A Tri-Modular Human-on-the-Loop Framework for Intelligent Smart Grid Cyber-Attack Visualization

Aditya Sundararajan, Tanwir Khan, Haneen Aburub, Arif I. Sarwat, and Shahinur Rahman
Department of Electrical and Computer Engineering
Florida International University
Miami, Florida 33174 USA
Email: {asund005, tkhan016, haburub, asarwat, srahm026}@fiu.edu

***Abstract*—To minimize the effort required by human security operators in understanding and resolving attacks on the smart grid cyber-physical system, automated detection, prevention and mitigation tools have been integrated into the infrastructure. However, existing visualization frameworks at command and control centers present information from such tools in a non-intuitive, non-contextual format, reducing the situation awareness and timeliness of decisions. There is a need for frameworks that can contextualize the data in a human-understandable format prior to visualizing. To this end, the paper conducts a high-level review of existing literature, and introduces a conceptual human-on-the-loop framework of three modules: data analyzer comprising Kafka, Apache Spark and R, classifier comprising a deep neural network, and situation-aware decision-maker comprising a learning-based cognitive model. Preliminary proof of concept is shown for data analyzer by applying it to contextualize alerts from multiple photovoltaic systems in Florida.**

***Index Terms*—smart grid, cyber-physical security, human-on-the-loop, situation awareness, Apache Spark, data processing.**

I. Introduction

Recent successful cyber-attacks involving the BlackEnergy3 malware on the Ukrainian smart grid in 2015 and Stuxnet worm on the Iranian nuclear power plant in 2009 have increasingly targeted the smart grid infrastructure [1]. Increased penetration of Renewable Energy Systems (RESs) like solar Photovoltaic (PV), and proliferation of ubiquitous IoT sensors capable of bidirectional communication have made the power grid more vulnerable to potential cyber-physical attacks (henceforth referred to as attacks in this paper) [2]–[4]. A successful attack on one realm (cyber or physical) has significant impacts on the other (physical or cyber, respectively) [5].

Extending from the National Institute of Standards and Technology (NIST) smart grid framework, five key logical, interdependent components can be identified, shown under the Physical Realm in Fig. 1. Correspondingly, logical components on the cyber realm can be broadly divided into three: availability, integrity and confidentiality. The North American Electric Reliability Corporation (NERC) has stipulated multiple standards including the Critical Infrastructure Protection (CIP) guidelines for physical security, personnel training, information protection, access control, and more. NERC also recommends the inclusion of human in the loop of cybersecurity design solutions in addition to an increasing implementation of automated tools [6]. However, the increasing performance and reliability of automated tools such as Intrusion Detection and Prevention Systems (IDS/IPS), firewalls, honeynets and honeypots, etc. imply more cyber-physical events for the humans to analyze and resolve in the same period of time. This places an unprecedented amount of stress on the operators, and increases the room for erroneous decisions. The non-intuitive visualization of such events without first generating any context has compounded the problem. Many visualization interfaces still rely on manual parsing and analysis of unstructured security data [7], [8]. Hence, there is a critical gap between the rate at which new events are visualized and the rate at which the human operators can capture all such visualized events before they lose their time-associated value. This gap creates a need to develop approaches that convert the machine-generated wealth of data into human-understandable format (called actionable information) that can yield well-informed decisions [9], [10].

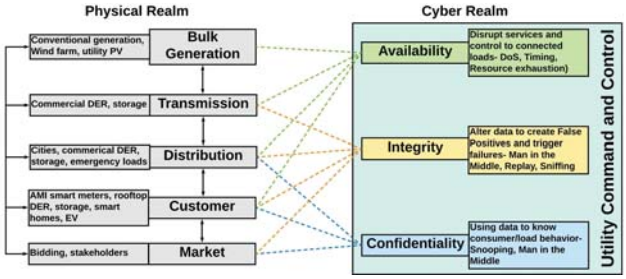

Fig. 1. Broad view of smart grid cyber-physical interdependencies

The main contributions of this paper are twofold: **1)** Steering the research towards proactive attack resolution by introducing a conceptual tri-modular, human-on-the-loop framework to transform useful, contextual data into actionable human-understandable format; and **2)** Demonstrating a preliminary proof-of-concept by applying one of the modules of the framework to protect real grid-tied PV systems across Florida from a specific class of data integrity attacks. The real-time raw alert logs from the systems' smart inverters and production meters are collected and aggregated using Kafka, structured

The material published is a result of the research supported by the U.S. Department of Energy under the Award number DE-OE0000779.

and managed using Apache Spark, and analyzed using R to provide contextual meaning. The objective of this paper is not to encourage a replacement of existing cybersecurity toolset, but to complement them with an equally powerful visualization framework to leverage better decision-making. As a future work, all the modules of the framework will be validated on the same system domain, and the complexity of attack scenarios will be increased.

The rest of this paper is organized as follows. Section II conducts a high-level review of related literature, discussing existing and emerging methods to secure smart grid. Section III introduces the proposed framework, defines briefly its related concepts, and elaborates on its three modules: Data, Classification and Action. Section IV applies the Data Module to real grid-tied PV systems in Florida and elaborates how the achieved results meet the module's objective. Finally, Section V gives a summary and highlights future work.

II. Related Work

The literature on cybersecurity technologies for smart grid can be broadly grouped into existing and emerging modes of protection. While a detailed review of each protection method is beyond the scope of this paper, a high-level summary of the protection modes is provided in this section.

A. Existing Modes of Protection

Standards detailing multiple guidelines and best practices have been published by various industry bodies such as IEEE, NERC, NIST and the Department of Energy (DOE), summarized in Table I [11]–[18]. The gap in these standards is discussed briefly below.

1) Cybersecurity Standards Gap

While security standards for Information Technology (IT) like ISO/IEC 27001 and 27002, and RFC 2196 have been streamlined over the years through revisions, the same is not true for Operational Technology (OT). With more interdependencies between IT and OT, the standards have focused on cyber-physical linkages, and how they could be exploited to create successful attacks. Further, they also prescribe cryptographic techniques, encrypted communication, end-to-end authentication and protocol-level security policies, all of which are resource-intensive and need frequent patches and upgrades [19]. Included in the standards such as NISTIR 7628 Revision 1 and NERC guidelines for Human Performance are methods to address the **human-in-the-loop** aspect of cybersecurity, which include disgruntled employees, human errors, awareness and training, access controls and certifications [20]. However, they ignore **human-on-the-loop**, which deals with the lack of Situation Awareness (SA) or a Common Operating Picture (COP) (terms defined in Section III), increased cognitive load and stress that contribute to lower attention span, and the difference in speed between technology and human cognition processes. The standards for smart grid cybersecurity reviewed in the literature largely fail to address this key aspect. Hence, there is certainly a need for a shift in security paradigms that not only consider human behavior but also human performance and human-machine-interaction as part of the problem.

2) Defense-in-Depth

When only a few layers of a system are secured, a successful attack on an insecure layer could propagate to secure layers [21]. The smart grid is one such system, with multiple interdependencies between constituent layers, as shown conceptually in Fig. 1. Hence, a defense-in-depth approach advocates defensive technologies at each layer with the aim to exhaust attacker resources, delay or dilute the impact of successful attacks, and give defenders more time to respond [22]. However, it has been shown in the literature both by reviewing past events and examining the model itself, that this approach, originally used in kinetic warfare, does not efficiently translate into cyber-warfare [23]. The security in each layer is provided by different vendors so that when an attacker breaks one layer, the same techniques cannot be exploited to break the next layer. However, issues with interoperability and information exchange have compounded the problem and increased system vulnerabilities at the expense of operational inefficiencies. Although the approach is touted to be one of the most progressive modes of protection, recent advances in smart grid like IoT, cloud and edge computing, and adoption of cellular communications, encourage its adoption in conjunction with other emerging approaches like defense-in-breadth.

B. Emerging Modes of Protection

Under an IoT-enabled smart grid, the components comprise heterogeneous nodes like smart meters, Phasor Measurement Units (PMUs) and Intelligent Electronic Devices (IEDs), each of which are either periodically or constantly online, sending bursts of data [24]–[28]. Unlike Defense-in-Depth which offers a layered security to the entire system, a Defense-in-Breadth approach implements multiple security methods at each layer of the system. Smart grid layers can be encapsulated using the GridWise Architecture Council-defined GWAC 9-layer Stack for interoperability and the ISO Open Systems Interconnect (OSI) 7-layer model for communications [29]–[31]. Both defense-in-depth and defense-in-breadth are human-in-the-loop approaches, considering human behavior as part of the problem. They account for appropriate defenses to address defective users, human errors and lack of technical awareness. However, they do not directly address the aspect of human performance and human-machine-interaction, and more specifically the gap between what is visualized and what they understand in order to make well-informed timely decisions.

III. The Proposed Framework

The primary objective of the proposed framework is to complement existing, automated cybersecurity tools at the utility CCCs by processing, contextualizing, classifying and rationalizing the data obtained from such tools to provide a need-to-know basis visualization of potential decisions that help human operators better understand and interpret the information presented. This in-turn contributes to an increased

TABLE I
BRIEF SUMMARY OF SMART GRID CYBERSECURITY STANDARDS.

Body	Standard	Core Contribution
NERC	CIP 002-011, 014	Guidelines to protect critical assets of Bulk Energy Systems (BESs) and train human personnel
NIST	FIPS series	Security requirements for cryptographic modules, digital signatures, encryption, and categorization of federal information systems
NIST	Executive Order 13636	Preliminary cybersecurity framework defining five functions: identify, protect, detect, respond, recover
NISTIR	7628 Rev 1	Guidelines for smart grid cybersecurity (includes human-in-the-loop security)
NISTIR	7823	Advanced Metering Infrastructure (AMI) smart meter upgradeability test framework
IEC	62351, 62443	Security for industrial automation and control systems; ensuring availability, integrity and confidentiality of power system protocols
IEC	TC57 WG15 Security Standards Version 14	Security of the communication protocols identified by series in IEC 60870-5,-6, IEC 61850, 61970 and 61968

SA that reduces the amount of time the operators spend manually interpreting information, and helps them make well-informed decisions in a timely manner [32].

Humans are at the end of the cybersecurity pipeline and their decisions must be backed by sound information associated with the context of the system's current situation [33]. They should reflect the growth of the decision-maker's knowledge due to experience, new inputs and dependencies and unprecedented events. Realization of such models is the highest evolution of security design for utility CCCs to parry attacks proactively. In order to better understand this approach, it is important to first look at a few associated concepts, summarized below:

Contextualized Information: To successfully counter a cyber-physical attack, it is important to view the threats and vulnerabilities of concerned systems not in isolation but in conjunction with external variables. For example, the security of PV systems is to be placed in the context of multiple external parameters such as location, weather, device calibration and measurement accuracy and human errors thereof, fault-tolerance of the communication infrastructure, and equipment specifications [34]. Then, the alerts from smart inverters due to communication failure during a stormy day could be ruled out. Hence, by giving context to data, previously unknown associations can be derived to gain more knowledge.

Situation Awareness: It is defined as the composition of three levels: perception (detection of crucial information about the system and its environment), comprehension (quantifying the data's significance and meaning in relation to the individual's goal), and projection (determining how this data will impact the future state of the system and its environment) [35]. The individual's stress, experience, missions and goals, long-term memory and working memory are captured by cognitive models, and the system's automation, complexity and interface are captured by mental models. The SA of a defender must also encompass the SA and objectives of the attacker. However, SA itself is a subjective concept, with certain individuals having overlapping aspects (called, Shared SA) and others having disjointed or conflicting aspects (called, Adversarial SA).

Common Operating Picture: It is important to contextualize data not only with dependent factors, but also with its impact on a defender's mission, his ongoing operations, cost and position in the general topography of the grid. A COP maps contextual data with its impact on mission and operations, and customizes that information to each operator to help them understand and act upon the information in a timely manner [36]. SA is a subjective concept that varies from one operator to the other, and so does COP. To this end, the proposed framework comprises three key modules, illustrated in Fig. 3 and conceptually briefed below.

A. Data Module ($\mathcal{DM}$)

The key goal of $\mathcal{DM}$ is to ingest, manage and process security data from various domains of smart grid using a data aggregation engine powered by Kafka, stream processing engine powered by Apache Spark, and a statistical engine

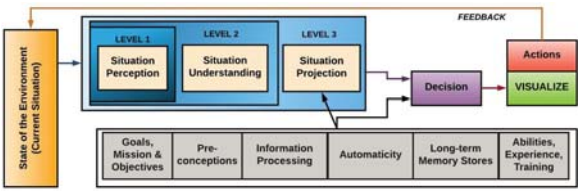

Fig. 2. Situation awareness conceptual model

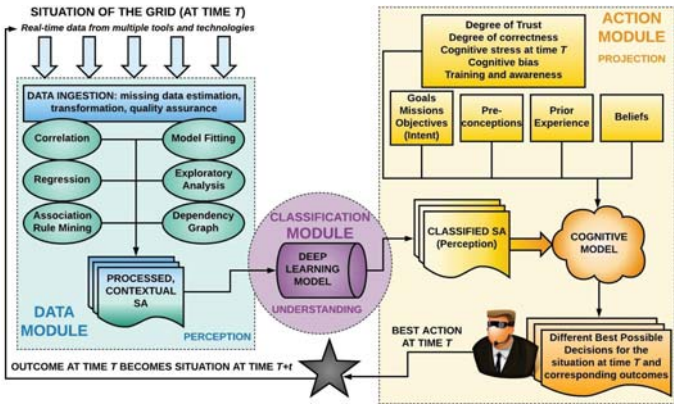

Fig. 3. The proposed framework and relation of its modules to SA levels

TABLE II
THE THREE MODULES OF THE PROPOSED FRAMEWORK.

Module Name	Algorithm(s) Included	Input(s)	Output(s)	Evaluation Methods
Data ($\mathcal{DM}$)	Multiple imputation, linear correlation, linear/polynomial regression, Maximum Likelihood Estimator (MLE), associative rule mining	Raw structured or unstructured alerts and logs from security tools or end-devices	Processed, contextualized dataset of acceptable quality	Null hypothesis testing, visual cues, comparing PDF of original and imputed data, Little's Test, Cohen's Distance test
Classification ($\mathcal{CM}$)	Long Short-Term Memories (LSTM) Recurrent Neural Network (RNN)	Processed, contextualized dataset of acceptable quality	Data classified into four categories based on defined rationales	Mean squared error, confusion matrix, precision, recall
Action ($\mathcal{AM}$)	Instance-based Learning and Bayesian inference	Data classified into four categories based on defined rationales	Actionable data: ranked vulnerable devices, likelihood of trust, optionally recommended action steps	precision, timeliness and recall, risk-tolerance

powered by R. It establishes context for that data by utilizing other environmental variables such as access control lists, external weather, employee information, equipment and tool logs, etc. $\mathcal{DM}$ is responsible to conduct descriptive analytics on cleaned, good quality data:

- Ingestion deals with collecting real-time data generated by tools like IDS, IPS, firewall and network analyzers, or alert logs and last gasps from field devices. The heterogeneous, often multi-dimensional, data is subjected to structuring and cleansing, including missing data estimation, splicing, and transformation, to get the data in required format. Optionally, quality assurance methods are also conducted to ensure logical consistency, accuracy, plausibility and origination.
- Different descriptive methods can be applied to derive context: correlation to understand attribute relationships, regression to determine the likely dependencies between dependent and independent variables, association rule mining to identify the most frequently occurring relationships between categorical attributes, and model fitting to determine the properties of numerical attributes such as Probability Density Function (PDF), skewness and corresponding statistical significance. Additionally, exploratory analysis to measure mean, variance and outliers can also be included in $\mathcal{DM}$.

B. Classification Module ($\mathcal{CM}$)

The $\mathcal{CM}$ constitutes Level 2 of SA. Its key aim is to discover more insights into the data from $\mathcal{DM}$. It uses a Long Short-Term Memory (LSTM) model to classify the data into four categories: ***normal*** (no abnormality), ***erroneous*** (due to device failure or faults), ***natural*** (due to inclement weather or extremities), or ***malicious*** (due to accidental or deliberate attacks internal or external to the system domain). LSTM is a Recurrent Neural Network (RNN) best suited to operate on time-series data [37]. Considering this study employs a streaming data that is tagged with timestamp information, it is a case of time-ordered dataset when accumulated. With the discrepancies due to missing, corrupted or unordered data points will be corrected by $\mathcal{DM}$, the data fed into $\mathcal{CM}$ will be a structured, time-series data. Studies in the literature have attempted to exploit the capability of LSTM to learn features on the fly by looking at the test data. However, it entails a risk of data classification in a manner that does not meet the objective. Hence, it is crucial to define classifier rationales, some of which are listed below [38], [39]:

- Malicious events use knowledge of the underlying environment, but errors lack that intrinsic intelligence
- Malicious events agent-driven, errors are event-driven
- Natural events are spurious and unprovoked while errors are persistent
- Natural events can be understood by associating with the weather data
- System errors of the same type have a tendency to bear the same signature/pattern, and the repetitive trends observed in historical data could be easily used to distinguish errors from attacks

C. Action Module ($\mathcal{AM}$)

This module forms the Level 3 of SA by first mapping the cognitive model of the operator it interacts with, and then feeding the information from $\mathcal{CM}$ into the model to derive projections (in the form of decisions or recommendations) which can be visualized for the operator. The module begins with a memory constructed based on the operator's responses to a detailed survey that captures their prior knowledge of and beliefs about the system. It then learns by capturing their responses to events, preconceptions to certain events, degree of trust on specific data (a probabilistic value), the level of criticality they assign to events, and cognitive stress. It employs a working memory and a long-term memory to enhance its performance in terms of recall, timeliness and precision. However, the final decision-making capability still rests with the operator. For instance, $\mathcal{AM}$ could, based on its assessment of the information it receives from $\mathcal{CM}$, understand that certain types of devices on the grid require a firmware update at the earliest to avoid a successful exploitation of specific vulnerabilities in the existing version. Depending on

the operator's priorities, $\mathcal{AM}$ could decide to either queue this message or preempt it for immediate attention. However, the final decision on whether to execute the requested firmware update rests with the operator himself. Since the decisions made by the operator modify the grid variables, this framework employs a feedback loop to account for such changes.

IV. Case Study: Protecting Grid-tied Photovoltaic System

In this section, the proposed Tri-Modular Framework is applied to the domain of grid-tied Photovoltaic (PV) systems. The domain comprises three geographically dispersed PV systems, located at Miami, Daytona and West Palm Beach in South Florida. The three systems are of similar generation capacities and employ a string inverter topology wherein multiple inverters are daisy-chained and the net aggregated energy is recorded by revenue grade production meters. This section delineates how the models of $\mathcal{DM}$ tabulated in Table II are used to derive context to the inverter alerts and local weather (Direct Normal Irradiance-DNI, Global Horizontal Irradiance-GHI, Diffuse Horizontal Irradiance-DHI, dew point, temperature, precipitation, and atmospheric pressure) recorded by the each system's Data Acquisition Unit (DAU) [40].

A. Attack Scenario

Utility dispatchers make crucial decisions based on the load patterns and power available to conduct demand response, peak load shaving, direct load control and other processes. When PV systems are a part of the local generation mix, the dispatchers monitor the real power injected into the grid by these systems to allocate the power distributed by a substation such that the loading on the segments and on the overall feeder do not exceed the thresholds [41]–[43]. An attacker, with the aim of triggering cascading failures, could inflict a data integrity attack which tampers with the settings of the smart inverters or data recorded by the meter. Falsified data leads the dispatchers to make erroneous decisions, potentially overloading a line [44], [45]. Specifically, the attacker could:

1) Tamper with the production meter to reflect a generation lower or greater than the actual
2) Exploit the communication protocol used by inverters to alter voltage values at the registers polled by DAU
3) Intercept the communication channel to steal the information transmitted and derive system generation behavior to cause a secondary attack
4) Tamper with inverter and meter settings, send false alarms and reduce operational efficiency

There are more types of attacks possible under this domain, but an elaboration on them is beyond the scope of this paper since it focuses on mining context from different data sources which then serves as a precursor for $\mathcal{CM}$ and $\mathcal{AM}$ to detect potential weaknesses in the system.

B. Apache Spark, Kafka and R Engines

As stated earlier, $\mathcal{DM}$ comprises three key engines: Kafka for data ingestion and aggregation, Apache Spark for processing and management, and R for statistical analytics on the processed data. The sequential interaction between these three engines is shown in Fig. 4. The data from field inverters and local weather stations from each PV system is fed into Kafka's producer-consumer engine at intervals of seconds. Kafka is a middleware distributed streaming platform that forms a real-time data pipeline [46]. It structures and standardizes the incoming data into topics, which are further divided into partitions. The custom Kafka consumer node used here decreases the latency of ingestion and aggregation, considering the high volume of incoming data. The spark streaming converts the aggregated data from Kafka into input discretized streams (Dstreams), which are functional Application Programming Interfaces (APIs) in Scala. The Dstreams are represented as Resilient Distributed Datasets (RDDs), to ensure primary data abstraction in Apache Spark [47]–[51]. The processed data could be pushed either directly into the dashboard, or exported as structured data to other modules. R, an open-source statistical analytics software, is used to conduct further analysis on the processed data obtained from Spark. The processes depicted by Fig. 3 were implemented using R and the results visualized as dashboards developed using Tableau software. Alternatively, the results could directly be fed into $\mathcal{CM}$ for further analysis.

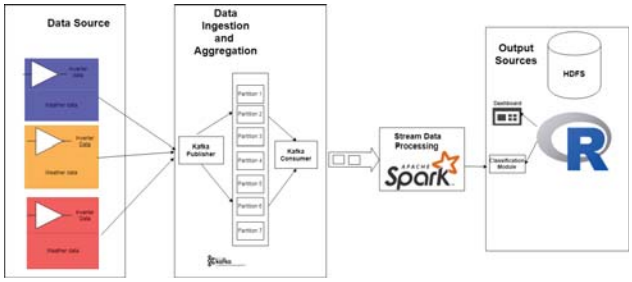

Fig. 4. The internal architecture of $\mathcal{DM}$

C. Discussion of Results

The front-end dashboard interface for $\mathcal{DM}$ comprises of four divisions which can be dynamically populated based on the availability and priority of the data. An example scenario is considered to demonstrate the results: three grid-tied PV systems are being monitored by the operator at this instance, and the $\mathcal{DM}$ has access to the alert logs and weather data from these sites for a period of one year. A view of the entire dashboard is shown in Fig. 5. The main goal of the operator here is to study the relationship between the number of inverter alerts and external weather conditions, with a hypothesis that certain weather conditions might contribute to an increased number of times an inverter malfunctions

Division $A1$ in the figure illustrates the total number of alerts received from the field inverters mapped against the various combination of weather parameters that the operator can select. In this case, he observes that the total number of alerts was exceptionally high (around 544) for a particular weather condition involving DHI, DNI, and dew point. This

could imply that the inverters, on an average, were likely to be more susceptible to malfunctions when specific weather conditions align. Upon clicking the histogram bar that represents this data in $A1$, the map in division $A3$ gets populated with the cities where these alerts were reported from, in this case, Miami, Daytona and West Palm Beach. At the same time, division $A2$ displays the number of alerts across all cities, organized month-wise. Clicking a specific histogram bar in $A2$ provides a more detailed view of the alerts as shown in division $B1$. This view can be filtered by the operator to obtain city-specific visualization of alerts.

To study the dependency of various weather parameters on the number of alerts, the operator can click on the histogram in division $C1$ to obtain the graphs illustrated by divisions $C2$ through $C4$, with $C1$ showing the bird's eye view of the dependency and $C2$, $C3$ and $C4$ showing number of alerts with respect to fluctuations in dew point, temperature and pressure, respectively. The division D is a dynamic menu item that shows to the operator a high-level summary statistics on how the site alerts changed over a period of time. The operator can change the period of observation to the right (shown here for 2 months) to view the increase or decrease in alerts from each location during this period. A magnified visual of $C3$ is shown in Fig. 6, which gives the operator a clear idea of the variations in the alert count trend with respect to changes in atmospheric pressure. The operator here gets a systematic, intuitive visualization of critical information on a need-to-know basis, which saves him from information clutter and instead guides him along a sequential, top-down visualization to drill into the greater details of the system state. Further, it also helps add context to the alert data. When viewed alone, it appeared that the significantly high number of alerts was anomalous. Closer inspection by associating the alerts with weather helped infer that the high number was more likely due to an environmental condition that was prevalent in April 2017. The operator, at this stage, could choose to further ascertain this deduction by feeding the data to $\mathcal{CM}$. Hence, the framework also adheres to the industrial standards' requirement that the final decision-making should rest with the human operator.

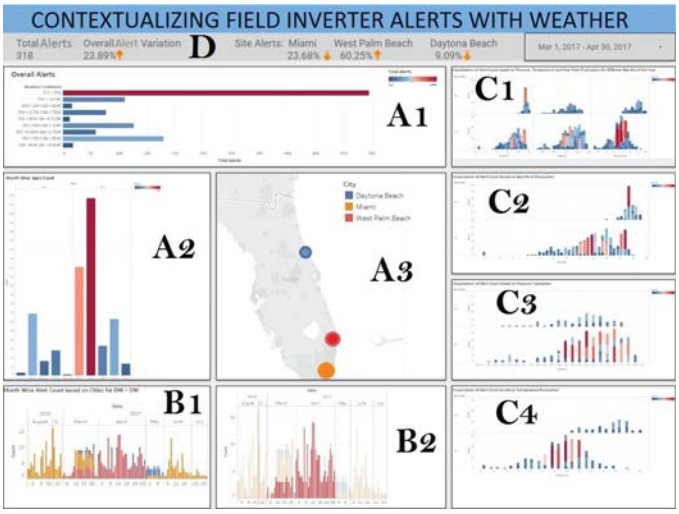

Fig. 5. Screenshot of the dashboard interface showing results from $\mathcal{DM}$

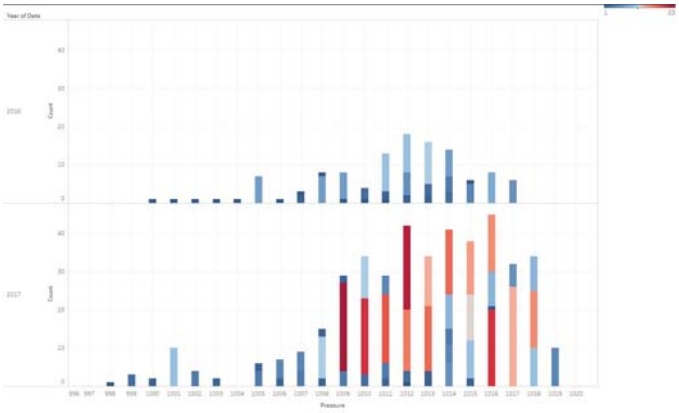

Fig. 6. Screenshot of graph shown in Division $B1$ of the dashboard

It is to be noted that the results from $\mathcal{DM}$ are interim, with the results feeding directly into the $\mathcal{CM}$. For the domain considered, it can be seen that the $\mathcal{DM}$ provides results that greatly elevate the value of the raw data collected from the PV systems of study by establishing context through mining previously unknown relationships between parameters (for example, the potential dependency between different types of inverter alerts and specific weather conditions as shown in the divisions $C1$ through $C4$). These results will help the operators better understand the data provided to them for further processing and analysis using $\mathcal{CM}$.

V. Conclusion and Future Work

This paper introduces a conceptual human-on-the-loop tri-modular framework for protecting critical assets of the smart grid. Designed to complement, but not replace, the existing cybersecurity technologies but not replace, the framework will enable utility CCC security operators to better understand and act upon machine-generated insights, and increase their overall SA. The framework's three modules: $\mathcal{DM}$, $\mathcal{CM}$ and $\mathcal{AM}$, respectively, deal with processing and providing context to raw data; classifying processed data into different categories; and providing likely projections to operators subject to their preferences and objectives. To show a proof-of-concept, $\mathcal{DM}$ was applied to real grid-tied PV systems located across Florida. Data from field inverters and production meters were collected for a period of 16 months, and the module was used to discover previously unknown properties and relationships by establishing a context with local weather data. The results showed that significant context can be mined from the data, achieving the intended goal of $\mathcal{DM}$. As a future work, the other two modules of this framework will be applied to grid-tied PV systems, and their performance will be evaluated. The complexity of attacks will be increased by considering scenarios that also affect confidentiality and availability of both data as well as power.

REFERENCES

[1] T. Rueters, "Cyberattack that crippled ukrainian power grid was highly coordinated," *CBC News*, vol. 11, 2016.

[2] I. Parvez, A. Sarwat, L. Wei, and A. Sundararajan, "Securing metering infrastructure of smart grid: A machine learning and localization-based key management approach," *Multidisciplinary Digital Publishing Institute Energies*, vol. 9, no. 9, p. 691, 2016.

[3] I. Parvez, J. pinto, and A. Sarwat, "A gossip algorithm based clock synchronization scheme for smart grid applications," in *IEEE NAPS Conference*. IEEE, 2017.

[4] I. Parvez, A. Sarwat, and F. Kaleem, "A key management-based two-level encryption method for ami," in *IEEE PES General Meeting Conference & Exposition*. IEEE, 2014.

[5] U. Ozgur, H. Nair, A. Sundararajan, K. Akkaya, and A. Sarwat, "An efficient mqtt framework for control and protection of networked cyber-physical systems," in *IEEE Conference on Communications and Network Security*. IEEE, 2017.

[6] C. Hawk and A. Kaushiva, "Cybersecurity and the smarter grid," *The Electricity Journal*, vol. 27, 2014.

[7] Y. Zhou, P. Li, Y. Xiao, A. Masood, Q. Yu, and B. Sheng, "Smart grid data mining and visualization," in *International Conference on Progress in Informatics and Computing (PIC)*, 2016.

[8] W. Matuszak, L. DiPippo, and Y. Sun, "Cybersave situational awareness visualization for cyber security of smart grid systems," in *Proceedings of the Tenth Workshop on Visualization for Cyber Security*, 2013.

[9] A. Kott, C. Wang, and R. Erbacher, "Advances in information security," *Cyber Defense and Situation Awareness*, 2014.

[10] F. Cleveland and A. Lee, "Cyber security for der systems version 1.0," *Electric Power Research Institute (EPRI)*, 2013.

[11] T. Phinney, "Iec 62443: Industrial network and system security," *Honeywell Integrated Security Technology Lab*.

[12] R. Schlegel, S. Obermeier, and J. Schneider, "A security evaluation of iec 62351," *Journal of Information Security and Applications*, pp. 197–204, 2017.

[13] F. Cleveland, "Iec tc57 wg15: Iec 62351 security standards for the power system information infrastructure," *IEC TC57 WG15 Security Standards Ver 14*, 2012.

[14] N.E.R.C., "North american electric reliability corporation (nerc) criticial infrastructure protection compliance standards," *NERC*, 2017. [Online]. Available: http://www.nerc.com/pa/Stand/Pages/CIPStandards.aspx

[15] F.E.R.C., "Federal energy regulatory commission (ferc) cyber & grid security," *FERC*, 2005. [Online]. Available: https://www.ferc.gov/industries/electric/indus-act/reliability/cybersecurity.asp

[16] N.I.S.T, "National institute of standards and technology (nist) framework for improving critical infrastructure cybersecurity," *NIST*, January 2017.

[17] N.I.S.T., "National institute of standards & technology improving critical infrastructure cybersecurity executive order 13686," *FERC*, 2005. [Online]. Available: https://www.ferc.gov/industries/electric/indus-act/reliability/cybersecurity.asp

[18] S.A.N.S., "Nerc cip standard mapping to the critical security controls-draft," *SANS- Securing the Human Report*, 2013.

[19] J. Benoit, "Making sense out of smart grid cyber security standards," *White Paper by Cooper Power Systems*, 2013.

[20] NIST, "Nistir guidelines for smart grid cybersecurity revision 1," *NIST*, 2014. [Online]. Available: http://nvlpubs.nist.gov/nistpubs/ir/2014/NIST.IR.7628r1.pdf

[21] I.N.L., "Control systems cyber security: Defense in depth strategies," *Idaho National Laboratory (INL) Control Systems Security Center Technical Report*, 2006.

[22] P. Small, "Defense in depth: An impractical strategy for cyber world," *SANS Institute InfoSec Reading Room Report*, 2011.

[23] S.A.N.S., "Defense in depth," *SANS Institute InfoSec Reading Room Report*, 2001.

[24] A. Anwar and A. Mahmood, "Cyber security of smart grid infrastructure," in *The State of the Art in Intrusion Prevention and Detection*, 2014.

[25] L. Wei, A. H. Moghadasi, A. Sundararajan, and A. I. Sarwat, "Defending mechanisms for protecting power systems against intelligent attacks," in *System of Systems Engineering Conference (SoSE), 2015 10th*, May 2015, pp. 12–17.

[26] D. Kundur, X. Feng, S. Mashayekh, S. Liu, T. Zourntos, and K. Butler-Purry, "Towards modelling the impact of cyber attacks on a smart grid," *International Journal of Security and Networks*, 2011.

[27] E. Rice and A. AlMajali, "Mitigating the risk of cyber attack on smart grid systems," in *Procedia Computer Science*. Elsevier, 2014.

[28] A. Sanjab, W. Saad, I. Guvenc, A. Sarwat, and S. Biswas, "Smart grid security: Threats, challenges, and solutions," *arXiv:1606.06992 [cs.IT]*, June 2016.

[29] E. Ibrahim, "Disruptive ideas for power grid security and resilience with der," in *NREL 2nd Annual Cyber Workshop*, 2017.

[30] N.R.E.L., "A layered solution to cybersecurity," *National Renewable Energy Laboratory (NREL) Technical Paper*.

[31] F. Cleveland, F. Small, and T. Brunetto, "Smart grid: Interoperability and standards an introductory review," *Utility Standards Board Technical Report*, 2008.

[32] R. Graf, F. Skopik, and K. Whitebloom, "A decision support model for situational awareness in national cyber operations centers," in *International Conference on Cyber Situational Awareness, Data Analytics and Assessment*, 2016.

[33] N.E.R.C., "Improving human performance: From individual to organization and sustaining the results," in *North American Electric Reliability Corporation (NERC) Technical Presentation*. NERC, 2012.

[34] A. Anzalchi and A. Sarwat, "A survey on security assessment of metering infrastructure in smart grid systems," in *SoutheastCon 2015*, April 2015, pp. 1–4.

[35] M. Endsley, "Situation awareness in the bulk power system," in *SA Technologies*, 2001.

[36] Intergraph, "Smart grid operations command-and-control center: Bringing a common operating picture to the control room," *Intergraph Technical Report: Solution Sheet*, 2010.

[37] W. Hardy, L. Chen, S. Hou, Y. Ye, and X. Li, "Dl4md: A deep learning framework for intelligent malware detection," in *International Conference on Data Mining*, 2016.

[38] R. Staudemeyer and C. Omlin, "Evaluating performance of long short-term memory recurrent neural networks on intrusion detection data," in *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference*, 2013.

[39] R. Staudemeyer, "Applying long short-term memory recurrent neural networks to intrusion detection," *SACJ*, 2015.

[40] A. Sundararajan and A. Sarwat, "Roadmap to prepare distribution grid-tied photovoltaic site data for performance monitoring," in *International Conference on Big Data, IoT & Data Analytics (BID)*, Dec 2017.

[41] A. Miranda and S. Goldsmith, "Cyber-physical risk management for pv photovoltaic plants," in *International Carnahan Conference on Security Technology (ICCST)*, 2017.

[42] J. Torres, "Cyber security for renewable energy systems," in *Asia Pacific Clean Energy Summit*, 2010.

[43] M. Shen, "Distributed solar photovoltaic grid integration system : A case study for performance," *Portland State University Dissertation Thesis*, 2012.

[44] J. Qi, A. Hahn, X. Lu, J. Wang, and C. Liu, "Cybersecurity for distributed energy resources and smart inverters," *EIET Cyber-Physical Systems: Theory & Applications*, 2016.

[45] L. Wei, A. Sundararajan, A. Sarwat, S. Biswas, and E. Ibrahim, "A distributed intelligent framework for electricity theft detection using benford's law and stackelberg game," in *Resilience Week*, Sep 2017, pp. 5–11.

[46] K. Goodhope, J. Koshy, J. Kreps, N. Narkhede, R. Park, J. Rao, and V. Y. Ye, "Building linkedin's real-time activity data pipeline," *IEEE Data Eng. Bull.*, vol. 35, pp. 33–45, 2012.

[47] P. M. Grulich, "Scalable real-time processing with spark streaming: implementation and design of a car information system," *CoRR*, vol. abs/1709.05197, 2017. [Online]. Available: http://arxiv.org/abs/1709.05197

[48] M. Zaharia, T. Das, H. Li, T. Hunter, S. Shenker, and I. Stoica, "Discretized streams: Fault-tolerant streaming computation at scale," in *Proceedings of the Twenty-Fourth ACM Symposium on Operating Systems Principles*, ser. SOSP '13. New York, NY, USA: ACM, 2013, pp. 423–438. [Online]. Available: http://doi.acm.org/10.1145/2517349.2522737

[49] T. Akidau, A. Balikov, K. Bekiroglu, S. Chernyak, J. Haberman, R. Lax, S. McVeety, D. Mills, P. Nordstrom, and S. Whittle, "Millwheel: Fault-tolerant stream processing at internet scale," in *Very Large Data Bases*, 2013, pp. 734–746.

[50] O. Backhoff and E. Ntoutsi, "Scalable online-offline stream clustering in apache spark," *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, pp. 37–44, 2016.
[51] R. C. Fernandez, P. R. Pietzuch, J. Kreps, N. Narkhede, J. Rao, J. Koshy, D. Lin, C. Riccomini, and G. Wang, "Liquid: Unifying nearline and offline big data integration," in *CIDR*, 2015.