Towards a Security Event Data Taxonomy

Gustavo Gonzalez-Granadillo[1], José Rubio-Hernán[2], and Joaquin Garcia-Alfaro[2(✉)]

[1] Atos Research & Innovation, Cybersecurity Laboratory, C/ Pere IV, 291-307, 08020 Barcelona, Spain
gustavo.gonzalezgranadillo.external@atos.net
[2] Institut Mines-Télécom, Télécom SudParis, CNRS UMR 5157 SAMOVAR, 9 Rue Charles Fourier, 91011 Evry, France
{jose.rubio_herman,joaquin.garcia_alfaro}@telecom-sudparis.eu

Abstract. The information required to build appropriate impact models depends directly on the nature of the system. The information dealt by health care systems, for instance, is particularly different from the information obtained by energy, telecommunication, transportation, or water supply systems. It is therefore important to properly classify the data of security events according to the nature of the system. This paper proposes an event data classification based on four main aspects: (i) the system's criticality, i.e., critical vs. non-critical; (ii) the geographical location of the target system, i.e., internal vs. external; (iii) the time at which the information is obtained and used by the attacker i.e., a priory vs. a posteriori; and (iv) the nature of the data, i.e., logical vs. physical. The ultimate goal of the proposed taxonomy is to help organizations in the assessment of their assets and events.

Keywords: Security event taxonomy · Data classification
Risk assessment · Countermeasure selection

1 Introduction

Visualization models have been widely proposed to help operators in the evaluation and selection of security countermeasures against cyber attacks [1–3]. Most of the approaches rely on statistical data and expert knowledge to fill the parameters composing the model. A great level of accuracy and detail is required to compute the impact of malicious actions detected on the target system and therefore, to determine the most suitable solution.

Geometrical models [4–6] have been previously proposed to represent graphically the impact of cyber security events (e.g., attacks, countermeasures), as geometrical instances (e.g., polygons, cubes, prisms). The approaches consider information of many kinds (e.g., logical, physical, internal, external, etc.) to fill up the model and compute the shape and size of the cyber event. As a result, it is possible to determine the impact (e.g., size, coverage, residual risk, collateral damage) of single and/or multiple events occurring on the target system through geometrical operations (e.g., union, intersection).

N. Cuppens et al. (Eds.): CRiSIS 2017, LNCS 10694, pp. 1–17, 2018.
https://doi.org/10.1007/978-3-319-76687-4_3

One issue that confronts the impact assessment of cyber security events is the identification of the type of information required to feed the model. Each system provides information according to the nature of the event (e.g., energy system provides data about power consumption, blackouts, voltage, etc.; Dam systems provide data related to the level of water, turbidity, volume, etc.). It is therefore important to properly classify the data of security events according to the nature of the system.

This paper is an attempt towards a security event data taxonomy. We propose to classify the information of events based on the criticality of the system (critical vs. non-critical), the time at which the information is obtained (a priory vs. a posteriori), the geographical location of the target system (internal vs. external), and the nature of the data itself (logical vs. physical). This classification is not intended to be exhaustive, but a guide to help organizations in the assessment of their assets and events.

The remaining of the paper is structured as follows: Sect. 2 defines security event data. Section 3 discusses about the information of critical and non-critical systems. Section 4 discusses about internal versus external data. Section 5 compares the a priori information versus the a posteriori information. Section 6 details logical versus physical data. Section 7 proposes a Security Event Data Matrix. Related work are presented in Sect. 8. Finally, conclusions and perspective for future work are presented in Sect. 9.

2 Security Event Data

Considering that an *event* is defined as any observable action in a system or network that indicates the occurrence of an incident; and *information* is defined as any communication or representation of knowledge (e.g., facts, data, opinions in any medium or form, including textual, numerical, graphic, cartographic, narrative, or audiovisual) [7], we define *Security event data* as all relevant information considered to have potential security implications to the system or network.

This article aims at organizing the information of security events based on their nature and usefulness. We consider any information that can potentially impact organizational operations (e.g., mission, functions, image, reputation), assets (physical or logical resources), or individuals (personnel, providers, customers) through an information system via unauthorized access, destruction, disclosure, modification of information, and/or denial of service.

Security event data are useful to identify threats, define risks, and determine the impact of malicious actions (e.g., attacks) and benign actions (e.g., countermeasures) in an information system. We identify relevant data for critical and non-critical systems. Information about critical systems is divided according to the system's nature (e.g., energy, water, telecommunications, finance, health, transportation), and further classified as *cyber systems* (based on ICT solutions); and *physical systems* (composed of physical processes managed by, e.g., control-theoretic solutions). Information about non-critical systems is divided into *internal information*, further classified as logical and physical data;

and *external information*, further classified as *a priori* and *a posteriori* data. The remaining of the paper details each type of data from our proposed classification.

3 Critical vs. Non-critical Systems Data

This section details the types of data required for critical and non-critical systems to analyze risks, assess events, draw conclusions, and select countermeasures.

3.1 Information About Critical Systems

Critical Infrastructures rely on the Supervisory Control And Data Acquisition (SCADA) technology to monitor industrial and complex infrastructures based on Networked Control Systems (NCSs). They include sectors that account for substantial portions of national income and employment such as energy, ICT, finance, health, food, water, transport, and government. Most of these sectors use Industrial Control Systems (ICSs), e.g., Supervisory Control And Data Acquisition (SCADA) in order to provide control of remote equipment (using typically one communication channel per remote station) [8]. For space constraints, we develop in this section the required and additional cyber and physical data for energy distribution and water supply infrastructures.

3.1.1 Energy Distribution

This category includes the production, storage, transportation, and refining of electrical power, gas and oil. The information used in the energy distribution process includes classification of losses as technical and non-technical. The former originates due to physical reasons and depend on the energy flowing through the network, the nature of transmission lines, and transformers. The latter includes measurement errors, recording errors, theft, and timing differences [9]. Examples of technical losses are underground cables and overhead lines. The information on this category includes the type of conductor (e.g., copper, aluminum); conductor temperature (e.g., 0 Celsius, resistance temperature, heating effect, losses due to heating); energy demand (e.g., 100 MWh/year); energy consumption (e.g., estimated annual consumption, real energy consumption, thresholds, kWh, kVAh); load (e.g., heating load, peak load, load factor); peak load times (e.g., winter afternoons). In addition, technical losses can be originated due to the fact that electricity is transported over long distances and the quality of records can be low. Examples of data retrieved in this category include transformer distance (e.g., Kms); transformer material (e.g., iron); power voltage (e.g., high voltage, medium voltage, 400/230 V, 132,000 V); transformer temperature (e.g., heating level, fixed losses, mean temperatures).

Examples of non technical losses include errors (e.g., reading errors, positive error, negative error, timeswitch errors); timing differences (e.g., meter reading period, meter reading frequency, absolute differences); profiling (e.g., profile coefficient, half hourly periods, street lighting profiles, domestic consumers profile,

business consumers profile); data collection frequency (e.g., monthly, quarterly, annually); reconciliation (e.g., reconciliation run, settlement reconciliation, post-final reconciliation run); service status (e.g., active, idle, energisation).

Other types of data found in energy distribution systems include meter identification (meter point administration service, meter point administration number); meter type (e.g., passive, dynamic); Calculation Factor (Group Correction Factor, Loss Factor, Peak Load Factor, Power Factor, Half Hourly Consumers Factor); agents (e.g., distributors, suppliers, collectors); wiring system for supplying electricity (e.g., three phase, single phase); sources (source of technical losses, potential source of error); electrical equipment (e.g., transformers, electrical switches); media type (e.g., fiber optics, leased line, Public Switch Telephone Network - PSTN, Global System for Mobile communications - GSM, General Packet Radio Service - GPRS, Terrestrial Trunked Radio - TETRA); communication protocols (e.g., Long-Term Evolution - LTE, High Performance Radio LAN - Hyperlan); Human Machine Interface (e.g., video wall, client console); switch brand (e.g., Cisco, HP, DIGI); Distribution Management System (e.g., high voltage, medium voltage, low voltage), security device (e.g., firewall, load balancer, IDS, IPS, anti-virus, SIEM).

3.1.2 Water Supply

This category includes services that maintain, store, pump, and process water used primarily for drinking.

Several parameters are monitored to assess the safety of a water supply infrastructure (i.e., dam) and foresee possible failures or anomalies [10,11]. Each parameter is measured using different sensors (e.g., Wireless Sensor Networks - WSN). The most common sensors used in monitoring applications are: inclinometers and tiltmeters, used for the measurement of lateral earth movements and wall tilt/rotation which could result in walls failures; crackmeters, used to monitor movement of cracks and joints on the dam surface and are installed on opposite sides of wall cracks to foresee cracks enlargements; jointmeters, deployed across joints to monitor expansion and contraction of a joint, e.g., between adjacent blocks of a concrete dam; earth pressure cells, used to measure the total pressure for embankment dams; piezometers, used to measure fluids pressure in the embankments or in the boreholes, as well as to monitor the seepage, measure uplift pressure and evaluate the shear strength; turbidimeters, used to measure the water turbidity and to identify signs of internal erosion and piping that can lead to the failure of the dam's walls; thermometers, used to measure water temperature and for environmental thermal monitoring to prevent damages to the water life habitat.

In addition to sensors, other components take part of a water supply infrastructure. Examples of such components are: Programmable Logic Controllers - PLCs (e.g., integrated, compact, modular, small, medium-sized, large); data collectors (e.g., human machine interaction interfaces, data storing units, command and data gateways, signal buses); control devices (e.g., workstation, database, Human Machine Interface, shared resource); monitoring device (e.g., Master

Control Unit - MCU, Remote Master Unit - RMU). These components use standard protocols (e.g., TCP/IP, Collection Tree Protocol - CTP, USB serial communication port, Modbus, Distributed Network Protocol - DNP3, Inter-Control Center Communications Protocol - ICCP); they are connected to a public network for exchanging information and data with remote sites a connecting links (e.g., satellite and radio links, telephone lines, Internet). They are protected using security mechanisms (e.g., Firewalls, VPNs, Intrusion Detection Systems, Intrusion Protection Systems); such mechanisms allow for software controls (e.g., patching, automatic updates, component changes).

3.2 Information About Non-critical Systems

The primary data needed for a risk assessment should include the organization's mission statement, a list of programs they have developed in support of that mission, a list of assets by classification that support the programs, the organization's functional organization chart, the relationship between the business functions and the physical property, existing countermeasures used to protect those assets, and any historical data relating to past security events [12].

The identification of methods in the system are proposed by Howard et al. [13] and further detailed by Manadhata and Wing [14]. An information system communicates with its environment through methods. These latter are entry/exit points that receive/send data directly or indirectly from/to the environment. Examples of a web server's direct entry points are the methods in the web server's API and the web server's methods that read configuration files. An example of exit points are methods that write to a log file.

Other types of data in non-critical systems include penetrating methods (e.g., password cracking, social engineering, masquerading); biometrics and physical tokens (e.g., fingerprint, iris, voice recognition, signatures); defeating mechanisms and policies (e.g., challenges related to authentication, authorization, access controls and policies); and malicious code (e.g., virus, bugs, coding problems) [15].

For events originating in Mobile Ad hoc Networks (MANETs), data can be defined based on the legitimacy of attacking node (e.g., internal, external node); based on the number of nodes involved (e.g., single, multiple), based on the exploited vulnerability (e.g., lack of security boundaries, lack of central management, scalability, cooperativeness); based on the targeted victim (e.g., host, network); based on the security violation (e.g., availability, confidentiality, integrity). More details on each type of data can be found in the work of Noureldien [16].

Information about non-critical systems is further classified as *internal* and *external* data.

4 Internal vs. External Data

Internal and external information are required to analyze the impact of a cyber security event. Internal information represents all logical and physical data from

the local network or from the information system, such as assets, vulnerabilities, defense mechanisms, etc. External information is related to entities outside the information system such as customers, providers, competitors, attackers. These latter can be identified according to their knowledge, motivation, and capabilities to exploit a given vulnerability from the target system. This section details both information from the target system and from outsiders.

4.1 Internal Data (Information About the Target)

Considering the characteristics of access control models [17], we identify three types of information associated to a particular event: User account - a unique identifier for users in the system that allows them to connect and interact with the system's environment (e.g., super admin, system admin, standard user, guest, internal user, nobody); Resource - either a physical component, (e.g., host, server, printer), or a logical component, (e.g., files, records, database), of limited availability within a computer system; and Channel - the way to execute actions, (e.g. connect, read, write, etc.). Channels can also regroup IP addresses, port numbers, protocols and all other kind of TCP/IP connections. More information about these data types are found in the research of Gonzalez-Granadillo et al. [4].

In addition, we consider the notion of contexts proposed in the Organization based Access Control (OrBAC) model [18], such as temporal conditions - granted privileges only during specific periods of time (working time, day time, night time, weekdays, weekends) or considering actions performed at a given time slot (e.g., connection time, detection time, time to react, time to completely mitigate the attack, recovery time, etc.); spatial conditions - granted privileges when connected within specific areas (e.g., user's location, security areas, specific buildings, a country, a network or sub-network); and historical conditions - granted privileges only if previous instances of the same equivalent events were already conducted. For instance, in order to access a web-server (resource) of a given organization, an external user (user account) connects remotely (spatial condition) to the system by providing his/her log-in and password (channel) at nights (temporal condition).

Information security properties (e.g., confidentiality, integrity, availability) are also a key aspect in the analysis of a cyber security event. An event can be associated to a particular issue compromising the system's confidentiality (e.g., unauthorized access to sensitive information, disclosure resources, etc.), integrity (e.g., unauthorized change of the data contents or properties, etc.), or availability (e.g., unavailable resources, denial of service, etc.).

Internal information is further classified as Logical and Physical. Section 6 details each type of data.

4.2 External Data (Information About the Attacker)

All information systems interact with people: internals, when they belong to the organization; and externals, otherwise. External people can have direct contact

to the organization (e.g., vendors, visitors, customers) or indirect contact with the organization (e.g., competitors, intruders, attackers). For people with direct contact with the organization, we need to identify their occupancies (where they work and interact), the hours of occupancy, tasks, uses of hazardous materials or equipment, their needs for access, and their frequency of access [12]. It is also important to note any classic or specific threats against these people. People with indirect contact to the organization are seen as adversaries.

According to Krautsevich et al. [19], adversaries can be either (i) omniscient, when they know all vulnerabilities and all possible patches of the system; (ii) deterministic, when they have a belief knowledge of the system and they choose the best possible action to break into the system; or (iii) adaptive, when they adapt the strategy to complete the attack, using updated knowledge about the system. In reality, attackers do not have the knowledge of all the system's vulnerabilities. We concentrate, therefore, in deterministic and adaptive attackers. Data coming from these type of entities are considered in Sect. 5 as a priori and a posteriori data.

5 A Priori vs. A Posteriori Data

This section discusses two types of information that can be used for a malicious entity in the execution of an attack. *A priori data*, which considers information before the attack is realized, and *a posteriori data*, which considers information discovered by the attacker once the attack is in place. The remaining of this section presents examples of each data type.

5.1 A Priori Data

This classification considers the set of information about the system, possessed by an attacker before exploiting a given vulnerability. If the attacker has a priori knowledge about the operation of the entire system, he/she would be able to inflict a much severe attack. We distinguish two types of a priori knowledge: the knowledge about the information system, and the knowledge about the attack. The former considers the understandings that the attacker has about the system, whereas the latter considers the skills and experience of the attacker in executing a given attack.

About the information system: Following the common vulnerability system scoring method (CVSS) [20], we consider in this category, the known vulnerabilities of the information system that can be exploited by an attacker to access the system (e.g., access vector, complexity, authentication type, required privilege, exploitability, report confidence, potential collateral damage, user interaction).

The *access vector* category considers the way a vulnerability can be exploited by an attacker in the system (e.g., physical, local access, adjacent network access, network access). The *access complexity* includes the complexity level required for an attacker to exploit a vulnerability once he/she has gained access to the target system (e.g., high, medium, low). The *authentication type* category considers the

number of times an entity must authenticate to a target in order to exploit a vulnerability (e.g., multiple, single, none). The *required privilege* category describes the level of privileges needed for an attacker to successfully exploit a vulnerability in the system (e.g., none, low, high). The *exploitability* category considers level of difficulty at which a vulnerability can be exploited (e.g., unproven, proof of concept, functional, high, not defined). The *report confidence* category identifies the degree of confidence in the existence of the vulnerability and the credibility of the known technical details (e.g., unconfirmed, uncorroborated, confirmed, not defined). The *potential collateral damage* category considers the potential for loss of life or physical assets through damage or theft of property or equipment (e.g., low, low-medium, medium, medium-high, high, not defined). The *user interaction* category considers the requirement for a user, other than the attacker, to participate in the successful exploitation of a vulnerability (e.g., none, required).

About the attack: Based on the taxonomy of cyber events proposed in [21], and the research proposed by Cayirci and Ghergherehchi [2], we consider in this category information about the attacker (e.g., type, location, quantity, motivation, technique, mobility), and the attack (e.g., cause, affected service, objective, impact).

The *attacker type* classification includes all threat agents that are primarily responsible for the cyber event (e.g., malicious agents, organizations, foreign governments, natural disasters, or human errors). In terms of *location*, attackers can be located within the network (i.e., insider), or outside the network (i.e., outsider). The *quantity* category defines three types of attackers: single, multiple, or coordinating multiple. These latter defines the case when multiple attackers collaborate with each other. The *attacker's motivation* as proposed by Bielecki and Quirchmayr [1], and Shinder [22] considers the different goals (motives) that can encourage an attacker to exploit a vulnerability on the system such as low (e.g., no motivation, just for fun), medium (e.g., political motives), and high (e.g., for monetary profit; anger, revenge and other emotional drivers; sexual impulses; psychiatric illness). The *technique* includes all types of actions used to achieve the attacker's objective (e.g., system compromise, protocol compromise, resource exhaustion, hardware failure, software crash). In terms of *mobility*, attackers can be fixed or mobile.

The *attack cause* classification differentiates between effects directly or indirectly caused by an event (e.g., disruption within service, cascade disruption from a service). The *affected services* classification considers the priority of service nodes (e.g., primary service node, intermediate service node, secondary service node). The *objective* of the attack considers how the malicious entity attempt to achieve its goal (e.g., data corruption, data fabrication, data destruction, data disclosure, data discovering, no objective). The *attack impact* considers the effects in terms of confidentiality, integrity and availability (e.g., none, low, medium, high, extreme).

5.2 A Posteriori Data

A set of information gained by the attacker after a successful exploitation of a system's vulnerability [19]. The system can release information that improves the attacker's knowledge to exploit vulnerabilities or to overcome the security controls set by the system, however, the adversary knowledge is generally incomplete. In this section we study the attacker's knowledge with respect to the system evolution (e.g., deployment of countermeasures).

About the countermeasures: From the adversary point of view, the ability to penetrate a system does not necessarily implies the ability to break into a system. Breaking a system means making the system to fail and keep on failing. It is more hostile, and more difficult than penetrating into the system, since it requires an understanding of what makes the system fail [23]. However, penetrating the system is the first step for an attacker to improve his/her knowledge about the system.

According to Krautsevich et al. [19], an attacker observes a system and can influence its behavior by making actions at a given moment. The system responds to an action probabilistically. Attackers do not make decisions about actions blindly. Instead, they take into account past, current, and possible future states of the system, as well as possible rewards that are connected with the actions. The goal of the attacker is to maximize the expected total reward according to a sole criterion.

We define the attacker's a posteriori knowledge based on the actions the defender performs to protect the system against a given attack (e.g., implementing security countermeasures). Security measures can be performed automatically by the system and can be soft (e.g., reducing credit limits, restarting the system, requesting password change), moderate (additional authentication method, temporal access denial, temporary fix, alarms) or aggressive (e.g., vulnerability patching, blocking user account, admin rights request). Depending on the decisions available to the attacker, he/she will be able to change its behavior and adapt to the system or quit his/her initial goal.

The Incident Object Description Exchange Format (IODEF) [24] classifies the actions taken a system as a defense mechanism. Examples of such actions are: nothing (i.e., no action is required); contact-source-site (i.e., contact the site identified as the source of the activity); investigate (i.e., investigate the systems listed in the event); block-host/network/port (i.e., block the host/network/port listed as sources in the event); status-triage (i.e., conveys receipts and the triaging of an incident).

In addition, physical countermeasures consider all security actions taken to prevent, protect, or react against a malicious physical event that originates in the system. Examples of physical countermeasures include blocking/opening doors, disabling/enabling hardware, disconnecting/connecting equipment, repairing/replacing hardware, turning on/off devices, posting banners and/or security messages within the organization's infrastructure, installing video surveillance and/or biometric systems.

6 Logical vs. Physical Data

As previously stated, internal information (i.e., related to the system and its entities) is classified according to its nature in *Logical* when the information is intangible (i.e., digital data) and *Physical* otherwise. This section details each type of data.

6.1 Logical Data

Logical information corresponds to all intangible data associated to the target system that can be used by an adversary to execute an attack. Examples of logical data are proposed by Howard et al. [13] as business records, application's information, and security issues. In terms of business records, we consider the organization's proprietary Information (e.g., proprietary business processes, strategic plans, customer lists, vital records, accounting records).

Application's information considers resource consumption (e.g., CPU cycles, memory capacity, storage capacity, and I/O bandwidth); communication channels (e.g., sockets, RPC connections, named pipes, files, directories, and registries); and process targets (e.g., browsers, mailers, and database servers).

Security issues consider alerts or alarm signals, access control violations, photo-ID alteration, noise in voice and video records. Examples of this category include the use of *security mechanisms* such as Transport Layer Security (TLS), expressing that the application uses HTTPS, or server side input validation; the use of *cookies* (considering the maximum number of cookies and the number of foreign cookies from other sites that the application sets during a session); the *access control method* required (e.g., unauthenticated, authenticated, or root); and the *access right* required (e.g., read, write, execute, root).

In addition, Howard et al. [13] have identified several attack vectors to determine a relative attack surface of different Windows applications. Examples of such vectors include open sockets (e.g., TCP or UDP sockets on which at least one service is listening), active web handlers (e.g., http, nntp), dynamic web pages (e.g., .exe files, .asp (Active Server Pages) files, and .pl (Perl script) files), VBScript enabled (whether applications, such as Internet Explorer and Outlook Express, are enabled to execute Visual Basic Script).

For event notification messages using the Syslog protocol [25], useful information is associated to the facility responsible of the message (e.g., kernel, user, mail system, clock daemon, log alert); to the severity associated to the message (e.g., emergency, alert, critical, error, warning, debug), to the identified machine that originally sent the message (e.g., Fully Qualified Domain Name, IP address, hostname), and to the time at which the message was originated (i.e., timestamp).

The Intrusion Detection Message Exchange Format (IDMEF) [26] identifies other fields of interest in the event data classification. The alert has been fired by an analyzer, from which we can derive the source, the target, the time at which the alert was created, the time at which the event was detected, the impact assessment, and information about the node or user that appears to be causing

the event. In addition, we can also consider the information about the completion of the event (e.g., failed, succeeded); the confidence on the evaluation of the event (e.g., low, medium, high); and the algorithm used for the computation of the checksum (e.g., MD4, MD5, SHA1, SHA2-256, Gost).

6.2 Physical Data

Physical information corresponds to all tangible elements that interact directly or indirectly with the target system and whose intrinsec vulnerabilities can be used by an adversary to execute an attack. Examples of physical data are proposed by Norman [12] as people, technical and non-technical devices.

People, represents all internal user accounts (e.g., Key Senior Management, Management and Employees, Contractors, Vendors, Visitors, Customers).

Hi-tech devices correspond to information technology systems (e.g., PCs, servers, laptops, tablettes, pads, mobile phones); office equipment (e.g., copiers, printers, furniture, cash registers); and security devices (e.g., sensors, intrusion detection systems, security information and event management systems, biometrical systems, physical access control systems).

Non-technical devices represent documents or equipment with low or no technical attributes. Examples of such devices are: lo-tech devices (e.g., Access-controlled and non-access-controlled gates, doors, and barriers, lighting, signage, property-marking system, key-control system); no-tech devices (e.g., Policies and procedures, guard patrols and posts, investigation programs, law enforcement liaison program, security awareness program, emergency preparedness program, disaster recovery program).

In addition, it is useful to identify the physical location of people (e.g., network administrator's room, employees offices, guests rooms), physical location of high-tech devices (e.g., server's room, control operation center's location), physical location of network elements (e.g., router location, sensor's physical location), information about the network topology (e.g., interconnection of network elements), location of lo-tech devices (e.g., printer's location, lighting control room), location of no-tech devices (e.g., drawer that stores disaster recovery programs, policies and procedures).

7 Security Event Data Matrix

Based on the information presented in previous sections, we propose a matrix that organizes the event information based on four main aspects: (i) system criticality, (ii) asset location, (iii) event time, and (iv) event nature. Table 1 shows a cyber and physical-based data classification of two critical infrastructure systems (i.e., energy production, water distribution). Table 2 shows a logical and physical-based data classification of internal and external sources of non-critical infrastructure systems.

In order to illustrate the applicability of the event data classification, we consider an issue originated in an infrastructure-less network that uses a Mobile

Table 1. Critical infrastructure systems classification

	Energy distribution		Water supply	
	Required	Additional	Required	Additional
Cyber systems	Technical losses (e.g., circuits, meters, transformers); non-technical losses (e.g., errors, profiling); type of conductor (e.g., copper); data collection frequency (e.g., annually); reconciliation (e.g., settlement reconciliation); protocols (e.g., DNP3, IEC-60870 101, IEC-60870 104)	Transformer material (e.g., iron); timing differences (e.g., absolute differences); profiling (e.g., profile coefficient); meter identification (meter point administration number); meter type (e.g., passive); media type (e.g., fiber optics); communication protocols (e.g., Long-Term Evolution - LTE); Human Machine Interface (e.g., client console); switch brand (e.g., Cisco); security device (firewall)	Security logs (e.g., logs provided by firewall); protocols (e.g., Modbus); resources (e.g., available bandwidth); virtual distribution map (e.g., virtual district metering area)	PLC type (integrated PLC); data collectors (e.g., data storing units); connecting elements (e.g., satellite links); security mechanisms (e.g., Firewall); software controls (e.g., patching)
Physical systems	Load (e.g., heating load); peak load times (e.g., winter afternoons); conductor temperature (e.g., Celsius); energy demand (e.g., 100 MWh/year); Calculation Factor (e.g., Loss Factor); energy consumption (e.g., KW/h); transformer distance (e.g., Kms); power voltage (e.g., high voltage); service status (e.g., idle); errors (e.g., reading errors); sources (e.g., potential source of error); transformer temperature (e.g., heating level)	Electrical equipment (e.g., transformers); Distribution Management System (e.g., medium voltage); wiring system for supplying electricity (e.g., three phase): agents (e.g., collectors); PMU (phasor measurement unit)	Sensors (e.g., WSN); inclinometer (e.g., lateral earth movements); tiltmeter (e.g., wall tilt/rotation level); crackmeter (e.g., movement of cracks and joints on the dam surface) jointmeter (e.g., expansion and contraction of a joint); earth pressure cell (e.g., total pressure for embankment dams); piezometer (e.g., fluids pressure in the embankments or in the boreholes); turbidimeter (e.g., water turbidity level); thermometer (e.g., water temperature)	Monitoring device (e.g., MCU); automated meter reading (ARM); acoustic measures (based on hydrophone sensors or on accelerometers, e.g., to determine leak positions); biosensors measures (e.g., behavior of living organisms in the water)

Table 2. Non-critical infrastructure systems classification

	Internal		External			
			A priori		A posteriori	
	Required	Additional	Required	Additional	Required	Additional
Logical	User account (e.g., admin); resource (e.g., file); Channel (e.g., IP address); confidentiality (e.g., unauthorized access); integrity (e.g., unauthorized change of data content); availability (e.g., denial of service); security mechanisms (e.g., TLS); access control method (e.g., authenticated); access right (e.g., read); event severity (e.g., alert)	Temporal conditions (e.g., detection time); spatial conditions (e.g., user's location); proprietary Information (e.g., accounting records); resource consumption (e.g., memory capacity); process targets (e.g., browsers); cookies (e.g., number of foreign cookies); open sockets (e.g., TCP), active web handlers (e.g., http); dynamic web pages (e.g., .exe files); facility (e.g., kernel); sender (e.g., Fully Qualified Domain Name); analyzer (e.g. source); event completion (e.g., failed); confidence (e.g., high); algorithm used (e.g., SHA1)	Access complexity (e.g., high); authentication type (e.g., multiple); required privilege (e.g., high); user interaction (e.g., required); attacker type (e.g., malicious agents); attacker's location (e.g., insider); quantity (e.g., multiple); technique (e.g., resource exhaustion); affected services (e.g., primary); objective (e.g., data corruption); attack impact (e.g., extreme)	Exploitability (e.g., proof of concept); report confidence (e.g., unconfirmed); potential collateral damage (e.g., high); attacker's motivation (e.g., monetary profit); mobility (e.g., fixed); attack cause (e.g., disruption within service)	Defense mechanism (e.g., block-host/ network/ port), confirmation about the access complexity, authentication type, required privilege and the user interaction required by the system	Soft countermeasures (e.g., restarting the system), moderate countermeasures (temporal access denial); aggressive countermeasures (e.g., blocking user account), confirmation about the exploitability of the system's vulnerabilities
Physical	People (e.g., employees); hi-tech devices (e.g., servers); hi-tech accessories (e.g., USB driver) office equipment (e.g., printers); security devices (e.g., Intrusion Detection Systems), physical access controls (e.g., fingerprint scanners)	Lo-tech devices (e.g., lighting systems); no-tech devices (e.g., disaster recovery program)	Access vector (e.g., local access), physical location of people (e.g., network administrator's room), physical location of high-tech devices (server's room), physical location of network elements (e.g., router location)	Network topology (e.g., interconnection of network elements), location of lo-tech devices (lighting control room), location of no-tech devices (e.g., drawer that stores the disaster recovery program)	Countermeasures in place (e.g., replace hardware), confirmation about access vectors, location of people, location of hi-tech devices, and location of network elements	Confirmation about the network topology, the physical location of lo-tech and no-tech devices

ad-hoc Network to connect devices wirelessly in a continuing self-configuring way. A malicious event has been detected on 2017-03-23 T 15:22 UTC, from an external node that compromised two internal nodes from the network (i.e., Node1: WEB_SRV03, ID 718bc323-9d78-4ada-9629-8176f42a9703; and Node2: FTP_SRV01, ID e470baab-5d88-4b20-ac28-61ea42b37da3). The malicious node exploits a resource exhaustion vulnerability to originate a DoS attack. The source IP address is unknown, and the target IP addresses are identified as 192.168.1.125, and 192.168.4.315.

- Internal logical data (Required): channel (IP address); node IP address (192.168.1.125, 192.168.4.315); node identification (718bc323-9d78-4ada-9629-8176f42a9703, e470baab-5d88-4b20-ac28-61ea42b37da3); security violation (availability);
- Internal logical data (Additional): number of nodes involved (multiple); detect time (2017-03-23 T 15:22 UTC); targeted victim (Node1, Node2).
- Internal physical data (Required): technical device (web server, ftp server);
- External logical a priori data (Required): legitimacy of attacking node (external node); exploited vulnerability (resource exhaustion); consequence (denial of service).

8 Related Work

Classification of cyber and physical security events has been widely researched in the past two decades. While some researches propose attack taxonomies, some others concentrate in countermeasure taxonomies, and some others present formats and standards for event messages. Classification of attacks is extensively proposed in the bibliography. Noureldien [16], for instance, proposes a taxonomy of MANET attacks. Such classifications, although well developed, they lack on information about security actions to mitigate the attacks.

The classification of security countermeasures have been studied by Norman [12] and Abbas et al. [15]. The former proposes a classification of assets for physical security countermeasure analysis; the latter proposes an approach to designing internet security taxonomies. Both researches concentrate on logical and physical security controls, leaving aside different attack scenarios.

Few research works have been dedicated to the classification of both benign and malicious events. Harrison and White [21], for instance, propose a taxonomy of cyber events affecting communities. The taxonomy classifies threats and countermeasures based on multiple criteria but it does not provide information on cyber-physical systems as a whole, nor they consider the time at which the information is detected and used by the attacker.

Howard et al. [13] propose an attack surface model with several attributes to be used in the analysis of the criticality of similar operating systems. The approach has been extended by Manadhata et al. [14] to compare different software systems based on entry points, methods, and channels. More recently, Gonzalez-Granadillo et al. [5] propose a geometrical approach to evaluate the impact of

security events based on a multi-dimensional tool. Even though the models are useful in the evaluation and analysis of the criticality of systems and events, they require to identify event relevant information to compute the results.

Based on the aforementioned limitations we propose an event data classification matrix that considers data formats, standards, and protocols (e.g., IDMEF [26], IODEF [24], Syslog [25], CVSS [20], as well as several other approaches used in the classification and assessments of cyber and physical events.

9 Conclusions and Future Work

We have proposed in this paper an event data taxonomy to be used in the identification of key axes and/or dimensions in the impact assessment of cyber security events. The taxonomy considers required and additional information about all entities involved in the identified event. As such, the proposed matrix separates critical from non-critical systems. The former details the useful data to model cyber and physical events in energy distribution systems and water supply infrastructures. The latter details the useful information related to internal and external entities affected to the events. The proposed matrix goes further by classifying the logical and physical data associated to internal entities (e.g., target system); as well as, the a-priori and a-posteriori data associated to external entities (e.g., attackers). As a result, it is possible to identify the main axes composing geometrical models to assess the impact of malicious and benign cyber security events.

Future work will focus on extending the classification matrix to other critical infrastructures (e.g., transportation, health, finance, etc.) and to use the outcome of this matrix to build and populate the axes of a geometrical model for impact assessment and countermeasure selection.

References

1. Bielecki, M., Quirchmayr, G.: A prototype for support of computer forensic analysis combined with the expected knowledge level of an attacker to more efficiently achieve investigation results. In: International Conference on Availability, Reliability and Security, pp. 696–701 (2010)
2. Cayirci, E., Ghergherehchi, R.: Modeling cyber attacks and their effects on decision process. In: Winter Simulation Conference (2011)
3. Kotenko, I., Doynikova, E.: Countermeasure selection in SIEM systems based on the integrated complex of security metrics. In: 23rd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (2015)
4. Granadillo, G.G., Garcia-Alfaro, J., Debar, H.: Using a 3D geometrical model to improve accuracy in the evaluation and selection of countermeasures against complex cyber attacks. In: Thuraisingham, B., Wang, X.F., Yegneswaran, V. (eds.) SecureComm 2015. LNICST, vol. 164, pp. 538–555. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-28865-9_29

5. Gonzalez-Granadillo, G., Rubio-Hernan, J., Garcia-Alfaro, J., Debar, H.: Considering internal vulnerabilities and the attacker's knowledge to model the impact of cyber events as geometrical prisms. In: Conference on Trust, Security and Privacy in Computing and Communications (2016)
6. Gonzalez-Granadillo, G., Garcia-Alfaro, J., Debar, H.: An n-sided polygonal model to calculate the impact of cyber security events. In: Cuppens, F., Cuppens, N., Lanet, J.-L., Legay, A. (eds.) CRiSIS 2016. LNCS, vol. 10158, pp. 87–102. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-54876-0_7
7. Kissel, R.: Glossary of key information security terms, Revision 2. National Institute of Standards and Technology. U.S. Department of Commerce (2013)
8. Gordon, K., Dion, M.: Protection of critical infrastructure and the role of investment policies relating to national security. OECD, White paper (2008)
9. Sohn Associates: Electricity Distribution System Losses. Non Technical Overview, White paper (2009)
10. Public Utilities Board Singapore: Managing the water distribution network with a smart water grid. Int. J. @qua - Smart ICT Water (Smart Water), 13 (2016)
11. Coppolino, L., D'Antonio, S., Formicola, V., Romano, L.: Integration of a system for critical infrastructure protection with the OSSIM SIEM Platform: a dam case study. In: Flammini, F., Bologna, S., Vittorini, V. (eds.) SAFECOMP 2011. LNCS, vol. 6894, pp. 199–212. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-24270-0_15
12. Norman, T.L.: Risk Analysis and Security Countermeasure Selection. CRC Press, Taylor & Francis Group, Boca Raton (2010)
13. Howard, M., Pincus, J., Wing, J.M.: Measuring relative attack surfaces. In: Computer Security in the 21st Century, pp. 109–137 (2005)
14. Manadhata, P.K., Wing, J.M.: An attack surface metric. IEEE Trans. Softw. Eng. **37**, 371–386 (2010)
15. Abbas, A., Saddik, A.E., Miri, A.: A comprehensive approach to designing internet security taxonomy. In: Proceedings of the Canadian Conference on Electrical and Computer Engineering, pp. 1316–1319 (2006)
16. Noureldien, A.: A novel taxonomy of MANET attacks. In: Conference on Electrical and Information Technologies ICEIT (2015)
17. Li, N., Tripunitara, M.: Security analysis in role-based access control. Trans. Inf. Syst. Secur. **9**(4), 391–420 (2006)
18. Cuppens, F., Cuppens-Boulahia, N.: Modeling contextual security policies. Int. J. Inf. Secur. **7**(4), 285–305 (2008)
19. Krautsevich, L., Martinelli, F., Yautsiukhin, A.: Towards modelling adaptive attacker's behaviour. In: Garcia-Alfaro, J., Cuppens, F., Cuppens-Boulahia, N., Miri, A., Tawbi, N. (eds.) FPS 2012. LNCS, vol. 7743, pp. 357–364. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37119-6_23
20. Mell, P., Scarfone, K., Romanosky, S.: Common vulnerability scoring system Version 2.0, Specification Document, June 2007
21. Harrison, K., White, G.: A taxonomy of cyber events affecting communities. In: Proceedings of the 44th Hawaii International Conference on System Sciences (2011)
22. Shinder, D.: Scenes of the Cybercrime. Computer Forensics Handbook. Syngress Publishing Inc. (2002)
23. Libicki, M.: Brandishing cyberattack capabilities. National Defense Research Institute, white paper (2013)
24. Danyliw, R., Meijer, J., Demchenko, Y.: The incident object description exchange format (IODEF), RFC5070, December 2007

25. Gerhards, R., Adiscon GmbH: The syslog protocol. Network Working Group (2009)
26. Debar, H., Curry, D., Feinstein, B.: The intrusion detection message exchange format (IDMEF), RFC4765 (2007)
27. Zhu, B., Joseph, A., Sastry, S.: A taxonomy of cyber attacks on SCADA systems. IEEE International Conferences on Internet of Things, and Cyber, Physical and Social Computing (2011)