17th International Conference on Environmental Science and Technology

Athens, Greece, 1 to 4 September 2021

An Agent-Based Modelling approach to assess risk in Cyber-Physical Systems (CPS)

Koutiva I.[1,*], Moraitis G.[1] and Makropoulos C.[1]

[1] Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical Univ. of Athens, Heroon Polytechneiou 5, Zografou GR-15780, Greece

*corresponding author:
e-mail: ikoutiva@mail.ntua.gr

Abstract. The classic approach to risk assessment in civil engineering infrastructure (incl. water systems) often takes an incomplete view of the socio-technical system and its cyber-physical extensions, thus confining the ability to properly quantify the level of risk. To tackle this limitation and enhance the water sector's preparedness, this work proposes the use of Agent-Based Modelling (ABM) to explore and derive alternative routes to quantify risk. ABM approaches carry the capacity to describe systems of complex adaptive nature that characterize behavioural rules (e.g., selection of target), socio-technical systems (e.g., water systems including human stakeholders), and their "real-world" interaction. This work takes advantage of those capabilities, to quantify risks through a generic, case independent approach where the cyber-physical attackers are treated as independent, autonomous agents (e.g. hackers, saboteurs) that follow various behavioural rules to decide on their targets and plan their attacks, and hence interact with an external environment simulating the critical nodes of water critical infrastructure (e.g. storage tanks, pumping stations). The ABM simulations can be used to provide the data sets required to derive probabilities for the cyber-physical events that allow the quantification of the risk in accordance with the classic approach to risk in infrastructure planning and natural hazards.

Keywords: Agent-Based Modelling, Water Security, Risk Management, Cyber-Physical Systems, CPRISK

1. Introduction

Urban water systems comprise of multiple technical, environmental, and social vectors which form complex adaptive systems, governed by goal-directed behaviours (Koutiva and Makropoulos, 2016). The digital transformation of water infrastructures and operations expanded (and continues to expand) the nexus between water systems and ICT, leading to more complex socio-technical schemes and dynamics, along with the additional exposure of water systems to threats of the cyber domain. Cyber-attack vectors exploit the intertwined cyber and physical layer operations and introduce new tactics, techniques and processes (TTPs) to infringe upon and compromise cyber-physical systems (CPS). Thus, further uncertainty is induced over the technical and behavioural mechanisms that characterise the modern cyber-physical water systems and their security. As a result, the limits of existing practices are challenged and risk assessors are invited to re-think the urban water systems under a more holistic, cyber-physical resilience view, aided by novel approaches and tools (Makropoulos and Savíc, 2019).

Customarily, risk approaches in infrastructure planning and natural hazards express the risk level as a combination of (i) the potential consequences of threat events, and (ii) the probability of occurrence (ISO, 2018; NIST, 2012). For the first component of risk level, various studies have been developed that aim to, inter alia, properly model the combined cyber-physical systems under stress (e.g. Nikolopoulos et al., 2020; Taormina et al., 2019) and quantitatively assess the potential consequences (e.g. Moraitis et al., 2020). The second element, that of threat likelihood, is a product of expert judgement and statistical analysis of event data, which are often incomplete, biased or debased (Florêncio and Herley, 2013; Wangen, 2019). In addition to data quality limitations, statistical analysis approaches lack the capacity to typify the purposive behaviour of malevolent human actions which are driven by motives, circumstantial opportunities, as well as available skills, resources and intelligence (Vidalis and Jones, 2005).

The PROCRUSTES approach tries to address the needs of contemporary water systems and provides a novel cyber-physical view of water system resilience under uncertainty (Moraitis et al., 2021). This paper presents the CPRISK ABM tool, created within the PROCRUSTES approach, which aims to render the behaviour of potential threat agents against key cyber assets with specific characteristics, as well as the effects a utility's actions can have on the threat landscape parameters. Overall, the model simulates the opportunities and vulnerability conditions exploited by threat actors to perform cyber-attacks, based on their preferences, skills and motives.

2. Method and ABM design

The CPRISK ABM utilizes a prey-predator approach or red team / blue team approach in cyber security language. The red team aims to compromise the water utility assets by using tactics, techniques and processes of real-world adversaries. On the other hand, the water utility applies protection measures and uses state-of-the-art cyber security practices to defend its assets and services.

The red team is comprised of independent, autonomous, moving agents, called *Attackers*, that follow various behavioural rules to interact with the critical cyber nodes of a system. Various taxonomies and parameters exist to group threat actors which can be used as a starting point, with appropriate tailoring (NIST, 2012). The attributes of resources, skills and intelligence are the principal enablers of an attack, under different motives and techniques associated to the threat agent profile. The CPRISK ABM Attackers are categorised in (i) amateurs, that have limited resources, mapping a lower expertise, and opportunities to support a successful attack, (ii) experts, that are experienced and skilled adversaries with access to resources and (iii) highly skilled– typically nation state affiliated (aliased Super). The latter is a profile with increased access to resources and intelligence, motivated to perform sophisticated attacks to key assets of a CPS. Thus, it describes a threat agent profile able to pursue and exploit zero-day vulnerabilities (Tuptuk et al., 2021). Each Attacker is assigned an attribute to jointly represent the range of technical skills, available resources, and access to intelligence. Both agent characteristics are assigned pseudo-randomly at the beginning of the simulation, with resources being linked to the type of the Attacker i.e. expert Attackers are assigned higher resources than amateurs. The distribution of the different types of Attackers is assigned upon the model initialization and remains the same throughout the simulation.

The blue team is comprised of independent, autonomous, static agents that are based on the ABM's grid offering protection to the cyber layer nodes, that are modelled using independent, autonomous, moving agents, called *Targets*. The Targets model the critical nodes that are susceptible to cyber-threats and represent the data communication and transmission connections of (i) sensors & (ii) actuators, as distributed field devices with lower protection protocols by design (iii) PLCs/RTUs in an intermediate level of communication and control, that are less accessible - both physically and through the implementation of firewalls, and (iv) SCADA, as the least accessible and most technically protected element of a utility's cyber layer. Hence, a cost attribute is assigned to depict the resources needed by Attackers to successfully carry out an attack against the relevant Target. Additionally, the Targets are separated to interesting or not, allowing the CPRISK ABM to model the influence that an asset's attractiveness (e.g., in terms of importance or recognizability) has in the process. The Targets are further categorized by a protection status variable which depicts whether the cyber node is protected i.e., is in a protected, private network, the highest protection protocols are implemented etc.

The CPRISK ABM initiates with a set number of agents representing Targets and Attackers, that move randomly in a predefined grid. The cells of the matrix represent the water utility (static agents). At each model step, Targets and Attackers (moving agents) take a random step. The number of agents at each time step is constant. There are three core procedures that govern the CPRISK ABM:

Attack procedure: If an Attacker meets a Target in its landing cell then they engage with two serial actions:

a. Compromise action

If the Target is unprotected, then all types of Attackers may compromise it and gain resources from this successful action. The increase in Attacker's resources represent benefits of knowledge and intelligence obtained from gaining access to the unprotected asset.

If the Target is protected only Attackers of type Expert and Super may compromise it and gain resources from this successful act. Amateur Attackers on the other hand will lose resources if they meet a protected Target. This penalty aims to capture the real-life disclosure of hackers attempts and their TTPs.

If the Target is successfully compromised by the Attacker, then the second step follows.

b. Attack action

The Attacker compares its available resources with the Target's cost. If the resource is higher than the cost then the Attacker performs the attack and gains resources and the Target is lost for the utility. Different types of Attackers are associated to different attack types.

If the resource is lower than the cost then the Attacker fails and loses resources and the Target remains compromised.

Resource procedure: An Attacker loses resources when it does not participate in a battle. If the resources reach zero, which means that the Attacker has lost many battles or that the agent has wondered without meeting a Target, then the Attacker is re-assigned resources according to the originally assigned type (i.e. dies and resurrects to satisfy the condition of a steady population of Attackers).

Protection procedure: A grid cell may offer protection to the Targets by increasing the cost of those Targets that land in a protection offering cell. There are four different scenarios of protection offering additional 0%, 10%, 20% and 30% protection, which corresponds to the percentage of the grid cells that offer protection to the Targets.

The flowcharts of Figure 1 present the core CPRISK ABM procedures from the viewpoint of the Attacker (Figure 1a) and the Target (Figure 1b). The CPRISK ABM has been created using the Mesa framework which is an Apache2 licensed agent-based modelling framework in Python. The model was influenced by two ABM models implemented in the Mesa framework: the NetLogo Wolf-Sheep Predation Model (Wilensky, 1997) and the NetLogo Virus on a Network Model (Stonedahl and Wilensky, 2008). The CPRISK ABM has a visualization user interface (UI), created using Mesa framework's own visualization capabilities, which can be used to (a) select the protection level offered by the utility agents, (b) enable access to the Darknet for the Attackers and (c) deploy honeynet deception technology, i.e. distributed honeypot network, for the blue team to gather valuable threat intelligence and protect against adversary's TTPs. The UI also presents the grid and visualizes the agents' movements, the success or not of attack processes through colour coding and charting.

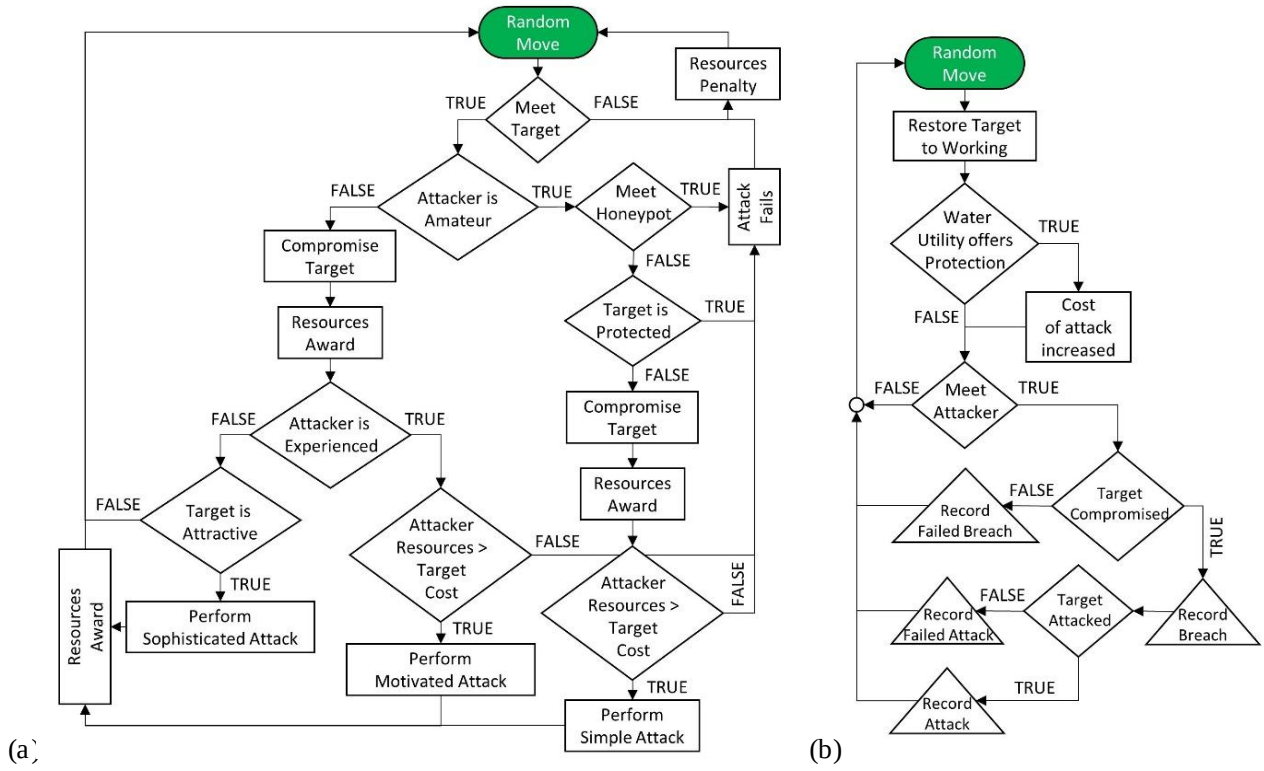

Figure 1. Generalised flowchart of a) Attacker procedure and b) Target procedure

3. Results

The CPRISK ABM tool is a generic behaviour-driven modelling approach for adversary actions, that is adjustable to case-specific CPS structures, the level of applied cyber-security and other relevant parameters, as previously discussed. The CPRISK ABM provides data to derive probabilities of attack to cyber elements of a water critical infrastructure. In view of their sensitive nature, and especially for the purposes of risk assessment processes, threat probabilities should be introduced in a semi-quantitative format (NIST, 2012).

For the purposes of this study, a simulation example is presented after 100 simulation steps. Table 1 presents the unprocessed results for a demo CPS where the generic CPRISK model contains 400 Targets [200 (50%) Sensors, 120 (30%) Actuators, 20 (5%) SCADA and 60 (15%) PLCs] and 200 Attackers [160 (80%) amateur, 30 (15%), experts, 10 (5%) supers]. The example does not include any type of proactive measures or threat intelligence for the utility, and no darknet interactions for the Attackers.

From Table 1, it is evident that a total of 71,09% of the utility's Targets is never found by any Attacker. A remaining 19,1% is found by Attackers which fails to compromise it mainly because of the Target's protection and the limited resources of amateur Attackers. Furthermore, a 1,69% of Targets gets compromised but the Attackers fail to complete the attack process against them. This leads to an 8.12% success rate of attacks of those Targets being found by attackers.

From these successful attacks, 32.88% represent a successful Simple attack performed by amateur Attackers, mainly against the less-protected, branched cyber assets of the utility. A 53.57% of the attacks is a successful Motivated attack performed by skilled Attackers, and is found almost the same to the percentage of confirmed disclosures achieved by organised criminal groups (51%) reported in Verizon's Data Breach Investigations Report (2017). The remaining 13.79% of attacks in the CPRISK ABM simulation is a successful Sophisticated attack by highly skilled adversaries against attractive assets of the utility. In comparison to the verified events reported, this ABM result captures the trend of targeted breaches conducted by state-affiliated actors that represent nearly 18% of events in the relevant database. The attack type and its characteristics could be further described according to the Attackers motives, the Target's type, and other parameters.

Table 1. Results of a CPRISK ABM simulation to assess the attack rates against a cyber-physical system (% per type action).

	Not found	Failed to compromise	Simple attack	Motivated attack	Sophisticated attack	Failed attacks
Sensors	35.63	9.48	1.36	2.22	0.57	0.75
SCADA	3.70	0.74	0.16	0.22	0.06	0.13
Actuators	21.22	5.88	0.74	1.31	0.33	0.51
PLC	10.54	3.00	0.41	0.60	0.16	0.30

4. Conclusions

This study presented an approach that simulates the key factors required by threat actors to exploit vulnerabilities and perform cyber-attacks against a CPS. The CPRISK ABM is a generic model which renders the goal-driven mechanisms that govern adversaries' behaviours against critical cyber assets of a CPS, and the effect that operator's protection actions can have on the outcome. The model builds on the underlying explicit taxonomy of entities and the core procedural relations to imbue real-life tactics, rationale and interactions to the autonomous agents' cell. The final product of CPRISK ABM provides the necessary data to infer probabilities of attack to cyber elements of a CPS, such as a water utility.

Through the showcased ABM simulation of threat actors acting against a CPS, it is evident that the CPRISK ABM approach holds the capacity to properly capture and impartially represent the broader image of the emerging cyber-physical threat landscape and yield credible risk data.

Considering the sensitive nature of those data, and partiality that may derive in risk assessment processes, the derived threat probabilities should be properly re-introduced in a semi-quantitative format.

CPRISK ABM is part of the novel PROCRUSTES approach that deals with inherent uncertainty in traditional risk assessment approaches. CPRISK ABM is a component of the process, which, coupled with cyber-physical risk analysis tools, will help derive the level of risk, as a function of probability and consequences.

Acknowledgment

The research work was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant" (Project Number: HFRI-FM17-2918).

References

Florêncio, D., Herley, C., 2013. Sex, Lies and Cyber-Crime Surveys, in: Schneier, B. (Ed.), Economics of Information Security and Privacy III. Springer New York, New York, NY, pp. 35–53. https://doi.org/10.1007/978-1-4614-1981-5_3

ISO, 2018. ISO 31000 Risk management - Principles and guidelines. Int. Organ. Stand. 34.

Koutiva, I., Makropoulos, C., 2016. Modelling domestic water demand: An agent based approach. Environ. Model. Softw. 79, 35–54. https://doi.org/10.1016/j.envsoft.2016.01.005

Koutiva, I., Makropoulos, C., 2011. Towards Adaptive Water Resources Management: Simulating The Complete Socio-Technical System Through Computational Intelligence, in: Proceedings of the 12th International Conference on Environmental Science and Technology. pp. A998–A1006.

Makropoulos, C., Savíc, D.A., 2019. Urban hydroinformatics: Past, present and future. Water (Switzerland) 11. https://doi.org/10.3390/w11101959

Moraitis, G., Nikolopoulos, D., Bouziotas, D., Lykou, A., Karavokiros, G., Makropoulos, C., 2020. Quantifying Failure for Critical Water Infrastructures under Cyber-Physical Threats. J. Environ. Eng. 146, 04020108. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001765

Moraitis, G., Nikolopoulos, D., Koutiva, I., Tsoukalas, I., Karavokyros, G., Makropoulos, C., 2021. The PROCRUSTES testbed: tackling cyber-physical risk for water systems, in: EGU General Assembly 2021. pp. EGU21-14903. https://doi.org/https://doi.org/10.5194/egusphere-egu21-14903

Nikolopoulos, D., Moraitis, G., Bouziotas, D., Lykou, A., Karavokiros, G., Makropoulos, C., 2020. RISKNOUGHT: Stress-testing platform for cyber-physical water distribution networks HS5.2.3-Water resources policy and management: digital water and interconnected urban infrastructure. https://doi.org/10.5194/egusphere-egu2020-19647

NIST, 2012. Guide for conducting risk assessments. NIST Spec. Publ. 95. https://doi.org/10.6028/NIST.SP.800-30r1

Stonedahl, F. and Wilensky, U. (2008). NetLogo Virus on a Network model. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Taormina, R., Galelli, S., Douglas, H.C., Tippenhauer, N.O., Salomons, E., Ostfeld, A., 2019. A toolbox for assessing the impacts of cyber-physical attacks on water distribution systems. Environ. Model. Softw. 112, 46–51. https://doi.org/10.1016/j.envsoft.2018.11.008

Tuptuk, N., Hazell, P., Watson, J., Hailes, S., 2021. A Systematic Review of the State of Cyber-Security in Water Systems. Water 13, 81. https://doi.org/10.3390/w13010081

Verizon, 2017. 2017 Data Breach Investigations Report Tips on Getting the Most from This Report. Verizon Bus. J. 1–48. https://doi.org/10.1017/CBO9781107415324.004

Vidalis, S., Jones, A., 2005. Analyzing Threat Agents and Their Attributes., in: Hutchinson, B. (Ed.), ECIW. Academic Conferences Ltd, pp. 369–380.

Wangen, G., 2019. Quantifying and Analyzing Information Security Risk from Incident Data. pp. 129–154. https://doi.org/10.1007/978-3-030-36537-0_7

Wilensky, U. (1997). NetLogo Wolf Sheep Predation model. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.