Modelling Information Assets for Security Risk Assessment in Industrial settings

Marcelo Masera & Igor Nai Fovino
Joint Research Centre

About Authors

Marcelo Masera is an Electronics & Electrical Engineer (1980), and an officer of the European Commission at the Joint Research Centre since November 2000. He is in charge of the “Information Security of Critical Networked Infrastructures” area within the Institute for the Security and Protection of the Citizen. His interests are on the dependability and security of complex socio-technical systems, and specifically those related to critical infrastructures, large-scale systems-of-systems, information and communication technologies and the information society. He has published more than 60 papers in the fields of dependability, security and risk.

Contacts Details: Institute for the Protection and the Security of the Citizen, Joint Research Centre Via E. Fermi 1, I-21020 Ispra (VA), Italy. Phone: +39-0332-789238 email: marcelo.masera@jrc.it

Igor Nai Fovino received the M.S. degree in Computer Science with full marks in 2002 and the Ph.D. with full marks in Computer Science in March 2006. He worked as at University of Milano in the field of privacy preserving data mining and computer security. In 2004 he was visiting researcher at CERIAS Research Centre (West-Lafayette, Indiana, USA) working on secure and survivable routing protocols. Currently he is Scientific Project officer at the Joint Research Centre of the European Commission. He serves as reviewer for some international journals. His main research activities are related to the computer security and, more specifically, three are the main interests: System Survivability, Secure Protocols and Privacy Preserving Data Mining. In these contexts, in the last two years he published on international journals and conference proceedings more than 12 scientific papers.

Contacts Details: Institute for the Protection and the Security of the Citizen, Joint Research Centre Via E. Fermi 1, I-21020 Ispra (VA), Italy. Phone: +39-0332-786541 email: igor.nai@jrc.it

Keywords

Information Assets, System modelling, Risk Assessment, Security Assessment
Modelling Information Assets for Security Risk Assessment in Industrial settings

Abstract

Industry has begun in the last years to take into consideration the use of Public Information Infrastructures (including the Internet) for remotely monitoring, managing and maintaining their technical systems. Concurrently, technical and business information systems are getting interconnected both through private and public networks. As a result, industry is exposed to internal and external cyber-threats, and the security assessment of the ICT infrastructures assumes a predominant relevance. However, underlying every useful security methodology there is a system description which decomposes the system in terms of services, component, relationships and assets. In this paper, we focus our attention on a particular type of system asset to which, to our knowledge, the usual security assessment methodologies do not pay sufficient attention, the information asset. Such an asset, in fact, represents the core of every ICT infrastructure (commands sent to components are information assets, data stored into databases are information assets, data flowing through the network are information assets); therefore we believe that its proper description and analysis is key for assuring reliable results for security assessments. Starting from some classical definitions of information and knowledge, we examine this type of asset aiming at identifying the more suitable representation with respect to its security attributes. In more detail, we identify as interesting properties the interdependence between information assets, their life cycles, their dynamics (i.e. the flows of the information assets within the system), their topological location (in term of subsystems that hosts the information assets) and the correlation between the information assets and the vulnerabilities affecting the components of the system. We provide then a formal modelling framework for describing the characteristics of the information assets under a security assessment perspective.

Introduction

Security threats are one of the main problems of this computer-based era. A recent trend (Paukatong, 2005; Bowen, 2005; Fulton, 2005) shows that also in the industrial context, the use of ICT infrastructures has reached a level that impose taking in consideration cyber failures and attacks as one of the main threats to the security of such systems. More in details, industrial systems combine typical ICT components (e.g. databases, communication protocols and networks, protocols, operating systems), with real-time elements (known as SCADA: supervisory control and data acquisition systems) where the control functions are implemented.

In such a scenario, it has become imperative for industrial actors to perform proper risk/vulnerability assessments, putting in evidence the main threats their critical systems are exposed to and eventually the effectiveness of the possible countermeasures. There exist in the scientific literature some interesting approaches to the risk assessment of ICT infrastructures [Alberts & Dorofee, 2002; Masera, Nai & Sgnaolin, 2005]. These methodologies have as core target the analysis of the system components, the interconnection between components and the set of correlated vulnerabilities, threats and attack mechanisms. However, they pay little attention to an important asset family of a system: the information assets. Although these methodologies have proven useful for zeroing in the security lacks of the analyzed systems, we believe that it is possible to improve the results of risk assessments by a more attentive and precise description of information assets.

Information permeates all aspects of industrial systems: based on the information stored and handled in a system, decisions are taken about its management and operation; information flows
within the system and through the connections with remote systems, link the different components and enable the command and observation of the system states; and both the discovery of vulnerabilities and the protection against threats heavily depend on the previous points.

It is therefore evident the necessity to consider the relevance of information assets to the failure of the system services, by means of their clear identification and characterisation, including the relevance of their security attributes, and a proper assessment of the related vulnerabilities and of the attack mechanisms that might assail them.

In this paper, we give a possible solution for this problem, presenting a framework for the description of the information assets of a target system, stating their relationships with the components of the system (at different level of details) and the vulnerabilities that can affect them. In order to do this we adopt as basis the risk assessment methodology presented in [Masera et al., 2005].

The paper is organized as follows: section 2 presents an overview of the risk assessment methodology. Section 3 presents a discussion of the characteristics of information assets under the risk assessment perspective. In section 4, our approach is presented and finally section 5 stresses the main conclusions.

The Security Risk Assessment Methodology: An Overview

The characterization of Information Assets (IA) strongly depends on the objectives the assessment process. In other words, an IA has several different interesting characteristics that can be modelled. However, only few of these characteristics are of interest for a particular scope. In our case, we are interested in the representation of the characteristics of an IA relevant to identify and evaluate the security of the system under analysis. It is, however, not sufficient to identify the characteristics to be modelled: in fact, different assessment methodologies put under analysis different aspects of a system, and thus require different types of models. For this reason, before describing the information assets, it is necessary to define the type of security risk assessment methodology for which the IA framework will be used.

As we have stated before, there exist in literature several types of risk assessment methodologies. We have chosen to adopt as reference the work of Masera et al. [Masera et al., 2005]. In such a work, the authors present a risk assessment methodology tailored to the analysis of the ICT infrastructure of complex industrial systems. In the remainder of this section, we give a brief overview of this methodology. More in details, this methodology foresees that in order to assess the security of a system, it is necessary firstly to provide a description of the system itself, of its components, of its assets, of the interaction and the relationships among the components, the assets and the external world.

Such a description (expressed analytically by tables) could be used to identify in a systematic way the vulnerabilities affecting the whole system. These vulnerabilities are then described by some significant parameters and used to identify the threat that can be associated to the components and to the whole system. From the analysis of this information, one can derive the evaluation of the possible damages to the components, their propagation to the system and the consequent attack pattern. All these operations are quantified in some risk related indexes that are then employed to perform the evaluation of the security failure risk and the countermeasures.

The approach adopted is based on five main steps. With regard to the topic of this paper, the information asset, are, in this methodology equated to any other type of asset relevant for the security of the system analyzed. Even if this can be a good choice in order to make easier the
analysis (i.e. it is not necessary to know a lot about the information asset, its flow etc.), we believe that a better and more detailed description of this particular asset, could give a great improvement in the analysis of a system.

**Information Asset Characterization**

As already claimed in the introduction, Information Assets have become extremely important elements of every technological system. Their relevance, however, is not due, as for other types of assets, only to their physical integrity or availability, but to a more subtle concept. In fact, if we imagine the IA as a black box in our system, the value, i.e. the relevance, of this black box is not associated to its physical attributes, but to the significance of the specific IA with reference to the particular state of the system at issue.

The significance of the data has therefore two facets: the value of the data per se, and the value of the data in light of possible security failures of the containing system (e.g. a system failure due to the lack of integrity or availability of the data).

In the first, the information asset connotes a type of knowledge whose contents and values are rather stable; in the second case, changes are relatively rapid. The approach to modelling has to fit both cases.

For this reason, before the characterization of Information Assets, we need to clarify the concept of information and the concept of knowledge contained by the information. Several definitions of Knowledge have been proposed in the scientific literature. As a first step, it is necessary to introduce the bricks by which Knowledge is built, data and information.

A data is a set of discrete, objective facts about events. In an organizational context, data is most usefully described as structured records of transactions. Drucker [Drucker 1999] said that information is “data endowed with relevance and purpose,” which of course suggests that isolated data by themselves have no relevance or purpose.

With regard to the concept of information, in the scientific literature, some people equate it with meaning [Miller 1987]. Hearing a statement is not enough to make an event an informative act; its meaning must be perceived to make the statement informative. Arguing against this approach, Bar Hillel points out that “it is psychologically almost impossible not to make the shift from the one sense of information… i.e. information = signal sequence, to the other sense, information = what is expressed by the signal sequences” [Bar-Hillel 1955]. In another approach, information may be understood as “that which occurs within the mind upon the absorption of a message” [Pratt 1982]. A useful analogy is with the structure of systems that is viewed by some as being equivalent to information. Thus, “information is what remains after one abstracts from the material aspects of physical reality” [Resnikoff 1989]. However, a good general definition of information is given by Losee in [Losee 1997]: information is produced by all processes (in every context) and the values associated to the processes output is the information.

The difference between knowledge and information may be extremely thin. Epistemologists have strived to understand what it means to know something; in any case, we can think of knowledge as a set of information explicitly and implicitly correlated. From a formal point of view, Frawley, Shapiro and Matheus in [Frawley Shapiro & Matheus, 1992] give an interesting definition of Knowledge:

**Definition 1:** Let $F$ a set of facts, $L$ a language, and $C$ a measure of certainty, a pattern is a statement $S \in L$ that describes relationships among a subset $FS$ of $F$ with a certainty $C$, such that $S$
is simpler (in some sense) than the enumeration of all facts in FS. A pattern that is interesting and certain enough is called knowledge.

We can deduce that the value of an Information Asset is represented by the value of the knowledge it contains. However, Frawley’s definition was developed in the context of datamining processes and not in the context of system descriptions; the IAs have some characteristics that make the mapping operation between the two worlds not immediate. In order to release an instantiation of definition 1 into the system description world, there is the needs to clarify these characteristics.

First, from the point of view of their nature, IA appears under the form of measurements, commands, alarms, actions, policies, etc. The set of facts that are relevant for the significance of an IA varies for the different types of IA.

Secondly, from their function, information assets can be generated by human operators, by sensors or by the same industrial computing systems. In the analysis of an IA, from a security perspective, we need to take into account of this aspect.

Third, information assets exist within the protected zones of an industrial system, or are communicated through networks, either private or public.

Starting from what we presented previously, we are then now able to identify some relevant characteristics of an information asset:

1. As of their nature, they are abstract. IAs are logical elements, implemented by means of software and hardware components, which encode and materialise them.

2. Their value can be the results of some computation based on the value of other information assets.

3. As objects, they are passive. They do not perform any action and they do not directly interact (we intend as interaction any type of operation in which an entity reacts to external stimulations or stimulates other entities) with any other entity. They can be assumed as simply token of information to be interpreted by components (e.g. a software) or human operator (they represent knowledge, and the other components interpret this knowledge giving it a meaning).

4. They have a life–cycle: they are generated, undergo several handlings, are temporally stored, and finally are deleted. Even if a single asset represents always the same information, they can change form of representation during their life–time.

5. They can flow through the system, and different instances of the same asset can be concurrently localised in more than one point of the system.

6. In the case of aggregate information, they can be distributed in a sparse manner over the whole system.

From the security viewpoint, one has to consider that information assets are entities containing data, encoded in a representation that only acquires meaning in a given context. For this reason, it is not possible to directly associate attacks, vulnerabilities and exploits to an Information Asset per se, without consideration of the context of reference: associated components, time, state of the system, etc. In fact, roughly speaking, a vulnerability, by definition [Alhazmi, Malaiya and Ray (2005), Bishop and Bailey (1996)], is a weakness in the architecture design/implementation of an application or a service, and an information is neither an application (e.g. hardware, software) nor a service.
Although it is true that an information asset can be the target of a threat, the weaknesses used by attackers do not pertain to the information asset itself, but to the services and components implement, interact with and manage the information asset.

Most information assets are mobile objects; they flow from one component to another within and outside the system. The security evaluation has to take into account all the vulnerabilities present in all the possible logical paths the information asset can follow.

The “mobility” feature introduced before, has another important effect. In fact, as the information can flow through the system, it can be kept and maintained at the same time in different point of the same system. In the evaluation of undesirable effects of an attack/failure, it is then important to be able to model even the consistency of the IA, or, more specifically, the mechanisms by which the IA is maintained consistent over all its representations in the system.

Our thesis is that the characterisation of information assets for supporting security assessments entails four sets of knowledge:

- Patterns (following Frawley’s meaning of the term) about the structural relationships among the information assets active in a given system.
- Patterns about the structural relationships between the information assets and the components of that system and their vulnerabilities.
- Patterns about the flows of information.
- Patterns about the dynamics of each information asset’s lifecycle.

We call this collection of knowledge connoted by an information asset its purport. For instance, in the case of a command information asset, its purport includes all the “interesting and certain enough” facts conveyed or implied by it: e.g. the source, the target, the numeric value of the command variable, the range of validity in terms of time and system states, the set and order of components that interact with it and the vulnerabilities that might affect it.

**Information Asset Description**

Due to the fact that generally the system under analysis is very complex and extended, it is completely unrealistic to try to describe every type of information generated in the system. On the other hand, the name itself of what we want to describe (Information Asset), implies that we want to describe only information items that, for the analyst or for who had commissioned the analysis, encompass some relevant value. This consideration directly implies that the analyst must know thoroughly enough the system.

We assume that the analyst has already performed a system description in term of subsystems, services and components (in the adopted methodology [Masera et al., 2005] this corresponds to phase 1 (pre-assessment)).

Moreover, as the scope of our description is to use the obtained knowledge in order to make a security assessment, it is relevant to introduce from the beginning of the IA description some classical security properties that will be used in the remaining of this paper:

- **Confidentiality**: it is the partial or total concealment of information or resource [Bishop 2004] with respect to the rights of the entity that accesses the resource. In our case, given an information asset $ia_i \in IA$ where $IA$ is the set of all the information asset of the system and considering an universe $U = \{u_1, \ldots, u_n\}$ of all the entities which can in some way interact with $ia_i$, we define a confidentiality level assigned to $u_k$ (i.e. the amount of information
contained by $ia_i$ accessible from $u_j$) as a function $f(ia_i, u_j): IA \times U \rightarrow \mathbb{N}$, which gives as result an integer $n$. As it is possible to see, conceptually, the approach is very similar to the approach presented in the Bell-LaPadula work [Bell and LaPadula 1997].

- **Integrity**: it is generally defined as the trustworthiness of data or resources [Bishop 2004] and it is related to the task of preventing their unauthorized modification. In our case, we say that the integrity property of an information asset $ia_i$ is preserved when only authorized users can modify it, and it is possible to identify any unauthorized modification. This presupposes the existence of some security mechanism. Given an asset $ia_i \in IA$ where $IA$ is the set of all the information assets of the system, a set $U=\{u_1, \ldots, u_n\}$ of the entities which are able in some way to interact with $ia_i$ and a set $O=\{o_1, \ldots, o_k\}$ of all the operations which can modify an information asset, we can represent the mechanisms which avoid the unauthorized modification of an information asset by a function $f(ia_i, u_j, o_k): IA \times U \times O \rightarrow \langle 0,1 \rangle$ where 0 means “no right to apply the modification” and 1 means “right granted”. Moreover, considering that there exists the risk that an entity $u_j$ claims to be $u_k$ in order to gain the right to modify an asset, we represent at high level the mechanism that guarantees the identity of any entity $u$ by a function $g(u, p): U \times Evidences \rightarrow \langle 0,1 \rangle$, where $Evidences$ is a set containing some evidences that can be used in order to proof the identity of $u$, and $\langle 0,1 \rangle$ mean false and true. In the scientific literature there exist a large number of different security mechanisms that can be used in order to guarantee the integrity property, with different assurance levels. Therefore, there is the need to specify for each information asset the guaranteed minimum level of integrity. As in the case of confidentiality, we define such a level by a function $i(ia_i): IA \rightarrow \mathbb{N}$ which returns an integer $n$ representing such required integrity level.

- **Availability**: it is related to the ability to have access to an information or resource [Bishop 2004]. In the case of information assets, we say that the availability property is preserved when accidental or malicious causes cannot prevent authorized entities to access the information asset. Each information asset has its own availability requirement. For example, for the access to an information asset $ia_i$ a certain entity $u_j$ may judge unacceptable a delay of 1 ms, while for another information asset $ia_k$ the same entity $u_j$ consider acceptable a delay of 10 min. Even the worst case of unavailability, that is the disruption of the information asset, can be represented as the case of a potentially infinite delay. The maximum access delay to an information asset $ia_i$ which an entity $u$ can tolerate can be then adopted as a measure of the availability level required. We can model the availability level as a function $al(ia_i, u_j, k): IA \times U \times K \rightarrow \mathbb{N}$ where $K$ is a generic set of some additional knowledge (provided by the analyst) about the requirements of the entities contained in $U$. The function $al$ gives as output a maximum delay threshold, which is the maximum acceptable value for a given entity with respect to a given information asset, beyond which the IA must be considered unavailable. The greater such threshold, the weaker are the availability requirements.

As it is simple to imagine, not all the IA in a system require the same level of integrity, confidentiality and availability. This means that information about these properties should be linked to every information asset we want to describe.

**Definition 2**: we define as Security Requirements a tuple $SR=<\text{Conf}, \text{Int}, \text{Av}>$, where:

- $\text{Conf}$ represents the weight (or relevance) of the confidentiality property for an IA.
- \( \text{Int} \) represents the weight (or relevance) of the integrity property for an IA.
- \( \text{Av} \) represents the weight (or relevance) of the availability property for an IA.

Keeping in mind the previous definitions and assumptions, we propose now a formal definition of information assets. In a preliminary analysis, and coherently with what we claimed in the previous section, we note that there exist two levels of description for an IA: a basic description, which is related to few, very evident characteristics of an IA (value, name etc.), and a more wide description, which we will call Extended Information Asset. This is the formal expression of the purport (as defined in the previous section) of an information asset, and it contains the knowledge implied by the IA in the context of its evolution in the system at issue.

The basic IA can be defined as follows:

**Definition 3:** an information asset is defined as a tuple \( \text{IA}_i = <\text{id}, \text{name}, \text{dsc}, \text{val}, \text{sr}> \), where

- \( \text{id} \): is the univocal identifier of the asset.
- \( \text{name} \): is the name of the asset.
- \( \text{dsc} \): is a free text description.
- \( \text{val} \): is the perceived value of the asset.
- \( \text{sr} \): is a security requirements tuple.

Definition 3 gives a very simple description of an IA at system level. However, this is not sufficient. In fact, let consider this example: in a system \( S \), \( \text{IA} = \{\text{ia}_1, \text{ia}_2, \text{ia}_3\} \) is the set of the IA judged relevant by the analyst. We know that the value of \( \text{ia}_3 \) is assigned by a human operator in light of the values of \( \text{ia}_1 \) and \( \text{ia}_2 \). In other words, we can think \( \text{ia}_3 = f(\text{ia}_1, \text{ia}_2) \) where \( f : \text{IA} \times \text{IA} \rightarrow \text{IA} \) is some logical or operational function. In this case, if anything happens at the lower levels of the system having a negative impact onto \( \text{ia}_1 \) or \( \text{ia}_2 \), the effects will be propagated to \( \text{ia}_3 \). Therefore, it is necessary to represent the eventual relationships existing among the different IAs.

**Definition 4:** Let \( \text{IA} = \{\text{ia}_1, \ldots, \text{ia}_n\} \) be the set of the information assets of a system \( S \), let \( L \) be a language, a statement \( S \in L \) is an Information Asset Dependence (IAD) knowledge if it represents, with a relevance \( c > K \) (where \( K \) is a lower bound chosen by the analyst), a relationship between two IAs (\( \text{ia}_j, \text{ia}_k \)).

Definition 4 gives a formal definition of the IAD knowledge. However, a similar definition is not useful if not supported by representation schema. Therefore, we introduce here the concept of IAD schema.

**Definition 5:** an IAD schema is an oriented graph with the following features:

- A set of nodes \( \text{IAN} \) where each node is a basic IA (see definition 3)
- A set of directed and weighted edges \( L_{\text{in},\text{ia}_k} : L_{\text{in},\text{ia}_k} \in (\text{IAN} \times \text{IAN}), \text{ia}_j = \text{ia}_k \), where each weight represents the strength of the relationship between \( \text{ia}_j \) and \( \text{ia}_k \).

The previous definitions allow us to represent a “structural knowledge” related to the interconnection between different IAs.

As explained in the previous section, an important characteristic of the IAs is their mobility. In fact, they are able to flow through and outside the target system. This feature, in a context in which we are interested to study the impact of threats and failure on the IAs is very interesting. However, an important question regarding the IA flows is to choose at which level of details they must be represented. Even if it is true that an IA can flow from a component to another (where for
component we identify, as in the adopted methodology, every hardware and software object in the system, in a complex system an exhaustive approach describing all these flows results unfeasible. We choose then to represent the IA flows at subsystem level. This approach allows maintaining the description simpler and at the same time extremely flexible (in fact, the analyst in this way is free to decide in each singular case the size of the different subsystems and therefore describe the flows with different level of details).

To represent the IAs flows, there is the need to describe the subsystems contained in the system S under analysis and then to associate to every information asset the subsystems which in some way interact with them.

Figure 1 gives a high level logical overview of the structures described in this section, highlighting the links between information assets, subsystems and components.

![Figure 1: High level overview of the IA description](image)

**Definition 6:** a subsystem $sb$ is defined as a tuple $<id, name, description, lia>$ where:

- $Id$ is the unique identifier for the subsystem
- $Name$ is the name of the subsystem
- $Description$ is a free text description of the subsystem
- $Lia$ is a list of tuple $<ia, sr>$ where $ia$ represents the id of an information asset which is managed or hosted by the subsystem $sb$ and $sr$ is a security requirement tuple magnifying the local security requirements for the information asset $ia$. 
Definition 7: Let $IA$ be the finite set of the information assets, we define the finite set of the subsystem through which an information asset $ia_k$ flows, as a set $SB_k = \{sb \in SB \mid ia_k \in sb.ia, ia_k \in IA\}$.

It is interesting to note that the number of sets $SB_k$ that can be described is equal to the size of the set $IA$ of the information assets. We can then define $SB_IA = \langle SB_1..SB_n \rangle$ where $n = |IA|$ as the collection of these sets. Each element of $SB_IA$ can be interpreted then as the description of the discrete relevant moments of an information asset life cycle. However, these sets do not capture any type of knowledge about the relationships of these moments. This type of knowledge can be described by the Information Asset Flow graph.

Definition 8: let $SB = \{sb_1...sb_n\}$ be the set of the subsystem of the system $S$, let $IA = \{ia_1...ia_n\}$ be the set of all the information assets of the system, let $IA_i = \{ia \in IA \mid ia \in sb_i.ia, sb_i \in SB\}$ be the set of the information asset related in some way with the information asset $sb_i$ and $SB_IA$ be the collection of the sets $SB_i$ as defined in definition 7, we define the Information Asset Flow (IAF) knowledge as an oriented, a-cyclic graph $IAF = (V, E)$ where:

- $V \subseteq SB$ is the set of nodes of $IAF$ (it is the set of the subsystem having some relation with the IAs judged relevant in the previous phases of the description process)
- $E$ is a set of oriented and labelled edges $L_{ia}(sb_k, sb_j) \in SB_{ia} \times SB_{ia}$, $SB_{ia} \in SB_{IA}$, where the label $ia$ means that the edge $L$ is related to the flow of an asset $ia : ia \in IA_k, ia \in IA_j$.

In some cases, an information flow could be cyclic. This, however, introduces an additional layer of complexity to the IA description and analysis. How to solve this problem in an elegant way will be a scope of our future studies. However, as first solution we can transform the graph IAF into a new graph $IAF'$ in which the edge between two nodes $sb_i$ and $sb_j$ which closes the cycle is removed and introducing a “virtual” information asset $ia_v$, which contains the original information asset, that flows from $sb_i$ to $sb_j$.

The IAF graph represents the knowledge about the dynamics of the information assets of the system under analysis. Moreover, it can be used in order to represent the life-cycle of an information asset. In fact given the $IAF=(V,E)$, it is possible to identify the set of the information asset generators $SBG$, which is the set of the subsystem which generates the information assets and at the same way it is possible to identify the set $SBS$ of the subsystem in which an information asset ends its life cycle. More formally, we define such sets as follows:

Definition 9: let $SB = \{sb_1...sb_n\}$ be the set of the subsystem of the system $S$ $IA_i = \{ia \in IA \mid ia \in sb_i.ia, sb_i \in SB\}$ be the set of the information asset related in some way with the information asset $sb_i$, let IAF=$(V,E)$ be the information asset flow graph of a target system $S$, we can define the set $SBG$ of IA generator as $SBG = \{sb_i \mid sb_i \in SB, \exists L_{ia}(sb_i, sb_j) \in E : \exists L_{ia}(sb_k, sb_i), ia \in IA_i\}$. At the same way, the set $SBS$ the subsystem in which the IA end its life cycle can be defined as $SBS = \{sb_j \mid sb_j \in SB, \exists L_{ia}(sb_i, sb_j) \in E : \exists L_{ia}(sb_i, sb_k), ia \in IA_i\}$.

The structure $LC=\langle SB_IA, IAF, SBG, SBS \rangle$ represent the knowledge about the dynamics of the information assets and the knowledge about their life cycle.

Finally, in order to complete our vision of the information asset under the security perspective, we need to describe the exposition of IAs to failures and vulnerabilities. As we claimed in the previous
section, an information asset is a passive object; this implies that it cannot fail and it is not directly affected by vulnerabilities. However, the components of a subsystem can fail, and an attacker can exploit them taking advantage of their vulnerabilities. Then we guess that an information asset can be damaged by the failures and the vulnerabilities of the components of a subsystem that in some way can interact or manipulate them. In other words, an IA inherits the vulnerabilities and the failures of these components, that, in some way, compromise the security requirements associated with the IAs. The relation between the information assets, security requirements and vulnerabilities/failures is a very useful knowledge to be used in the risk assessment process. In order to represent this knowledge, we need at first instance to give a definition of component.

**Definition 10:** a component $c_i$ is defined as a tuple $<\text{name}, \text{desc}, \text{lov}, \text{sb}_i>$ where:

- **Name** is the name of a component
- **Desc** is a free text describing the component
- **Lov** is the list of the known vulnerabilities affecting the target component
- **Sb_id** is the id of the subsystem containing the component

It is important to underline that the component descriptions and the link between components and vulnerabilities is directly provided by the methodology presented in [Masera et al., 2005].

We need now to express the relation between components and information assets.

**Definition 11:** let $SB = \{sb_1,...,sb_n\}$ be the set of the subsystem of the system S and $IA$ be the set of all the information assets of the system, we define the set $IAC = \{<ia_i,c_j,\text{sr}>|ia_i \in sb_i, lia_i,sb_i, \text{id} = c_j, \text{sb}_i \text{id}\}$ where $\text{sr}$ is a local security requirement tuple, as the set representing the relationships between information assets and components and then indirectly, with the vulnerabilities affecting the components.

However, not all the vulnerabilities of a component $c$ may be of interest for every related information asset. For this reason, the analyst needs to make a pruning operation on the space $V$ of the vulnerabilities, removing from the list of vulnerabilities related to every component $c$, the vulnerability not interesting for a target information asset $ai$.

The result of this operation is a set representing the Information Asset Vulnerabilities knowledge ($IAV$)

**Definition 12:** let be $IAC$ the information asset-components set, we define the Information Asset-Vulnerabilities set as:

$IAV = \{<ia_i,c_j,\text{sr}>, \text{list}_\text{of}_\text{vul} >|ia_i,c_j,\text{sr} >\in IAC, \text{list}_\text{of}_\text{vul} \subseteq c_j, \text{lov}\}$ where the subset $\text{list}_\text{of}_\text{vul}$ contains only the vulnerabilities judged of interest by the analyst.

The set of the four knowledge patterns $IAD$, $LC$, $IAC$ and $IAV$ defined in the previous points, constitute the representation of the Extended Information Asset (EIA), and of its purport forms the security viewpoint (bearing in mind that the concept of “purport” refers to the set of “interesting and certain enough” facts that correspond to the knowledge conveyed by the information asset).

As a conclusion, we state the purport of the Extended Information Asset, in brief $EIAP$, as the superset $EIAP = \{IAD, IAC, LAV, LC\}$.
Conclusion and Future Works

In this paper, we have proposed a characterization of the information assets, underlining the characteristics which are relevant in the context of the security analysis. Moreover, we have defined a formal description framework for information assets that can be synthesized in the extended information asset description that formalizes the concept of information asset purport, expressed as

\[ EIAP = \{IAD, IAC, IAV, LC\} \]

where:

- IAD captures the concept of interdependence between information assets.
- IAC describes the set of security requirements of the information assets, with respect to the system components they interact with.
- IAV correlates the vulnerabilities of the system components with the information assets potentially damaged by their exploitations.
- LC represents the life-cycle of the information asset under consideration, and includes knowledge on the flows of the information assets (IAF).

This formal approach is being tried in our design framework (as represented in Figure 1) for supporting the analysis of security features. This approach allows the specification of the security requirements both at high level (by means of IAD) and at low level (by means of IAC), providing two security perspectives on the information asset: a global view and a local view.

As a part of future work, we plan to integrate the information asset description framework into the security assessment methodology we have developed [Masera et. al., 2005]. Moreover, in the context of such integration, we plan to elaborate synthetic indexes for measuring the assurance for the security attributes, to be computed with the elements of the model presented in the current paper. Finally, we plan to validate the integration of our security assessment framework and the information asset description presented here in some study cases.
References


