

Trace-Based Assessment of Self-Regulation in Distributed Cyber-Physical Control Systems: Disturbance Rejection and Recovery. An Empirical Study

Iliyan Vasilev (i.vasilidis@uctm.edu)

University of Chemical Technology and Metallurgy, 8 Kliment Ohridski Blvd., 1000 Sofia, Bulgaria

Abstract - A return to the operating set-point does not necessarily demonstrate that a cyber-physical system has rejected a disturbance. The output may recover while the disturbance is still active or only after the normal command path is restored. This study develops a trace-based assessment that separates these two situations. The empirical part of the study is based on an open KIOS cyber-physical sandbox dataset describing battery set-point manipulation. The record includes 205 observations collected over 1115 s. Four command-mismatch episodes are examined separately from the output error. Their behaviour is compared using tracking error, time spent within the tolerance band, recovery time, and the proposed Self-Regulation Index (SRI). With a 1% error tolerance, the SRI values during the active disturbance are 0.0070, 0.8558, 0.0134, and 0.4090 for the four episodes. The additive disturbance is corrected within one recorded update. The fixed-command case is more problematic: after the imposed command is removed, the deviation becomes 7.08 times larger than the highest error observed during the disturbance itself. In the held-command case, the system remains close to the operating point at first, but this changes once the legitimate reference starts to move. The results show descriptively that disturbance rejection, recovery after restoration, and the ability to follow a changing reference should not be treated as the same property. The proposed measures describe what is observed in these traces, but they do not by themselves prove stability, autonomy, or resilience under other disturbances.

Keywords - Cyber-physical systems, disturbance rejection, process automation, resilience, self-regulation.

INTRODUCTION

A controller can respond well to one type of command disturbance and poorly to another, even when the software and communication structure remain unchanged. For this reason, a trace that eventually returns to the set-point does not tell the whole story. The system may recover while the disturbance is still present, or only after the interference has been removed. This difference is important when assessing whether a distributed cyber-physical system can genuinely be described as self-regulating.

Distributed secondary control is already well established in both theory and experiment. Simpson-Porco et al. [1], for example, study distributed averaging methods for frequency restoration, voltage regulation and power sharing, while Ge et al. [2] combine state estimation with distributed control under uncertain microgrid dynamics. Those studies are concerned mainly with controller design and performance. The focus here is different. Rather than proposing a new

control law, this study asks what can reasonably be inferred about regulation from an independently published operational trace alone?

Resilience also depends on the event and operating context. Mishra et al. [3] explicitly account for the conditions under which a microgrid must absorb, restore and adapt. An error score averaged across a convenient recording window cannot represent all these capabilities. A satisfactory operating point under an unchanged reference may conceal an inability to respond to a new command.

Self-regulation denotes the observed maintenance or restoration of the intended process output under the system's existing feedback arrangements. It does not imply that the controller diagnoses attacks, changes its own gains or operates without supervisory coordination. "Distributed" refers to the networked control of distributed energy resources in the source use case [4]; the analysis does not establish a fully decentralized control topology.

The contribution is an event-based estimation procedure applied to a complete and directly inspectable trace. Command anomalies are used first to identify the intervals of interest, after which output error is evaluated within those intervals. This makes it possible to separate compensation that occurs while interference is still present from recovery that begins only after the command path has been restored. The method also keeps legitimate reference in the analysis when that reference changes and checks how the results vary with different tolerance levels and observation windows. The aim is therefore not to introduce a new controller, but to provide a reproducible way of assessing regulatory behaviour from recorded operational data.

OPEN DATA AND STUDY DESIGN

Provenance and analytical scope

The empirical source is the KIOS CoE SUC1 collection on DER energy management and control [4]. Its scenarios concern interference with measurements or battery commands in a cyber-physical sandbox. The present analysis uses only the S3 battery set-point file, `Electron_SUC1_S3_Pbss_ref_Attack_Measurements.csv` [5]. The evidence is a published sandbox trace, not a field trial in a commercial electricity network. Other files in the collection are not treated as replications.

The working extract was transcribed from the publicly rendered CSV and retains all 205 timestamps and five

source fields: `time_datetime`, `Phres_real`, `Phres_ref`, `Pbss_ref` and `Pbss_bus8_ref_real`. These represent time, the aggregate process output, its reference, the issued battery command and the command recorded at the battery side. Numerical values in the accompanying extract are rounded to six decimal places. The extract is not a byte-identical copy of the original CSV; source row identifiers and a raw-file verification utility are supplied for audit.

The analysed file covers 1115 s, from 15:45:50 to 16:04:25. The sampling is close to 5 s, but it is not perfectly regular. Of the recorded intervals, 110 are 5 s long, 93 are 6 s, and one is 7 s. Because of this variation, the calculations use the actual time difference between consecutive records rather than assuming a constant step. No missing values were filled in, and the data were not resampled to create additional points. Another limitation is that the CSV header does not specify the physical unit of the power variables. To avoid introducing an unsupported assumption, the original numerical values are kept as recorded. The analysis therefore relies mainly on relative errors, which allows the different episodes to be compared without treating the values as watts or kilowatts.

In the analysis, y represents `Phres_real`, r is `Phres_ref`, u is `Pbss_ref`, and v is `Pbss_bus8_ref_real`. The aggregate reference remains at 8000 until 16:01:15, after which it changes to 4000. The first recorded point shows a relatively large initial mismatch, so it is not used when calculating the pre-event baseline. It is nevertheless retained in the published data extract and in the overview figure. The baseline is therefore based on source rows 2–24 and covers 123 s under the left-held convention. This adjustment is used only for the baseline statistics and has no effect on how the disturbance events are identified.

Command alignment and event reconstruction

The issued command and the battery-side command are first aligned before their difference is evaluated. In the normal parts of the trace, the value of v at a given timestamp corresponds to u from the preceding row, not to u recorded in the same row. This pattern is treated as a feature of the logging sequence, not as evidence of communication delay. The aligned command mismatch is therefore defined as

$$d_k = v_k - u_{k-1}. \quad (1)$$

After the first row, contiguous observations with $|d| > 1$ native unit define a mismatch episode. Table 1 summarises the resulting intervals.

TABLE 1
COMMAND-MISMATCH EPISODES RECONSTRUCTED FROM THE S3 TRACE

Event	Recorded mapping	Mismatch interval*	N
E1	$10 \times \text{previous } u$	15:48:00–15:50:04	23
E2	$\text{previous } u + 10000$	15:52:08–15:54:01	21
E3	fixed at 10000	15:56:01–15:58:02	22
E4	held at 3948.340332	16:00:20–16:02:05	19

*The endpoint is the first restored-command timestamp and is excluded from the active interval. E4 begins at the first observable mismatch, not a known injection time.

The largest mismatch outside these intervals is 0.002594, whereas the first mismatch in the held-command episode is 14.216. The threshold is used only to separate the close numerical agreement seen in normal operation from the larger discrepancies in the trace. It should therefore be understood as a retrospective segmentation rule, not as a validated online detector. The four event intervals are identified from the command channels alone.

In E1, v is approximately ten times the preceding issued command, with a maximum residual of 0.016782 native units from this relation. In E2, v is about 10000 units above the preceding command, and the largest residual is 0.000153. During E3, v remains fixed at 10000. In E4, it is held at 3948.340332 while the issued command continues to change. These patterns describe what is present in the recorded data; they should not be interpreted as independently confirmed attack labels.

In E1, v is approximately ten times the preceding issued command, with a maximum residual of 0.016782 native units from this relation. In E2, v is about 10000 units above the preceding command, and the largest residual is 0.000153. During E3, v remains fixed at 10000. In E4, it is held at 3948.340332 while the issued command continues to change. These patterns describe what is present in the recorded data; they should not be interpreted as independently confirmed attack labels.

E4 is taken to begin at the first point where a mismatch can actually be observed in the log. This does not necessarily mark the true start of the command hold, because repeated identical commands make any earlier onset impossible to identify from these data. An episode is considered to end at the first row where the mismatch falls back below the chosen threshold. That row is still included in the recovery analysis. As a result, a large output deviation at the moment normal command alignment returns is not discarded, even when the output remains far from the legitimate reference.

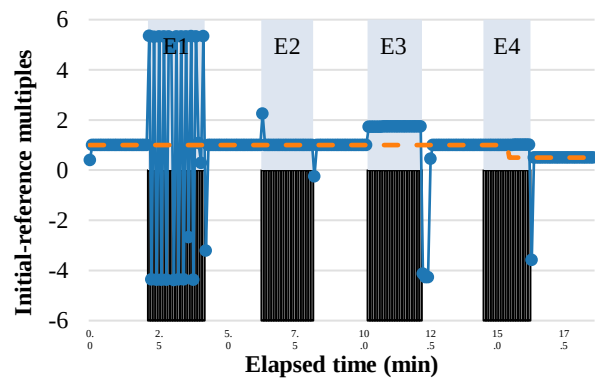


Fig. 1. Recorded output and legitimate reference, divided by the initial reference of 8000. Shading marks the four command-mismatch intervals; all 205 observations are retained.

ESTIMATION METHOD

Tracking error and the self-regulation score

The legitimate aggregate reference, not the potentially corrupted battery command, defines successful regulation. For every recorded observation, the dimensionless absolute error is

$$z_k = \frac{|y_k - r_k|}{|r_k|}, \quad r_k \neq 0. \quad (2)$$

Both reference values used in this record are above zero, so relative error can be calculated directly. The same approach would not be suitable for a zero or sign-changing reference, where a separate engineering scale would have to be defined. Here, the main tolerance is set to $\tau = 0.01$, meaning that an output is treated as within tolerance when its error is no more than 1% of the reference at that time. This 1% band is used only as an analytical criterion for comparing the recorded episodes. It should not be read as a grid-code requirement, protection threshold, or equipment-safety limit. For window W , observation k is assigned the duration w of its overlap with the interval from its timestamp to the next timestamp. T is the sum of these durations. This left-hold convention is used for numerical integration only and does not claim that the physical output is constant between samples. Normalised mean absolute error is reported in percent:

$$\text{NMAE}(W) = \frac{100}{T_W} \sum_k w_k z_k. \quad (3)$$

The peak error is 100 times the largest z in the window. Tolerance-band occupancy is the percentage of its duration for which the left-held observations satisfy $z \leq \tau$. Occupancy records persistence but ignores how far an out-of-band observation lies from the target. The complementary SRI is defined by

$$\text{SRI}_\tau(W) = \frac{1}{T_W} \sum_k \frac{w_k}{1 + z_k/\tau}. \quad (4)$$

The score lies in $(0, 1]$, equals one for zero error and equals one half when every observation is exactly at the selected tolerance. A high value indicates small observed errors relative to that tolerance. It does not establish stability, latent adaptive capacity or safe operation. The reciprocal form avoids negative scores and does not require normalisation by the largest error in this particular dataset. Its choice remains a modelling decision, which is examined through sensitivity analysis.

SRI is treated here as an additional indicator, not as a substitute for the usual measures of tracking error and maximum deviation. The four events are too few, and too closely related in time, to justify fitting data-driven weights. For that reason, the index is not presented as a statistically established property of the system. Any attempt to tune the weights from these episodes would risk making the result look more exact than the available evidence allows.

Compensation, recovery and observation windows

Compensation is recorded when the command mismatch is still active and the output returns to the tolerance band. Recovery after restoration is considered separately and starts from the first timestamp at which normal command alignment is re-established. A single in-band observation is not enough to count as recovery. Instead, three consecutive observations must remain within the tolerance band. The return time is taken from the first of these three points, while the confirmation time is measured to the third.

These results are tied to the sampling of the recorded data and should not be read as continuous-time settling times. For example, if the first in-band value appears 6 s after

command restoration, the actual return may have happened at any point between those two samples. The same limitation applies to the three-sample confirmation rule, since an excursion between recorded points cannot be ruled out. If the output does not return to the tolerance band during an active episode, the result is reported as “not observed in that interval” rather than as an infinite recovery time.

Each episode is examined over two time ranges. One covers only the period in which the command mismatch is present. The other adds 60 s after normal command alignment returns; this is called the complete-event window. The longer window is included because some of the largest deviations appear only after restoration. Its SRI is calculated in exactly the same way as for the active interval, but over a longer period that includes both disturbance and recovery. To check whether the result depends strongly on this choice, the same calculation is repeated with 30 s and 120 s post-restoration windows, provided that the extended window does not reach the next detected episode.

Disturbance-response ratio and implementation

An error-to-command-mismatch ratio is included as a diagnostic:

$$G(W) = \frac{\sum_k w_k |y_k - r_k|}{\sum_k w_k |d_k|}. \quad (5)$$

It is evaluated only where the denominator represents a material mismatch. G is not an induced system gain: command corruption can depend on the controller’s own output, and the record contains changing operating conditions. Nor does $1 - G$ necessarily lie between zero and one. A large mismatch denominator can make this quantity look favourable even when the intended process output is poorly regulated.

The method uses only a few time-stamped signals: the reference, the measured output, the command reaching the battery side, and the command issued in the previous record. After aligning the two command signals, the procedure marks the mismatch periods and updates the time-weighted error and SRI values as the trace progresses. It also keeps track of consecutive samples that fall inside the tolerance band. This means that SRI can be followed during an event, while the final event length and confirmed recovery time are known only once later samples have been recorded. The approach also assumes that the reference signal is reliable and that the output measurement has not been corrupted. If either of these conditions is not met, the interpretation used here would need to be changed.

All four episodes come from the same recorded run and were handled by the same controller. For that reason, the 205 samples are not treated as separate experiments. The results are used to describe the behaviour seen in this trace, not to make broader statistical claims. No significance tests are applied, no classifier is trained, and no comparison is made between different controller designs. The sensitivity analysis has a narrower role: it checks how much the reported results depend on choices made in the analysis, such as the tolerance level and the length of the recovery window.

EMPIRICAL RESULTS

Disturbance-specific responses

Before E1, the baseline normalized mean absolute error is 0.183%, every left-held observation is within the 1% band, and SRI is 0.8547. The value below one reflects small residual errors rather than an inferred defect in autonomy. Figure 1 shows the full recorded response, while Table 2 gives the active-interval estimates.

TABLE 2
ACTIVE-INTERVAL REGULATION AT A 1% RELATIVE-ERROR TOLERANCE

Event	NMAE, %	Peak, %	In band, %	SRI
B0	0.183	0.416	100.000	0.8547
E1	432.652	537.960	0.000	0.0070
E2	5.638	124.782	95.575	0.8558
E3	73.769	74.490	0.000	0.0134
E4	48.518	101.797	52.381	0.4090

B0 is the 123 s pre-event baseline (source rows 2–24). Means, occupancy and SRI use recorded interval durations; peaks use the observed samples.

E1 shows the clearest loss of regulation. Under the multiplicative disturbance, the output swings strongly above and below the reference, giving a mean absolute relative error of 432.652%. None of the recorded points falls inside the 1% tolerance band, and the SRI drops to 0.0070. The fact that the output later returns to normal after the command path is restored should therefore not be taken as evidence that the disturbance was rejected during the 124 s active interval. E3 also has no in-band observations, but its behaviour is different. The error remains much more stable, with an average of 73.769%, and the corresponding SRI is slightly higher at 0.0134.

E2 shows a very different pattern behaviour. At 15:52:08, the output is 17,982.57 while the reference is 8,000. By the next recorded point, 5 s later, it has moved to 7,994.89, even though the lag-aligned command offset is still close to 10,000. From that point onward, every observation during the active interval remains inside the 1% tolerance band. This gives a band occupancy of 95.575% and an SRI of 0.8558. After the first disturbed sample, the error stays below 0.305% of the reference for the rest of the episode.

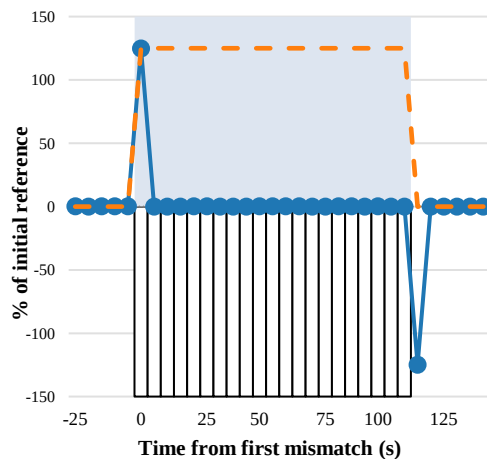


Fig. 2. E2 output error and command mismatch relative to 8000. The offset persists while the output returns to target. [Points are logged observations; connecting lines are visual guides].

This is evidence that the output can recover even when the command mismatch is still present. It does not mean that the command channel itself has returned to normal. During E2, the issued command becomes negative while the battery-side command stays positive, which is consistent with a compensating controller response. The trace does not, however, reveal the controller's internal state. Without additional logs, the exact mechanism cannot be identified, and there is no basis for claiming a specific attack-response algorithm.

Lost responsiveness under a held command

The held-command episode illustrates a different limitation of aggregate scoring. During its first 55 s, before the legitimate reference changes, normalized mean absolute error is only 0.312%, band occupancy is 100%, and SRI is 0.7720. At 16:01:15 the legitimate reference falls from 8000 to 4000, but the output remains near its former level. Over the remaining 50 s of command mismatch, mean absolute relative error rises to 101.545%, occupancy falls to zero, and SRI becomes 0.0098.

The active-period SRI for E4 is 0.4090, but this single value combines two very different parts of the episode. Before the reference changes, the held-command is still close enough to support the existing operating point. After the step, the same command can no longer produce the requested response. For this reason, the score should not be read as evidence of generally moderate resilience. Using a fixed reference of 8000 throughout would hide the loss of responsiveness that appears after the set-point change.

Restoration can produce the larger excursion

TABLE 3
RECOVERY AND PERFORMANCE INCLUDING 60 S AFTER COMMAND RESTORATION

Event	Return / confirm, s	Post peak, %	Complete-event SRI
E1	6 / 16	420.873	0.2595
E2	6 / 17	124.928	0.8499
E3	22 / 33	527.099	0.1935
E4	6 / 17	817.415	0.5333

Return is the first of three consecutive in-band observations; confirmation is the third. Post peak uses the 60 s recovery window. Complete-event SRI includes the active interval and the following 60 s.

At the 1% tolerance, the first three-sample-qualified returns occur 6, 6, 22 and 6 s after the respective restored-command timestamps. Their confirmation delays are 16, 17, 33 and 17 s. These observations establish recovery within the available log. They do not establish recovery without an external change: each post-restoration clock begins when the previously observed command mismatch disappears.

The most consequential restoration transient follows E3. Its largest active-period error is 74.490% of the reference. After the fixed command is released, the recorded output reaches −34167.88 against the target of 8000, giving a 527.099% error. The post-restoration peak is therefore 7.08 times the active-period peak. The issued command has meanwhile drifted to −125702.42 by the final active sample. This is consistent with accumulated corrective action during the imposed command, but integral windup remains an

interpretation rather than a directly observed controller state.

E4 also incurs a larger deviation when the held command is released: the post-restoration peak is 817.415% of its then-current 4000 reference, compared with 101.797% during the active interval. In E2, removing the compensated offset causes a second transient of opposite sign. The physical consequences of restoration are therefore part of the event, not a reason to end the evaluation.

Adding the 60 s recovery period changes the comparison between E1 and E3. During the active disturbance, E3 has the higher SRI, but over the complete event the scores become 0.2595 for E1 and 0.1935 for E3. Both episodes show poor regulation while the mismatch is active, but E3 is followed by a larger and longer recovery transient, which lowers its overall score. The same change in ordering appears when 30 s and 120 s recovery windows are used. This comparison refers only to the two recorded events and should not be taken as a general ranking of the control architectures.

Sensitivity and the limits of a single ratio

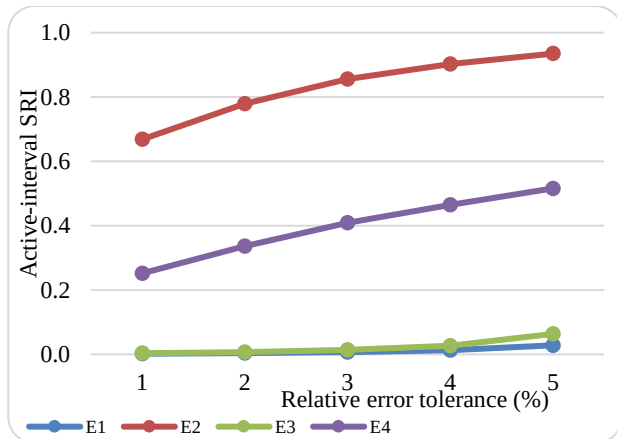


Fig. 3. Active-interval SRI at five declared error tolerances.

The active-interval ordering does not change across the tested tolerances from 0.25% to 5%: E2 remains highest, followed by E4, E3 and E1. The actual SRI values do change, however. E2 varies from 0.6682 to 0.9348, while E4 ranges from 0.2518 to 0.5158. Any reported SRI should therefore be read together with the tolerance used to calculate it. Using equal weights for all samples instead of duration-based weights changes the active scores by at most 0.0042 in this trace. Even so, the timestamp-based calculation is retained because the sampling intervals are not perfectly uniform.

Recovery is more sensitive when the tolerance band is very tight. At 0.25%, the return delays are 11, 6, 22 and 34 s for E1–E4. Under this condition, changing the requirement from two consecutive in-band observations to four can noticeably shift the reported recovery time. At the 1% tolerance used in the main analysis, however, the entry delays remain the same. Varying the command-mismatch threshold between 0.01 and 10 native units also identifies the same four episodes, although a stricter threshold makes the start of E4 appear slightly later.

The diagnostic G values are 0.2306, 0.0451, 0.0873 and 0.2509 for E1–E4. Used on its own, 1-G would give E3 a

high value of 0.9127, even though none of its observations falls inside the tolerance band. This happens because the denominator increases as the issued command moves farther from the imposed value of 10000. For that reason, disturbance-normalized ratios are useful only as supporting diagnostics and should not replace performance measures based on the actual reference.

DISCUSSION

The results point to three different behaviors. In E2, the output returns to target while the command mismatch is still active. The other episodes do not maintain good regulation throughout the disturbance. All four eventually return after command restoration, but the largest deviation can occur during that recovery period. Calling this simply “autonomous recovery” would hide these differences. For DCS applications, this method can be used as a local monitoring measure when a reliable reference and process signal are available. Similar local scores could later be combined across nodes, but this study does not implement a multi-node consensus estimator.

A low SRI indicates that a subsystem needs attention, but it is not a controller update rule. Gain adaptation or topology changes would require a plant model, operating constraints and separate closed-loop validation.

The main limitations are the small number of events, the single recorded trajectory and the relatively coarse sampling. The episodes are not independent experiments, and the data cannot prove continuous-time stability, restored communications or absence of hidden internal effects. SRI should therefore be read as a practical trace-based indicator, not as a direct measure of system intelligence or adaptive capability.

CONCLUSION

This study looks at self-regulation in a public cyber-physical control trace and treats the disturbance period and the recovery period separately. Four command-mismatch episodes are taken from 205 observations. Their limits are set from the lag-aligned command signals, not from the output error. This matters because the four episodes do not behave in the same way. In one of the cases, the output returns to its target even though the additive command offset is still present. In another, the system stays close to the target while the command is held, but this changes once the reference itself moves. The fixed-command episode is different again: its largest deviation occurs after normal command operation has already been restored. A single recovery score would not show these differences clearly. The main practical result is not that one universal index characterises resilience. It is that a bounded, tolerance-referenced score becomes useful when accompanied by error magnitude, band occupancy, declared exposure windows and sample-resolved return times.

DATA AND CODE AVAILABILITY

The original S3 CSV is publicly available in Zenodo record 12773981 [4], [5], accessed 18 September 2026. The accompanying reproducibility package contains the five-field, 205-row extract with source row identifiers, the

analysis and figure-generation scripts, numerical result tables, sensitivity outputs, and a utility for comparison against the original CSV. No synthetic measurements are included.

REFERENCES

- [1] Simpson-Porco, J. W., Q. Shafiee, F. Dörfler, J. C. Vasquez, J. M. Guerrero, F. Bullo. Secondary frequency and voltage control of islanded microgrids via distributed averaging. *IEEE Transactions on Industrial Electronics*, Vol. 62, No. 11, 2015, pp. 7025–7038. doi:10.1109/TIE.2015.2436879. Open manuscript: arXiv:1504.06784.
- [2] Ge, P., Y. Zhu, T. C. Green, F. Teng. Resilient secondary voltage control of islanded microgrids: An ESKBF-based distributed fast terminal sliding mode control approach. 2020. doi:10.1109/TPWRS.2020.3012026. Open manuscript: arXiv:2007.11494.
- [3] Mishra, S., T. Kwasnik, K. Anderson. Microgrid resilience: A holistic and context-aware resilience metric. 2021. arXiv:2106.09640.
- [4] Hadjidemetriou, L., M. Asprou, I. Ciornei, C. Charalambous, E. Tekki. Dataset SUC1/S1–S4: Cyberattack scenarios on DER energy management and control. Zenodo, Version v1.0 author version, 2024. doi:10.5281/zenodo.12773981.
- [5] KIOS CoE SUC1 data collection. *Electron_SUC1_S3_Pbss_ref_Attack_Measurements.csv*. Raw S3 file in [4], 2024. Public file accessed 18 September 2026. File-specific access is provided in the accompanying code package.
- [6] Kia, S. S., B. Van Scoy, J. Cortés, R. A. Freeman, K. M. Lynch, S. Martinez. Tutorial on dynamic average consensus: The problem, its applications, and the algorithms. 2018. arXiv:1803.04628v2.
- [7] Tekki, E., L. Hadjidemetriou, M. Asprou, I. Ciornei. Dataset for KIOS CoE sandboxing use-case SUC5 corresponding to cyber attacks affecting the control of active distribution grids and microgrids. Zenodo, Version v1, 2024. doi:10.5281/zenodo.12967154.
- [8] Sen, Ö., L. Bader, M. Emara, M. Serror, B. Schwalm, J. Tsurumi, T. Oba, T. Izawa, Y. Ujiie. Data4Cyber: A labeled cyber-physical dataset for distribution-grid management systems under representative OT attacks. Zenodo, Version v2, 2026. doi:10.5281/zenodo.20004612.