field:

- An attack surface metric [91].
- The Open Source Testing Methodology Manual (OSSTMM) 3 [92].
- Common Criteria Evaluation Methodology (CEM) [93].

The integration is the merging of 3 theory that assures the expression of SPD metric by a cardinal number.

An attack surface is a set of modes in which an attacker can entry in contact with a system and cause a disaster or a failure.

The work is on two different issues: the dependability and security. The former concerns non-malicious faults, the latter concerns malicious attacks or faults.

The nSHIELD proposals integrate the dependability and security concept. It consider the threat as the origin of the fault chain (fault -> errors -> failures) for the dependability and as the potential for abuse of protected assets by the system for security.
The malicious human activity or non-malicious event of an attacker are addresses to the entry and exit point of a system. The entry and exit point are characterized by 3 factors: Porosity, Controls, and Limitations [92]. The characteristics of entry and exit point defines the likelihood of being used by an attacker. The measurement is the total contribution of Porosity, Controls and Limitations.

5.3 SPD level

A threat has the capacity to subvert the security or the dependability of a system and in order to be effective, it shall interact directly or indirectly with the asset. So the aim is to separate the threat from the asset in order to avoid the interaction the total separation means SPD level=100. Otherwise the protection and increase of SPD level can be assured by Control can be applied on asset in order to lessen the impact and the interaction of a threat. The Controls are different, but they can increase the interaction so often more controls introduce more threats. So it is important to separate controls by what they do in operations.

5.3.1 Porosity

The separation between an asset and a threat exists or it does not. There are 3 logical and proactive modes to create this separation:

1. Create a physical or logical barrier between the asset and the threats
2. Change the threat to a harmless state
3. Remove the threat.

During the analysis of a system, the important is to identify the possible interaction. This parameter id called “Porosity” [92]. The porosity reduces the separation between a threat and an access. It is characterized by three elements: Complexity, Access and Trust.
Each point of interaction (Access) reduces the security and then the SPD Level. The increase of porosity is the decrease in SPD and each *pore* is a Complexity, Access, or Trust. In detail:

- **Complexity**: number of critical components for the dependability and security of the system, which failure might not be tolerated.
- **Access**: since the SPD level is the separation of a threat and an asset then the ability to interact with the asset directly is to access it. Access is the number of possible interaction with the system. Removing direct interaction with an asset will halve the number of ways it can be taken away.
- **Trust:** is a trust Access which don't menace the security of the system, it is an access but trust.

For each access pore identified, it's calculate the *damage potential-effort ratio* to have a consistent measure of the lack of separation that introduces. Not all access pores contribute equally to system's porosity measurement because not all access pores are equally likely to be used by an attacker [91]. From an attacker's point of view, however, damage potential and effort are related; if the attacker has higher privilege by using a method in an attack, then the attacker also gains the access rights of a larger set of methods. The attacker spends more effort to gain a higher privilege level that then enables the attacker to cause damage as well as gain more access rights. The ratio is similar to a cost-benefit ratio; the damage potential is the benefit to the attacker in using a resource in an attack and the effort is the cost to the attacker in using the resource.

5.3.2 Controls

Controls reduce the interaction between threat and assets. There are 2 main categories of controls in which are declared 12 kind of controls.

Interactive Controls

The Interactive Controls are directly related complexity, access, or trust interactions and they influence them. The categories are the following:

- **Authentication** is a control through the challenge of credentials based on identification and authorization.

- **Indemnification** is a control through a contract between the asset owner and the interacting party. This contract may be in the form of a visible warning as a precursor to legal action if posted rules are not followed, specific, public legislative protection, or with a third-party assurance provider in case of damages like an insurance company.
- **Resilience** is a control over all interactions to maintain the protection of assets in the event of corruption or failure.
- **Subjugation** is a control assuring that interactions occur only according to defined processes. The asset owner defines how the interaction occurs which removes the freedom of choice but also the liability of loss from the interacting party.
- **Continuity** is a control over all interactions to maintain interactivity with assets in the event of corruption or failure.

Process Controls

The Process Controls define defensive processes. These controls do not directly influence interactions rather they protect the assets once the threat is present.
The categories are the following:

- **Non-repudiation** is a control which prevents the interacting party from denying its role in any interactivity.
- **Confidentiality** is a control for assuring an asset displayed or exchanged between interacting parties cannot be known outside of those parties.
- **Privacy** is a control for assuring the means of how an asset is accessed, displayed, or exchanged between parties cannot be known outside of those parties.
- **Integrity** is a control to assure that interacting parties know when assets and processes have changed.
- **Alarm** is a control to notify that an interaction is occurring or has occurred.

The Idemnification and Authorization are related and they have not sense as single presence. So the Operational Security (OpSec), the porosity, has ten controls that a system Analyst will need to identify and define.
This two controls cannot be expressed by an operation, they are a process of identification and verification respectively. The process can be corrupted or circumvented. For example, a person authorized to enter a room can be authorized with an identification of an other

person, but this doesn't mean that the person is authorized, because this concepts cannot be transferred, this is a limitation. So Authentication control combines identification and authorization to map Access.
While controls are a positive influence in OpSec, minimizing the attack surface, they can introduces a new access point to surface so it's called they have limitations. The use of controls shall assure that they do not add new attack vectors into the target.

5.3.3 Limitations

The corruption of controls is their limitations. Therefore the state of security in regard to known flaws and restrictions within the operations scope is called Limitation. Limitations has five categories that define the type of vulnerability, mistake, misconfiguration or deficiency by operation.
The controls can be attacked by different threats, an Analyst shall know and understand the mechanism of a control failure in order to be aware of necessary level of protection of the system under test. So this is useful in order to plan a precise planning in case of disaster and contingencies.

The Limitation classifications are the following:

- **Vulnerability** is the error that can deny access to assets for authorized people or processes, allow for privileged access for unauthorized people or processes, or allow unauthorized people or processes to hide assets or themselves within the scope.
- **Weakness** is the error that abuses, or nullifies specifically the effects of the interactivity controls.
- **Concern** is the error that disrupts or reduces the effects of the process controls.
- **Exposure** is an unjustifiable action or error that provides direct or indirect complexity of targets or assets within the chosen scope channel.
- **Anomaly** is unexpected error or flaw.

To better understand how the connection with Controls, is possible to see Table 1.

Table 1 Controls and Limitations

Category		OpSec	Limitations
Operations		Complexity	Exposure
		Access	Vulnerability
		Trust	
Controls	Interactive	Authentication	Weakness
		Idemnification	
		Resilience	
		Subjugation	
		Continuity	
	Process	Non- Repudiation	Concern
		Confidentiality	
		Privacy	
		Integrity	
		Alarm	
Anomaly			

5.4 Mathematical model of SPD Level

The SPD level derives from three metrics defined in the previous sections: Operational Security, Controls and Limitations. The input information will be aggregated and associated in the input categories. The SPD level formula assigns to each categories a logarithmic base value in order to scale the three factors: Porosity, Control and limitations. In breaf:

Porosity (OpSec)

Porosity, called also Operational security defined as:

$$OpSec = Complexity + Access + Trust = Pc + Pa + Pt$$

The logarithmic base is defined as:

$$OpSec_{base} = \log^{2}(1 + 100 * OpSec)$$

Controls (OpSec)

The Control formula is defines as a Loss of control and it is :

$$OpSec = Complexity + Access + Trust = Pc + Pa + Pt$$

Thus the Loss Control sum *LCsum* is given as:

$$LCsum = LCAu + LCId + LC\,\mathrm{Re} + LCSu + LCCt + LCNr + LCCf + LC\,\mathrm{Pr} + LCIt + LCAL$$

Now, it is necessary to determine the amount of Missing Controls, *MCsum*, in order to assess the level of the Security Limitations. This shall be done for each Loss Control categories. For example, to determine the Missing Controls for Authentication (*MCAu*) we must subtract the sum of Authentication Controls (*LCAu*) of the scope from the *OpSecsum*. The Missing Controls can never be less than zero however.

The resulting Missing Control totals for each of the 10 Loss Controls must then be added to arrive at the total Missing Control value (*MCsum*).

True Controls (*TCsum*) is the inverse of Missing Controls which means the True Controls for each individual control also need to be calculated before the results can be tallied into *TCsum*. The resulting True Control totals for each of the 10 Loss Controls must then be added to arrive at the total True Control value (*TCsum*).

True Controls are used to measure the ideal placement of controls. The base value also helps to eliminate the influence of a disproportionate placement of controls on security. The True Controls base (TCbase) value is given as:

$$TCbase = log_2(1 + 100 \times (OpSecsum - MCsum \times 0.1))$$

True Coverage (TCvg) can be used to measure the percentage of controls in place regarding the optimal amount and placement of controls. True Coverage is then derived using the Missing Control totals.

Full Controls, on the other hand, take into account all controls in place regardless of a balanced distribution. This value is important for measuring the worth of two-factor authentication, for example, and other instances of defense in depth for the same complexity, access or trust. The Full Controls base (*FCbase*) value is given as:

$$FCbase = log^2(1 + 10 \times LCsum)$$

The Limitations Formula

The Limitations are individually weighted. The weighting of the Vulnerabilities, Weaknesses and Concerns are based on a relationship between the Porosity or *OpSecsum*, the Loss Controls and in the case of Exposures and Anomaly the existence of other Limitations also plays a role. An Exposure or Anomaly poses no problems alone unless a Vulnerability, Weakness or Concern is also present. Think of an Exposure like a pointer. If there is a pointer that goes nowhere, or in this case doesn't lead to anything exploitable (Vulnerability, Weakness, Concern) and all Controls are accounted for, then at the time of the test the Exposure has no effect on security and thus has no value in the SPD level.

Table 2 is used to calculate the *SecLimsum* variable, as an intermediate step between the Security Limitation inputs and the *SecLimbase* variable.

Table 2 SecLimsum variable calculating

Input	Weighted Value	Variables
Vulnerability L_V	$\frac{OpSec_{sum} + MC_{sum}}{OpSec_{sum}}$	MC_{sum}: sum of Missing Controls
Weakness L_W	$\frac{OpSec_{sum} + MC_A}{OpSec_{sum}}$	MC_A: sum of Missing Controls Class A
Concern L_C	$\frac{OpSec_{sum} + MC_B}{OpSec_{sum}}$	MC_B: sum of Missing Controls Class B
Exposure L_E	$\frac{((P_C + P_A) \times MCvg + L_V + L_W + L_C)}{OpSec_{sum}}$	P_C : sum of Complexity P_A : sum of Accesses MCvg: Percent Missing Coverage
Anomaly L_A	$\frac{(P_C \times MCvg + L_V + L_W + L_C)}{OpSec_{sum}}$	P_C: sum of Complexity MCvg:Percent Missing Coverage

SecLimsum is then calculated as the aggregated total of each input multiplied by its corresponding weighted value as defined in the table 2.

SPD Level Formula

The Actual SPD level Delta is useful for comparing products and solutions:

$$ActSPDL\Delta = FCbase - OpSecbase - SecLimbase$$

To measure the current state of operations with applied controls and discovered limitations, a final calculation is required to define Actual SPD level.

$$ActSPDL = 100 + ActSPDL\Delta - (1\backslash 100) \times (OpSecbase \times FCbase - OpSecbase \times SecLimbase + FCbase \times SecLimbase)$$

5.5 The nShield attack surface metrics ontology

As previous described, the nSHIELD framework trough a specific middleware is able to elaborate and define in machine understandable manner the calculate metrics. This is trough and ontology description of metrics information coming by appropriate device installed in the monitoring application who have the responsibility to study and control the SPD level of the system.

The system is modelled as a surface (definition by metrics methodology). Surface has amount of interfaces to the external world (**access**), interactions between components (**complexity**) and internal/external interactions with no direct impact on security (**trust**). These three concepts are represented by a number. In an ontology model these attributes can be represented in Figure 12.

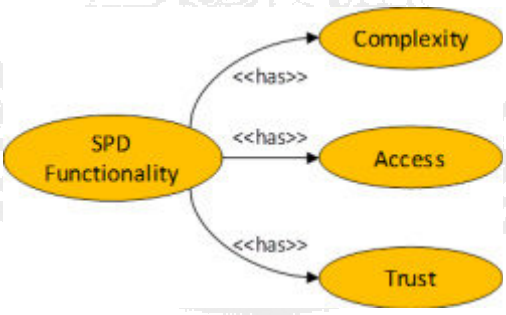


Figure 12 Porosity Ontology

Controls counteracts vulnerability (identified by the number of "accesses") The controls can be classified in interactive (Class A) and process (Class B) controls (Figure 13).

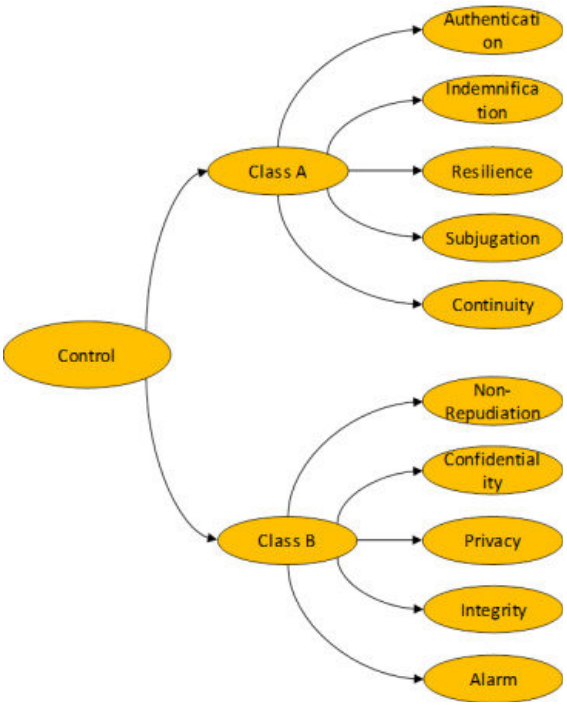


Figure 13 Control Ontology

Each control can be affected by a set of limitations, Figure 14

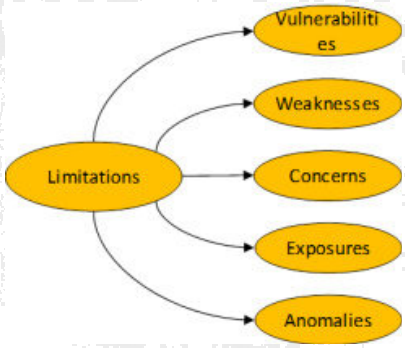


Figure 14 Limitation Ontology

These attribute, as described in the previous section, are composed by a proper algebra, to obtain the metric value for the whole system. Then a proper domain data base is needed to tailor the result to the specific application scenario.

The Figure 15 represents better the relation between different values and attributes, by an E-R diagrams.

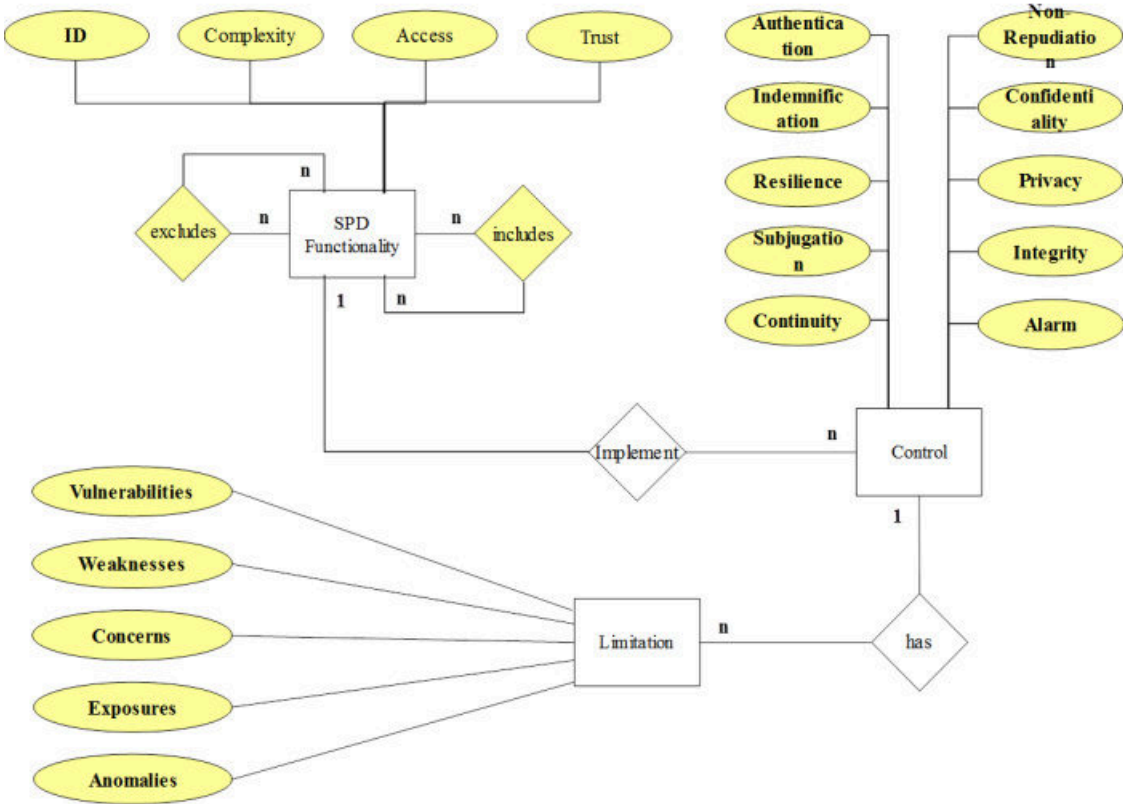


Figure 15 SPD attribute integration

Chapter 6

Case study

In this section will be presented some application of the methodology shown in the previous section both physical security and logical. Furthermore, the section will present a particular case study in which the Post Reasoner analysis it will be essential for the configuration assessment.

6.1 Physical Security a WSN Application: Post Reasoner application

In this section we will present a experimentation results conducted with real data gathered from a Wireless Sensor Network (WSN). The real date was taken from the Intel Lab Set of Berkeley [105]. The dataset contains data collected from 54 sensors deployed in the Intel Berkeley Research Lab between February 28th and April 5th, 2004. The nodes are Mica2Dot sensors with weather boards collected topology information, along with humidity, temperature, light and voltage values with timestamp information for each measurement, once every 31 seconds . The format of the dataset is: date, time, epoch, mote ID, temperature, humidity, light, and voltage.

The sensor ids range from 1-54. Data from some sensor motes may be missing or truncated. Temperature is measured in degrees Celsius. Humidity is temperature corrected relative humidity, ranging from 0-100%. Light is in Lux (a value of 1 Lux corresponds to moonlight, 400 Lux to a bright office, and 100,000 Lux to full sunlight.) Voltage is expressed in volts, ranging from 2-3.

The data was in CVS format, in order to be compliant with our classifier system, we have translate the data in XML format, according to a specific data model represented in Listing 9.

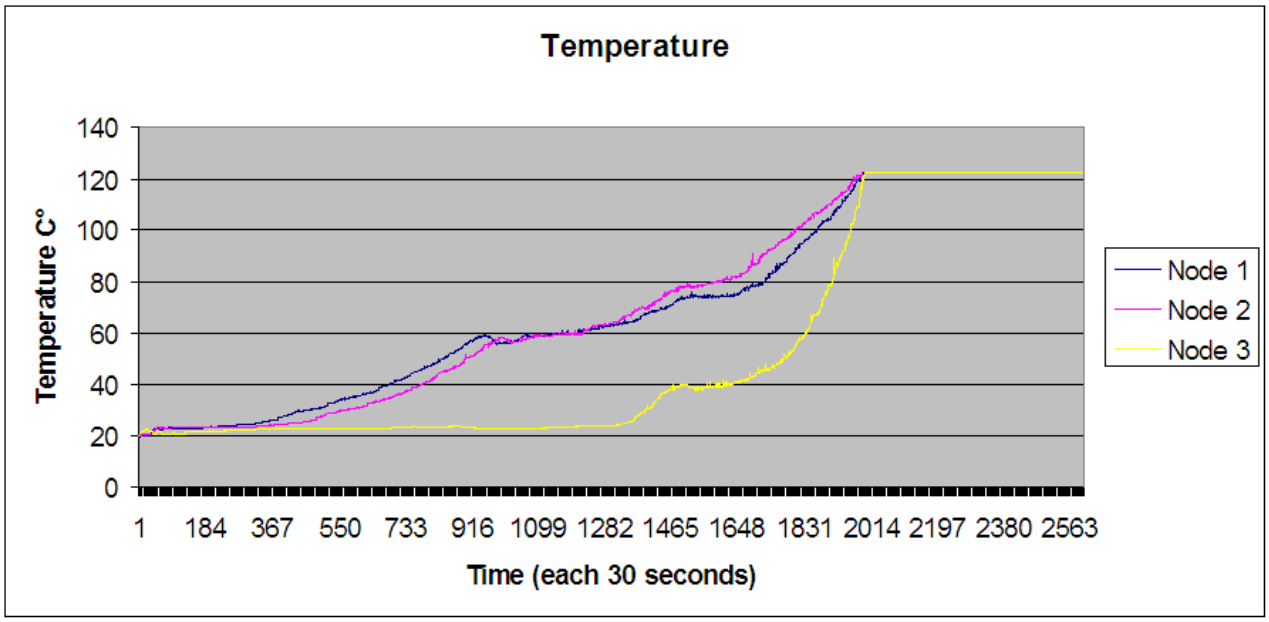
```
<?xml version="1.0" encoding="UTF-8"?>
<!--Sample XML file generated by XMLSpy v2008 (http://www.altova.com)-->

<Sensor1 xmlns:xsi=http://www.w3.org/2001/XMLSchema-instance

       name="XXX"
       NET="XXX"
       ID="XXX"
       Measure_type="XXX"
      Measure_type="XXX"
      position="XXX">

   <Humidity
      name="XXX"
      values="XXX"
      Time="timestamp"
   />


   <Temperature
      name="XXX"
      values="XXX"
      Time="timestamp"
   />
<Voltage…./>
.........
```

Listing 9 XML example of data model

In order to create a rule for a real case study, it is chosen a range of temperature and humidity, which, when they was exceeded an alarm was raised. The case study is a monitoring of technical room, such as a server room. The critical value in a server room are: temperature over 24° and Humidity over 55%. When this values exceed the environment condition are not compliant for the server installed in a room. And when the temperature reach the values of 40° and Humidity about 30%, this values reveal a critical situation.

The experimentation have noticed that in normal operation the sensor reveal a value of temperature between (19° - 23,9°) and Humidity between (50%-55%). But after a more days of observation the temperature increase in uncontrolled manner and an critical alarm was raising (according to the configured rules). After the increment temperature sensors get stuck at value (122◦C). In a Post Reasoner Analysis with query SPARQL and other methodology of data analysis form the database, it is notice that the stuck at was correlated to the low battery. The batteries in this case were lithium ion cells, which keep almost a constant voltage over their duration; note that variations in voltage are highly correlated with temperature (Figure 16 and Figure 17). Analysing the medium value of the voltage it is notice that the battery can

be considered low when it is under 2,25 volts. In this manner is possible to update and re-configure the detection rules in order to deny next false alarms.



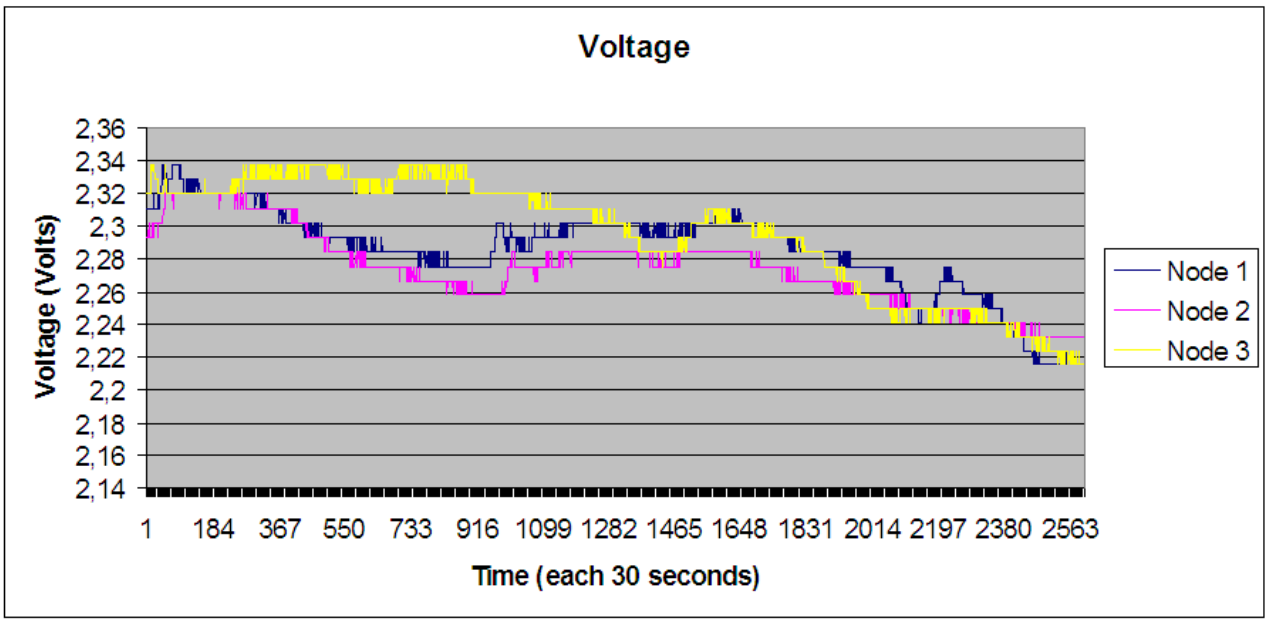


Figure 16 Temperature, Voltage results

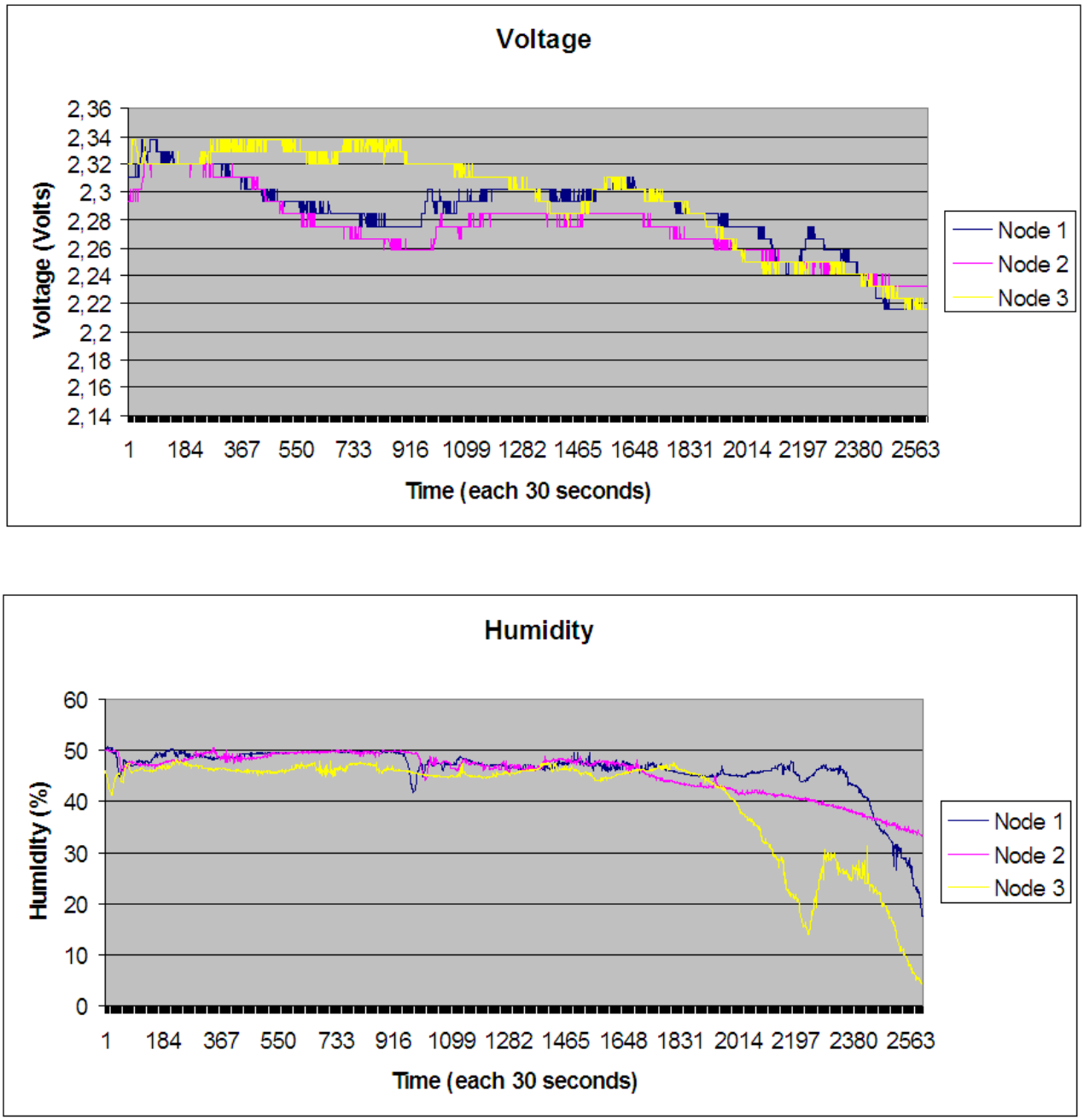


Figure 17 Voltage, humidity results

6.2 Joint estimation of physical and logical security

Approaches analysed, for the joint estimation, are different and they can be go from event, correlation to the integration of SPD metrics for logical security. In this section we will describe and explain with simple example the two approaches.

6.2.1 Integration as event correlation

In this kind of approach to the convergence and evaluation of physical and logical security, the estimation and detection is made by event correlation.
In this case we have needed the use of physical sensors and logical sensors with specific function for logical security detection. For this kind of scenario, it is possible refer to ontology presented in the section 4, naturally to be specialized for the domain.

The scenario consist in a detection of logical security attack. A typical scenario can be the following:

1. A person enters in a room server. The servers manage the video surveillance system.
2. During the presence of this people, it is detected an high number of failed access to the server
3. After a successful log in, the Intrusion detection system for the network reveals some malicious action and alert the operator.
4. In the same time there is a video loss signal from some cameras.
5. The attacker has made an information attack to the video surveillance system.

The sensors are:

- S1. Card Reader (Physical Access)
- S2. File Log Server (logic security sensor)
- S3. Intrusion detection on network
- S4. cameras

The event detected by sensors are:

- E1. Violated Physical Access of X by S1
- E2. Anomaly number of failed Log in with a UserName “pippo” by S2
- E3. successful Log in of “pippo” by S2
- E4. activity of intrusion detection on network
- E5. Video loss by S4
- C1. “Possible attack to server ” (E1,E2)
- C2. “Possible Information Attack” (E3,E4)

The event “Attack to network” can be detected by two different modes.:

1. If C1 and C2
2. If C2 and E5

This kind of approach gives a clear situation awareness, detection of attack in time.
On the contrary, it requires elaboration of server, specific sensor and countermeasure for detection of security logic attack.

6.2.2 Integration with SPD metrics elaboration

This approach tries to detect a possible situation of a malicious attack to the information system starting from event of physical security and the elaboration of primitive data of sensor.
The scenario consist in a monitoring of Access to a server room of a “Data Center”. In this scenario the objective is to detect an attack to logical security, in other words an attack to the server and its information system. A typical scenario can be the following:

1. An user X enters in the server room with its card reader. It could be for the maintenance service or a simple employer, he has specific privileges to enter in the server room.
1. after that it is registered a log in on server with credential of user Y (which does not have done physical access in the room).
2. X is different from Y, so it indicates a possible attack: log in not authorized by an user.

In the room can be present other persons or people.

We take in care as value the SPD metrics. As described in section 5, the SPD metrics derive from the Attack surface metrics and they take account of different kind of variables, it even an aggregated information.

Sensors are:

- S1. Physical Log (Card Reader) this sensor gives information about: User, Entry, Exit, privileges
- S2. Logic Log (File Log Server)

- E1. Physical Access X, R\W, by S1
- E2. Physical Access Y, R\W by S1
- E3. Physical Access Z, -\-. By S1
- E4. Logic Access X by S2
- E5. Logic Access Y by S2
- E6. Logic Access T by S2
- E7. Not Physical Access T by S1 (it is translated in the semantic classifier with a simple verification of the attribute of user T (Entry=false).

The composed event are:

- C1. "SPD Medium" it is verified E3
- C2. "Possible malicious logic Access" (E6, E7)

The alarm "Malicious log in by an attacker" it is detected if:

- if C1 and C2 -> "SPD Low"

This kind of approach don't request elaboration on server but only the analysis of log file, the log file of server is considered as a logic sensor. There is non need of addiction devices for the detection of security logic intrusion, the inference engine trough the rules elaborate the information and gives the alert. The effort is in configuration phase. In Figure 18 it is represented an ontology model for the scenario.
The additional information of SPD metrics it is important and gives to the operator an awareness of the situation. The escalation of SPD level give an early warning, this is most

important in particular for the information protection. It elaborates a complex information in a simples syntactic information with an important semantic.

The fusion and integration of this two different information, logical and physical security, given by different kind of sensors it was possible trough semantic application model. The SPD metrics has already their semantic model in order to be machine-understandable.

The dynamic nature of semantic model, gives the possibility of fusion of information that can be appear uncorrelated.

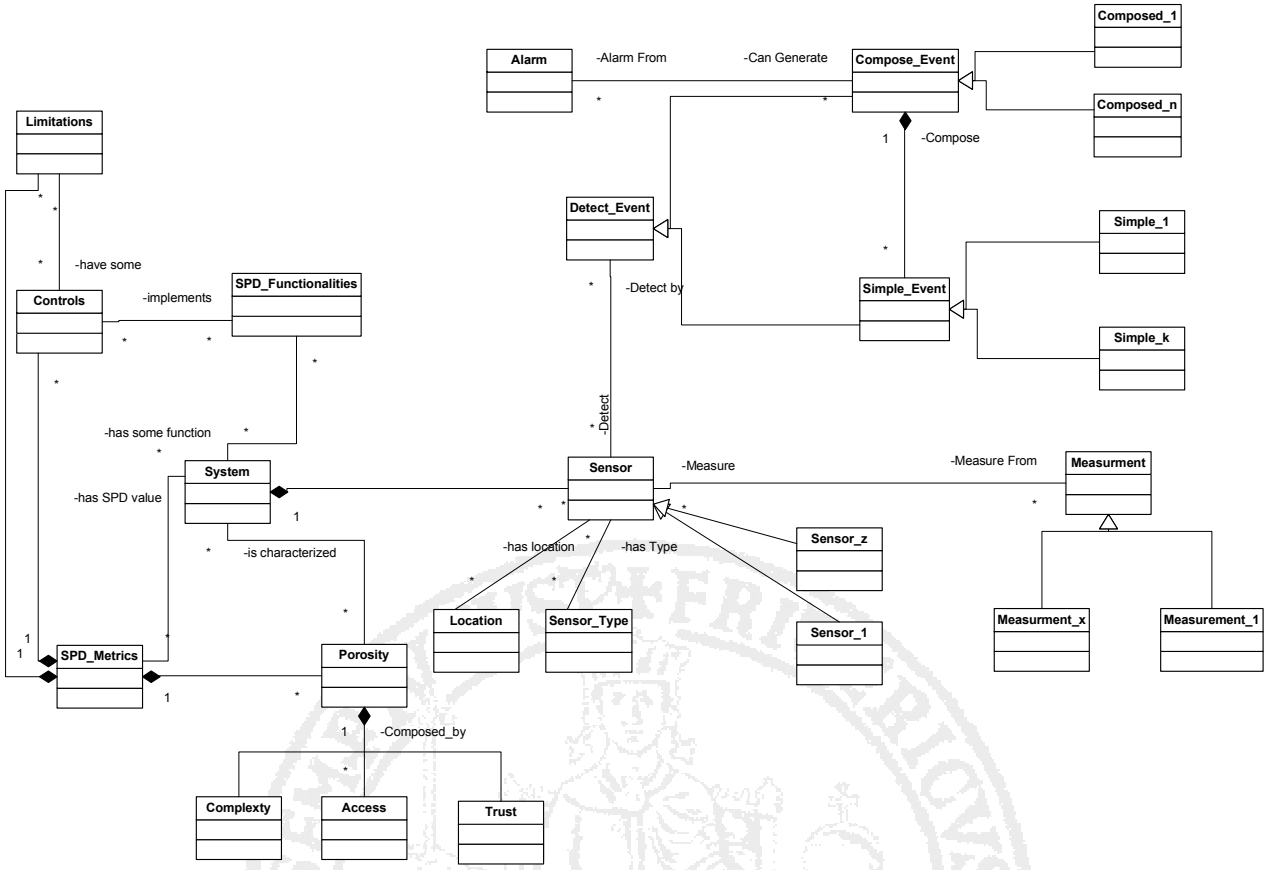


Figure 18 Physical and Logical Security Ontology

Conclusions

The main objective of this thesis is the proposal of an approach for physical and logical security convergence. This is a problem that had a diffusion in the last ten years due to the development of new technologies.

The work starting from the general problem of information integration for the critical infrastructure protection. It is been described the problem both from a physical security perspective and logical security point of view. Main issues on this topic are been introduced. The works has given a whole panoramic view on the state of art of:

- Semantic model: solution used in this thesis for the modelling of information integration ad management for different source. This solution was adopted even for the possibility to apply reasoning and produces inferences on the information source and in order to ensure a correct and shared information interpretation into events and situation assessment before raising an alarm.
- Complex Event processing: the theory of CEP is used to model the detection model in the ontology, used to do inferences.
- Decision support system: is the base of whole development systems. As matter of fact the system it is based on DSS with semantic and ontological processing. The semantic enrichment process is automatically performed to build a knowledge base, which will be inferred on-line by a light smart classifier that will raise an alarm in case of risk detection. The decision approach will be based on two steps: (1) a smart in-line classifier based on the semantic model to raise an alarm, in case of threat event detection, (2) a post reasoner offline inference engine, in order to further comprehend the event and its causes.

The thesis has proposed and implemented a methodology model simple data and atomic events coming from different sensors, it allows to model the complexity of the sensors and the

correlation among different events to define composed events to improve the knowledge about the system and locate critical scenarios. The proposed monitoring system models the detection system as a two steps process

- a in-line reaction by means of a fast classifier
- an offline activity through a semantic post reasoner.

The former aiming at providing proper alarms when dangerous events occur, the latter aiming at providing a complete and detailed picture of the situation, useful for operators both in understanding the situation and for decision supporting.
It is been presented an experimentation of real data from WSN in which is possible to observe the effectiveness of the inline smart classifier and the work of semantic post reasoner.

After it is deled with the question of physical and security convergence. From the results obtained in the nSHIELD European project it is adopted the methodology of Attack surface metrics in order to estimate the logical security of a system under observation. This issued it is treated in two different modes:

- for a pure event correlation
- from SPD metrics solutions.

The latter was a most suitable solution for our objective, because it declines the use of specific sensor for the logical security event detection and the adoption of metrics gives a detailed modelling on the situation and take in care different variable that influence the logical security of the system.

References

[1] Official website of Department of Homeland Security: https://www.dhs.gov.com

[2] U.S. DoD. The Department of Defense Critical Infrastructure Protection (CIP) Plan: *A Plan in Response to Presidential Decision Directive 63 'Critical Infrastructure Protection'*, 1998. URL: http://www.fas.org/irp/offdocs/pdd/DOD-CIP-Plan.htm

[3] White paper Carney J. *"Why Integrate Physical and Logical Security?"* https://www.cisco.com/web/strategy/docs/gov/pl-security.pdf"

[4] Crowell, W. P., Contos, B. T., DeRodeff, C., & Dunkel, D. (2011). *Physical and Logical Security Convergence: Powered By Enterprise Security Management: Powered By Enterprise Security Management.* Syngress.

[5] Yahya Mehdizadeh, *"Convergence of Physichal and Logical Security* "http://www.sans.org/reading-room/whitepapers/authentication/convergence-logical physical-security-1308

[6] Bocchetti, G., Flammini, F., Pragliola, C., & Pappalardo, A. (2009, August). Dependable integrated surveillance systems for the physical security of metro railways. *In Distributed Smart Cameras, 2009. ICDSC 2009. Third ACM/IEEE International Conference on (pp. 1-7). IEEE.*

[7] U.S. Department of Transportation, *Transit Security Design Considerations*. *Federal Transit Administration, Final Report, 2004.*

[8] BUCHMANN, Alejandro; KOLDEHOFE, Boris. *Complex event processing*. I Information Technology, 2009, 51.5: 241-242.

[9] Müller, A. Event Correlation Engine. Computer engineering and networks laboratory, *TIK-Institut fürTechnische Informatik und Kommunikationsnetze, 2009.*

[10] Tiffany, Michael. "A survey of event correlation techniques and related topics." *Research paper, Georgia Institute of Technology (2002).*

[11] Hopcroft, J. E., Motwani, R., and Ullman, J. D. *Introduction to automata theory, languages, and computation, vol. 3. Addison-wesley Reading, MA, 1979*

[12] Zeigler, Bernard P., Herbert Praehofer, and Tag Gon Kim. Theory of modeling and simulation: integrating discrete event and continuous complex dynamic systems. *Academic press, 2000.*

[13] Sergio Zamarripa Lóp ez, María del Carmen Calle Villanueva, Emitzá Guzmán, Tobias Röhm, Benoit Gaudin, Newres Al Haide "D4.1: State-of-the-art of event correlation and event processing". (2010).

[14] Kim, H. C., Pang, S., Je, H. M., Kim, D., & Yang Bang, S. (2003). Constructing support vector machine ensemble. Pattern recognition, 36(12), 2757-2767.

[15] Kruegel, C., Mutz, D., Robertson, W., & Valeur, F. (2003, December). Bayesian event classification for intrusion detection. *In Computer Security Applications Conference, 2003. Proceedings. 19th Annual (pp. 14-23). IEEE.*

[16] Debar, H., Becker, M., & Siboni, D. (1992, May). A neural network component for an intrusion detection system. *In Research in Security and Privacy, 1992. Proceedings., 1992 IEEE Computer Society Symposium on (pp. 240-250). IEEE*.

[17] Ntalampiras, S., Audio Surveillance. *In Critical Infrastructure Security: Assessment, Prevention, Detection, Response, WIT Press, 2012: pp. 191-205.*

[18] Ntalampiras, S., Potamitis, I., Fakotakis, N., An Adaptive Framework for Acoustic Monitoring of Potential Hazards. *In EURASIP J. Audio, Speech and Music Processing, 2009.*

[19] Ntalampiras, S., Potamitis, I., Fakotakis, N., On acoustic surveillance of hazardous situations*. In Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP'09), Taiwan, Taipei, 19-24 April 2009.*

[20] Casola, V.; Gaglione, A. & Mazzeo A. 2009. A Reference Architecture for Sensor Networks Integration and Management. *Proc. of the 3rd International Conference on Geosensor Networks (GSN 2009), 2009*

[21] Flammini12, F., Gaglione, A., Mazzocca, N., Moscato, V., & Pragliola, C. (2009). *On-line integration and reasoning of multi-sensor data to enhance infrastructure surveillance.*

[22] Hong, K. S., Chi, Y. P., Chao, L. R., & Tang, J. H. (2003). An integrated system theory of information security management. *Information Management & Computer Security, 11(5), 243-248.*

[23] Roadnight, J., Will Physical Security Information Management (PSIM) Systems change the Global Security World?. *CornerStone GRG Ltd Whitepaper, February 2011*.

[24] Crowell, W. P., Contos, B. T., DeRodeff, C., & Dunkel, D. (2011). Physical and Logical Security Convergence: *Powered By Enterprise Security Management: Powered By Enterprise Security Management. Syngress.*

[25] Eugene Schultz, E. "Risks due to convergence of physical security systems and information technology environments." *Information Security Technical Report 12.2 (2007): 80-84.*

[26] National Research Council. Making the nation safer: the role of science and technology in countering terrorism. *Washington, DC: National Academies Press; 2002.*

[27] Mehdizadeh Yahya. Convergence of logical and physical security. Available from: <www.sans.org/reading_room/whitepapers/authentication/1308.php>; 2003

[28] Andrew P. Sage. Decision Support Systems Engineering. *John Wiley & Sons, Inc., New York, 1991.*

[29] GML 3.1 specification. Open Geospatial Consortium, Inc. http://portal.opengeospatial.org/

[30] A. Sheth and C. Henson, S. Sahoo, Semantic Sensor Web, *IEEE Internet Computing, vol. 12, no. 4, 2008, pp. 78-83*

[31] A. Sheth and M. Perry, Traveling the Semantic Web through Space, Time, and Theme, *IEEE Internet Computing, vol. 12, no. 2, 2008, pp. 81 86.*

[32] M. G. Ceruti, Ontology for Level-One Sensor Fusion and Knowledge Discovery. *Knowledge Discovery and Ontologies Workshop at ECML PKDD, Pisa, Italy, September 2004*

[33] M. Eid, R. Liscano, A. El Saddik. A Novel Ontology for Sensor Networks Data. CIMSA 2006 *IEEE International Conference on Computational Intelligence for Measurement Systems and Applications La Coruna - Spain, 12-14 July 2006*

[34] W3C (1998). Extensible markup language (xml) 1.0. W3C Recommendation. [W3C, 1999] W3C (1999*). Resource description framework (rdf) schema specification. W3C Proposed Recommendation.*

[35] Wache, H., Scholz, T., Stieghahn, H., and K¨onig-Ries, B. (1999). An integration method for the specification of rule–oriented mediators. *In Kambayashi, Y. and Takakura, H., editors, Proceedings of the International Symposium on Database Applications in Non-Traditional Environments (DANTE'99), pages 109–112, Kyoto, Japan.*

[36] Battle, S. (2004, November). Round-tripping between XML and RDF. *In International Semantic Web Conference (ISWC).*

[37] Stuckenschmidt, H., & Van Harmelen, F. (2001, October). Ontology-based metadata generation from semi-structured information. *In Proceedings of the 1st international conference on Knowledge capture (pp. 163-170). ACM.*

[38] Visser, U., Stuckenschmidt, H., Wache, H., & Vögele, T. (2000). Enabling technologies for interoperability. *In Workshop on the 14th International Symposium of Computer Science for Environmental Protection (pp. 35-46). Bonn, Germany.*

[39] Patig, S. (2001). Environmental information systems in industry and public administration. C. Rautenstrauch (Ed.). *IGI Global.*

[40] Stuckenschmidt, H., & van Harmelen, F. (2005). Ontology-based information sharing. *Information Sharing on the Semantic Web, 25-44.*

[41] Gruber, T. (1993). A translation approach to portable ontology specifications. Knowledge Acquisition, 5(2).

[42] Lenat, D. (1998). The dimensions of context space. *Available on the web-site of the Cycorp Corporation. (http://www.cyc.com/publications).*

[43] Jasper, R. and Uschold, M. (1999). A framework for understanding and classifying ontoogy applications. *In Proceedings of the 12th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop. University of Calgary/Stanford University.*

[44] van Harmelen, F. and Fensel, D. (1999). Practical knowledge representation for the web. *In Fensel, D., editor, Proceedings of the IJCAI'99 Workshop on Intelligent Information Integration.*

[45] Arens, Y., Hsu, C.-N., and Knoblock, C. A. (1996). Query processing in the sims information mediator. *In Advanced Planning Technology, California, USA. AAAI Press.*

[46] Mitra, P., Wiederhold, G., and Kersten, M. (2000). A graph-oriented model for articulation of ontology interdependencies. *In Proc. Extending DataBase Technologies, EDBT 2000, volume Lecture Notes on Computer Science, Konstanz, Germany. Springer Verlag.*

[47] Mitra, P., Wiederhold, G., and Jannink, J. (1999). Semi-automatic integration of knowledge sources. *In Fusion '99, Sunnyvale CA.*

[48] Wache, H., Scholz, T., Stieghahn, H., and K¨onig-Ries, B. (1999). An integration method for the specification of rule–oriented mediators. *In Kambayashi, Y. and Takakura, H., editors, Proceedings of the International Symposium on Database Applications in Non-Traditional Environments (DANTE'99), pages 109–112, Kyoto, Japan.*

[49]] Berners-Lee T., Hendler J., Lassila O.: "The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities". *Scientific American, 284 (5), pp. 34-43, May 2001. http://www.sciam.com/*

[50] Semantic Technology. TopQuadrant Technology Briefing, March 2004, http://www.topquadrant.com/tq_white_papers.htm

[51] Corcho, O., Fernandez-Lopez, M., Gomez-Perez, A., "Methodologies, tools and languages for building ontologies. Where is the meeting point?" *Data&Knowledge Engineering 46, pp. 41-64, 2003.*

[52] Genesereth, M.R., Fikes, R.E., "Knowledge Interchange Format. Version 3.0. *Reference Manual." Stanford University, 1992.*

[53] MacGregor, R., Bates, R., "The Loom Knowledge Representation Language". *Technical Report ISIRS- 87-188, USC Information Sciences Institute, Marina del Rey, CA, 1987.*

[54] Not used.

[55] McGuinness, D., van Harmelen, F., "OWL Web Ontology Language Overview". *W3C Recommendation, 2004.*

[56] Kifer, M., Lausen, G., F-Logic: "A Higher-Order language for Reasoning about Objects, Inheritance, and Scheme", *in Proc. of SIGMOD Conference 1989, pp. 134-146, 1989.*

[57] Fensel, D., Decker, S., Erdmann, M., Studer, R., Ontobroker: "The Very High Idea*", in Proceedings of the 11th International Flairs Conference (FLAIRS-98), Florida, 1998.*

[58] Knutilla, A., et. al., "Process Specification Language: Analysis of Existing Representations", *NISTIR 6133, National Institute of Standards and Technology, Gaithersburg, MD, 1998.*

[59] Gruber, T.R., A Translation Approach to Portable Ontology Specifications, in Knowledge Acquisition, 5(2), 1993.

[60] Motta E., Reusable Components for Knowledge Modelling. *IOS Press, Amsterdam, Netherlands, 1999*

[61] Chaudhri, V., Farquhar, A., Fikes, R., Karp, P., Rice, J. (1998), OKBC: A programming foundation for knowledge base interoperability, *in Proceedings of AAAI'98, pp. 600-607, 1998.*

[62] Karp, R., Chaudhri, V., Thomere, J., XOL: An XML-Based Ontology Exchange Language, Technical report, 1999.

[63] Sowa, J.F., Semantic networks, available online http://www.jfsowa.com/pubs/semnet.htm, 2002.

[64] Sowa, J.F., Knowledge Representation: Logical, Philosophical, and Computational Foundations, *Brooks Cole Publishing Co., Pacific Grove, CA, 2000.*

[65] R.E. Kent., Conceptual Knowledge Markup Language: The Central Core', *in: Twelfth Workshop on Knowledge Acquisition, Modeling and Management, 1999.*

[66] TopicMaps.Org Authoring Group, XML Topic Maps (XTM) 1.0 *TopicMaps.Org specification, 2001*

[67] Yergeau, F., Bray, T., Paoli, J., Sperberg-McQueen, J.M., Maler, E., Extensible Markup Language (XML) 1.0 (Third Edition). *W3C Recommendation, 2004.*

[68] Manola, F., Miller, E., McBride, B., RDF Primer. *W3C Recommendation, 2004.*

[69] Fensel, D., Horrocks, I., van Harmelen, F., McGuiness, D., Patel-Schneider, P.F., OIL: Ontology Infrastructure to Enable the Semantic Web*. IEEE Intelligent Systems, 16 (2), 2001.*

[70] Stein, L.A., Connolly, D., McGuinness, D.,Annotated DAML Ontology Markup. *Initial draft, http://www.daml.org/2000/10/daml-walkthru, 2000.*

[71] Connolly, D., van Harmelen, F., Horrocks, I., McGuinness, D., Patel-Schneider, P.F., Stein, L.A., DAML+OIL (March 2001) Reference Description. W3C Note, 2001.

[73] Martin, D., et al., Bringing Semantics to Web Services: The OWL-S Approach, in Proceedings of the First International Workshop on Semantic Web Services and Web Process Composition (SWSWPC 2004), July 6-9, 2004, San Diego, California, USA.

[74] Berners-Lee T., Hendler J., Lassila O.: "The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities". *Scientific American, 284 (5), pp. 34-43, May 2001. http://www.sciam.com/*

[75] Semantic Technology. TopQuadrant Technology Briefing, March 2004, http://www.topquadrant.com/tq_white_papers.htm

[76] Casola, V., Esposito, M., Mazzocca, N. and Flammini, F. (2012) 'Freight train monitoring: a case-study for the pSHIELD project', *in Proceedings of the 2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS '12).*

[77] Amato, F., Casola, V., Gaglione, A. and Mazzeo, A. (2011) 'A semantic enriched data model for sensor network interoperability', *Journal of Simulation Modelling Practice and Theory, Vol. 19, No. 8, pp.1745–1757, Elsevier.*

[78] Amato, F., Casola, V., Gaglione, A. and Mazzeo, A. (2010a) 'A common data model for sensor network integration', *4th International Conference on Complex, Intelligent and Software Intensive Systems, CISIS-2010, IEEE Publishing, Krakow, 15–18 February 2010.*

[79] Amato, F., Casola, V., Mazzeo, A. and Romano, S. (2010b) 'A semantic based methodology to classify and protect sensitive data in medical records', *in 2010 Sixth International Conference on Information Assurance and Security (IAS), IEEE, pp.240–246.*

[80] Horrocks, I., Patel-Schneider, P.F., Boley, H., Tabet, S., Grosof, B. and Dean, M. (2004) 'SWRL: a semantic web rule language combining OWL and RuleML', *W3C Member Submission, World Wide Web Consortium, May.*

[81] Liu, B., Ma, Y. and Wong, C.K. (2000) 'Improving an association rule based classifier', *in Principles of Data Mining and Knowledge Discovery, Vol. 1910, pp.504–509, Springer Berlin Heidelberg.*

[82] Wittig, R.D. and Chow, P. (1996) 'OneChip: an FPGA processor with reconfigurable logic', *Conference Proceedings of IEEE Symposium on FPGAs for Custom Computing Machines, pp.126–135.*

[83] Pérez, J., Arenas, M. and Gutierrez, C. (2006) 'Semantics and complexity of SPARQL', The Semantic Web-ISWC, *Springer Berlin Heidelberg, pp.30–43.*

[84] Sirin, E., Parsia, B., Cuenca Grau, B., Kalyanpur, A. and Katz, Y. (2007) 'Pellet: a practical OWL-DL reasoner', *Web Semantics: Science, Services and Agents on the World Wide Web, Vol. 5, No. 2, pp.51–53*.

[85] Broekstra, J., Kampman, A. and van Harmelen, F. (2002) 'Sesame: a generic architecture for storing and querying RDF and RDF schema'*, in ISWC.*

[86] Prud'hommeaux, E. and Seaborne, A. (2008) 'SPARQL query language for RDF', *W3C Recommendation, 15.*

[87] Kaplanski, P. (2010) 'Description logic based generator of data centric applications', *Conference Proceedings of 2nd International Conference on Information Technology (ICIT), IEEE Publishing, pp.53–56.*

[88] Flammini, Francesco, et al. "Augmenting Surveillance System Capabilities by Exploiting Event Correlation and Distributed Attack Detection." *Availability, Reliability and Security for Business, Enterprise and Health Information Systems* (2011): 191.

[89] Esposito, M., Fiaschetti, A., & Flammini, F. (2013). The New SHIELd Architectural framework. *Mobile Computing*, 53.

[90] Casola, Valentina, et al. "Freight train monitoring: a case-study for the pSHIELD project." *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2012 Sixth International Conference on*. IEEE, 2012.

[91] An Attack Surface Metric - Pratyusa K. Manadhata, Member, IEEE, and Jeannette M. Wing, *- IEEE Transactions on Software Engineering, 2010*

[92] OSSTMM 3 The Open Source security Methodology Manual – *Contemporary security Testing and Analysis – created by Pete Herzog – Developed by ISECOM – 2010*

[93] *Common Criteria – Common Methodology for Information Technology Security Evaluation – Evaluation methodology, September 2012, Version 3.1, Revision 4.*

[94] Brachman, Ronald J. "What's in a concept: structural foundations for semantic networks." *International Journal of Man-Machine Studies 9.2 (1977): 127-152.*

[95] Brickley, Dan, and Ramanathan V. Guha. "Resource Description Framework (RDF) Schema Specification 1.0: *W3C Candidate Recommendation 27 March 2000." (2000).*

[96] Borgida, Alexander, and Peter F. Patel-Schneider. "A semantics and complete algorithm for subsumption *in the CLASSIC description logic." arXiv preprint cs/9406101 (1994).*

[97] Rumbaugh, James, Ivar Jacobson, and Grady Booch. Unified Modeling Language Reference Manual, *The. Pearson Higher Education, 2004.*

[98] Deokar A.V. and El-Gayar O.F.(2013), 'On semantic annotation of decision models', *Inf. Syst. E-bus. Manag. 11, 1. pgg 93-117.*

[99] Jarupadung S. (2012), 'Distributed Event Detection and Semantic Event Processing', *In The 6th ACM International Conference on Distributed Event-Based Systems (DEBS 2012) (Doctoral Symposium), July 16-20, 2012, Freie Universitaet Berlin, Berlin, Germany*

[100] Huang V. and Javed M., (2008), 'Semantic sensor information description and processing', *In 2nd International Conference on Sensor Technologies and Applications.*

[101] Gomez, L., and Laube, A. (2009), 'Ontological Middleware for Dynamic Wireless Sensor Data Processing', *In Proceedings of the 2009 Third International Conference on Sensor Technologies and Applications IEEE Computer Society, Washington, DC, USA, pp. 145-151.*

[102] Amato, F., Casola, V., Gaglione, A., Mazzeo, A. (2010). 'A common data model for sensor network integration ', *4th International Conference on Complex, Intelligent and Software Intensive Systems, CISIS-2010;Krakow;15-18 February 2010; IEEE Publishing.*

[103] Konstantinou N., Solidakis E., Zoi S., Zafeiropoulos A., Stathopoulos P., Mitrou N., (2007) 'Priamos: A Middleware Architecture for Real Time Semantic Annotation of Context Features', *3rd IET International Conference on Intelligent Environments.*

[104] Konstantinou N., Solidakis E.,. Zafeiropoulos A, Stathopoulos P., Mitrou N., (2010) 'A Context aware Middleware for Real-Time Semantic Enrichment of Distributed Multimedia Metadata' *International Journal of Multimedia Tools and Applications (MTAP), Springer, special issue on Data Semantics for Multimedia Systems, 46(2): 425-461.*

[105] Intel Lab Set: http://db.csail.mit.edu/labdata/labdata.html.

[106] Janjua, Naeem Khalid, Farookh Khadeer Hussain, and Omar Khadeer Hussain. "Semantic information and knowledge integration through argumentative reasoning to support intelligent decision making." *Information Systems Frontiers 15.2 (2013): 167-192.*

[107] Power, D.J. (2002). Decision support systems: Concepts and resources for managers. *Greenwood Publishing Group.*

.