2024 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT) | 979-8-3503-1360-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISGT59692.2024.10454158

Data Integrity and Cyberattack Detection using Dynamic Watermarking for Resilient Microgrids

Pradeep Kumar Mallaiah, *Member, IEEE,* Ankit yadav, *Member, IEEE,* Gelli Ravikumar, *Member, IEEE,*
Department of Electrical and Computer Engineering, Iowa State University, Ames, IA, United States, 50010
Email:pradi@iastate.edu, ankity@iastate.edu, gelli@iastate.edu

***Abstract*—Cyber-attacks on microgrid systems, especially data manipulation attacks such as replay attack and Denial-of-Service (DoS), causes communication delay and unstable responses. Even though control strategies such as Consensus Control (CC) are able to coordinate electric current and voltage flow, they are at risk of malicious attacks. Communication delay leads to undetected changes in line current, and voltage leads to incorrect responses from the consensus controller, which overloads the microgrid in milliseconds. To address these challenges, this paper presents an Observer System (OS) based Dynamic Watermark (DW) detection model that detects delay-induced cyber-attacks during steady states and load fluctuations. We have developed a Grid-Specific Dynamic Watermarking (GSDW) signal that enhances real-time detection capabilities, resulting in a real-time non-zero residual showing cyber attack dynamics in the proposed observer system. Our detailed case study demonstrates real-time attack detection and prevention, ensuring the stability and integrity of Microgrid (MG) systems under challenging cyber threat conditions. Comprehensive simulations and validation demonstrate the practicality and efficacy of our approach in mitigating risks posed by delay-induced cyber attacks in MG systems.**

***Index Terms*—Communication delay, Grid-Specific Dynamic Watermarking, Real-time detection, Steady state analysis.**

I. Introduction

In the last eight years, there has been a significant rise in the deployment of renewable energy, which has increased the number of island MG [1]. These MGs have received significant research attention, particularly those based on DC technology like DC microgrids (DCMGs) [2]. To make MGs stable and ensure cyber security, different hierarchical control architectures have been explored in existing literature [3], [4]. Control architectures have primary layers for stability and secondary/tertiary layers for load sharing. Incorporating communication networks raises security concerns and requires monitoring for abnormal behavior.

MGs are vulnerable to cyber attacks that disrupt operations and cause improper power distribution. One such attack is a replay attack, which involves recording and replaying data transmitted over the communication network. This kind of attack poses a significant challenge for monitoring efforts [5]. Detecting such cyber attacks is complicated as they mimic the statistical characteristics of normal behavior [6]. Developing effective countermeasures to mitigate replay attacks and other cyber threats is challenging [7]. Ensuring the seamless integration of cybersecurity measures within the complex and interconnected environment of Cyber-Physical Systems (CPS), especially microgrids, presents a challenge [8]. The challenge lies in maintaining the system's operational efficiency with integrity, performance, and stability while maintaining security [9]. The limitations of CCs are evident in their inability to recognize alterations in line current and voltages. Their vulnerability to cyber attacks ultimately results in improper data exchange and instability in the MG [10]. In addition, real-time communication delay disrupts the coordinated current distribution, leading to voltage deviations and system instability. The traditional CCs are also limited in adapting to changing network conditions or unexpected disturbances. Computational and communication challenges arise when scaling up CCs for complex MG systems.

To address the above significant challenges and to enhance CC-based MG resilience, we propose an observer-based GSDW technique for cyber-attack detection in MGs. Unlike conventional methods that use a simple non-zero power sawtooth waveform as a watermark that can be easily predicted [11], our GSDW signal allows for real-time detection and precise differentiation between regular fluctuations and intended cyber-attacks in real-time by remaining resilient to load and system condition changes. Our proposed architecture includes an interconnected grid system with a Proportional-Integral-Derivative (PID) controller to maintain a stable voltage and a detection method that involves Watermarking System (WS) to embed GSDW and De-Watermarking System (DWS) to extract and make precise voltage adjustments. The proposed architecture supports a multi-microgrid system and improves system integrity monitoring without additional costs.

The key contributions of the proposed work are:

- Designed a unique GSDW signal to improve cyber-attack detection in power measurements and generate non-zero residuals reflective of attack dynamics.
- Developed a real-time, OS-based detection system to detect malicious cyber-attacks
- Proposed an efficient watermarking subsystem (WS) for power measurements without affecting average power and a De-watermarking subsystem (DWS) for precise voltage adjustment.
- Modeled and performed a reply attack to validate the performance of the proposed GSDW.

II. Proposed Observer-based GSDW for cyber attack detection

This section introduces the proposed framework with its essential features, including modeling approaches, representations of dynamic behavior, and a watermarking model for MG.

A. Proposed Architecture of OS-based GSDW system

Fig. 1 represents the proposed architecture. It consists of an interconnected MG system, a detection system, and a CC located between the communication channels connected to the integrated grid system and the proposed detection system, forming a closed loop in the CPS. The interconnected MG

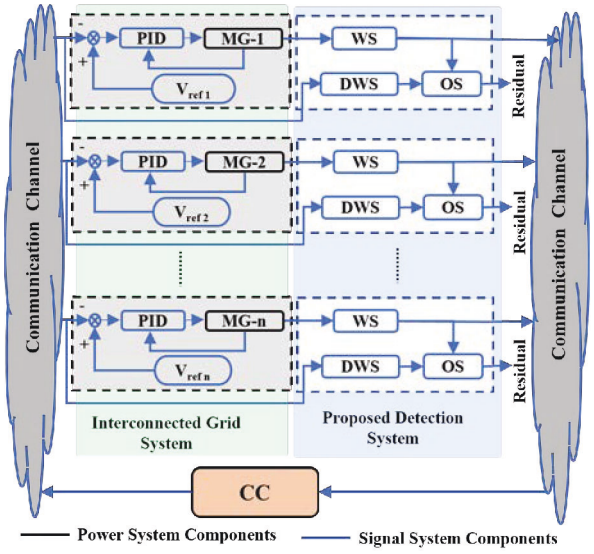

Fig. 1. Architecture of proposed OS-based GSDW

system connects the multiple MGs represented by equivalent models labeled MG-1 through MG-n. Each MG implements a PID controller to maintain stable voltage levels. The proposed detection method has two integral components: WS and the DWS. The WS embeds an authentic GSDW into the system. The GSDW is characterized by a zero-average power and periodicity(0.02sec), which transforms the attack into corresponding residuals or discrepancies, reflecting the dynamics of the attack in real-time. The DWS then subtracts the incorporated GSDW before calculating the voltage reference adjustments V_{ref} to the MG. The modular architecture can support a range of MGs from 1 to n, making it practical for use across multiple grid systems. The OS generates residuals to detect deviations between actual and expected measurements, providing an early detection against potential threats and maintaining the security and efficiency of the MG model.

B. DC Microgrid Model

The proposed concept has been implemented on an interconnected DC microgrid (DCMG) power system. Fig. 2 presents an equivalent model of DCMG. It consists of an equivalent DC-DC converter, PID controller, energy source, and load. The load connected to the DC-DC converter output is denoted by L. The converter allows bidirectional power flow following the PID controller commands.PID controller are widely adopted in MG control due to their proven effectiveness in regulating voltage and ensuring system stability at the primary level. The PID controller regulates the output voltage V_{L} in accordance with V_{ref}. This regulation is important for stable MG operation. The main objective of the MG operation

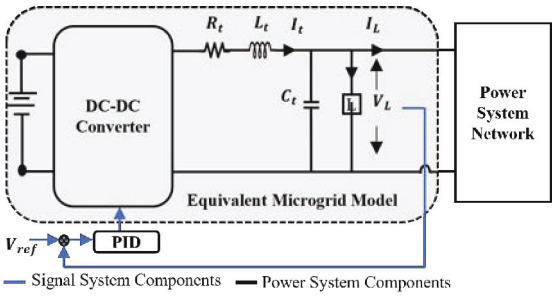

Fig. 2. Equivalent DCMG connected to the power system network

is maintaining a stable and accurate output voltage while the loads, disturbances, and other operating conditions varies continuously. However, in cases of cyber-attacks, the PID controllers may fail to contribute to stability. The significant changes in the connected load cause a variation in output voltage.

C. Design of Proposed Zero-average Power GSDW signal

The methodology for generating a secure zero-average power GSDW signal is depicted in Fig. 3. A signature-based image is used to create a unique and authentic identifier that ensures the signal's integrity and prevents unauthorized alterations. The input is a 550x1280 pixel signature image, processed using DWT to extract the feature and to form a time series GSDW signal with a 0.02-second periodicity.

Mapping a signature authentic image to a time series involves converting its pixel values into a sequential representation over time using an Eq. (1).

$$X(t) = \sum_{i=1}^{N} \sum_{j=1}^{M} (P_{ij} \times f_{type}(I) \times f_{length}(T)) \qquad (1)$$

The Eq. (1) defines a time series represented by $X(t)$. N and M are the image dimensions, and P_{ij} is the pixel value at row i and column j. I is the image type, and T is the desired time series length. Two functions, $f_{type}(I)$ and $f_{length}(T)$, are used to convert the image type to a scaling factor and to scale pixel values based on the desired time series length.

The functions $f_{type}(I)$ and $f_{length}(T)$ vary based on the application and image conversion to time series.$f_{type}(I)$ is a constant factor for grayscale images, and it is a factor depends on the number of color channels for RGB images. $f_{length}(T)$ is a linear scaling function that divides pixel values by T to normalize longer time series.

The designed GSDW is then incorporated into transmitted voltage and current measurements, enhancing robustness against attacks and minimizing signal vulnerability.

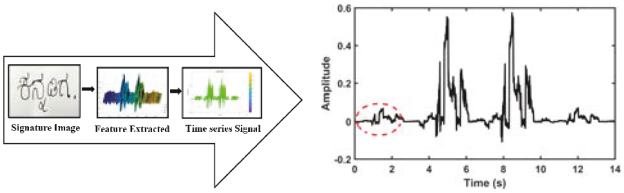

Fig. 3. Design procedure of GSDW signal and its plot in MATLAB plot

979-8-3503-1360-4/24/$31.00 ©2024 IEEE

We used a DWT for feature extraction and grayscale conversion to generate a time-series signal associated with computational complexity. These combined techniques produced the GSDW signal with 61568 samples, shown in Fig. 3. A portion of the signal is shown in the window with a dotted circle (red) with an amplitude of 1 and frequency of 2 Hz, suitable for Phasor Measurement Unit [PMU] data rates (practical application), which were used to incorporate voltage and currents. In the proposed GSDW, we can use the entire time series or a portion of it in the CPS loop (depending on the robustness level). GSDW is made with MATLAB R2023a. It is static, which maintains the authenticity and stability of the watermark. Any alteration attempt would be quickly detected. The security and reliability of information rely on an authentic GSDW, which is challenging to replicate or tamper. The GSDW is represented as,

$$w(t) = \beta \times \alpha \times \text{GSDW}(\omega t) \tag{2}$$

Where, α, β, and ω determine the watermark signal's strength, scaling, and frequency, and t represents the time step and modifying any of these parameters would result in a fundamentally different watermark.

D. Watermarking Subsystem and Communication Security

The WS enhances the system security by injecting a GSDW signal into the voltage and current measurements before these measurements are communicated to the CC. The GSDW serves as a unique identifier and a reference signal for anomaly detection. The WS helps convert the delay induced by attacks into perceptible anomalies. The $GSDW(t)$ is then added to the original voltage $V_{\text{i}}(t)$ and current $I_{\text{i}}(t)$ measurements, resulting the modified measurements $V_{\text{i}}{}'(t)$ and $I_{\text{i}}{}'(t)$:

$$V_{\text{i}}{}'(t) = V_{\text{i}}(t) + GSDW(t) \tag{3}$$

$$I_{\text{i}}{}'(t) = I_{\text{i}}(t) + GSDW(t) \tag{4}$$

The GSDW is characterized by zero-average power and periodicity(0.02sec). Thus it doesn't interfere with the normal operation of the MG as shown in Fig. 9.

E. De-watermarking Subsystem and Control Integrity

The DWS maintains the integrity of the control actions within the distributed control system. This subsystem subtracts the previously injected GSDW from the signal obtained at the communication channel after the CC and before calculating voltage reference adjustments to the MG. The removal of the GSDW from the voltage and current measurements is represented as:

$$V_{\text{i}}(t) = V_{\text{i}}{}'(t) - GSDW(t) \tag{5}$$

$$I_{\text{i}}(t) = I_{\text{i}}{}'(t) - GSDW(t) \tag{6}$$

Where, $V_{\text{i}}(t)$ and $I_{\text{i}}(t)$ represent the corrected voltage and current measurements after watermark removal. $V_{\text{i}}{}'(t)$ and $I_{\text{i}}{}'(t)$ are the modified measurements in the presence of $GSDW(t)$ signal at time t. The DWS removes the watermark to ensure the power measurements remain unaffected by the GSDW. This helps to maintain the operational integrity of the control system. The WS and DWS work together with the OS to form anomaly detectors to effectively detect attacks.

F. Observer System with Kalman Filter Residual Generation

The observer system estimates MG model states (voltage $V_{\text{out,i}}(t)$ and current $I_{\text{out,i}}(t)$) even in the presence of GSDW. The primary function of the OS is to generate residuals, which are the discrepancies between the actual measurements and the estimated states as predicted by the Kalman filter. These residuals are a measurable way to identify inconsistencies, discrepancies, or anomalies in the system that may result from an attack.

$$V_{\text{err,i}}(t) = V_{\text{out,i}}(t) - \hat{V}_{\text{out,i}}(t) \tag{7}$$

$$I_{\text{err,i}}(t) = I_{\text{out,i}}(t) - \hat{I}_{\text{out,i}}(t) \tag{8}$$

where, i is the subsystem/component index. If the residuals surpass the integrity margin thresholds of the valuable measurements, it indicates the presence of a potential attack.

Using the kalman filter in each OS helps the system to estimate the state based on the available measurements. Thus os generates the non-zero residuals to detect anomalies or attacks, ensuring system integrity and security. By analyzing the behavior measurements of each microgrid, including communication measurements and rated limited capacity, it's possible to determine the threshold of disturbance and attack bounds.

III. Simulation Setup and Implementation Environment

This section investigates the effectiveness of the proposed OS-based GSDW as a defensive mechanism for CPS against cyber attacks.

Case-Study: We conducted a test-bed implementation of replay attacks on DCMGs using Control System Toolbox, Simulink, and Simscape Electrical. Our proposed model system includes two interconnected DCMGs that operate under a CC loop for load sharing. The communication channel transmits line currents among DCMGs and is susceptible to potential replay attacks, as illustrated in Fig. 4. We conducted simulations for a sample time (T_s) of 50 microseconds.

At t = 3sec (T_{a}), a replay attack was introduced with a delay of 0.1sec (t_0), resembles the practical scenario of the attacker introducing the modified delay signal which mimics the statistical characteristics of normal behavior which was unable to detect by the monitoring system. This resulted in line current overshoots and damage to the MG when the workload changed at t = 4sec.

Reply attack model: We conducted a replay attack that negatively impacts signal integrity in a system by delaying signals in the communication channel. This can cause discrepancies in the control loop, especially during steady-state conditions. The replay attack is mathematically represented as :

$$y_a(t) = y(t) + \beta(t - T_a)[-y(t) + y(t - t_0)] \tag{9}$$

where $y_a(t)$ is the modified signal after the attack and $y(t)$ is the original signal. $\beta(t - T_a)$ is an activation function.This introduces a delay at time $T_a = 3$ sec, and $t_o = 0.1$ sec, which denotes the delay of the attack. The modified signal is an undetectable anomaly created by adding a transformed version

979-8-3503-1360-4/24/$31.00 ©2024 IEEE

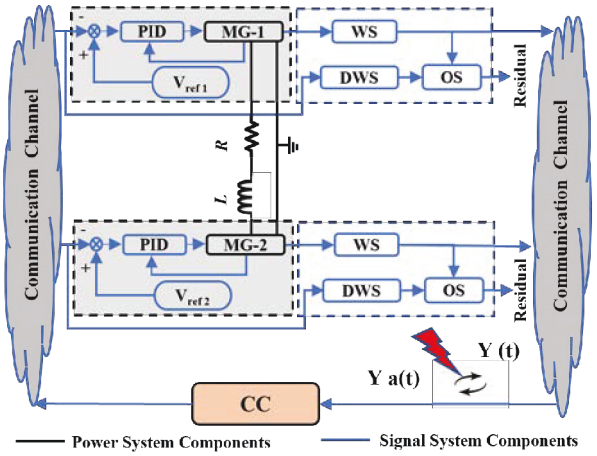

Fig. 4. Proposed OS-based GSDW detection system with two interconnected DCMGs in the presence of Replay Attack

of the original signal to a delayed and inverted portion of itself that mimics normal behavior along with original measurements [6]. This compromises the system's decision-making, leading to incorrect responses, instability, and potential consequences within CPS.

IV. Result and Discussion

This section evaluates and validates the performance of the proposed system on two interconnected DC microgrids in three scenarios (A, B, C). *Scenario-A* examines the MG load distribution and balance during regular MG network operation without external threats, *Scenario-B* examines a replay attack as a case study, and *Scenario-C* uses GSDW to detect anomalies and demonstrates how proposed technique enhance attack dynamics and improve real-time detection capabilities. In all scenarios (A, B, C), sub-figure (a) shows voltage values: $V_{\text{err},1}$ (black dotted) for MG-1 and $V_{\text{err},2}$ (red continuous) for MG-2. Sub-figure (b) displays current values: $I_{\text{err},1}$ (black dotted) for MG-1 and $I_{\text{err},2}$ (red continuous) for MG-2.

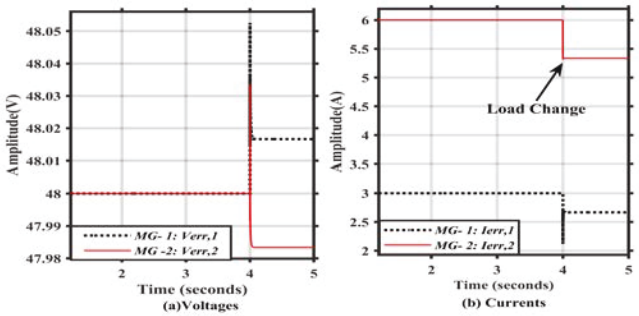

Fig. 5. Balanced load distribution in MG1 and MG2 during normal operation without attack and GSDW Signal, with load change point at t = 4 seconds

Scenario-A: Initially, the MG network operates normally without any attacks. The load sharing between MGs is well-distributed. A notable load change occurs at t = 4 sec. Both MGs (1, 2) exhibit currents proportionate to their rated capacities, reflecting balanced load distribution without attack. as shown in Fig. 5.

Scenario-B: We examined a replay attack with a delay of 0.1 sec (t_0) and an inverted portion of itself introduced at t = 3 sec

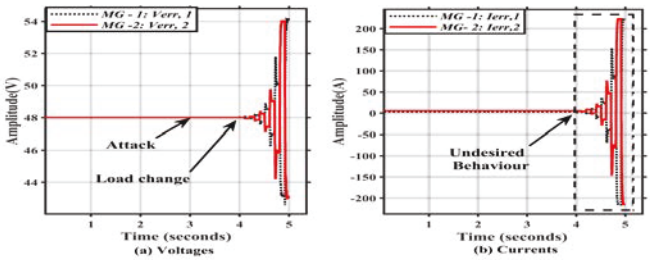

Fig. 6. Existing system's inability to detect Replay attack dynamics at t = 3 seconds, followed by undesired behavior after load change in the absence of GSDW

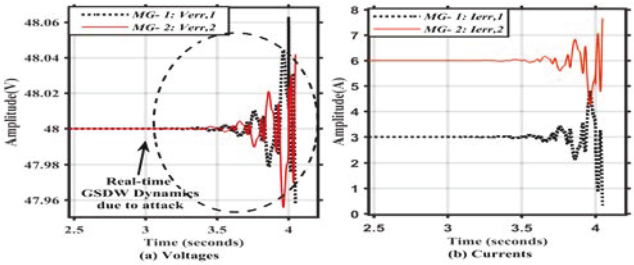

Fig. 7. Real-time detection of Replay Attack inducing Non-zero residual (discrepancy) at t = 3 seconds using proposed GSDW signal

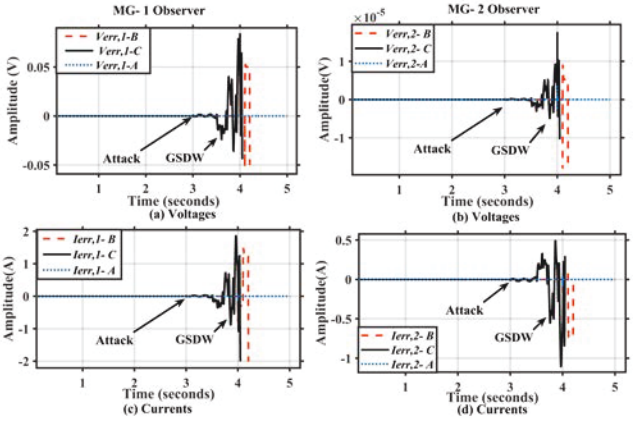

Fig. 8. Observer System behavior for DCMG-1 and DCMG-2 across Various Scenarios

(T_a) in the communication channel. This reply attack caused the consensus control to misinterpret line current changes. This misinterpretation occurred especially during steady-state operation, resulting in uncontrolled currents exceeding their capacities within 300 milliseconds of a load change. As a result, drastic uncontrolled dynamics in voltage were observed after the load change at t = 4 seconds, leading to instability. This unstable state is indicated by a rectangular box (in a dotted black line) in Fig.6 (b), indicating MG instability.

Scenario-C: Upon introducing the proposed GSDW with a signal amplitude of 0.1 and a frequency of 2 hertz. GSDW successfully detected delay as a disturbance in real-time, particularly during steady state operation. As shown by the dotted circle in Fig. 7 (b), the replay attack was transformed into a detectable anomaly. The proposed GSDW signal amplified the attack dynamics, which improves the system's ability to perceive and detect the anomaly. The discrepancy and its pattern began at the point of attack at t = 3 sec, demonstrating that the proposed GSDW effectively detects the attack exactly

979-8-3503-1360-4/24/$31.00 ©2024 IEEE

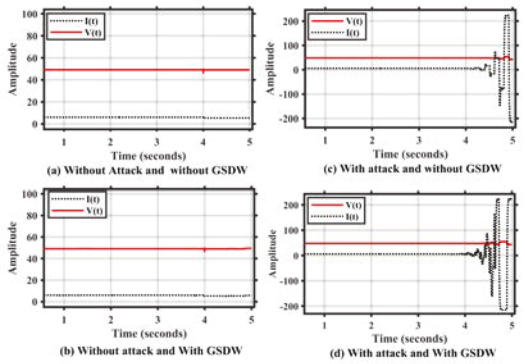

Fig. 9. Voltage and current status in various scenarios at Consensus Controller

at the time it occurs (real-time), shown as dotted circle (in black) in Fig. 7 (b).

A. Analysis of V(t) and I(t) status in various scenarios at Consensus Controller

Fig. 9 represents the voltage (red continuous lines) and current dynamics (black dotted lines) at CC. Subfigures (a) and (b) show that GSDW does not interfere with power signal measurements and preserves data integrity. In subfigure (c), a replay attack leads to uncontrolled currents surpassing capacity limits within just 300 milliseconds of a load change. Subfigure (d) demonstrates the effectiveness of GSDW in detecting attacks in real-time and reducing uncontrolled currents.

B. Comparitive analysis of all scenarios in OS window

The observer system window shown in Fig. 8 concisely presents the results of scenarios A, B, and C. Subfigures (a) and (b) display the voltage residuals, while subfigures (c) and (d) show the current residuals for MG-1 and MG-2.

TABLE I
COMPARISON OF CONVENTIONAL WATERMARK AND GSDW

Property	**Conventional-DW**	**GSDW**
Amplitude	0.198	0.4786
Rise Time	39.6 µs	12.884 µs
Slew Rate (Up)	4,000 V/ms	2.972 V/s
Slew Rate (Down)	-400 V/ms	-1.912 kV/s
Max Value	+0.2	+0.4491
Min Value	-0.09998	-0.06905
Peak to Peak	0.2	0.114
Mean	0.00505	0.0001156
Median	0.00051	0.0001824
RMS	0.08802	0.02654
Real-time Detection	No	Yes
Capturing Dynamics	No	Yes
Detection at Steady State	No	Yes
Distortion Rate	Medium	Low
Impact of Attack on signal	High	Low

The blue dotted line represents a balanced load distribution with zero residuals, implying the system operates safely without an attack (scenario A), while the red dashed line shows an undesirable behavior where the current limit is exceeded within 300 milliseconds (scenario B). The continuous black line confirms the GSDW's effectiveness in detecting attacks in real-time (scenario C). Based on the data presented in Table I, it is clear that the proposed GSDW offers numerous advantages over conventional watermarks [11], including smaller transitions, higher amplitude, and robustness. GSDW is also associated with faster rise and fall times to minimize disruption, and alignment with zero mean and median values to maintain signal baseline and a lower Root Mean Square (RMS) value, resulting in minimal distortion.

V. CONCLUSION

In this paper, we presented the OS-based GSDW detection techniques to ensure the integrity and stability of DCMG. The proposed WS, DWS, and OS enhanced the detection capabilities. The case study outlined the drastic system disruptions due to failure or late detection of attacks. The proposed GSDW signal, injecting a known signal into the communication channel, was presented as an effective strategy for real-time attack identification and mitigation, significantly enhancing the cybersecurity of DCMGs. This technique can be extended to Alternating Current microgrid (ACMG) systems. It brings benefits such as streamlined power flow management, reduced vulnerability to disturbances, and improved resilience against synchronization complexities and voltage-related issues in ACMG and DCMG environments. Further research is needed to adapt and test the technique for different system setups and attack scenarios to ensure reliability in practical applications.

ACKNOWLEDGMENT

This research is funded partly by US NSF Grant #CNS 2105269, US DOE CESER Grant DE-CR000016, and Iowa Energy Centre Grant #21-IEC-009.

REFERENCES

[1] A. J. Gallo, M. S. Turan, F. Boem, G. Ferrari-Trecate, and T. Parisini, "Distributed watermarking for secure control of microgrids under replay attacks," vol. 51. Elsevier B.V., 1 2018, pp. 182–187.

[2] B. Abdolmaleki and G. Bergna-Diaz, "Distributed control and optimization of dc microgrids: A port-hamiltonian approach," *IEEE Access*, vol. 10, pp. 64 222–64 233, 2022.

[3] A. J. Gallo, M. S. Turan, F. Boem, T. Parisini, and G. Ferrari-Trecate, "A distributed cyber-attack detection scheme with application to dc microgrids," *IEEE Transactions on Automatic Control*, vol. 65, pp. 3800–3815, 9 2020.

[4] P. Cheng, L. Shi, and B. Sinopoli, "Guest editorial special issue on secure control of cyber-physical systems," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 1–3, 2017.

[5] D. I. Urbina, D. I. Urbina, J. Giraldo, A. A. Cardenas, J. Valente, M. Faisal, N. O. Tippenhauer, J. Ruths, R. Candell, and H. Sandberg, *Survey and new directions for physics-based attack detection in control systems*. US Department of Commerce, National Institute of Standards and Technology, 2016.

[6] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 93–109, 2015.

[7] R. M. Ferrari and A. M. Teixeira, "Detection and isolation of replay attacks through sensor watermarking," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 7363–7368, 2017.

[8] C. De Persis, E. Weitenberg, and F. Dörfler, "A power consensus algorithm for dc microgrids," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 10 009–10 014, 2017.

[9] L. Meng, Q. Shafiee, G. F. Trecate, H. Karimi, D. Fulwani, X. Lu, and J. M. Guerrero, "Review on control of dc microgrids and multiple microgrid clusters," *IEEE journal of emerging and selected topics in power electronics*, vol. 5, no. 3, pp. 928–948, 2017.

[10] S. Islam, S. De, S. Anand, and S. R. Sahoo, "Consensus based ideal current sharing controller for dc microgrid," in *2020 IEEE 14th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG)*, vol. 1. IEEE, 2020, pp. 200–205.

[11] MathWorks, "Detect replay attacks in dc microgrid using watermarking," https://www.mathworks.com/help/control/ug/detect-replay-attacks-in-dc-microgrid-using-watermarking.html, 2023.