

*Citation for published version:*

Taormina, R, Galelli, S, Tippenhauer, NO, Salomons, E, Ostfeld, A, Eliades, DG, Aghashahi, M, Sundararajan, R, Pourahmadi, M, Banks, MK, Brentan, BM, Campbell, E, Lima, G, Manzi, D, Ayala-Cabrera, D, Herrera, M, Montalvo, I, Izquierdo, J, Luvizotto, E, Chandy, SE, Rasekh, A, Barker, ZA, Campbell, B, Shafiee, ME, Giacomoni, M, Gatsis, N, Taha, A, Abokifa, AA, Haddad, K, Lo, CS, Biswas, P, Fayzul, M, Kc, B, Somasundaram, SL, Housh, M & Ohar, Z 2018, 'Battle of the Attack Detection Algorithms: Disclosing cyber attacks on water distribution networks', *Journal of Water Resources Planning and Management*, vol. 144, no. 8, 04018048. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000969](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000969)

DOI:

[10.1061/\(ASCE\)WR.1943-5452.0000969](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000969)

Publication date:

2018

Document Version

Peer reviewed version

[Link to publication](#)

© ASCE 2018.

University of Bath

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

1 **THE BATTLE OF THE ATTACK DETECTION ALGORITHMS:**
2 **DISCLOSING CYBER ATTACKS ON WATER DISTRIBUTION**
3 **NETWORKS**

4 Riccardo Taormina¹, Stefano Galelli², Member, ASCE, Nils Ole Tippenhauer³, Elad Salomons⁴,
 Avi Ostfeld⁵, Fellow, ASCE, Demetrios G. Eliades⁶, Mohsen Aghashahi⁷, S.M., ASCE,
 Raanju Sundararajan⁸, Mohsen Pourahmadi⁹, M. Katherine Banks¹⁰, Fellow, ASCE,
 B. M. Brentan¹¹, Enrique Campbell¹², G. Lima¹³, D. Manzi¹⁴, D. Ayala-Cabrera¹⁵,
 M. Herrera¹⁶, I. Montalvo¹⁷, J. Izquierdo¹⁸, E. Luvizotto Jr.¹⁹, Sarin E. Chandy²⁰,
Amin Rasekh²¹, Member, ASCE, Zachary A. Barker²², Bruce Campbell²³, M. Ehsan Shafiee²⁴,
 Marcio Giacomoni²⁵, Nikolaos Gatsis²⁶, Ahmad Taha²⁷, Ahmed A. Abokifa²⁸, S.M., ASCE,
 Kelsey Haddad²⁹, Cynthia S. Lo³⁰, Pratim Biswas³¹, M. Fayzul K. Pasha³², Bijay Kc³³,
 Saravanakumar Lakshmanan Somasundaram³⁴, Mashor Housh³⁵, Ziv Ohar³⁶

5 **ABSTRACT**

6 The BATtle of the Attack Detection ALgorithms (BATADAL) is the most recent competition
7 on planning and management of water networks undertaken within the Water Distribution

¹Singapore University of Technology and Design, 8 Somapah Road, Singapore, 487372.

²Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372. E-mail: stefano_galelli@sutd.edu.sg

³Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372.

⁴OptiWater, 6 Amikam Israel St., Haifa 3438561, Israel.

⁵Faculty of Civil and Environmental Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel.

⁶KIOS Research and Innovation Center of Excellence, University of Cyprus, 75 Kallipoleos Avenue, CY-1678, Nicosia, Cyprus.

⁷Zachry Dept. of Civil Engineering, Texas A&M Univ., College Station, TX.

⁸Dept. of Statistics, Texas A&M Univ., College Station, TX.

⁹Dept. of Statistics, Texas A&M Univ., College Station, TX.

¹⁰Zachry Dept. of Civil Engineering, Texas A&M Univ., College Station, TX.

¹¹CRAN, Universitet de Lorraine, Nancy, France.

¹²Universitat Politècnica de València, Valencia, Spain, Berliner Wasserbetriebe, Berlin, Germany.

¹³Universidade Estadual de Campinas, Campinas, Brazil.

¹⁴Universidade Estadual de Campinas, Campinas, Brazil.

¹⁵Irstea, Cestas, France.

¹⁶Univ. of Bath, Bath, U.K.

¹⁷Ingeniousware GmbH, Karlsruhe, Germany.

¹⁸Universitat Politècnica de València, Valencia, Spain.

¹⁹Universidade Estadual de Campinas, Campinas, Brazil.

²⁰Sensus Inc., 8601 Six Forks Rd., Suite 700, Raleigh, NC 27615.

²¹Sensus Inc., 8601 Six Forks Rd., Suite 700, Raleigh, NC 27615.

²²Sensus Inc., 8601 Six Forks Rd., Suite 700, Raleigh, NC 27615.

²³Sensus Inc., 8601 Six Forks Rd., Suite 700, Raleigh, NC 27615.

²⁴Sensus Inc., 8601 Six Forks Rd., Suite 700, Raleigh, NC 27615.

²⁵Dept. of Civil and Environmental Engineering, Univ. of Texas at San Antonio, San Antonio, TX 78249.

²⁶Dept. of Electrical and Computer Engineering, Univ. of Texas at San Antonio, San Antonio, TX 78249.

²⁷Dept. of Electrical and Computer Engineering, Univ. of Texas at San Antonio, San Antonio, TX 78249.

²⁸Dept. of Energy, Environmental, and Chemical Engineering, Washington Univ., St. Louis.

²⁹Dept. of Energy, Environmental, and Chemical Engineering, Washington Univ., St. Louis.

³⁰Dept. of Energy, Environmental, and Chemical Engineering, Washington Univ., St. Louis.

³¹Dept. of Energy, Environmental, and Chemical Engineering, Washington Univ., St. Louis.

³²Dept. of Civil and Geomatics Engineering, California State Univ., Fresno, CA 93740.

³³Dept. of Civil and Geomatics Engineering, California State Univ., Fresno, CA 93740.

³⁴Dept. of Civil and Geomatics Engineering, California State Univ., Fresno, CA 93740.

³⁵Faculty of Management, Dept. of Nat. Res. and Environmental Manag., Univ. of Haifa, Haifa, Israel.

³⁶Faculty of Management, Dept. of Nat. Res. and Environmental Manag., Univ. of Haifa, Haifa, Israel.

8 Systems Analysis Symposium. The goal of the battle was to compare the performance of
9 algorithms for the detection of cyber-physical attacks, whose frequency increased in the past
10 few years along with the adoption of smart water technologies. The design challenge was set
11 for C-Town network, a real-world, medium-sized water distribution system operated through
12 Programmable Logic Controllers and a Supervisory Control And Data Acquisition (SCADA)
13 system. Participants were provided with datasets containing (simulated) SCADA observa-
14 tions, and challenged with the design of an attack detection algorithm. The effectiveness of
15 all submitted algorithms was evaluated in terms of time-to-detection and classification accu-
16 racy. Seven teams participated in the battle and proposed a variety of successful approaches
17 leveraging data analysis, model-based detection mechanisms, and rule checking. Results were
18 presented at the Water Distribution Systems Analysis Symposium (World Environmental &
19 Water Resources Congress), in Sacramento, on May 21-25, 2017. This paper summarizes the
20 BATADAL problem, proposed algorithms, results, and future research directions.

21 **Keywords:** Water distribution systems, Cyber-physical attacks, Cyber security, EPANET,
22 Smart water networks, Attack detection

23 INTRODUCTION

24 The past decades witnessed the transition of water distribution systems from traditional
25 physical infrastructures to *cyber-physical systems* that combine physical processes with com-
26 putation and networking: physical assets—such as pipes, pumps, and valves—work in unison
27 with networked devices that monitor and coordinate the operations of the entire system.
28 These devices include Programmable Logic Controllers (PLCs), Supervisory Control And
29 Data Acquisition (SCADA) systems, Remote Terminal Units (RTUs), static and mobile
30 sensor networks, and smart meters (Hill et al. 2014; Gong et al. 2016; Sønderlund et al.
31 2016). The adoption of such smart water technologies plays a pivotal role in enhancing the
32 automation and reliability of water distribution systems, but simultaneously exposes them
33 to cyber-physical attacks (Rasekh et al. 2016)—namely the deliberate exploitation of com-
34 puter systems aimed at accessing sensitive information or compromising the operations of

35 the underlying physical system. Water (and wastewater) systems represent one of the sixteen
36 critical infrastructure sectors identified by the U.S. Department of Homeland Security (U.S.
37 Department of Homeland Security 2017), according to which the number of reported attacks
38 on water infrastructures has been growing steadily (ICS-CERT 2014; ICS-CERT 2015; ICS-
39 CERT 2016)—making them the third highest targeted sector after critical manufacturing
40 and energy (ICS-CERT 2016). To take remedial actions, several countries are establishing
41 research centres and international collaborations, such as the Israel–New York collaboration
42 to defend water systems from "infrastructure terrorists" (The Times of Israel 2018).

43
44 Protecting water distribution systems from cyber attacks requires (as with other cyber-
45 physical systems) a combination of proactive and reactive mechanisms (Cardenas et al. 2008).
46 Proactive mechanisms comprise all tools that reduce the chances to penetrate the system,
47 such as appropriate measures for traffic authentication and confidentiality protection, access
48 control, and device hardening (Graham et al. 2016; Adepu et al. 2017). Since it is not pos-
49 sible to rule out all attacks, cyber-physical systems should also be equipped with intrusion
50 detection schemes that assist with the recovery phase (Anderson 2010). Disclosing cyber
51 attacks—without issuing false alarms—is thus crucial. Unfortunately, this does not come
52 without some system-specific challenges. First, the definition of anomalous behaviours should
53 not only be related to point, or content, anomalies—i.e., data points lying beyond some
54 specific thresholds—since cyber-physical attacks can tamper one or multiple network com-
55 ponents while keeping the performance characteristics within the historical bounds (Abokifa
56 et al. 2017). This implies that detection schemes should be capable of disclosing both content
57 and contextual anomalies, namely, data points that are considered abnormal when viewed
58 against meta-information associated with the data points (Hayes and Capretz 2015). For
59 example, unaccounted high volumes of water leaving tanks during nighttime, when demand
60 is generally low, may be seen as a contextual anomaly revealed by looking at the flow data in
61 the context of time. Second, the same hydraulic response of a water network (e.g., low water

62 levels in a tank) can be obtained through different attacks (Taormina et al. 2017). There-
63 fore, detection schemes should also identify the cyber components that have been attacked;
64 a non-negligible challenge in large water networks. Third, all networked devices, including
65 SCADA systems, represent potential targets. This means that the information provided by
66 SCADA systems may not be fully reliable.

67
68 As the field of intrusion detection continues to grow, so too does the need of an objective
69 comparison of attack detection algorithms for water distribution systems. The BATtle of
70 the Attack Detection ALgorithms (BATADAL) was organized for this purpose. Participants
71 were provided with datasets containing (simulated) SCADA data for a water distribution
72 system victim of cyber attacks, and were tasked with the design of an attack detection
73 mechanism. The design goals of a detection algorithm were to: (1) disclose the presence
74 of an ongoing attack in the minimum time possible, (2) avoid issuing false alarms, and (3)
75 identify which components of the system have been compromised (optional). Seven teams,
76 from both academia and industry, contributed with novel solutions, which were evaluated
77 using specific evaluation criteria—i.e., time-to-detection and classification accuracy. The
78 BATADAL results were presented at a special session of the Water Distribution Systems
79 Analysis Symposium (World Environmental & Water resources Congress), in Sacramento,
80 on May 21-25, 2017.

81
82 This paper summarizes the main solutions and outcomes of the BATADAL, and proposes
83 future research directions for event detection in the realm of cyber-physical security. The
84 remainder of the paper describes: (1) the BATADAL problem, data, and evaluation criteria;
85 (2) a synopsis of the proposed attack detection algorithms; (3) an analysis of the results;
86 and (4) conclusions and future research directions.

87 **PROBLEM DESCRIPTION**

88 The operators of C-Town water distribution system have observed anomalous behaviors

89 in some hydraulic components, e.g., tank overflows, reduction in pump speed, anomalous
90 activation/deactivation of pumps. They suspect that the anomalies are attributable to cyber-
91 physical attacks that interfered with the system operations and tampered with the readings
92 recorded by the SCADA system. The aim of the participants was to develop an attack
93 detection mechanism that detects the presence of attacks—in the shortest amount of time—
94 from the available hourly SCADA data. In particular, attack detection algorithms must
95 classify the system state as either ‘safe’ or ‘under attack’. A summary description of C-Town
96 is provided below, along with the development data and evaluation criteria. BATADAL rules,
97 problem details, and data are available in the supplemental material of the paper.

98 **C-Town Network**

99 C-Town water distribution system is based on a real-world, medium-sized network, first
100 introduced for the *Battle of the Water Calibration Network* (Ostfeld et al. 2011). The network
101 consists of 429 pipes, 388 junctions, 7 storage tanks, 11 pumps (distributed across 5 pumping
102 stations), 5 valves, and a single reservoir (see Figure 1). Water consumption is fairly regular
103 throughout the year. These physical assets were augmented with a network of nine PLCs,
104 which are located in proximity of pumps, storage tanks, and valves. As shown in Table 1,
105 most of the PLCs controlling the pumps receive the information needed by the control logic
106 from other PLCs—for instance, PLC1 controls pump PU1 and PU2 on the basis of tank
107 T1 water level, which is monitored by PLC2. PLCs controlling pumps and valves record
108 information on the device status (ON/OFF or OPEN/CLOSED), the flow passing through
109 it, and the inlet and outlet pressure of pumping stations. The cyber network includes a
110 SCADA system, whose role is to coordinate the operations and store the readings provided
111 by the PLCs. All information regarding the distribution system were incorporated into the
112 EPANET2 (Rossman 2000) input file *C-Town.inp*, which was provided to the participants.
113 Water demand in all nodes of C-Town was not shared, meaning that participants could not
114 run the model for the same period and then compare the results with the provided SCADA
115 data.

116 Development data

117 Participants were provided with three datasets containing SCADA readings for 43 sys-
118 tem variables, i.e., tank water levels (7 variables, denoted as $L_{\langle \text{tank id} \rangle}$), inlet and
119 outlet pressure for one actuated valve and all pumping stations (12 variables, denoted as
120 $P_{\langle \text{junction id} \rangle}$), as well as their flow and status (24 variables, denoted as $F_{\langle \text{actuator}$
121 $\text{id} \rangle}$ and $S_{\langle \text{actuator id} \rangle}$, respectively). All variables are continuous, with the excep-
122 tion of the status of valve and pumps, represented by binary variables. The datasets were
123 generated via simulation with *epanetCPA*, a Matlab toolbox that allows to design a va-
124 riety of cyber attacks and simulate, with EPANET2 (version 2.0.12), the hydraulic re-
125 sponse of a water distribution network (Taormina et al. 2017). The toolbox is available
126 at <https://github.com/rtaormina/epanetCPA>. The hydraulic time step was set to 15
127 minutes, while the SCADA data reported to the participants were sampled with fixed hourly
128 intervals. The first two datasets, hereafter named *Training dataset 1* and *Training dataset*
129 *2*, were provided at the beginning of the competition, while the third one (*Test dataset*) was
130 subsequently used to evaluate and rank the attack detection algorithms.

- 131 • *Training dataset 1* was generated with a simulation horizon of 365 days. A key aspect
132 of the dataset is the absence of cyber attacks, which made it suitable for studying the
133 operations of the water distribution system under normal operating conditions.
- 134 • *Training dataset 2* contains seven attacks, spanning over 492 hourly time steps. One
135 attack was entirely revealed to the participants (by appropriately labelling the cor-
136 responding time steps), while the remaining attacks were either partially revealed or
137 hidden; see Table 2 for additional details. This corresponds to a post-attack scenario,
138 in which forensics experts carry out an investigation to determine whether, when, and
139 where the water distribution system has been affected.
- 140 • *Test dataset* contains seven additional attacks, spanning over 407 hourly time steps
141 (see Table 3). Naturally, no information regarding the attacks was revealed. Partici-
142 pants were required to run the detection algorithms on the *Test dataset* and to submit

143 a detection report containing the following information: number of attacks detected,
144 start and end time of each attack (in *DD-MM-YYYY hh* format), and the label of
145 the attacked device(s) (optional).

146 The operations of the water system were altered through malicious activation of hydraulic
147 actuators, change of actuator settings, and *deception* attacks—amongst the most common
148 for cyber-physical systems (Cardenas et al. 2009). The latter were aimed at manipulating
149 the information sent or received by sensors and PLCs, with the ultimate goal of affecting the
150 operations of an actuator (Urbina et al. 2016). Note that deception attacks were also used to
151 alter the information received by SCADA, therefore concealing the real, physical outcomes
152 of the attacks. SCADA concealment was performed by either adding an offset to the trans-
153 mitted sensor readings or by replacing actual traffic information between PLCs and SCADA
154 with previously-recorded data, a type of manipulation known as *replay attack* (Urbina et al.
155 2016). The replay attacks featured in the BATADAL consisted in replacing data for a given
156 hour of the day with those recorded during the same hour one or two days before. Figure 2
157 illustrates attack #3 (Training dataset 2), where both pump operations and SCADA data
158 are compromised. In this case, a deception attack manipulates Tank T1 water level readings
159 sent by PLC2 to PLC1. PLC1 receives a reading equal to 0.5 meters, which is below the
160 low level thresholds that activate pumps PU1 and PU2 (4 and 1 meter, respectively). This
161 results in both pumps working for the entire period of the attack, which lasts for 60 hours.
162 Consequently, the water level in Tank T1 reaches the full tank level (6.5 meters), with the
163 excess water being spilled. The adversary tries to conceal the surge in T1 water level with a
164 second deception attack that alters the signal sent by PLC2 to SCADA with a time-varying
165 offset.

166 **Evaluation criteria**

167 The attack detection algorithms were evaluated by comparing the detection report submitted
168 by each team against the provided Test dataset. The assessment was based on two scores

169 that account for (1) the time taken to detect an attack, and (2) the classification accuracy.
170 The two scores were eventually combined into an overall ranking score, as explained next.

171 *Time-to-detection*

172 The time-to-detection (TTD) is the time needed by an algorithm to disclose a threat. It is
173 defined as the difference between the time t_d at which the attack is detected and the time t_0
174 at which the attack started:

$$175 \quad TTD = t_d - t_0. \quad (1)$$

176 The value of t_d is inferred from the detection report, and it corresponds to the first time
177 stamp flagged as ‘under attack’ while the attack is ongoing. The lower the value of TTD ,
178 the better the algorithm performs. If an attack is detected, we then have:

$$179 \quad 0 \leq TTD \leq \Delta t, \quad (2)$$

180 where Δt is the total duration of the attack. If the attack is not detected while it is ongoing
181 (or at all), we set $TTD = \Delta t$. To facilitate the comparison of all algorithms under different
182 attack scenarios, the following performance score (S_{TTD}) was computed:

$$183 \quad S_{TTD} = 1 - \frac{1}{n_a} \sum_i^{n_a} \frac{TTD_i}{\Delta t_i}, \quad (3)$$

184 where n_a is the number of attacks contained in a dataset, TTD_i the time-to-detection relative
185 to the i -th attack, and Δt_i the corresponding duration. S_{TTD} varies between 0 and 1, with
186 $S_{TTD} = 1$ being the ideal case in which all attacks are immediately detected, and $S_{TTD} = 0$
187 the case in which none of the attacks is detected.

188 *Classification performance*

189 We determined the accuracy of an algorithm as its ability to disclose threats without raising
190 false alarms. In the context of binary classification problems—like the BATADAL—the
191 ability to identify threats is generally assessed with the *True Positive Rate* (TPR , also

192 known as *recall* or *sensitivity*), which is defined as:

$$193 \quad TPR = \frac{TP}{TP + FN}, \quad (4)$$

194 where TP and FN represent the number of True Positives and False Negatives, respectively.
195 In other words, the True Positive Rate is the ratio between the number of time steps cor-
196 rectly classified as under attack and the total number of time steps during which the system
197 is under attack.

198
199 The ability to avoid false alarms is measured with the *True Negative Rate* (TNR , or *speci-*
200 *ficity*), defined as

$$201 \quad TNR = \frac{TN}{FP + TN}, \quad (5)$$

202 where FP and TN represent the number of False Positives and True Negatives, respectively.
203 The True Negative Rate is thus the ratio between the number of time steps correctly classi-
204 fied as safe conditions and the total number of time steps during which the system is in safe
205 conditions.

206
207 To ease the comparison across all algorithms, the True Positive and True Negative Rate were
208 combined into a single classification performance score (S_{CLF}), defined as the mean between
209 TPR and TNR , namely:

$$210 \quad S_{CLF} = \frac{TPR + TNR}{2}. \quad (6)$$

211 This score accounts for both correct detection and false alarms, so it is suited for binary
212 classification problems in which the sample distribution is biased towards one of the two
213 classes—i.e., safe conditions, in the BATADAL. The value of S_{CLF} varies between 0 and 1,
214 with 1 representing a perfect classification.

215 *Ranking score*

216 The time-to-detection and accuracy scores were finally merged into an overall ranking score
217 (S), defined as:

$$218 \quad S = \gamma \cdot S_{TTD} + (1 - \gamma) \cdot S_{CLF}, \quad (7)$$

219 where γ ($0 \leq \gamma \leq 1$) determines the relative importance of the two evaluation scores. The
220 coefficient γ was set to 0.5 for the analysis reported below; so, early detection and accurate
221 classification were equally weighed. Note that a naïve detection mechanism that predicts the
222 system to be always in safe conditions gets a score S equal to 0.25 ($S_{TTD} = 0$, $S_{CLF} = 0.5$).
223 On the other hand, flagging the system as always under attack yields a value of S equal to
224 0.75 ($S_{TTD} = 1$, $S_{CLF} = 0.5$). This reflects the fact that S is intrinsically biased towards
225 attack identification, since the the consequences of failing to disclose an attack are deemed
226 more costly than issuing false alarms. These naïve detection methods have the same value
227 of S_{CLF} (equal to 0.5); yet, TPR and TNR are equal to 0 and 1 in the first case, and to
228 1 and 0 in the second case. This highlights the contrasting nature of the two components
229 of S_{CLF} , and suggests how increased sensitivity may come at the cost of issuing more false
230 alarms (and vice versa). Similarly, a potential conflict seems to exist between ensuring a
231 timely detection of the attacks (high S_{TTD}) and issuing few false alarms, as recently pointed
232 out by Housh and Ohar (2017c).

233 **ATTACK DETECTION ALGORITHMS**

234 Seven teams participated in the BATADAL. Here, we provide a brief description of each
235 team’s attack detection algorithm.

- 236 • Aghashahi et al. (2017) adopted a two-stage method that first extracts a four-
237 dimensional feature vector from the observed (multi-dimensional) time series data,
238 and then constructs a classifier to detect attacks. In the first stage, the time periods
239 of attack/no attack were used to extract four features that captured information on
240 the covariance and mean structure. Here, for every time instance, a local neighbor-

241 hood is utilized to construct estimates of mean and covariance. In the second stage, a
242 supervised classification technique (i.e., Random Forests, Breiman (2001)) was used
243 to classify the system state as safe or under attack.

- 244 • Brentan et al. (2017) reduced the dimensionality of the problem by exploiting the
245 division of C-Town network in District Metered Areas (DMAs). For each DMA, the
246 authors used data on normal operating conditions to create Recurrent Neural Net-
247 works that forecast tank water levels as a function of pump flow, upstream pressure
248 (of the corresponding pump station), and hour of the day (Díaz et al. 2016). A statis-
249 tical control process was finally used to identify abrupt changes in the neural networks
250 error time series when the latter were applied to data containing cyber attacks (Gu-
251 ralnik and Srivastava 1999). The rationale behind this approach is that it is plausible
252 to expect an increase in the error time series when the system is under attack, since
253 all neural networks are trained with data pertaining to normal operations.
- 254 • Chandy et al. (2017) developed two detection models running sequentially. The first
255 one uses features of the SCADA data (e.g., combined flow of pump stations, volume
256 pumped and stored) to check whether physical and/or operating rules have been
257 violated (e.g., tank levels within the bounds, hydraulic relationships between nodes
258 hold). The outcome of this model is a set of flagged events, which are confirmed by the
259 second model. The latter is a Convolutional Variational Auto-Encoder—belonging to
260 the family of deep learning methods (Kingma and Welling 2013; Doersch 2016)—that
261 calculates the reconstruction probability of the data: the lower the probability, the
262 higher the chance of the data being anomalous.
- 263 • Giacomoni et al. (2017) proposed two detection methods. The first one verifies the
264 integrity of the actuator rules and SCADA data—by (1) checking whether the SCADA
265 readings are consistent with the actuator rules defined for the water distribution
266 system, and (2) comparing the data for all variables to identify values falling below
267 or above thresholds created by analyzing data corresponding to normal operating

268 conditions. The second method builds on unveiling low-dimensionality components
269 in the available data as well as the sparse nature of anomalies, thereby facilitating the
270 separation of anomalies from the overall data. The separation of data into normal and
271 anomalous components can be performed using principal component analysis (PCA)
272 (Lakhina et al. 2004) or a convex optimization routine (Mardani et al. 2013). (The
273 results reported below for Giacomoni et al. (2017) correspond to the second detection
274 method based on PCA.)

- 275 • Abokifa et al. (2017) introduced a three-stage detection method, with each stage tar-
276 geting a specific class of anomalies. The first step features outlier detection techniques
277 to find statistical outliers in the data, thereby focusing on local anomalies that affect
278 each sensor individually. The second stage employs an Artificial Neural Network—in
279 the form of a Multi-Layer Perceptron—to detect contextual anomalies that do not
280 conform to normal operating conditions. The third stage targets global anomalies
281 that simultaneously affect multiple sensors. To disclose these anomalies, the layer
282 uses Principal Component Analysis to decompose the high-dimensional datasets of
283 sensor measurements into two sub-spaces representing normal and anomalous condi-
284 tions (Lee et al. 2013).
- 285 • Pasha et al. (2017) presented an algorithm consisting of three main interconnected
286 modules working on control rules and consistency checks, pattern recognition, and
287 hydraulic and system relationships. The first module checks the consistency of the
288 data against the set of control rules characterizing the water system, while the second
289 one uses statistical analysis to identify patterns for single hydraulic parameters and
290 combination thereof. The idea is that patterns under cyber attacks may not follow the
291 original ones. The anomalous behaviors detected by the first two modules are finally
292 confirmed by the third one, which develops relationships for some physical quantities
293 (e.g., tank levels, flows) and compares their estimates against those reported by the
294 first two modules.

- Housh and Ohar (2017b) proposed a model-based approach that employs EPANET to simulate the hydraulic processes of the water distribution systems, and then uses the error between EPANET simulated values and the available SCADA readings to detect anomalous behaviors. The approach consists of three main steps: first, available SCADA readings are used in a Mixed-Integer Linear Program to estimate the water demand in all nodes of C-Town; second, EPANET is used to generate reference values for the SCADA readings which are used to produce simulation errors when compared to actual readings; and third, a multi-level classification approach is implemented to classify the obtained simulation errors into event and normal conditions. A similar approach was successfully developed by Housh and Ohar (2017a) to detect contamination events in water distribution systems.

RESULTS

Algorithms performance

Table 4 reports the values of the ranking, time-to-detection, and classification score (S , S_{TTD} , and S_{CLF}) obtained by the competing algorithms on the test dataset. The table also reports the number of attacks detected, the values of TPR and TNR yielding the classification score, and the elements of the confusion matrix (i.e., TP , FP , TN , and FN). A visual comparison of S , S_{TTD} , and S_{CLF} is given in the scatter plot of Figure 3.

Figure 3 highlights a cluster of four high-performing algorithms, all achieving a ranking score S higher than (or close to) 0.90. The group is led by the algorithm proposed by Housh and Ohar (2017b), which shows the best overall performance ($S = 0.970$). Note that this algorithm is the top scorer in terms of both time-to-detection S_{TTD} and classification score S_{CLF} . Indeed, the detection trajectory depicted in Figure 4(a) shows that all attacks were immediately detected, with the exception of the last one, which was disclosed a few hours after its starting time. The algorithm of Abokifa et al. (2017) comes a close second, with S

321 equal to 0.949. This method was almost as quick as Housh and Ohar (2017b) in identifying
322 the attacks, but it was more prone to false alarms. As shown in Figure 4(b), Abokifa et al.
323 (2017) algorithm disclosed Attack #10 and #11 as a single continuous episode, erroneously
324 flagging the system as under attack for the period in between. The algorithm proposed by Gi-
325 acomoni et al. (2017) has the same TNR as that of Housh and Ohar (2017b)—meaning that
326 both algorithms were the most successful in avoiding false alarms. However, Giacomoni
327 et al. (2017) algorithm is less sensitive, resulting in lower TPR and minor timing errors (see
328 Figure 4(c)) that led to a score S equal to 0.927. With a value of S equal to 0.896, the
329 algorithm proposed by Brentan et al. (2017) can also be regarded as a strong performer. As
330 shown in Figure 4(d), this algorithm was able to consistently and accurately detect most of
331 the attacks, but it failed to identify the last one.

332
333 Although outdistanced by the leading group, the contributions of Chandy et al. (2017)
334 and Pasha et al. (2017) are still sensibly better than the naïve detection mechanisms de-
335 scribed in the second section. Their score S is equal to 0.802 and 0.773, respectively. As
336 illustrated in Figure 4(e,f), these two detection algorithms appear to suffer from opposite
337 problems. The algorithm of Chandy et al. (2017) turned out to be over-sensitive—meaning
338 that it was able to identify most of the attack instances, but at the cost of issuing numerous
339 false alarms. This is reflected on a relatively high value of the TPR , which, however, coin-
340 cides with the lowest overall value of the TNR . On the other hand, the algorithm of Pasha
341 et al. (2017) issued just a few false alarms, but it lacked sensitivity, thus failing to flag the
342 system as under attack for the entire duration of the events. This resulted in a very high
343 value of the TNR and the overall lowest TPR . Finally, the contribution of Aghashahi et al.
344 (2017) detected only three attacks, leading to a score S equal to 0.534.

345 **General Observations**

346 The main insights from the results presented above can be summarized as follows:

- 347 • All algorithms but one achieved a ranking score S larger than 0.75, meaning that
348 they performed better than naïve detection mechanisms. Yet, we observed a large
349 variability in the algorithm performance.
- 350 • Both time-to-detection and classification score are important aspects of performance.
351 Logically, the algorithms that performed consistently well for both metrics achieved
352 a higher ranking score. There appears to be a strong correlation between these two
353 metrics for most of the proposed algorithms (see Figure 3).
- 354 • Interestingly, the BATADAL was won by the only model-based approach. The idea
355 of estimating the water demands to simulate system dynamics with EPANET, and
356 then measure the errors with respect to the SCADA readings, proved successful. In
357 this regard, it is important to note that the BATADAL demand patterns were fairly
358 regular and consistent across the three datasets. Similarly, the participants were given
359 the same computational model of the C-Town network that was used to generate the
360 SCADA data (i.e., the input file *C-Town.inp*). Therefore, successful application of
361 this approach in real-world settings might be hindered by various factors, such as
362 the intrinsic variability of demand patterns, key uncertainties in the hydraulic model
363 (e.g., actual status of each component, pipe roughness, or pump performance curves),
364 or the unavailability of a reliable system model.
- 365 • Three data-driven algorithms belong to the cluster of high-performing detection mech-
366 anisms. This indicates that both model-based and data-driven approaches may be
367 suitable for attack detection problems, although their performance would probably
368 vary with the modelling context at hand.
- 369 • Only a few algorithms provided information on the attacked devices. Among these,
370 the algorithms proposed by Brentan et al. (2017) and Giacomoni et al. (2017) were
371 the most accurate.
- 372 • Most teams presented multi-stage detection methods. Comparing and confirming the
373 detection issued by different modules can help decrease classification errors.

- 374 • Detection algorithms adopting a ‘multivariate’ approach may be best suited than al-
375 gorithms analyzing a single time series per time. The inherent interdependence of the
376 elements in the water network should theoretically allow for the detection of anoma-
377 lies, even when the adversaries try to conceal their actions by altering the SCADA
378 readings of one or a few deployed sensors. Note that such interdependence generally
379 presents a nonlinear nature, which can be well described by nonlinear models—such
380 as those belonging to the class of Artificial Neural Networks.
- 381 • The adoption of supervised classification algorithms that learn how to classify the
382 system state (as either safe or under attack) may not be ideal, since the number of
383 attacks in the available data is generally limited. Supervised classification algorithms
384 should always be combined with cross-validation schemes.
- 385 • It appears that consistency checks and the analysis of control rules should lead to the
386 identification of the simplest attacks.

387 We note that the results described above were obtained on three specific datasets, which
388 represent only a small portion of the entire set of cyber-attacks that could threaten a water
389 distribution system. Hence, the generation of different attacks is likely to produce different
390 results—a limitation observed in other battles (e.g., Ostfeld et al. (2008)).

391 Another factor that influences the BATADAL results relates to the evaluation criteria. First,
392 the time-to-detection score S_{TTD} is based on the ratio between the time taken to detect an
393 attack and the attack duration; this implies that a 2-hour attack detected within 1 hour
394 would have the same score as a 10-hour attack detected on hour 5. Some operators may
395 prefer to define scores that account explicitly for the absolute value of the attack duration
396 or its corresponding damage. Second, the classification performance score S_{CLF} is based on
397 TPR and TNR , which are common metrics for classification problems. Yet, one may adopt
398 other metrics, such as the $F1$ score (Sokolova and Lapalme 2009). Third, time-to-detection
399 and classification performance score were given the same importance (the coefficient γ is
400 equal to 0.50 in Eq. (7)). Depending on the problem at hand, one may want to outweigh

401 the time-to-detection (or the classification accuracy).

402 FUTURE RESEARCH DIRECTIONS

403 The BATADAL highlighted the following gaps that may need additional research efforts:

- 404 • *Robustness analysis.* As mentioned above, the performance of an attack detection
405 algorithm may depend—to a certain extent—on the data used during the calibration
406 and validation process. To limit the impact of data when evaluating the robustness of
407 an algorithm, it is thus advisable to generate stochastic simulation scenarios compris-
408 ing varying hydraulic conditions (i.e., water demand, initial tank levels) and multiple
409 attack sequences.
- 410 • *Use of real SCADA data.* A major limitation of the current research on cyber-security
411 is the absence of detailed information on cyber attacks to water utilities (e.g., timing,
412 compromised devices, hydraulic response of the system). Access to such information
413 and to the corresponding SCADA data—perhaps, in some anonymized forms—would
414 drastically enhance our understanding on skills and limitations of detection algo-
415 rithms. Another challenge with SCADA data is that they often contain noise and
416 measurement errors, so attack detection algorithms should be coupled with data pre-
417 processing techniques.
- 418 • *Pressure deficient conditions and water quality problems.* A limitation of this battle
419 is its reliance of data generated with a demand-driven engine (Taormina et al. 2017).
420 The range of attacks should be thus extended to include pressure-deficient conditions,
421 water quality problems, and adversarial attempts aimed at threatening emergency re-
422 sponses, such as firefighting operations. In the absence of real SCADA data, sim-
423 ulated data could be generated by combining *epanetCPA* with more sophisticated
424 hydraulic engines (e.g., Sayyed et al. (2015)) or water quality models (e.g., EPANET-
425 MSX, Shang et al. (2007)).
- 426 • *Sensitivity analysis.* The definition of the cut-off criteria defining outliers regulates

427 the trade-off between TPR and TNR for most of the algorithms, so there is a need
428 to adopt or develop sensitivity analysis tools that draw the appropriate line between
429 normal and anomalous data (Abokifa et al. 2017). This step should always precede
430 the application of an algorithm to new datasets—or its deployment in a SCADA
431 system.

- 432 • *Computational requirements and scalability to large networks.* The algorithms pre-
433 sented in this paper were applied to a medium-sized water distribution system com-
434 prising one SCADA system and nine PLCs. Since attack detection algorithm are
435 meant to run in real-time, it is necessary to evaluate their computational require-
436 ments as well as their scalability to larger networks.
- 437 • *Attack localization.* To facilitate and hasten incident resolution, an ideal detection
438 mechanism should be able to identify which components of the network are being
439 attacked. This is a rather challenging task due to the intrinsic correlation among the
440 hydraulic variables. For data-driven detection mechanisms, the task may be solved
441 with variable (or feature) selection algorithms (Galelli et al. 2014; Karakaya et al.
442 2016), which identify the variables that are strongly related to the detected anomalies.
- 443 • *Integration with other fault detection mechanisms.* Since attack detection mechanisms
444 aim to disclose outliers and contextual anomalies in the system behavior, they may
445 accidentally disclose anomalous behaviors that are not necessarily caused by cyber at-
446 tacks (e.g., a water level sensor reporting wrong readings or a malfunctioning pump).
447 Hence, there is a need to disclose the nature of each problem being identified—for ex-
448 ample, by combining the attack detection algorithms with fault detection mechanisms
449 that monitor the operations of PLCs.
- 450 • *Cost effectiveness of attack detection.* In the BATADAL, the different algorithms were
451 evaluated based on their responsiveness and classification performance. Although
452 these metrics provide some insights on the potential benefits of deploying an attack
453 detection mechanism, a more comprehensive evaluation is needed. For example, one

454 could try to estimate the damage or cost associated to each cyber-physical attack and
455 the corresponding cost savings guaranteed by a detection algorithm.

456 CLOSURE

457 The BATADAL is the first *battle competition* dealing with the emerging topic of cyber-
458 physical security of water distribution systems. This battle gave an opportunity to develop,
459 test, and compare attack detection algorithms for SCADA data. The solutions provided by
460 seven teams suggest that timely and accurate detection can be obtained by both model-
461 based and data-driven approaches, usually made of multiple sequential stages. While the
462 data and algorithms presented here provide a first step towards an objective comparison of
463 attack detection algorithms for water distribution systems, they do not represent the entire
464 spectrum of modelling contexts that practitioners and researchers would encounter. Hence,
465 we hope that the availability of a dedicated website (www.batadal.net) will help share more
466 datasets and case studies.

467 SUPPLEMENTAL DATA

468 The supplemental data include the following files, which are available online in the ASCE
469 Library (www.ascelibrary.org):

- 470 • *BATADAL rules.pdf*—competition rules, available to participants;
- 471 • *C-Town.inp*—EPANET input file, version 2.00.12, available to participants;
- 472 • *Training dataset 1.csv*—data without attacks, available to participants;
- 473 • *Training dataset 2.csv*—data with attacks and corresponding labels, available to the
474 participants with partial labels;
- 475 • *Test dataset.csv*—data with attacks and corresponding labels, available to the partic-
476 ipants without labels;
- 477 • *Detection Reports.zip*—detection reports submitted by the participants.

478 Additional details about BATADAL are available at www.batadal.net. *epanetCPA* is avail-
479 able at <https://github.com/rtaormina/epanetCPA>.

ACKNOWLEDGEMENTS

Riccardo Taormina, Stefano Galelli, and Nils Ole Tippenhauer are supported by the National Research Foundation (NRF), Singapore, under its National Cybersecurity R&D Programme (Award No. NRF2014NCR-NCR001-40). Demetrios Eliades is supported by the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 739551 (KIOS CoE). Mohsen Aghashahi and M. Katherine Banks are supported by Qatar National Research Fund (QNRF) under the grant NPRP8-1292-2-548. B. M. Brentan, Enrique Campbell, G. Lima, D. Manzi, D. Ayala-Cabrera, M. Herrera, I. Montalvo, J. Izquierdo, and E. Luvizotto Jr. are supported CAPES and CNPq founding agencies. Ahmed Abokifa, Kelsey Haddad, Cynthia Lo, and Pratim Biswas’ work was carried out with the partial support from the Lucy and Stanley Lopata Endowment at Washington University in St. Louis.

REFERENCES

- Abokifa, A. A., Haddad, K., Lo, C. S., and Biswas, P. (2017). “Detection of cyber physical attacks on water distribution systems via principal component analysis and artificial neural networks.” *World Environmental and Water Resources Congress 2017*, 676–691, <<http://ascelibrary.org/doi/abs/10.1061/9780784480625.063>>.
- Adepu, S., Mishra, G., and Mathur, A. (2017). “Access control in water distribution networks: A case study.” *Software Quality, Reliability and Security (QRS), 2017 IEEE International Conference on*, IEEE, 184–191.
- Aghashahi, M., Sundararajan, R., Pourahmadi, M., and Banks, M. K. (2017). “Water distribution systems analysis symposium; battle of the attack detection algorithms (BATADAL).” *World Environmental and Water Resources Congress 2017*, 101–108, <<http://ascelibrary.org/doi/abs/10.1061/9780784480595.010>>.
- Anderson, R. J. (2010). *Security engineering: a guide to building dependable distributed systems*. John Wiley & Sons.
- Breiman, L. (2001). “Random forests.” *Machine Learning*, 45(1), 5–32.
- Brentan, B. M., Campbell, E., Lima, G., Manzi, D., Ayala-Cabrera, D., Herrera, M., Montalvo, I., Izquierdo, J., and Luvizotto, E. (2017). “On-line cyber attack detection in water networks through state forecasting and control by pattern recognition.” *World Environmental and Water Resources Congress 2017*, 583–592, <<http://ascelibrary.org/doi/abs/10.1061/9780784480625.054>>.
- Cardenas, A., Amin, S., Sinopoli, B., Giani, A., Perrig, A., and Sastry, S. (2009). “Challenges for securing cyber physical systems.” *Proceedings of Workshop on future directions in cyber-physical systems security*, Vol. 5.
- Cardenas, A. A., Amin, S., and Sastry, S. (2008). “Secure control: Towards survivable cyber-physical systems.” *Proceedings of Conference on Distributed Computing Systems Workshops (ICDCS)*, IEEE, 495–500.
- Chandy, S. E., Rasekh, A., Barker, Z. A., Campbell, B., and Shafiee, M. E. (2017). “De-

518 tection of cyber-attacks to water systems through machine-learning-based anomaly detec-
519 tion in scada data.” *World Environmental and Water Resources Congress 2017*, 611–616,
520 <<http://ascelibrary.org/doi/abs/10.1061/9780784480625.057>>.

521 Díaz, S., González, J., and Mínguez, R. (2016). “Uncertainty evaluation for constrained
522 state estimation in water distribution systems.” *Journal of Water Resources Planning and
523 Management*, 142(12), 06016004.

524 Doersch, C. (2016). “Tutorial on variational autoencoders.” *arXiv preprint:1606.05908*.

525 Galelli, S., Humphrey, G. B., Maier, H. R., Castelletti, A., Dandy, G. C., and Gibbs, M. S.
526 (2014). “An evaluation framework for input variable selection algorithms for environmental
527 data-driven models.” *Environmental Modelling & Software*, 62, 33–51.

528 Giacomoni, M., Gatsis, N., and Taha, A. (2017). “Identification of cyber at-
529 tacks on water distribution systems by unveiling low-dimensionality in the sen-
530 sory data.” *World Environmental and Water Resources Congress 2017*, 660–675,
531 <<http://ascelibrary.org/doi/abs/10.1061/9780784480625.062>>.

532 Gong, W., Suresh, M. A., Smith, L., Ostfeld, A., Stoleru, R., Rasekh, A., and Banks, M. K.
533 (2016). “Mobile sensor networks for optimal leak and backflow detection and localization
534 in municipal water networks.” *Environmental Modelling & Software*, 80, 306–321.

535 Graham, J., Olson, R., and Howard, R. (2016). *Cyber security essentials*. CRC Press.

536 Guralnik, V. and Srivastava, J. (1999). “Event detection from time series data.” *Proceedings
537 of the fifth ACM SIGKDD international conference on Knowledge discovery and data
538 mining*, ACM, 33–42.

539 Hayes, M. A. and Capretz, M. A. (2015). “Contextual anomaly detection framework for big
540 sensor data.” *Journal of Big Data*, 2(1), 2.

541 Hill, D., Kerkez, B., Rasekh, A., Ostfeld, A., Minsker, B., and Banks, M. K. (2014). “Sensing
542 and cyberinfrastructure for smarter water management: the promise and challenge of
543 ubiquity.” *Journal of Water Resources Planning and Management*, 140(7), 01814002.

544 Housh, M. and Ohar, Z. (2017a). “Integrating physically based simulators with event detec-

545 tion systems: Multi-site detection approach.” *Water Research*, 110, 180–191.

546 Housh, M. and Ohar, Z. (2017b). “Model based approach for cyber-physical attacks detection
547 in water distribution systems.” *World Environmental and Water Resources Congress 2017*,
548 727–736, <<http://ascelibrary.org/doi/abs/10.1061/9780784480625.067>>.

549 Housh, M. and Ohar, Z. (2017c). “Multiobjective calibration of event-detection systems.”
550 *Journal of Water Resources Planning and Management*, 143(8), 06017004.

551 ICS-CERT (2014). “NCCIC/ICS-CERT year in review: FY 2013.” *Report No. 13-50369*,
552 U.S. Department of Homeland Security – Industrial Control Systems-Cyber Emergency
553 Response Team, Washington, D.C.

554 ICS-CERT (2015). “NCCIC/ICS-CERT year in review: FY 2014.” *Report No. 14-50426*,
555 U.S. Department of Homeland Security – Industrial Control Systems-Cyber Emergency
556 Response Team, Washington, D.C.

557 ICS-CERT (2016). “NCCIC/ICS-CERT year in review: FY 2015.” *Report No. 15-50569*,
558 U.S. Department of Homeland Security – Industrial Control Systems-Cyber Emergency
559 Response Team, Washington, D.C.

560 Karakaya, G., Galelli, S., Ahipaşaoğlu, S. D., and Taormina, R. (2016). “Identifying (quasi)
561 equally informative subsets in feature selection problems for classification: a max-relevance
562 min-redundancy approach.” *IEEE Transactions on Cybernetics*, 46(6), 1424–1437.

563 Kingma, D. P. and Welling, M. (2013). “Auto-encoding variational bayes.” *arXiv*
564 *preprint:1312.6114*.

565 Lakhina, A., Crovella, M., and Diot, C. (2004). “Diagnosing network-wide traffic
566 anomalies.” *Proceedings of the 2004 Conference on Applications, Technologies, Ar-*
567 *chitectures, and Protocols for Computer Communications*, SIGCOMM ’04, 219–230,
568 <<http://doi.acm.org/10.1145/1015467.1015492>>.

569 Lee, Y.-J., Yeh, Y.-R., and Wang, Y.-C. F. (2013). “Anomaly detection via online oversam-
570 pling principal component analysis.” *IEEE Transactions on Knowledge and Data Engi-*
571 *neering*, 25(7), 1460–1470.

572 Mardani, M., Mateos, G., and Giannakis, G. B. (2013). “Recovery of low-rank plus com-
573 pressed sparse matrices with application to unveiling traffic anomalies.” *IEEE Transactions*
574 *on Information Theory*, 59(8), 5186–5205.

575 Ostfeld, A., Salomons, E., Ormsbee, L., Uber, J. G., Bros, C. M., Kalungi, P., Burd, R.,
576 Zazula-Coetzee, B., Belrain, T., Kang, D., et al. (2011). “Battle of the water calibration
577 networks.” *Journal of Water Resources Planning and Management*, 138(5), 523–532.

578 Ostfeld, A., Uber, J. G., Salomons, E., Berry, J. W., Hart, W. E., Phillips, C. A., Watson,
579 J.-P., Dorini, G., Jonkergouw, P., Kapelan, Z., et al. (2008). “The battle of the water
580 sensor networks (bwsn): A design challenge for engineers and algorithms.” *Journal of*
581 *Water Resources Planning and Management*, 134(6), 556–568.

582 Pasha, M. F. K., Kc, B., and Somasundaram, S. L. (2017). “An approach to detect the cyber-
583 physical attack on water distribution system.” *World Environmental and Water Resources*
584 *Congress 2017*, 703–711, <<http://ascelibrary.org/doi/abs/10.1061/9780784480625.065>>.

585 Rasekh, A., Hassanzadeh, A., Mulchandani, S., Modi, S., and Banks, M. K. (2016). “Smart
586 water networks and cyber security.” *Journal of Water Resources Planning and Manage-*
587 *ment*, 142.

588 Rossman, L. A. (2000). *EPANET 2 Users Manual*. U.S. Environmental Protection Agency,
589 Washington, D.C., EPA/600/R-00/057 edition.

590 Sayyed, M. A. H. A., Gupta, R., and Tanyimboh, T. T. (2015). “Noniterative application of
591 epanet for pressure dependent modelling of water distribution systems.” *Water Resources*
592 *Management*, 29(9), 3227–3242.

593 Shang, F., Uber, J. G., and Rossman, L. A. (2007). “Modeling reaction and transport of mul-
594 tiple species in water distribution systems.” *Environmental Science & Technology*, 42(3),
595 808–814.

596 Sokolova, M. and Lapalme, G. (2009). “A systematic analysis of performance measures for
597 classification tasks.” *Information Processing & Management*, 45(4), 427–437.

598 Sønderlund, A. L., Smith, J. R., Hutton, C. J., Kapelan, Z., and Savic, D. (2016). “Ef-

599 fectiveness of smart meter-based consumption feedback in curbing household water use:
600 Knowns and unknowns.” *Journal of Water Resources Planning and Management*, 142(12),
601 04016060.

602 Taormina, R., Galelli, S., Tippenhauer, N. O., Salomons, E., and Ostfeld, A. (2017). “Charac-
603 terizing cyber-physical attacks on water distribution systems.” *Journal of Water Resources*
604 *Planning and Management*, 143(5), 04017009.

605 The Times of Israel (2018). “Israel tech to protect NY water systems from
606 cyberattacks, <[https://www.timesofisrael.com/israel-tech-to-protect-ny-water-systems-](https://www.timesofisrael.com/israel-tech-to-protect-ny-water-systems-from-attack/)
607 from-attack/> (January).

608 Urbina, D., Giraldo, J., Tippenhauer, N. O., and Cárdenas, A. (2016). “Attacking fieldbus
609 communications in ICS: Applications to the SWaT testbed.” *Proceedings of Singapore*
610 *Cyber Security Conference (SG-CRC)* (January).

611 U.S. Department of Homeland Security (2017). “Critical infrastructure sectors,
612 <<https://www.dhs.gov/critical-infrastructure-sectors>> (September).

613 **List of Tables**

614	1	Sensors and actuators monitored/controlled by the PLCs	28
615	2	Attacks featured in Training dataset 2.	29
616	3	Attacks featured in the Test dataset.	30
617	4	Performance of all attack detection algorithms	31

TABLE 1. Sensors and actuators (pumps, valves) monitored/controlled by the PLCs. For each PLC, we also report the corresponding controlling sensor, which provides the information needed to operate the actuators. Note that a PLC-to-PLC connection is established whenever an actuator and the corresponding control sensor are connected to two different PLCs.

PLC	Sensor	Actuators (Controlling sensor)
PLC1	-	PU1(T1), PU2(T1)
PLC2	T1	-
PLC3	T2	V2(T2), PU4(T3), PU5(T3), PU6(T4), PU7(T4)
PLC4	T3	-
PLC5	-	PU8(T5), PU9(-), PU10(T7), PU11(T7)
PLC6	T4	-
PLC7	T5	-
PLC8	T6	-
PLC9	T7	-

TABLE 2. Attacks featured in Training dataset 2.

ID	Starting time [dd/mm/YY HH]	Ending time [dd/mm/YY HH]	Duration [hours]	Attack description	SCADA concealment	Labeled [hours]
1	13/09/2016 23	16/09/2016 00	50	Attacker alters SCADA transmission to PLC9 and changes the L_T7 thresholds determining when pumps PU10/PU11 are switched ON/OFF. Low levels in T7.	Replay attack on L_T7 .	42
2	26/09/2016 11	27/09/2016 10	24	Like Attack #1.	Like Attack #1 but replay attack extended on PU10/PU11 flow and status.	0
3	09/10/2016 09	11/10/2016 20	60	Attack alters L_T1 readings sent by PLC2 to PLC1, which reads a constant low level and keeps pumps PU1/PU2 ON. Overflow in T1.	Polyline to offset L_T1 increase.	60
4	29/10/2016 19	02/11/2016 16	94	Like Attack #3.	Replay attack on L_T1, PU1/PU2 flow and status, as well as on pressure at pumps outlet (P_J269).	37
5	26/11/2016 17	29/11/2016 04	60	Working speed of PU7 reduced to 0.9 of nominal speed. Lower water levels in T4.		7
6	06/12/2016 07	10/12/2016 04	94	Like Attack #5, but speed reduced to 0.7.	Replay attack on L_T4.	73
7	14/12/2016 15	19/12/2016 04	110	Like Attack #6.	Replay attack on L_T4, as well as on PU6/PU7 flow and status.	0

TABLE 3. Attacks featured in the Test dataset.

ID	Starting time [dd/mm/YY HH]	Ending time [dd/mm/YY HH]	Duration [hours]	Attack description	SCADA concealment
8	16/01/2017 09	19/01/2017 06	70	Attacker gains control of PLC3 and changes the L_T3 thresholds determining when pumps PU4/PU5 are switched ON/OFF. Low levels in T3.	Replay attack on L_T3, as well as on PU4/PU5 flow and status.
9	30/01/2017 08	02/02/2017 00	65	Attack alters L_T2 readings arriving to PLC3, which reads a low level and keeps valve V2 OPEN. The attack leads T2 to overflow.	Polyline to offset L_T2 increase.
10	09/02/2017 03	10/02/2017 09	31	Malicious activation of pump PU3	
11	12/02/2017 01	13/02/2017 07	31	Similar to Attack #10	
12	24/02/2017 05	28/02/2017 08	100	Similar to Attack #9	Replay attack on L_T2, V2 flow and status, as well as on V2 inlet and outlet pressure readings (P_J14, P_J422)
13	10/03/2017 14	13/03/2017 21	80	Attacker gains control of PLC5 and changes the L_T7 thresholds determining when pumps PU10/PU11 are switched ON/OFF. The pumps are forced to switch ON/OFF continuously during the attack.	Replay attack on L_T7, PU10/PU11 flow and status, as well as on pumps inlet and outlet pressure readings (P_J14, P_J422). Inlet pressure concealment terminates before that of other variables.
14	25/03/2017 20	27/03/2017 01	30	Alteration of T4 signal arriving to PLC6. Overflow in T6.	

TABLE 4. Performance of all attack detection algorithms, assessed in terms of number of attacks detected, overall ranking score (S), time-to-detection (S_{TTD}), accuracy (S_{CLF}), True Positive Ratio (TPR), True Negative Ratio (TNR), and number of True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN). The algorithms are ranked according to the their overall ranking score.

Rank	Team	# Attacks detected	S	S_{TTD}	S_{CLF}	TPR	TNR	TP	FP	TN	FN
1	Housh and Ohar	7	0.970	0.965	0.975	0.953	0.997	388	5	1677	19
2	Abokifa et al.	7	0.949	0.958	0.940	0.921	0.959	375	69	1613	32
3	Giacomoni et al.	7	0.927	0.936	0.917	0.838	0.997	341	5	1677	66
4	Brentan et al.	6	0.894	0.857	0.931	0.889	0.973	362	45	1637	45
5	Chandy et al.	7	0.802	0.835	0.768	0.857	0.678	349	541	1141	58
6	Pasha et al.	7	0.773	0.885	0.660	0.329	0.992	134	14	1668	273
7	Aghashahi et al.	3	0.534	0.429	0.640	0.396	0.884	161	195	1487	246

618	List of Figures	
619	1	Graphical representation of C-Town water distribution system 33
620	2	Illustration of attack #3 (from Training dataset 2). The attacker alters Tank
621		T1 water level readings (continuous black line) sent by PLC2 to PLC1, which
622		reads a constant low level (dotted black line) and keeps Pumps PU1/PU2 ON.
623		This causes an overflow in Tank T1 (thick gray line). To conceal the action,
624		the attacker alters the signal sent by PLC2 to SCADA (dashed black line)
625		by adding a time-varying offset (continuous gray line). The duration of the
626		entire attack is highlighted by the light gray line on the horizontal axis. . . . 34
627	3	Graphical representation of the algorithm performance 35
628	4	Comparison between actual and detected attacks 36

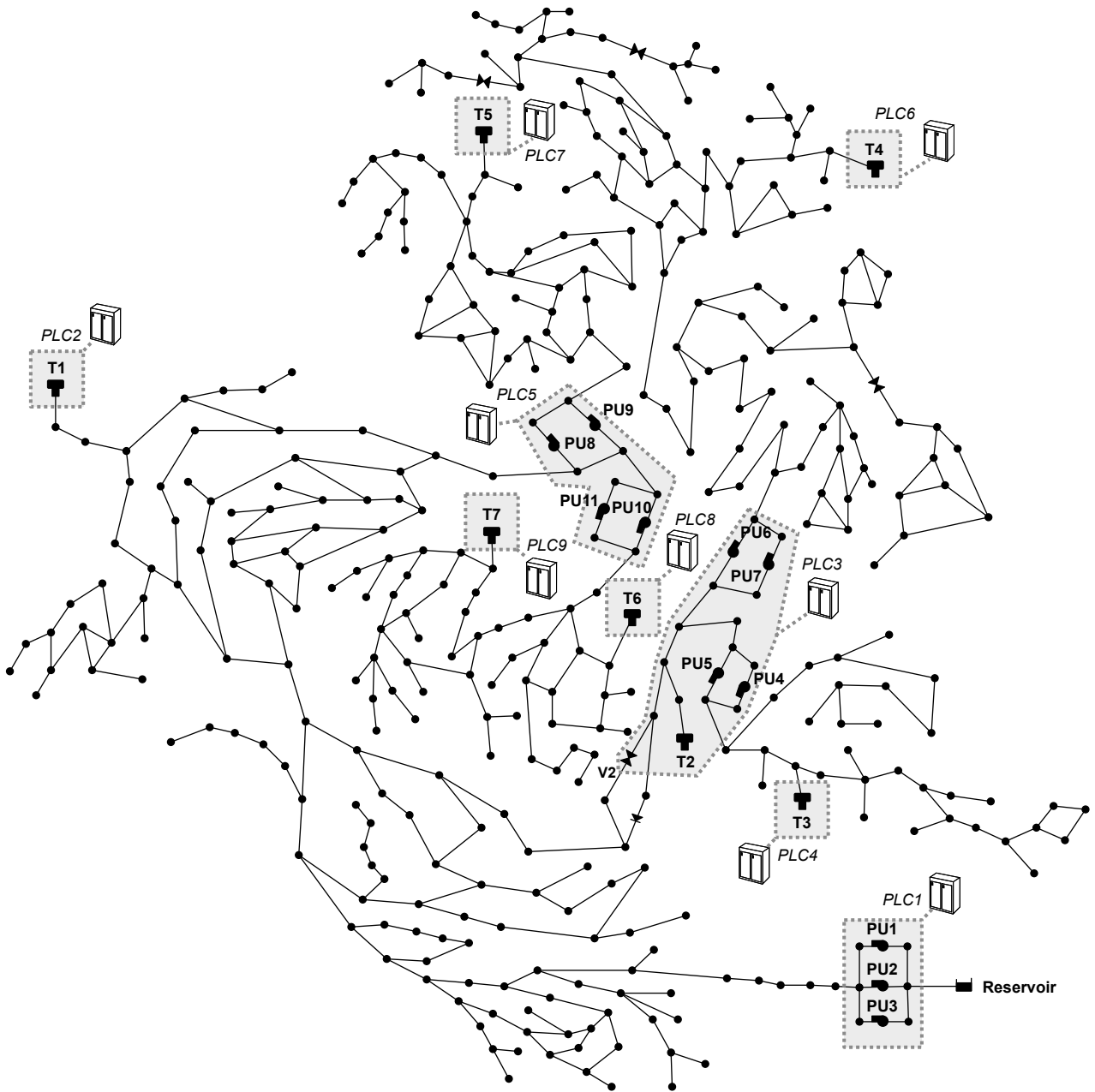


FIG. 1. Graphical representation of C-Town water distribution system (adapted from Taormina et al. 2017).

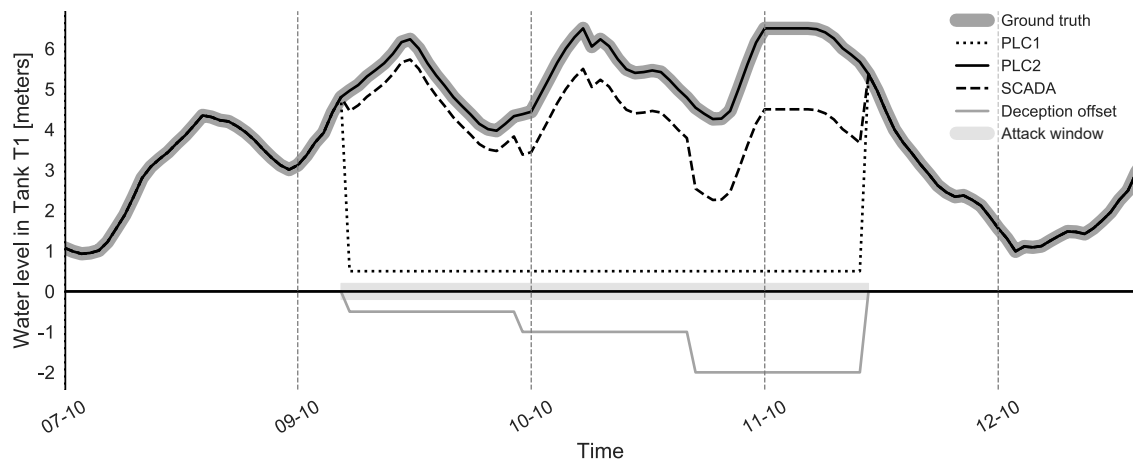


FIG. 2. Illustration of attack #3 (from Training dataset 2). The attacker alters Tank T1 water level readings (continuous black line) sent by PLC2 to PLC1, which reads a constant low level (dotted black line) and keeps Pumps PU1/PU2 ON. This causes an overflow in Tank T1 (thick gray line). To conceal the action, the attacker alters the signal sent by PLC2 to SCADA (dashed black line) by adding a time-varying offset (continuous gray line). The duration of the entire attack is highlighted by the light gray line on the horizontal axis.

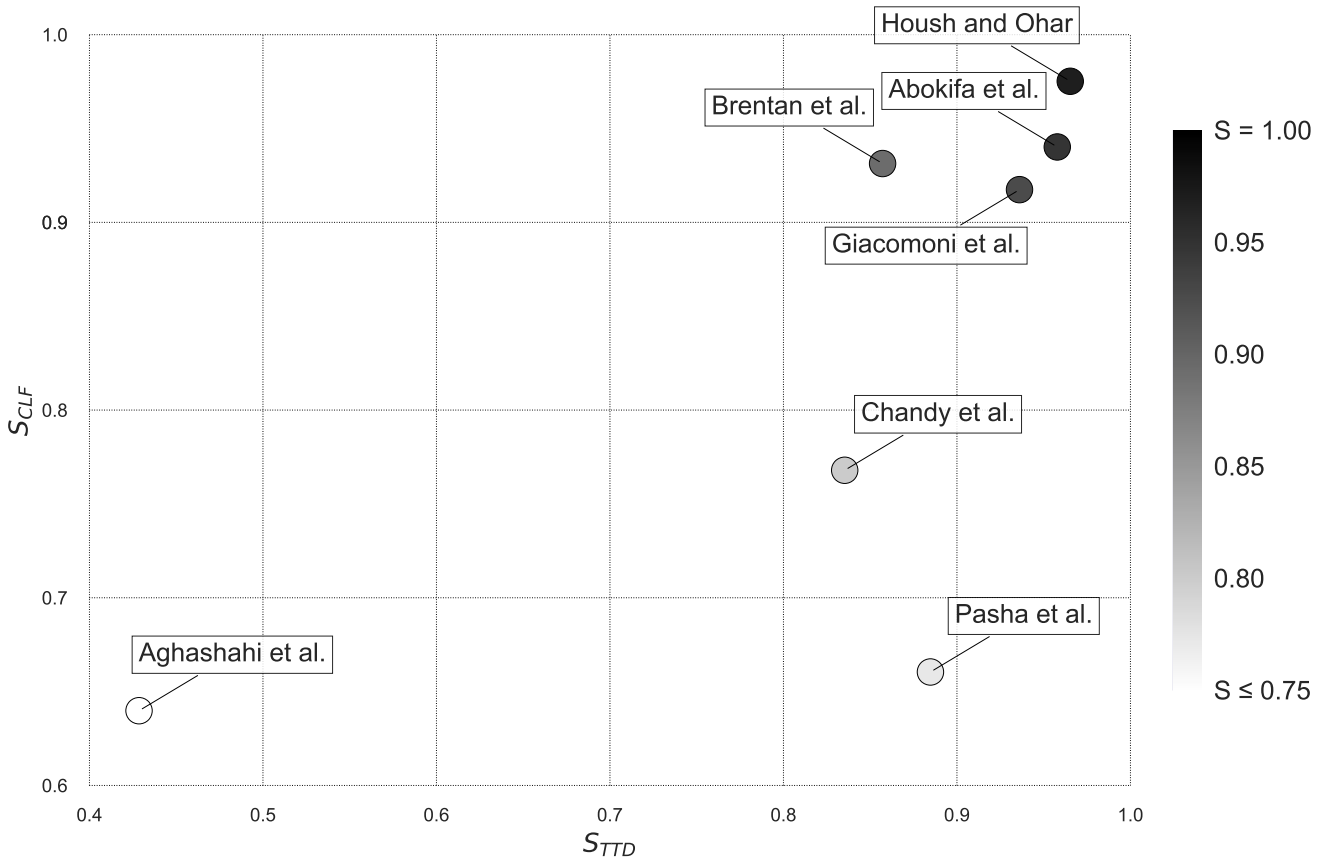


FIG. 3. Graphical representation of the algorithm performance, measured in terms of time-to-detection (S_{TTD} , horizontal axis), classification performance (S_{CLF} , vertical axis), and overall ranking score (S , color-bar).