



# Using an event data taxonomy to represent the impact of cyber events as geometrical instances

Gustavo Daniel Gonzalez Granadillo, Jose Rubio-Hernan, Joaquin Garcia-Alfaro

## ► To cite this version:

Gustavo Daniel Gonzalez Granadillo, Jose Rubio-Hernan, Joaquin Garcia-Alfaro. Using an event data taxonomy to represent the impact of cyber events as geometrical instances. IEEE Access, 2018, 6, pp.8810 - 8828. 10.1109/ACCESS.2017.2740402 . hal-01729364

**HAL Id: hal-01729364**

**<https://hal.science/hal-01729364>**

Submitted on 12 Mar 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Using an Event Data Taxonomy to Represent the Impact of Cyber Events as Geometrical Instances

Gustavo Gonzalez-Granadillo, *Member, IEEE*, Jose Rubio-Hernan, *Member, IEEE*,

and Joaquin Garcia-Alfaro, *Senior Member, IEEE*

**Abstract**—Visualization and simulation models used for the evaluation and selection of security countermeasures need accurate data to compute the impact of cyber events (e.g., malicious and benign actions). The information required to build appropriate impact models depends directly on the nature of the system. The information dealt by water supply systems, for instance, is particularly different from the information obtained by energy, telecommunication, transportation, or finance 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 taxonomy based on the system's criticality, the geographical location of the target, the time at which the information is obtained by the attacker, and the nature of the data. A use case on the impact assessment of events originated in a critical infrastructure is presented to show the applicability of the proposed taxonomy.

**Index Terms**—Security, Detection Algorithms, Data visualization, Event Detection, Security Events, Event Data Taxonomies, Geometrical Models, Impact Assessment, Critical Infrastructures.

## I. INTRODUCTION

Information is a valuable asset that needs a particular protection. Its value changes from organization to organization. For some kinds of information, e.g., in the medical sector, where a single event results into a life or death, the value of information cannot only be measured in terms of monetary values [1]. Several other parameters must be considered to assess such events.

Furthermore, information comes from different sources in different formats. Current information systems deal with huge amount of data coming from multiple sources. Security Information and Event Management Systems (SIEMs) are proposed as an excellent option to process such volumes of information within reasonable periods of time. The SIEM proposes a wider and complete view of the malicious entities, as well as the target system, making it possible to analyze and select the actions to be taken in order to mitigate the effects of the detected intrusions or attacks.

One question that security operators are frequently confronted to, is how to make sure that the deployed security action is the most suitable for the detected malicious event.

G. Gonzalez-Granadillo is with ATOS Spain, Atos Research & Innovation, Cybersecurity Laboratory, C/ Pere IV, 291-307, 08020 Barcelona, Spain. E-mail: gdgg@ieee.org

J. Rubio-Hernan and J. Garcia-Alfaro are with the Department of Networks and Telecommunication Services (RST), of Télécom SudParis, Paris-Saclay University, CNRS SAMOVAR, France. E-mail: josemrubiohernan@ieee.org, jgalfaro@ieee.org

Manuscript received July 13, 2017; revised xxxx.

Visualization models have been widely proposed to help operators in the evaluation and selection of security countermeasures against cyber attacks [2]–[7]. 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 [6]–[10] have been previously proposed to represent graphically the impact of cyber security events (e.g., attacks, countermeasures), as geometrical instances (e.g., polygons, polytopes, 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. All geometrical models have been widely used and implemented in real case scenarios [11]–[13]. However, 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.

In this paper 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 priori vs. a posteriori), the geographical location of the target system (internal vs. external), and the nature of the data itself (logical vs. physical). This taxonomy is not intended to be exhaustive, but it can be considered as a guide to help organizations in the assessment of their assets and events. The ultimate goal of this paper is to provide a security event data matrix that classifies the information of each type of system and provides the guidelines that operators and practitioners need in the appropriate assessment of cyber events and the use of simulation and visualization tools to evaluate and select security countermeasures.

The contributions on this article are summarized as follows: (i) An analysis on the required and optional information to feed risk assessment models; (ii) A taxonomy of security event data based on the criticality of the system, the time at which the data is obtained, the geographical location of the target elements, and the nature of the data; (iii) The deployment of the approach over a critical infrastructure with multiple events; and (iv) A security event data matrix that

provides the guidelines for the appropriate assessment of cyber events in the process of evaluation and selection of security countermeasures.

The remaining of the paper is structured as follows: Section II introduces the state of the art in visualization models for countermeasure selection III defines the proposed security event data classification and gives examples of the required and optional information to be used in the impact analysis process. Section V discusses about the impact assessment process for the evaluation and selection of security countermeasures. Section VI illustrates the applicability of the model by implementing a use case example from a critical infrastructure scenario. Related work are presented in Section VII. Finally, conclusions and perspective for future work are presented in Section VIII.

## II. VISUALIZATION MODELS FOR COUNTERMEASURE SELECTION

The current state of the art in visualization tools propose a wide range of models (e.g., geometrical models [6]–[13] and graphical models [14]–[16]) to estimate the impact of cyber security events and to select countermeasures accordingly. This section presents the different visualization models that use geometry as a tool to compute the impact of cyber attacks and security countermeasures within an information system.

### A. Polygonal Models

Polygonal models have been proposed to calculate the impact of cyber events in a 2-dimensional system. The approach considers information about all entities composing an information system (e.g., users, IP addresses, communication protocols, physical and logical resources, etc.), as well as contextual information (e.g., temporal, spacial, historical conditions) to plot cyber attacks and countermeasures as polygons of  $n$  sides. A variety of geometrical instances (e.g., regular and irregular polygons such as: line segments, triangles, squares, pentagons, etc.) results from the analysis of the entities' information included in a system, attack and/or countermeasure [9].

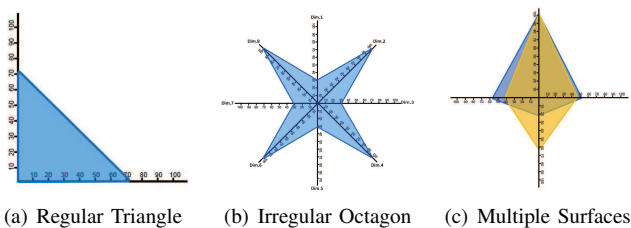


Figure 1: Visualization of Events as Polygons

Each side of the polygon is computed as the contribution of the entity in the execution of an event. The contribution for the user account dimension, for instance, can be evaluated as the number of users affected by a given attack over the total number of active users from the system. Examples of visualization of attacks and countermeasures in the polygonal system are shown in Figure 1.

### B. 3D Models

We identified three main dimensions that contribute directly to the execution of a given attack: User account (subject), Resource (object), and Channel (the way to execute actions, e.g., connect, read, write, etc). This latter is represented as the transitions between subjects and objects. For instance, in order to access a web-server (object) of a given organization, a user (subject) connects to the system by providing his/her log-in and password (action) [8].

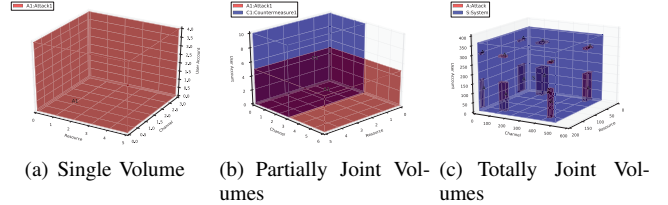


Figure 2: Visualization of Events as Parallelepipeds

The projection of the three axis in our coordinate system generates a parallelepiped in three dimensions. The volume of this parallelepiped is equal to the absolute value of the scalar triple product of all three vectors. The volume calculation requires the computation of the contribution of each axis represented in the coordinate system. This contribution is determined as the sum of each set of axis entities (e.g., user account type, port class, resource type) times its associated weighting factor. Examples of visualization of attacks and countermeasures in a 3D system are shown in Figure 2.

### C. Polytopic Models

An extension of the 3D model into an  $n$ -dimensional model is proposed to represent attacks and countermeasures as polytope instances [12]. The model considers information about user accounts (subjects), resource (objects), and channels (the way subjects have access to objects), as well as, temporal contexts (e.g., granted privileges only during working hours), spatial contexts (e.g., granted privileges when connected within the company premises), and historical contexts (e.g., granted privileges only if previous instances of the same equivalent events were already conducted).

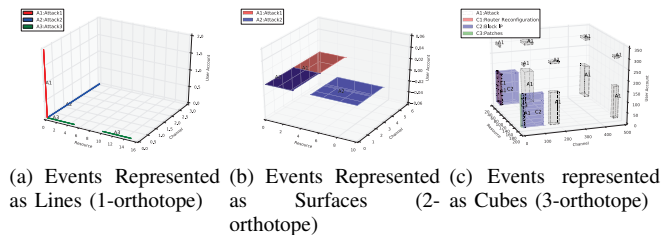


Figure 3: Visualization of Events as Polytopic Instances

Using the polytope-based model, it is possible to represent the impact of cyber events in one dimension (i.e., lines), in two dimensions (i.e., surfaces), in three dimensions (i.e., volumes), and in  $n$ -dimensions (i.e.,  $n$ -orthotopes). It is important to

note, however, that the orthographic projection of instances in more than three dimensions results into complex figures that only provides a general overview of the size and density of the represented event. Unions and intersections are therefore difficult to be plotted from the projected events. Examples of visualization of events as Polytopic instances are shown in Figure 3.

#### D. Prismatic Models

A prism-based model is proposed to represent cyber security events (e.g., attacks, countermeasures) as prismatic instances of  $n$ -sides. The base of the prism integrates the information from the target's side (internal entities), whereas the height of the prism integrates the information from the attacker's side (external entities).

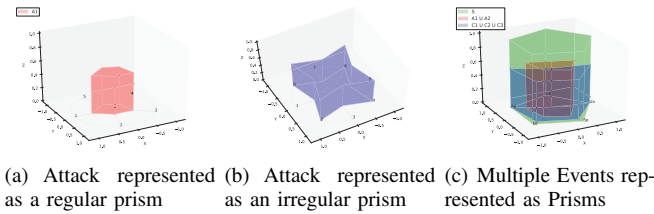


Figure 4: Visualization of Events as Prismatic Instances

The approach considers information about all entities composing an information system (e.g., users, IP addresses, communication protocols, physical and logical resources, etc.), as well as contextual information (e.g., temporal, spacial, historical conditions), and the attacker's information (e.g., knowledge, motivation, skills, etc.), to plot cyber attacks and countermeasures in a geometrical system. The ultimate goal of our model is to help organizations make the most cost-effective decisions in minimizing the risk of the studied cyber events [7].

A variety of geometrical instances (e.g., regular and irregular prisms) results from the analysis of the internal and external information related to a given cyber security event. Examples of visualization of attacks and countermeasures in an  $N$ -Prismatic system are shown in Figure 4.

#### E. Other Visualization Models

Graph-based visualization models are widely used and proposed in the literature. They are based on elements of Visual Grammar e.g., objects (dot, line, plane), abstract structures (gradation, concentric radial, centrifugal), specific objects (shape, size, color), acts (repeat, mirror reflection, rotation) and relationships (symmetry, balance, cluster). The combinations of these elements create different graphical models that are used in the representation of security actions.

Examples of graph models are charts, parallel coordinates, trilinear coordinates, tree maps, geo-maps, as seen in Figure 5. More details about graph-based models can be found in [17].

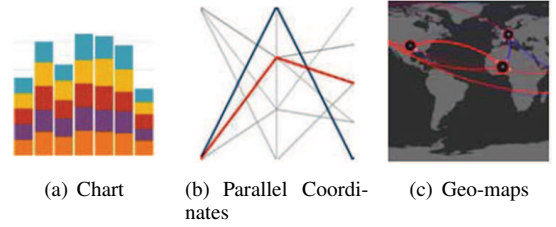


Figure 5: Visualization of Events as Graph Instances

### III. SECURITY EVENT DATA TAXONOMY

*Security Event Data* is defined as all relevant information considered to have potential security implications to the system or network that may require further action (e.g., analysis, trace-back, reaction). For the scope of this article, we use the terms events, actions, and incidents indistinctly. A set of individual actions performed either by the attacker (i.e., malicious actions executed in order to exploit a system's vulnerability) or by the target system (i.e., benign actions executed as a response to an adversary) is considered as a cyber security event. In addition, faults, errors, and failures, whether intentionally or not, are equally considered as cyber security events. These latter generate a wide range of information that, if treated properly, can be useful to analyze the source and consequences of such events.

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. Our proposed taxonomy is depicted in Figure 6, in which we identify relevant information for critical and non-critical systems.

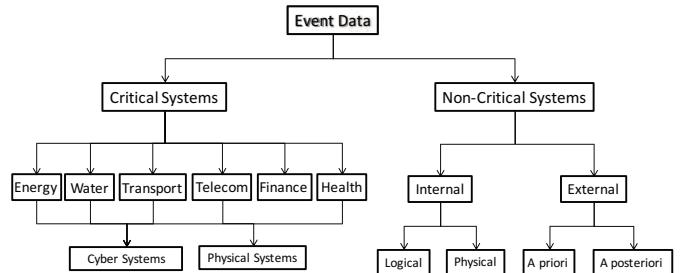


Figure 6: Security Event Data Taxonomy

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). It is worth noting that the interaction between the cyber systems and the physical systems is more and more common moving them to the *cyber-physical systems* (traditional control systems upgraded by novel ICT solutions, and combining cross-layer data). 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.

#### A. Information about Critical Infrastructures

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 (ICS), in instance to provide control of remote equipment (using typically one communication channel per remote station) [18].

Recently a coined cyber-physical system (CPS) term integrates a physical infrastructure and a cyber framework of these critical systems [19]. The technology evolution brings the traditional systems towards a combination of two systems [20]: (i) The *physical systems* which encompasses the physical framework (composed of sensors, actuators, and other devices that interact with the physical processes); and (ii) The *cyber systems*, encompasses the communication and computation framework, which host the SCADA technology. These two types of systems form the two layers of the cyber-physical systems. These latter combine the control strategies, traditionally used to manage the physical layer, with the cyber strategies, in order to monitor the physical layer.

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 [21], i.e., information generated from external action or conditions that are not taken into account in the computation of the control of the physical systems. 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, losses due to heating); energy demand (e.g., 100 MWh/year); energy consumption (e.g., kWh, kVAh); load (e.g., heating load, peak load); 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, 132,000 volts); transformer temperature (e.g., heating level, mean temperatures).

Examples of non technical losses include errors (e.g., reading errors, timeswitch errors); timing differences (e.g., meter reading period, absolute differences); profiling (e.g., profile coefficient, street lighting profiles); data collection frequency (e.g., monthly, annually); reconciliation (e.g., 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, Power 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, Global System for Mobile communications, General Packet Radio Service, Terrestrial Trunked Radio); communication protocols (e.g., Long-Term Evolution, High Performance Radio LAN); Human Machine Interface (e.g., video wall, client console); switch brand (e.g., Cisco, HP, DIGI); Distribution Management System (e.g., high, medium, or low voltage), security device (e.g., firewall, load balancer, IDS, IPS, anti-virus, SIEM).

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 [22], [23]. Each parameter is measured using different sensors (e.g., Wireless Sensor Networks). The most common sensors used in monitoring applications are: inclinometers and tiltmeters, used for the measurement of lateral earth movements and wall tilt/rotation; 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 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, large); data collectors (e.g., data storing units, command and data gateways, signal buses); control devices (e.g., workstation, database, Human Machine Interface); monitoring device (e.g., Master Control Unit, Remote Master Unit). These components use standard protocols (e.g., TCP/IP, Collection Tree Protocol, USB serial communication port, Modbus, Dis-

tributed Network Protocol, Inter-Control Center Communications Protocol); 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); such mechanisms allow for software controls (e.g., patching, automatic updates, component changes).

3) *Transportation*:: this category includes all the transportation system services [25] divided into four main sub-categories: (i) Aviation Service, which includes e.g., air traffic control systems, commercial and recreational aircraft, heliport, aircraft repair stations, navigation aids, military airports; (ii) Highway and Motor Carrier, which include the roadways, bridges, tunnels, vehicles (e.g., school buses, motorcoaches, commercial and personal vehicles, trucks of dangerous material, police vehicles, emergency vehicles, etc.), traffic management system, driver licensing systems, operational management systems; (iii) Maritime Transport System, including waterway, ports, freight and passenger ships, cruise ship, fishing ships; (iv) Passenger and Freight Rail, including light rail, passenger rail, subway, high speed trains, tramways. Each of these subsectors have specific plans to manage and coordinate passengers and freight within the subsector and with the rest of subsectors. Some authors add also Pipeline Systems to carry e.g., natural gas as another subsector.

The transportation sector includes also general parameters (e.g., from computers, communication or management point of views); communication technologies e.g., WiMAX, Wi-Fi, Bluetooth, ZigBee, Wireless Access for Vehicular Environment (WAVE), cellular, satellite, fiber optic, copper file; exchange information protocols (e.g. Vehicle-to-Vehicle (V2V), or Vehicle-to-Roadside (V2R)) [24]; network architecture (e.g., public cloud, private cloud, collaborative cloud using clusters); important network parameters for safety transportation (e.g. bandwidth loss, latency, delay, fast access communication technology, channels, protocols, transmission range), and for security transportation (e.g., resilient security techniques, privacy techniques); storage (e.g., centralized or distributed database, public and private data); network management (e.g., security planning, communication and coordination with industry stakeholder, government and private industry collaboration planning) [25]. In addition, we identify the following physical components: vehicles, ships, trains, infrastructures, human drivers, sensors, motors, surveillance cameras, traffic lights, railways, locals, devices materials, cables, remote workstation (nodes).

4) *Financial Services*:: this category includes services that manage monetary assets e.g., banks, stock exchanges, ATMs. The sources of financial data include: transactions, log data, events, emails, social media, sensors, external feeds, RFID scans or POS data, Free-form text, Geo-spatial, Audio, Still images/video [26]. Examples of financial data are: assets (e.g., current, non-current); liabilities (e.g., current, non-current); Equity; Financial facts (e.g., income, currency, budget); and cash flows (e.g., from operating activities, from investment activities, from financing activities).

Current Assets (e.g., cash, cash equivalents, trade, other

receivables, inventories, short-term financial assets, income taxes, other current assets); Non-current assets (e.g., property, plant, equipment, goodwill, intangible assets, investments accounted for using the equity method, long term financial assets, asset for retirement benefits, other non-current assets, deferred tax assets). Current Liabilities (e.g., trade and other payable, other short-term financial liabilities, income taxes payable, provisions, advances received, accrued expenses, accrued bonuses, other current liabilities); Non-current liabilities (e.g., long-term financial liabilities, liability for retirement benefits, provisions, other non-current liabilities, deferred tax liabilities); Equity (e.g., equity attributable to owners of the parent, capital stock, capital surplus, retained earnings, treasury stock, other components of equity, non-controlling interests).

Income (e.g., net income, cost of sales, gross profit, general selling and administrative expenses, research and development expenses, other operating income, other operating expenses, operating profit, financial income, financial expenses, share of profit of associates accounted for using the equity method, foreign exchange gain, profit before tax, income taxes expenses, profit attributable to owners of the parent, profit attributable to non-controlling interests, basic earnings per share, diluted earnings per share); currency (e.g., USD, EUR, CNY); net sales (e.g., by product type, by country, by department, by period of time); budget (e.g., per account, per unit, per year).

Cash flows from operating activities (e.g., profit before tax, depreciation and amortization, decrease/increase in trade receivable, in inventories, in trade payable, in consumption taxes receivable and payable, in a trade payable other, in advances received, in accrued bonuses, interest and dividend received, interest paid, income taxes paid); cash flows from investing activities (e.g., purchases of property, plant and equipment, purchases of intangible assets, purchases of investments in equity instruments, acquisitions of subsidiaries or other businesses, net decrease/increase in short-term loans receivable); cash flow from financing activities (e.g., dividends paid, other, net cash provided by financing activities); General Cash flows (e.g., effects on exchange rate changes on cash and cash equivalents, net increase/decrease in cash and cash equivalents, cash and cash equivalents at the beginning of the term, cash and cash equivalents at the end of the term).

Additional information in this category include : organization (e.g., business unit, district, corporation, parent company), Customer (e.g., ID, name, address, parent company, status), vendor (e.g., ID, description, type); invoice (invoice date, invoice item, due date, payment date); calendar time (e.g., date, month, quarter, year); company's account category (e.g., cash, benefit allocation, business insurance, payroll tax); Account classification (e.g., cost, revenue, liability, asset, equity); account type (e.g., salaries, commission, rent, insurance); sub-account type (e.g., asset insurance, travel insurance, laptop insurance); transaction code (e.g., revenue, income, tax, amortization); Employee (ID, name, description, active flag, title, rank).

Financial services also include technical equipment (e.g., ATM, database, servers, personal computers, hard drives, printers); physical storage (e.g., drawers, organizers, racks,



cabinets); physical documents (e.g., paper work, accounting books, financial statement printed documents); physical security devices (e.g., video surveillance, smoke detectors, access control devices, biometric devices); banking infrastructure (e.g., building, furniture, security rooms, counselor cabins, teller desks).

5) *Health Services*:: this category includes medical care providers and the health care data at the individual patient level or on a particular population. Examples of health-care data used by a health organization<sup>1</sup> include: Health Services (e.g., percentage of staff employed in medical/dental, nursing, health and social care professions, management/administration, consultant doctors employed in public services, non-consultant hospital doctors, health expenditure in millions of Euro, percentage gross non-capital voted expenditure by program, current health care expenditure); Primary Care Services (e.g., reimbursement service scheme, percentage of population with medical cards by age group, prescription items dispensed under general medical service, claims, payments, community schemes); Private Health Insurance (e.g., percentage of population covered by private health insurance, by gender, by age group, by county); Disability (e.g., number of people with physical and sensory disability, number of people with intellectual disability, degree of disability).

Hospital care provides demographic, administrative and clinical data on all inpatient and day case discharges from publicly funded hospitals. Examples of these data are: Hospital activity (e.g., discharges, bed days, length of stay by county, by gender, by age; number of discharges by principal diagnoses, by principal procedure; emergency department, outpatient attendances, average length of stay for inpatients in public hospitals by age group, number of hospital discharges for selected conditions, rate of discharges); weighting list (e.g., adult inpatient waiting list by hospital, monthly child inpatient waiting list, outpatient waiting list by specialty); Psychiatric data (e.g., psychiatric hospitals, psychiatric units, admission rate per population by age group, number of inpatient admissions per year by gender, admission change percentage).

For comparative studies, enhancing forecasts, devising strategies to reduce medical expenses or evaluating attempts to change patient or provider behavior, additional data are required. Examples of these data are: hospital inpatient (e.g., discharge, diagnose, procedure, length of stay, age, sex, charges), hospitalized patients (e.g., pediatric inpatient discharges, ambulatory surgery patients, emergency patients); claims (e.g., inpatient claim, outpatient claim, medical eligibility, utilization, payment); medicare service (e.g., medical equipment, home health, hospice, inpatient, outpatient, physician, skilled nursing facility); pharmaceutical dispenses (e.g., Drug dispense event information dispense date, pick-up date, cancellation date, Patient, prescriber, dispenser, and facility information); Drug information details (e.g., Drug Identification Number (DIN), Anatomic Therapeutic Classification (ATC) code, quantity, compound components).

Information about the population of a city, a region, or

a country is useful in developing statistics and health care strategies. Examples of these data are: Demographics (e.g., population estimates for regional authority areas by age group, cumulative percentage increase in population, actual population by age group, regional population projection); Life expectancy (e.g., life expectancy by location, by age, by gender, life expectancy at birth, healthy life years); Birth (e.g., numbers, crude rates, total fertility rates, percent outside marriage, annual number of births, birth rates by mother's age, by marital status, by sex of infant, perinatal data); Mortality (e.g., principal causes of death numbers, aged standardized death per population, percentage distribution, mortality by county, by age, by gender, single year age standardized mortality, potential years of life lost, infant mortality).

Optional information include data about productivity (e.g., productivity of physicians in group practices, number of visits in group practices by specialist); hospital statistics (e.g., name, address, phone number, teaching programs, financial status, services offered, number of admissions and outpatient visits); codes (e.g., international classification of diseases, codes of diseases used by hospitals for inpatient and outpatient care, codes for procedures performed by physicians); provider information (e.g., identifier, specialty, role, location, referring provider identifier and discipline); service additional information (e.g., health service code, date of service, amount paid, role, trays, time surcharges, total), service units paid and calls claimed (e.g., facility, diagnostic codes, shadow billed claims).

In addition, health services include infrastructure (e.g., building, furniture, location, physical areas, room dimensions); essential medical devices (e.g., Protective breathing masks, head covers, protective goggles, gloves, coats / gowns, shoe covers), respiration medical devices (e.g., laryngoscope blade, endobronchial and endotracheal intubation tube, tracheo-bronchial aspiration tube, respirator circuit, high-concentration oxygen masks, Oxygen goggles); blood circulation devices (short peripheral catheter, infusion, long central catheter); digestive devices (e.g., aspiration tubing, enteral feeding tubing, lavage and bladder irrigation tubing); ; emergency devices (e.g., drones); robot-assisted surgical machines; personal real-time devices (e.g. real-time EKG device) [27].

6) *Telecommunications*:: this category includes services that facilitate the electronic transfer of information, e.g., telephone companies, Internet service providers, and their hardware infrastructure. This sector, managed by Information and Communication Technology (ICT) companies, is classified as critical because it provides the facilities necessities to manage and enable the communications among the different devices in other critical infrastructure sectors. We can separate five different areas [28]: (i) Broadcast medium frequency (e.g., radio), very high frequency -VHF (e.g. radio and TV), and ultra-high frequency -UHF (e.g., TV); (ii) Cables, e.g., copper, fiber, coaxial cables and physical networks (e.g., hybrid fiber/coaxial (HFC) network); (iii) Satellites and antennas in terrestrial stations (e.g., navigation, bidirectional transmission of data, voice and video); (iv) Wireless, e.g., WiFi, unlicensed wireless, high-frequency radio, WIMAX, GSM, GPRS, EDGE, UMTS, LTE; (v) Wireline, e.g., via coaxial, copper, fiber transmission media, circuit and packet-switched

<sup>1</sup>An Roinn Slainte, Ireland Department of Health, available at <http://health.gov.ie/>

networks (e.g., private companies data, telephone networks, public switched telephone network -PSTN, the Internet) .

The telecommunication sector involves both physical and cyber infrastructures. Physical telecommunication infrastructures are composed of buildings, antennas, satellites, switch, towers, stations, cables (e.g., fiber, copper, coaxial), user terminals (e.g., PC, tablet, smartphone), among others. Cyber telecommunication infrastructures are composed of network protocols (e.g., protocol layering and software layering); network security devices (e.g., firewall, antivirus); network technologies and software e.g., Ethernet, Token ring, Asynchronous Transfer Mode (ATM), Fiber Distributed Data Interface (FDDI), Frame Relay, Synchronous Optical Networking (SONET), Multiprotocol Label Switching (MPLS), Software Defined Networking (SDN); user applications (e.g., mobile and computer applications).

This sector has few significant dependencies on other critical infrastructures, however, most of the critical infrastructures have a high dependency on telecommunications. For this reason, we can conclude that physical devices as sensors or actuators used in other critical infrastructures may be included in this category as terminals of the infrastructure. In addition, new technologies and facilities are developed daily in this sector, for example, related to The Internet of Things (IoT) as smart home technologies or smart cities.

#### B. Information about Non-Critical Infrastructures

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 [29].

The identification of methods in the system are proposed by Howard et al. [30] and further detailed by Manadhata and Wing [31]. 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.

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 [32].

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

1) *Internal Data*: : Considering the characteristics of access control models [34], [35], 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. [6], [8], [12].

In addition, we consider the notion of contexts proposed in the Organization based Access Control (OrBAC) model [36], [37], 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 login 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.

a) *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. [30] 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. [30] 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, Active Server Pages files, and 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 [38], 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 (ID-MEF) [39] 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).

*b) Physical Data::* Physical information corresponds to all tangible elements that interact directly or indirectly with the target system and whose intrinsic vulnerabilities can be used by an adversary to execute an attack. Examples of physical data are proposed by Norman [29] 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).

*2) External Data: :* 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 [29]. 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. [40], [41], 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 the remaining of this section as a priori and a posteriori data.

*a) 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) [42], [43], 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 [48], and the research proposed by Cayirci and Ghergherehchi [3], 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 [2], and Shinder [44] 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).

*b) A Posteriori Data:* : A set of information gained by the attacker after a successful exploitation of a system's vulnerability [40]. 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 [41]. 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 [45]. However, penetrating the system is the first step for an attacker to improve his/her knowledge about the system.

According to Krautsevich et al. [40], 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) [46] 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 orga-

nization's infrastructure, installing video surveillance and/or biometric systems.

#### IV. SECURITY EVENT DATA MATRIX

Based on the information presented in previous sections, we propose a matrix that organizes the event data based on the required and optional information of critical and non-critical infrastructures. Table I organizes the information of critical infrastructures (e.g., energy distribution, water supply, telecommunications, health, finance, transportation) and provides examples of the data to be used in particular systems (cyber, physical, and cyber-physical). Information labeled as *Required* is proposed for the identification of important axes of the geometrical instances, whereas information labeled as *Optional* provides additional information about the target and the security events that could be useful in the construction of the geometrical instances. Optional information is not always easy to obtain but it can be useful to develop other axes of impact for the studied events.

In Table I, it is worth noting that cables, equipment, and other physical devices belong to the physical system but the definition of physical systems also involves physical processes and the data obtained from the industrial system or environments, for example, a voltage peak, because it has repercussions in the physical systems and in the industrial processes (energy distribution, water supply, etc.). These data are normally analogical data then transformed into digital data for communicating them through the network.

Table II organizes the information of non-critical infrastructures based on three main aspects: (i) asset location (i.e., internal, external), (ii) event time (i.e., a priori, a posteriori), and (iii) event nature (i.e., physical, logical).

#### V. IMPACT ASSESSMENT USING GEOMETRICAL MODELS

This section details the process for the impact assessment of cyber security events using geometrical tools.

##### A. Data Collection

It refers to obtaining information that can be useful during the evaluation process of cyber security events. This includes information about subjects (e.g., users, customers, potential clients, competitors, providers), objects (e.g., files, devices, servers, audio, video), and the actions that subjects perform on objects (e.g., read, write, execute). Depending on the target system, the information can be obtained from multiple sources.

Examples of such sources are databases (e.g., data warehouse, cloud database, knowledge database), logs (e.g., access log, error log, warning log), access control lists (e.g., permissions, prohibitions), directories (e.g., home directory, parent directory, LDAP), repositories (e.g., local repository, SVN), traffic flow (e.g., processing delay, queuing). The type of target system is crucial in the identification of the data to be collected. For instance, if the target is a critical system for energy distribution, data must focus on technical and non-technical losses, as presented in Table I. If the target is a non-critical system, from which only internal information is available, data will focus on the logical and physical elements integrating the system, as detailed in Table II.

##### B. Data Normalization

The information is obtained in different formats and requires to be normalized for further treatment. The second step is therefore related to the data normalization process. This latter includes data organization, data cleaning, and data-type construction. The organization of data corresponds to the process of grouping elements with similarities in a data slot (e.g., all users with admin privileges are grouped in a slot called administrator). Duplicated data assigned to the same slot are deleted in the data cleaning process. Then, similar slots are merged to create data-types (e.g., internal users, system admin, super admin, guess, and external users can be all merged into a single data-type called USER).

##### C. Axes Identification and Population

Each data-type defined in the previous steps represents one side/axis of the geometrical instance. Examples of axes are given in Tables I to II which include, users (e.g., all user accounts having interaction with the information system), resources (e.g., logical and physical resources from the target system), channels (e.g., all actions performed by users to have access to resources). The more data-types we obtain from step 2, the more number of axes in the geometrical instance, and thus, the more detailed information about the impact of the security event.

Once the axes are identified, we need to populate them. This process requires to assign a weight to every data slot composing the axis, based on its contribution to the realization of a particular event. Contributions are either absolute or relative. The former considers every single element of the information system that is affected to a security event, whereas the latter considers the percentage of all affected elements. Absolute contributions are computed as proposed in Definition 1 whereas relative contributions are computed as proposed in Definition 2.

**Definition 1.** *Absolute Contribution.* Let  $X = x_1, x_2, \dots, x_i$  be a finite set of size  $i$  composed by the total number of elements of entity  $T$ ; and let  $Y = y_1, y_2, \dots, y_j$  be a finite set of size  $j$ , composed by the affected elements of entity  $T$ . Knowing that the set  $Y$  is a subset of  $X$ , thus  $Y \in X$ , then, the absolute contribution  $Co_A$  of entity  $T$  in the execution of event  $E$  is computed using Equation 1.

$$Co_A(T, E) = \sum_n^{i=1} Y_j \times WF(Y_j) \forall j \in Y \quad (1)$$

Note that  $WF$  corresponds to the weighting factor associated to each data-type.  $WF$  is a value between 1 and 10 that is assigned based on the CARVER methodology (i.e., criticality, accessibility, recuperability, vulnerability, effect, and recognizability) [29], [47].

**Definition 2.** *Relative Contribution.* Let  $X = x_1, x_2, \dots, x_i$  be a finite set of size  $i$  composed by the total number of elements integrating the entity  $T$ ; and let  $Y = y_1, y_2, \dots, y_j$  be a finite set of size  $j$  composed by the affected elements of the entity  $T$ . Knowing that the set  $Y$  is a subset of  $X$ , thus  $Y \in X$ , then,

Table I: Critical Infrastructure Systems Taxonomy

	Cyber Systems		Physical Systems	
	Required	Additional	Required	Additional
Energy Distribution	Technical losses (e.g., overhead line); non-technical losses (e.g., error); type of conductor (e.g., copper); data collection frequency (e.g., annually); reconciliation (e.g., settlement reconciliation); protocols (e.g., DNP3); service status (e.g., active).	Transformer material (e.g., iron); timing differences (e.g., absolute differences); profiling (e.g., profile coefficient); meter characteristics (e.g., identification); communication protocols (e.g., LTE); HMI (e.g., client console); switch brand (e.g., Cisco); security software (firewall).	Load (e.g., heating load); peak load times (e.g., winter); conductor temperature (e.g., Celsius); energy demand and consumption (e.g., MWh/year, KW/h); Calculation Factor (e.g., Loss Factor); power voltage (e.g., high voltage); errors; transformer characteristics (e.g., temperature, distance).	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); media type (e.g., fiber optics); security device.
Water Supply	Security logs (e.g., logs provided by firewall); communication protocols (e.g., Modbus); resources (e.g., available bandwidth); virtual distribution map (e.g., virtual district metering area).	PLC type (e.g., 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).	Inclinometer (e.g., lateral earth movements); tiltmeter (e.g., wall tilt); crackmeter (e.g., movement of cracks) jointmeter (e.g., expansion of a joint); earth pressure cell (e.g., pressure of the dams); piezometer (e.g., fluids pressure); turbidimeter; thermometer.	Sensors (e.g., WSN); 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).
Financial Services	Current assets (e.g., cash); Non-current assets (e.g., property); current and non-current liabilities (e.g., income taxes payable); equity (e.g., capital stock); income (e.g., cost of sales); currency (e.g., USD); net sales (e.g., by product type); budget (e.g., per account); cash flows (e.g., from operating activities); general cash flows (e.g., net in-crease /decrease in cash).	Organization (e.g., business unit); customer (e.g., address); vendor (e.g., type); invoice (e.g., payment date); calendar time (e.g., date); company's account category (e.g., cash); account classification (e.g., cost); account or sub-account type (e.g., salaries); transaction code (e.g., revenue); employee (e.g., name).	Technical equipment (e.g., ATM, database); physical storage (e.g., cabinets); physical documents (e.g., paper work); banking infrastructure (e.g., buildings); video surveillance (e.g., cameras).	Physical security devices (e.g., biometric devices).
Health Services	Health Services (e.g., nursing); hospital inpatient (e.g., age); primary care services (e.g., community schemes); private health insurance (e.g., people by country); disability (e.g., number of people with disability); hospital activity (e.g., emergency service); weighting list (e.g., waiting list by hospital); psychiatric data (e.g., psychiatric hospitals); hospitalized patients (e.g., emergency patients); life expectancy (e.g., healthy life years); pharmaceutical dispenses (e.g., drug dispense date); demographics (e.g., number of births).	Claims (e.g., payments); productivity (e.g., productivity of physicians in group practices); hospital statistics (e.g., name, address); codes (e.g., international classification of diseases); provider information (e.g., identifier); service additional information (e.g., date of service); service units paid and calls claimed (e.g., diagnostic); medical wireless network development.	Medicare service (e.g., medical equipment, home health); essential medical devices (e.g., Protective breathing masks); respiration medical devices (e.g., laryngoscope blade, endobronchial and endotracheal intubation tube); blood circulation devices (e.g., short peripheral catheter); digestive devices (e.g., aspiration tubing); infrastructure (e.g., emergency number infrastructure).	Robot-assisted surgical machines; personal real-time devices (e.g., real-time EKG device); emergency devices.

**Table 1 (Cont).** Critical Infrastructure Systems Taxonomy

Cyber Systems		Physical Systems		
	Required	Additional	Required	Additional
Transportation	Communication technologies (e.g., WiMAX); network architecture (e.g., public cloud) ; safety communication (e.g., protocols); security software (e.g., antivirus); storage (e.g., private data); network management (e.g. security planning).	Collaborative network (eg. between public and private network, or among railway, roadway and emergency transport networks); exchange information protocols (e.g., V2V).	Vehicles; ships; trains; infrastructures (e.g., roads) ; human drivers; sensors; motors; surveillance cameras; traffic lights; railways; security devices; safety devices (e.g., speed control).	Locals (e.g., garages); cables; devices materials: remote workstation (nodes); automatic transports (e.g., smart car); intelligent traffic control devices (e.g., traffic lights in smart cities).
	Communication system model (e.g., OSI model); transmission technologies (e.g., modulation); network protocols (e.g., protocol layering and software layering); routing and switching software (e.g., public switched telephone network -PSTN); telephone companies; Internet service providers ; network security (e.g., IDSs).	User applications, Over the Top (OTT) companies (e.g., used technologies, type of market, business).	Buildings; antennas; satellites; switch; towers; stations; hardware infrastructure (e.g., radio receiver); cables (e.g., optical fiber, copper, coaxial); user terminals (e.g., smart-phone), wireless systems (e.g., radio transmitters, radio waves, acoustic signals).	IoT (Internet of Thing) devices (e.g., smart home, smart city (multiple sensors, video cameras), connected watches).

**Table II:** Non-Critical Infrastructure Systems Taxonomy

		Logical	Physical	
Internal	Required	Additional	Required	Additional
	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).	Vehicles; ships; trains; infrastructures (e.g., roads) ; human drivers; sensors; motors; surveillance cameras; traffic lights; railways; security devices; safety devices (e.g., speed control).	Locals (e.g., garages); cables; devices materials: remote workstation (nodes); automatic transports (e.g., smart car); intelligent traffic control devices (e.g., traffic lights in smart cities).
External	A Priori	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).	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).
	A Posteriori	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.	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.

the contribution  $Co_R$  of entity  $T$  in the execution of event  $E$  is computed using Equation 2.

$$Co_R(T, E) = \frac{\sum_{j=1}^n Y_j \times WF(Y_j) \forall j \in Y}{\sum_{i=1}^n X_i \times WF(X_i) \forall i \in X} \quad (2)$$

Applying Equation 2 in a practical case, we evaluate, for instance, the contribution of the user account axis as the number of users affected by a given attack over the total number of active users in the system. For special contexts, we can evaluate the number of incidents occurring in a given location over the total number of reported incidents within a period of time.

#### D. Event Definition and Graphical Representation

The fourth step in the impact assessment of cyber security events corresponds to the definition of malicious and benign events along with their graphical representation. This process is limited to known events. The definition of unknown events is out of the scope of this research work.

Based on the axes modeled in the previous step and taking into account the system's vulnerabilities, a list of possible malicious actions against the system are defined. These actions include threats, errors, faults, natural disasters, and attacks. In order to mitigate malicious actions, countermeasures are also defined. They consider the system's topology, security mechanisms, access control rules, policy enforcement points, capabilities of current security equipment, and conflicts that may arise due to their implementation.

The definition of events considers not only the number of elements affected in each axis but also its corresponding weighting factor. The projection of each axis results into a geometrical representation of the event. Taking for instance, the 3D representation presented in Section II, we need to define at least three main axes: user accounts (e.g., internal user, admin, guest), resources (e.g., PCs, files, printers), and channels (e.g., IP addresses, credentials, port numbers). The projection of all the axes in a three-dimensional coordinate system results into a parallelepiped. Each event is therefore transformed into a geometrical instance, whose size indicates its impact on the system.

As stated by Kolomeec et al. [17], visualization models have different advantages that depend on the metrics used in the model construction and the context in which the model is used. In addition, it is important to note that there is no universal visualization model that can represent all stages of risk analysis process. It is therefore important to define the model that best captures the different data-types obtained in the studied scenario.

Table III compares the visualization models discussed in Section II based on the following properties: P1 (allows mono-dimensional representation of events), P2 (allows bi-dimensional representation of events), P3 (allows three-dimensional representation of events), P4 (allows representation of events in more than three dimensions), P5 (considers contextual data in the impact representation), P6 (considers attacker's data in the impact representation), P7 (requires the

Table III: Comparison of Visualization Models

Properties	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Polygonal Model		✓	✓	✓	✓		✓			
3D Model			✓					✓	✓	
Polytope Model	✓	✓	✓		✓			✓	✓	
Prismatic Model			✓	✓	✓	✓	✓			
Graph Model		✓	✓		✓	✓		✓	✓	✓

calculation of relative contributions), P8 (requires the calculation of absolute contributions), P9 (considers dependencies in the impact representation), and P10 (considers high abstraction level of events).

As seen in Table III each model has features that are considered to select the visualization model that best adapts to the system and attack scenario. For instance, if the selected model is the 3D-based, we must be able to identify three axes to be represented in the coordinate system, and we will consider dependencies among entities while computing their contribution to the studied event. If we identify more than three axes, we can either switch to another model (e.g., polygonal, prismatic) or merge several data-types in order to obtain the three axes to be modeled in the coordinate system. The choice will depend on the security analyst and on the case study to be developed.

#### E. Impact Computation and Analysis

This step corresponds to the computation and analysis of the impact of cyber security events. Such impact measures the size of the geometrical instance based on geometrical operations (e.g., length, area, volume, hyper-volume). It is then possible to compare multiple events and to propose countermeasures accordingly. Examples of the different measurements propose to compute the impact of cyber events are given in Table IV.

Results are expressed in  $units^n$ ,  $n$  being the number of axes that contain the affected elements. Comparing the impact of two events based on their size is possible, as long as they are based in the same scale of units. For instance, if attack  $A_1$  is a cube, whose size is expressed in  $units^3$ , and attack  $A_2$  is a triangle whose size is expressed in  $units^2$ , we must transform both attacks measurements into the same unit system (e.g., compute the area of  $A_1$ ) before analyzing them.

By quantifying the size of cyber security events it is possible to: (i) compute the impact of individual and multiple events; (ii) analyze simultaneous events and assign priorities to those with higher impact; (iii) set up reaction strategies that may include the selection of multiple countermeasures against a given attack scenario; (iv) compute the union and intersection of multiple events, (v) accurately identify the size of the system under attack, the residual risk, and the potential collateral damage, as described in [7].

### VI. USE CASE: COMPUTING THE IMPACT OF EVENTS IN CRITICAL INFRASTRUCTURES

This section provides a case study of a railway infrastructure with multiple attacks and countermeasures. We define two attacks and four countermeasures to be modeled and analyzed using a geometrical tool for impact assessment.

Table IV: Examples of Geometrical Operations

Geometrical Instance/ Operation	Length $L$ ( $unit$ )	Area $A$ ( $unit^2$ )	Volume $V$ ( $unit^3$ )	Hyper-volume $H$ ( $unit^n$ )
Line	$L$	-	-	-
Equilateral Triangle	$3 \times L$	$\frac{1}{2}b \times h$	-	-
Isosceles Triangle	$2 \times L_1 + L_2$	$\frac{1}{2}b \times h$	-	-
Scalene Triangle	$L_1 + L_2 + L_3$	$\frac{1}{2}b \times h$	-	-
Square	$4L$	$L^2$	-	-
Kite	$2L_1 + 2L_2$	$L_1 \times L_2$	-	-
Rhomboid	$L_1 + L_2 + L_3 + L_4$	$b \times h$	-	-
Trapezoid	$L_1 + L_2 + L_3 + L_4$	$\frac{1}{2}b_1 \times b_2 \times h$	-	-
Rectangle	$2L_1 + 2L_2$	$L_1 \times L_2$	-	-
Equilateral Polygon	$n \times L$	$\frac{1}{4}n \times L^2 \cot(\frac{\pi}{n})$	-	-
Irregular Polygon	$L_1 + \dots + L_n$	$\frac{1}{2} \sum_{i=1}^{n-1} (X_{i+1} \times Y_i - (Y_{i+1} \times X_i))$	-	-
Cube	$12L$	$6L^2$	$L^3$	$L^3$
Parallelepiped	$4(L_1 + L_2 + L_3)$	$2(L_1 \times L_2) + 2(L_1 \times L_3) + 2(L_2 \times L_3)$	$L_1 \times L_2 \times L_3$	-
Regular Prism	$2(n \times L) + n \times h$	$(2 \times A) + S$	$A \times h$	-
Irregular Prism	$2 \times (L_1 + \dots + L_n) + n \times h$	$n \times L(a + h)$	$A \times h$	-
Hyper-cube	$En \times L$	$En \times L^2$	$Cn \times L^3$	$L^n$
n-orthotope	$4(L_1 + L_2 + \dots + L_n)$	$2(L_1 \times L_2 + L_1 \times L_3 + \dots + L_1 \times L_n + L_2 \times L_3 + \dots + L_{n-1} \times L_n)$	$L_1 \times L_2 \times L_3 + L_1 \times L_2 \times L_4 + \dots + L_1 \times L_2 \times L_n + L_2 \times L_3 \times L_4 + \dots + L_{n-2} \times L_{n-1} \times L_n$	$L_1 \times L_2 \times \dots \times L_n$

$L$ : side length;  $b$ : base;  $h$ : height;  $n$ : number of sides;  $a$ : apothem length;  $A$ : base area;  $S$ : lateral surface area;  $X_i, Y_i$ : Cartesian coordinate data pairs;  $En$ : hyper-cube edges;  $Cn$ : hyper-cube faces;  $Fn$ : hyper-cube cells;  $En$ : hyper-cube edges

## A. General Description

Three main block elements take part in the infrastructure: the controller, the system itself, and the sensors. The controller reads data from the sensors, computes new information and transmits new commands to the system (i.e., the system control input). This latter is generated by the controller with the purpose of correcting the behavior of the system, under some previously established limits. The system is the entity under control. The sensors are the feedback links between the system and the controller. Their purpose is to quantify the output and provide the necessary information to the controller, in order to compare and, if necessary, correct the behavior of the system.

All the aforementioned elements are distributed across several nodes in a shared network combining Distributed Network Protocols (DNP3) and Modbus protocols. From a software standpoint, the controller never connects directly to the sensors. Instead, it is integrated in the architecture as a SCADA PLC node, with eventual connections to some other intermediary nodes. Such nodes are able to translate the controller commands into SCADA commands (e.g., Modbus, DNP3). The architecture is able to handle several industrial protocols and connect to complementary SCADA elements, such as additional PLCs and RTUs.

The implementation of the SCADA testbed consists on *Lego Mindstorms* EV3 bricks [50] and Raspberry Pi [51] boards as PLCs to control some representative sensors (e.g., distance sensors) and actuators (e.g., speed actuators). We refer the reader to <http://j.mp/legoscada> for additional information and video captures of the testbed.

Figure 7 shows the components of the autonomous industrial agents composing the use case. It consists of a controller (Personal Computer), an RTU (Raspberry Pi) and a PLC (Lego EV3 Brick). The controller corrects the speed of the car by polling the distance between the car and an obstacle. One single controller and one single RTU can control various PLCs. To start the testbed, it is necessary to execute the Java automaton deployed over each EV3 bricks [52], as well as the automata deployed over the Raspberry Pi boards. Once started, the controller verifies and controls the dynamics of the car, i.e., the car behavior is continually modified by the controller, hence varying its speed according to the controller's commands.

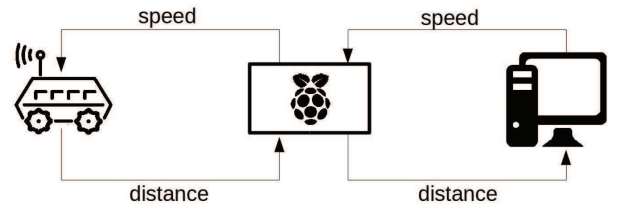


Figure 7: Railway Use Case Component Overview

In order to represent the impact of attacks and countermeasures as geometrical instances, we use the attack volume model [8] which identifies three main dimensions of information (i.e., users, channels, and resources). With the help of the event data taxonomy proposed in Section III, we have identified entities



to each dimension composing our model. Each entity has an associated weighting factor (assigned based on the CARVER methodology) which allows to define its coordinate in the system.

Table V: Entities Information

Dimension	Range	Description	N	WF	Coord.
Resource	R1	Server	1	3	0:3
	R2	Router	1	3	3:6
	R3:R4	RTU	2	3	6:12
	R5	Front End	1	2	12:14
	R6	Database	1	3	14:17
	R7:R8	Controller	2	4	17:25
	R9:R10	Workstation	2	4	25:33
	R11:R13	Infrastructure PLC	3	4	33:45
	R14	Back End	1	4	45:49
	R15:R16	Train PLC	2	4	49:59
	R17:R24	Actuator	8	5	59:99
	R25:R33	Firewall	9	5	99:144
	R34:R43	Sensor	10	5	144:194
Channel	Ch1:Ch20	UDP Port	20	2	0:40
	Ch21:Ch29	User Credentials	9	3	40:67
	Ch30:Ch283	Private IP	254	3	67:829
	Ch284:Ch288	Public IP	5	4	829:849
	Ch289:Ch318	TCP Port	30	4	849:969
	Ch319	DNP3	1	4	969:973
	Ch320	Modbus	1	4	973:977
	Ch321	Admin Credentials	1	5	977:982
User	U1:U2	External	2	1	0:2
	U3:U10	Std. Operator	8	2	2:18
	U11:U13	Business Owner	3	3	18:27
	U14	Op. Stakeholder	1	3	27:30
	U15:U18	SCADA Operator	4	4	30:46
	U19	System Admin	1	5	46:51

Table V summarizes this information by providing the list of affected entities for each category (*range*), the type of entity (*description*), the total number of entities per category (*N*), the weighting factor value assigned to each category of entities (*WF*), and the initial and final position of the entity within the coordinate system (*Coordinate*).

### B. Attack Scenarios and Countermeasures

This section describes the events (i.e., attacks and countermeasures) to be modeled and analyzed in the railway infrastructure scenario.

1) *Attack Scenarios*: We assume that attackers are able to intercept any communication exchanges between ends, and thus can alter, store, analyze, replay and forge false data from and towards the communication channels. Address Resoulution Protocol (ARP) poisoning [53] is used by the attacker to intercept the channels and eavesdrop the communications. The attacker has a passive and active mode of operation. During the *passive mode*, the attacker eavesdrops, processes, and analyzes the data without modifying the information contained in the payload of the messages. During the *active mode*, the attacker injects data to the hijacked communication. Two attacks are considered for modeling and analysis: (i) Replay attack, and (ii) Injection Attack.

**Replay Attack**: The attackers use ARP poisoning to start eavesdropping the connection (passive mode, from the physical-layer standpoint). After capturing enough data, the *active mode* starts. The attackers inject the old captured data

following the stream of packets of the previous capture. Before starting to disrupt the system, the attacker conducts the attack between the sensors and the controller, forging only the TCP headers that correspond to the opened TCP sessions. Once replayed the packets, the system gets disrupted by forging data between the controller and the PLCs.

**Injection Attack**: After the attacker eavesdrops connections using the *physical-layer passive mode*, and analyzes the data in order to infer the dynamics of the system, he/she injects correct data in the communication channel. To evade the detector, the attacker calculates the effect of the system's countermeasures and cancels the ability of the detector to send the changes in the feedback signal. Two different techniques are implemented: 1) a non-parametric adaptive filter (e.g., a *non-parametric cyber-physical attack*); and 2) autoregressive methods (e.g., ARX and ARMAX, a *parametric cyber-physical attack*).

2) *Countermeasures*:: Security actions to protect a railway infrastructure from replay and/or inject attacks are based on improving the detection mechanisms. Examples of such countermeasures include the following:

- C.1 Watermark-based detection by adapting traditional failure detection mechanisms in order to detect replay attacks. (replay attack)
- C.2 Signal-based and model-based intrusion detection used to detect and identify random signal data-injections attacks using Wavelet analysis. (data injection)
- C.3 State relation-based detection using correlation of system states together with system behavior to identify anomalies. (data injection)
- C.4 Message counter used in the MODBUS/TCP messages in order to increase the communication security against replay attacks. (replay attack)

It is worth noting that other countermeasures can be also used to mitigate each of the aforementioned attacks. However, for demonstration purposes, we have selected only two countermeasures for each attack. Their impact values and graphical representations are presented in the remaining of this section.

### C. Impact Computation and Graphical Representation

Based on statistical data and expert knowledge, we have identified the affected entities for each Attack and countermeasure. The impact of such events is represented as a parallelepiped that covers a set of user accounts, channels and resources from the system  $S$ . Table VI summarizes this information.

In order to compute the volume of each geometrical instance, we refer to Table IV. The graphical representation of each attack and countermeasure with respect to the system is depicted in Figure 8. Each attack is represented as the union of multiple parallelepipeds that cover affected users, channels and resources from the system. Attack  $A_1$  has a volume of  $80,400 \text{ units}^3$ , whereas attack  $A_2$  has a volume of  $455,469 \text{ units}^3$ . The impact of the injection attack is more than five times bigger than the impact of the reply attack.

For each attack, a set of countermeasures have been evaluated. In this use case, we have identified two countermeasures

Table VI: System, Attack and Countermeasure Data

	Resource	Channel	User	Volume (units <sup>3</sup> )
S	R1:R43	Ch1:Ch321	U1:U19	8,718,748
A <sub>1</sub>	R2:R3,	Ch1:Ch3,	U3:U5,	80,400
	R9:R11,	Ch30:Ch35,	U11:U12,	
	R14:R18,	Ch284:Ch285,	U15:U16	
	R25:R27,	Ch289:Ch293,		
A <sub>2</sub>	R34:R35	Ch319:Ch320		455,469
	R2:R4,	Ch1:Ch3,	U3:U13,	
	R7:R28,	Ch30:Ch38,	U15:U18	
	R34:R43	Ch284:Ch285,		
C <sub>1</sub>		Ch289:Ch293,		148,500
		Ch319:Ch320		
	R9:R10,	Ch30:Ch38,	U11:U13,	
	R15:R24,	Ch289:Ch293,	U15:U18	
C <sub>2</sub>	R34:R43	Ch319, Ch320		173,900
	R7:R13,	Ch30:Ch38,	U11:U13,	
	R15:R28,	Ch289:Ch293	U15:U18	
	R34:R43			
C <sub>3</sub>	R7:R28,	Ch30:Ch38,	U3:U13,	392,616
	R34:R43	Ch284:Ch285,	U15:U18	
		Ch289:Ch293,		
		Ch319:Ch320		
C <sub>4</sub>	R2:R4,	Ch1:Ch20,	U11:U13,	943,850
	R11:R13,	Ch30:Ch150,	U15:U18	
	R25:R27,	Ch284:Ch285,		
	R34:R43	Ch289:Ch293,		
		Ch319:Ch320		

against each attack. Each countermeasure affects a set of users, resources and channels, whose volume range from 148,500 units<sup>3</sup> to 943,850 units<sup>3</sup> as shown in Figure 8(c) and 8(d).

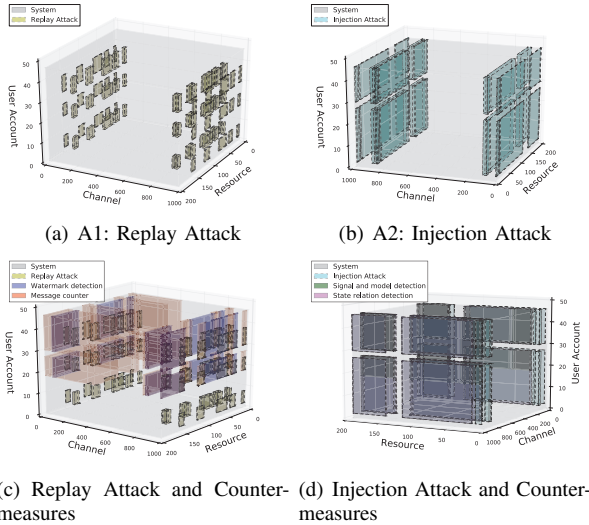


Figure 8: Visualization of Events as Polygons

## VII. 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. Shameli-Sendi et al. [1], for instance, propose a taxonomy of information security risk assessment; Noureldien [32] propose a taxonomy of MANET attacks; Zhu et al. [49] propose a taxonomy of cyber attacks on SCADA systems; and Teixeira et al. [54] propose a taxonomy of cyber attacks on cyber-physical systems taking into account the adversary's knowledge and resources. 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 [29] and Abbas et al. [33]. 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 [48], 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. [30] 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. [31] to compare different software systems based on entry points, methods, and channels. More recently, Gonzalez-Granadillo et al. [7], [10] 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. Examples of such formats include the Intrusion Detection Message Exchange Format the Incident Object Description Exchange Format ( and the Syslog protocol [38]. In addition, we consider the Common Vulnerability Scoring System (CVSS) [43] to assess the impact of security vulnerabilities.

## VIII. CONCLUSION

In this paper we have proposed an event data taxonomy to represent the impact of cyber security events as geometrical instances. We review several visualization models for countermeasure selection and we discuss about the identification of the type of information required to feed the model and plot the results. Results show that the appropriate identification and classification of the data associated to cyber security events is crucial in the analysis and selection of security countermeasures.

## REFERENCES

- [1] A. Shameli-Sendi, R. Aghababaei-Barzegar, M. Cheriet, "Taxonomy of Information Security Risk Assessment (ISRA)", *Computer & Security*, vol. 57, pp. 14-30, 2015.
- [2] M. Bielecki and G. Quirchmayr, *A prototype for support of computer forensic analysis combined with the expected knowledge level of an attacker to more efficiently achieve investigation results*. International Conference on Availability, Reliability and Security, pp. 696-701, 2010.
- [3] E. Cayirci, R. Ghergherehchi, "Modeling cyber attacks and their effects on decision process", *Winter Simulation Conference*, 2011.
- [4] I. Kutenko, A. Chechulin *A Cyber Attack Modeling and Impact Assessment Framework*, 5th International Conference on Cyber Conflict, 2013.
- [5] I. Kutenko, E. Doynikova, "Countermeasure Selection in SIEM systems based on the integrated complex of security metrics", *Conference on Parallel, Distributed, and Network-Based Processing*, 2015.
- [6] G. Gonzalez-Granadillo, J. Garcia-Alfaro, H. Debar, "Using a 3D geometrical model to improve accuracy in the evaluation and selection of countermeasures against complex cyber attacks", *Conference on Security and Privacy in Communication Networks*, pp. 538-555, 2015.
- [7] G. Gonzalez-Granadillo, J. Rubio-Hernan, J. Garcia-Alfaro, H. Debar, "Considering internal vulnerabilities and the attacker's knowledge to model the impact of cyber events as geometrical prisms", *Conference on Trust, Security and Privacy in Computing and Communications*, 2016.
- [8] G. Gonzalez-Granadillo, G. Jacob, and H. Debar, "Attack Volume Model: Geometrical Approach and Application", *10th International Conference on Risks and Security of Internet and Systems (Crisis)*, pp. 242-257, 2015.
- [9] G. Gonzalez-Granadillo, J. Garcia-Alfaro, and H. Debar, "An n-sided polygonal model to calculate the impact of cyber security events", *11th Conference on Risks and Security of Internet and Systems*, 2016.
- [10] G. Gonzalez-Granadillo, J. Garcia-Alfaro, H. Debar, "A Polytope-based approach to measure the impact of events against critical infrastructures", *Journal of Computer and System Sciences*, Vol. 83(1), pp. 3-21, 2016.
- [11] G. Gonzalez-Granadillo, M. Belhaouane, H. Debar, G. Jacob, "RORI-based countermeasure selection using the OrBAC formalism", *International Journal of Information Security*, Vol. 13(1), pp. 63-79, 2014.
- [12] G. Gonzalez-Granadillo, J. Garcia-Alfaro, E. Alvarez, M. El-Barbori, H. Debar, "Selecting optimal countermeasures for attacks against critical systems using the Attack Volume model and the RORI index", *Computers and Electrical Engineering*, Vol. 47, pp. 13-34, 2015.
- [13] G. Gonzalez-Granadillo, A. Motzek, J. Garcia-Alfaro, H. Debar, "Selection of Mitigation Actions Based on Financial and Operational Impact Assessments", *Conference on Availability, Reliability and Security*, 2016.
- [14] M. Kolomeets, A. Chechulin, I. Kutenko, "Visualization Model for Monitoring of Computer Networks Security Based on the Analogue of Voronoi Diagrams", *Cross-Domain Conference, and Workshop on Privacy Aware Machine Learning for Health Data Science*, 2016.
- [15] E.V. Doynikova, I.V. Kutenko, "Countermeasure Selection Based on the Attack and Service Dependency Graphs for Security Incident Management", *International Conference on Risks and Security of Internet and Systems*, vol. 9572, Springer, pp. 107-124, 2016.
- [16] M.V. Kolomeec, A.A. Chechulin, I.V. Kutenko, "Methodological Primitives for Phased Construction of Data Visualization Models", *Journal of Internet Services and Information Security*, Vol.5(4), pp.60-84, 2015.
- [17] M. Kolomeec, G. Gonzalez-Granadillo, E. Doynikova, A. Chechulin, I. Kutenko, H. Debar, *Choosing Models for Security Metrics Visualization*, 7th International Conference on Mathematical Methods, Models and Architectures for Computer Networks Security MMM-ACNS, 2017.
- [18] K. Gordon, M. Dion, *Protection of Critical Infrastructure and the role of investment policies relating to National Security* Whitepaper, 2008.
- [19] R. Rajkumar, I. Lee, L. Sha and J. Stankovic, *Cyber-physical systems: The next computing revolution*, Design Automation Conference, Anaheim, CA, 2010, pp. 731-736. doi: 10.1145/1837274.1837461
- [20] B. Genge, I. Kiss, and P. Haller *A system dynamics approach for assessing the impact of cyber attacks on critical infrastructures* International Journal of Critical Infrastructure Protection, vol. 10, pp. 3-17, 2015.
- [21] Sohn Associates, "Electricity Distribution System Losses", *Non Technical Overview*, Whitepaper, 2009.
- [22] L. Coppolino, S. D'Antonio, V. Formicola L. Romano, "Integration of a System for Critical Infrastructure Protection with the OSSIM SIEM Platform: A dam case study", *International Conference on Computer Safety, Reliability, and Security SAFECOMP*, pp. 199-212, 2011.
- [23] Public Utilities Board Singapore *Managing the water distribution network with a Smart Water Grid*, International Journal for @qua - Smart ICT for Water (Smart Water), vol. 1(4), pp. 1-13, 2016.
- [24] D.B. Rawat, C. Bajracharya, G. Yan, *Towards intelligent transportation Cyber-Physical Systems: Real-time computing and communications perspectives*, In SoutheastCon 2015. IEEE, pp. 1-6, 2015.
- [25] J. P. Sammon, and R. J. Caverly, *Transportation Systems: Critical Infrastructure and Key Resources Sector-Specific Plan as Input to the National Infrastructure Protection Plan*, Homeland security dept., 2007.
- [26] D. Turner, M. Schroeck, R. Shockley, Analytics, "The real-world use of big data in financial services", *Executive Report IBM Institute for Business Value in collaboration with Saïd Business School*, 2013.
- [27] L. Sha, S. Gopalakrishnan, X. Liu and Q. Wang, *Cyber-Physical Systems: A New Frontier*, International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing, pp. 1-9, 2008.
- [28] K. Condello, A. Ozment and C. Durkovich, *Communications Sector-Specific Plan, An Annex to the NIPP 2013* Homeland security dept., 2015.
- [29] T. L. Norman, "Risk Analysis and Security Countermeasure Selection", *CRC Press, Taylor & Francis Group*, 2010.
- [30] M. Howard, J. Pincus, J.M. Wing, "Measuring Relative Attack Surfaces", *Computer Security in the 21st Century*, pp. 109-137, 2005.
- [31] P.K. Manadhata, J.M. Wing, "An Attack Surface Metric", *IEEE Transactions on Software Engineering*, 2010.
- [32] A. Noureldien, "A Novel Taxonomy of MANET Attacks", *Conference on Electrical and Information Technologies ICEIT*, 2015.
- [33] A. Abbas, A.E. Saddik, A. Miri, "A comprehensive approach to designing Internet security taxonomy", *Proceedings of the Canadian Conference on Electrical and Computer Engineering*, pp. 1316-1319, 2006.
- [34] A. A. E. Kalam, R. E. Baida, P. Balbiani, S. Benferhat, F. Cuppens, Y. Deswarte, A. Mieke, C. Saurel, and G. Trouessin, *Organization based access control*. International Workshop on Policies for Distributed Systems and Networks, 2003.
- [35] N. Li and M. Tripunitara, *Security analysis in role-based access control*. Transactions on Information and System Security, 9(4):391-420, 2006.
- [36] F. Cuppens, N. Cuppens-Bouahia, and A. Mieke, *Modelling contexts in the or-bac model*. Computer Security Applications Conference, 2003.
- [37] F. Cuppens and N. Cuppens-Bouahia, *Modeling contextual security policies*. Int. Journal of Information Security, 7(4):285-305, 2008.
- [38] R. Gerhards, A. GmbH, "The Syslog Protocol", *Network Working Group*, 2009.
- [39] H. Debar, D. Curry, B. Feinstein, "The Intrusion Detection Message Exchange Format (IDMEF)", *RFC4765*, 2007.
- [40] L. Krautsevich, F. Martinelli, and A. Yautsiukhin, *Towards modelling adaptive attacker's behaviour*. FPS Symposium, pp. 357-364, 2013.
- [41] L. Krautsevich, A. Yautsiukhin, *Evaluation of Adaptive Attacker Models*. ESSoS Doctoral Symposium, 2014.
- [42] P. Mell, K. Scarfone, S. Romanosky, "Common Vulnerability Scoring System Version 2.0", *Specification Document*, June 2007.
- [43] Forum of Incident Response and Security Teams, *Common Vulnerability Scoring System v3.0: Specification Document*. Technical Paper, 2015.
- [44] D. Shinder, *Scenes of the cybercrime*. Computer forensics handbook. Syngress Publishing Inc., 2002.
- [45] M. Libicki, *Brandishing Cyberattack Capabilities*. National Defense Research Institute, white paper, 2013.
- [46] R. Danyliw, J. Meijer, Y. Demchenko, "The Incident Object Description Exchange Format (IODEF)", *RFC5070*, December 2007.
- [47] Federation of American Scientists, "Special Operations Forces Intelligence and Electronic Warfare Operations", *Technical report, Appendix D: Target Analysis Process*, Available at: <http://www.fas.org/irp/dodd/irp/army/fm34-36/toc.htm>, 1991.
- [48] K. Harrison, G. White, "A Taxonomy of Cyber Events Affecting Communities", *Proceedings of the 44th Hawaii International Conference on System Sciences*, 2011.
- [49] B. Zhu, A. Joseph, S. Sastry, "A Taxonomy of Cyber Attacks on SCADA Systems", *IEEE International Conferences on Internet of Things, and Cyber, Physical and Social Computing*, 2011.
- [50] R. Mark *Beginning LEGO MINDSTORMS EV3*, Apress, 2014.
- [51] S. S. Lagu, S. B. Deshmukh *Raspberry Pi for Automation of Water Treatment Plant*, International Conference on Computing Communication Control and Automation (ICCUBEA), pp. 532-536, 2015.
- [52] D. Wimberger, J. Charlton *Java Modbus Library*, <http://jamod.sourceforge.net>, 2004.
- [53] S. Y. Nam, D. Kim, J. Kim *Enhanced ARP: Preventing ARP poisoning-based man-in-the-middle attacks*, IEEE Communications Letters, vol. 14(2), pp. 187-189, 2010.
- [54] A. Teixeira, I. Shames, H. Sandberg, K. H. Johansson. *A secure control framework for resource-limited adversaries*. Automatica, vol. 51, pp. 135-148, 2015.