Networked Loads in the Distribution Grid

Zhifang Wang*, Xiao Li†, Vishak Muthukumar‡, Anna Scaglione†, Sean Peisert‡§, Chuck McParland§
* Virginia Commonwealth University, Electrical and Computer Engineering, Richmond, VA, USA
† University of California Davis, Electrical and Computer Engineering, Davis, CA, USA
Contact Author E-mail: ascaglione@ucdavis.edu
‡ University of California Davis, Computer Science, Davis, CA, USA
§ Lawrence Berkeley National Laboratory, Berkeley, CA, USA

***Abstract*—Central utility services are increasingly networked systems that use an interconnection of sensors and programmable logic controllers, and feed data to servers and human-machine interfaces. These systems are connected to the Internet so that they can be accessed remotely, and the network in these plants is structured according to the SCADA model. Although the physical systems themselves are generally designed with high degrees of safety in mind, and designers of computer systems are well advised to incorporate computer security principles, a combined framework for supervisory control of the physical and cyber architectures in these systems is still lacking. Often absent are provisions to defend against external and internal attacks, and even operator errors that might bypass currently standalone security measures to cause undesirable consequences. In this paper we examine a prototypical instance of SCADA network in the distribution network that handles central cooling and heating for a set of buildings. The electrical loads are networked through programmable logic controllers (PLCs), electrical meters, and networks that deliver data to and from servers that are part of a SCADA system, which has grown in size and complexity over many years.**

I. Introduction

There are many physical systems today that impact or are impacted by networked computers. The growing trend is to embed intelligence onto physical devices, especially critical assets, and then network them, for added convenience of the operators who monitor and interact with them.

In the past, information networks and controls of this type were limited to the electricity transmission infrastructure and generation facilities. The operations and management of these facilities are characterized by strict scrutiny and include contingency planning. But, increasingly, these cyber-physical networked systems are used in relatively small plants, with network configurations that are ad hoc and designed opportunistically to lower cost and meet local needs efficiently. In these small plants, inevitably, the safety management model is extremely streamlined. In one prototypical central heating and cooling plant, large electrical loads are networked through a handful of PLCs, connected with one main PLC cabinet and to a wide area network.

As the desire to harness these networked cyber-physical systems to perform demand response grows, the open questions about their safety becomes increasingly compelling. The physical systems themselves have always been designed with extremely high degrees of safety in mind, using a technique called *safety engineering* [Lev11]. Safety engineering sets the requirements and best practices for how systems should be operated by human operators, failure scenarios such as *fail safe*, *fail fast*, and *fail stop*. Similarly, designers of computer systems are well-advised to consider computer security principles [SS75].

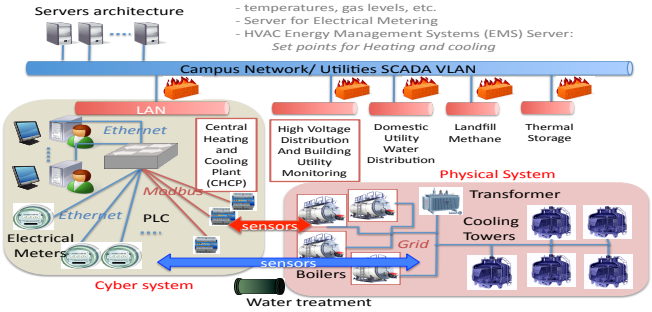

Fig. 1. Utility Network.

Analyzing one such plant in detail, we have seen an emerging trend: the newest physical assets have expanded and improved their networking capabilities and, in parallel, they have also strengthen their local controls, limiting unsafe use of the individual machines. But, as much as the operators are gratified with these safety advances and improved capabilities, they recognize that there are several older assets in these plants that were never intended to be connected to any network. Furthermore, these improved controls are local, which means there is no mechanism in place, nor test or certification process, for the networked system, that would ensure to a certain degree that machines cannot have a collective behavior that is damaging. One of the key reasons why the collective physical actions matter, is that all these physical systems, newer and older, are drawing energy from the same electric grid. It is the combination of the two networks—the data network, and the electrical network—that, even under the assumption of perfectly robust local control, can leave vulnerabilities.

In fact, much existing effort in research and development relating to the security of these systems focuses on the two elements in tight compartments, where either the cyber infrastructure or the physical one work in an ideal manner and cannot be the root cause or trigger of security violations in the other domain [Bau10], [KHLF10]. The intersection of safety engineering and computer security is one of the most significant sources of concern for cyber-physical systems, however. Specifically, where are the gaps left by the designers

of such systems in which unsafe assumptions are made about which particular system among the "cyber" and "physical" systems are responsible for safety and security? As the targets of the *Stuxnet* [Sym11] worm now know, systems often *do* allow behavior that is damaging to individual devices since computer systems frequently can be misinformed or ignore tolerances of the physical systems they control, the serial protocols that the devices use to receive commands, or other physical operating constraints. Even more insidious is the eventuality of a perfectly tolerable local behavior which results in collective network actions that are damaging.

Of course key failures that the system must defend against need not be limited to external attacks, but also threats from operator error and malicious insiders [BEP+08]. Even well-meaning attempts to secure systems can be problematic: consider the election official who installed a virus scanner onto his electronic voting machines in attempt to make them more secure, despite the fact that these machines were never connected to any network, and therefore could not have been contaminated by malware. The installation rightly invalidated the certification of the voting machine.

II. A CASE STUDY

A diagram of the networked components in the system is shown in Figure 1, including some of the elements that appear in the cyber system and in the physical system. It is evident that its complex needs are handled in a centralized fashion with the aid of a network and sensors on the ground that are supposed to provide situational awareness to the operators.

Given the large size of heaters, chillers, gas exhausts, etc..., plants are typically divided in several sub-plants, plus a control room, which often is the site of central networking equipment and of human operators, who operated the machinery through a Human Machine Interface (HMI) but are on-site because they often physically verify the results of their control. As the diagram indicates, the networked system includes three categories of nodes: the physical devices (e.g., the boilers, the chilling towers, the transformer, the water alkalizer etc.), the computational and data elements (e.g., computers, Human Machine Interfaces, Programmable Logic Controllers, and various sensors), and the human operators. Programmable Logic Controllers (PLCs) are the center of a star network that includes both analog and digital sensors, typically communicating through wired links. There are two main categories of networks as well: the communication and computers network, and the physical supply networks including electrical, water and gas. There is also a human network of operators.

Within each class of nodes there is wide heterogeneity of functions, permissions and capabilities. The overall utility, as the diagram in Figure 1 suggests, is a set of local networks which are all connected through a VLAN SCADA backbone, which is part of the campus network. Not all the facilities connected to the SCADA network are likely to have the variety of nodes and functions that the main utility plant has, and generally several plants do not have control rooms with human operators because their desks are concentrated in the control room in the main plant facility. This is why networking is of paramount importance: to provide situation awareness in a complex landscape of distributed systems. The inventory of the elements contained in the facility is reported in Figure 2.

System Name	Devices	Transformers MV/LV(4.2 kv/ 286 v)	MCC (Motor Control Center)		Electric Meters		Communication Wires
A	Boiler 1, 2, 3 Furnace 1, 2, 3	System E, A, D transformers, with main switch: MCC-1A, 2A, 3A	MCC-6	MCC-1	2M(3710)	3M(7350)	3M has a separate com wire
D				MCC-7			2M, 1M wired together with COM wires
E	Boiler 4, Furnace 4					1M(7350)	
B	Chiller 1, 2 , 3	Chiller transformers: 1, 2, 3	MCC-4		1M(3170)		All the meters in B &C systems wired together by COM then output through Fiber Optic Cable to MH 5.24 SW
C	Chiller T3C_A, T2C_B, T1C_C, C1, C2, C3, C4	Chiller transformers: TA, TB, TC, CH-1 C1&C2, CH-2 C3&C4	MCC-2 MCC-3		1M(3170) 3M(7650) 7M(7650)		
Storage Container	domestic water storage	n/a			Not shown in the map (but must have some meter with data shown in the web)		Not shown in the map (but must has some communication wires)
Exhaust	Waste gas exhaust	n/a					
VFD room	VFD (Variable-Freq Drive)	n/a					Connected with a Fiber Optic Cable to MH 4.27 SW

Fig. 2. Inventory of elements in the plant.

A. Physical Elements

A recurring feature of these networked systems is that the infrastructure is heavily layered and evolves over many years. Typically new facilities are tacked on to the old ones.

1) Heating and Cooling Plant Devices: The multiple heating and cooling devices in such a plant and, because of their large size boilers and chillers are usually arranged into different subsystems (or plants). There include chillers, boilers, furnaces, waste gas exhausts, variable-frequency drive (VFD) room, and motor control centers (MCC), located at a number of sub-plants of the plant we studied.

2) Grid: The plant we studied receives electricity from a substation dedicated to the complex of buildings that are served by it. The electric transformers, feeding from the 4160/480V feeder, supply electricity to the heating and cooling devices at the plant such as the boilers, the chillers, and the motor control centers (MCC). The plant adopts three-phase AC grid, and has voltage levels: 4160V, 480V, and 280V.

B. Computational and Sensing Elements network

In the typical SCADA reference model, dispersed plants like the utility we analyzed, have one or more single point of convergence for a network of sensors on the field (i.e., on the actual machines). These aggregation points are typically referred to as Remote Terminal Units (RTU). In Figure 1 the RTU role is logically played by the several PLC in the facility, each connected via ethernet cables to a subnetwork of sensors. The ratio of PLC to assets is quite high, since the individual boilers and chillers, the exhaust etc. have several sensors within each unit. It seems that any expansion of the facility has come with a new set of PLC (one or two). They have been then connected to the same point of access to the SCADA network. The network in the facility is structured as a forest with a few trees connecting to the campus network. A number of PLCs, meters and sensors communication links converge towards a single network relay located in the control

room, in the same cabinet of one important PLC in the plant. A separate network trees is the electrical meters network, directly connected to the campus network, without a specific PLC associated to them. So there is no RTU in the plant for the electrical metering infrastructure.

This configuration reflects a hierarchy typical of the SCADA model with a single RTU concentrating information. However, it appears that there is no data aggregation and processing performed anywhere within the plant, or in the close proximity of the physical assets. The information and command streams are routed to, processed, aggregated, and interpreted in a number of remote servers maintained as part of the campus IT infrastructure. In fact, the command stream coming from the operators desk keyboards is not sent directly to the local PLC. Rather, it is sent to the appropriate server in the SCADA network that then issues the commands back to the physical device in the appropriate format for the PLC. As we discuss next, there are graphical user interfaces on the machines that can bypass this and allow the operators to directly change the set points. Architecturally, this choice of mapping functionalities to servers and use of them as centers for the network traffic seems more justified by historical artifacts rather than by deep technical thinking behind the design of cyber-physical systems. It is the connection to the databases in the server which is also used for the immediate feedback and control in the infrastructure, whenever the operators are not in close contact with the machine. This means that, if the servers are away from the plant and the plant loses its connection to the external network that links them from to the servers, the operations of the plant cannot be performed through the control room, but have to be performed manually, interacting with the graphical interface on the machines.

Furthermore, we noticed that each server's typical configuration has its own specific data function and, interestingly, there is a preferred server to accrue data for water heaters and chillers that is used as historian and to obtain the system analytics. This server transfers data to another server that is used to inform the human-machine interface (HMI). The software is provided by two different vendors that specialize in HMI software for SCADA systems. Also, electrical metering is on a separate cloud server and not displayed concurrently with the water system.

1) Control and Communication Devices: Programmable Logic controllers (PLC) are commonly used in modern cyber-physical systems to provide automation of electromechanical processes. A PLC is a digital computer especially designed for multiple inputs and output arrangements and endure extreme environmental conditions. Human-Machine interfaces (HMI) that we observed are typically either a computer screen or graphical user interface panels at the side of the physical machines, allowing the operators to examine and adjust the configuration and condition of a device. The built-in communication ports in these devices can be RS-232 or Ethernet ports. The digital communications between these devices adopt the industry automation protocol like Modbus or Fieldbus which enables the PLC communicate over a over a network to other

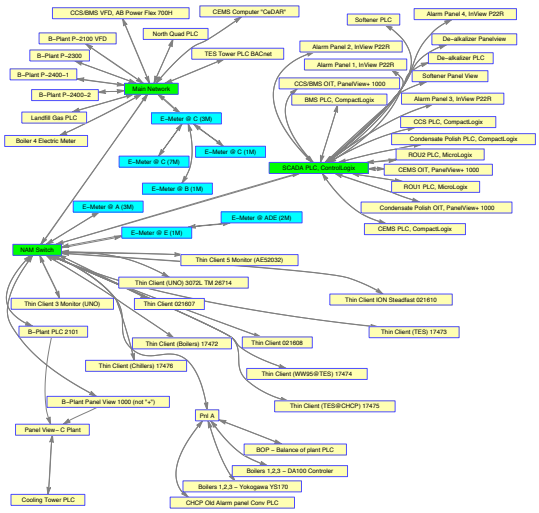

Fig. 3. The PLC and the electrical meter maps

systems. Modbus or Fieldbus belong to the de facto standard protocols widely adopted in industry PLCs. These protocols use plain text commands and have no security features for either integrity (e.g., checksums) or confidentiality (e.g., encryption). Security also requires strictly administered access, but there is no security mechanism in the devices interface to identify authorized users. Another protocol we found in the utility plant is the NetBios protocol (for printer/file sharing) used in the Ethernet/IP network, which, if not implemented correctly, can expose the data to the entire Internet.

In Figure 3 on the top and bottom we show respectively the network of PLCs and electrical meters. The PLC network has, for the most part, a tree structure emanating from the NAM switch, with depth 4. The SCADA cabinet in the control room is by far the most networked, and is directly linked to a number of PLC in Plant E as well as an alarm system in Plant A. Furthermore it is the convergence site for most of the electrical meters.

2) Data and Traffic analysis: In our investigation we realized first hand how no feedback or monitoring for the cyber system is provided in these deployments. One of the interests and concerns the utility manager had was gaining a better idea on how to control the system holistically. An example discussed was that of an old boiler whose serial communication port was connected to the network through an interface allowing to transfer data from an obsolete serial port to the Ethernet link, and that continued to create problems for the operators. Occasionally, due to the imperfect interface, the communications to the boiler would be interrupted and the monitors of the operators would simply freeze the image of the sensor measurements, without giving any clue whatsoever to warn the operators of the cyber failure. This exemplifies the ad hoc nature of the connection of sensor interface to the network that resulted in the issues with monitoring. The operators decided to address the problem by maintaining a manual log and periodically inspecting the elements. In this

case, the operators gained insight over time on a defect of the communication network simply by observing inconsistencies between physical facts and the sensor readings on the screen. The fact that the technology was outdated gave them a clue of what could be the source of the problem. But the shared concern was the rest of the network status.

One of the activities we performed as part of our preliminary study, is to examine a trace of traffic in the network by connecting a laptop with a network protocol analyzer (Wireshark) to the main PLC Cabinet. This step is necessary to develop mechanisms that would measure and enhance the security of such systems. Some of the questions that we are seeking to answer include: what is the network really doing, anyhow? What does the network reveal about the operation of the cyber physical system? How will we evaluate the intrusion detection system (IDS) that we are constructing? What kinds of tests (e.g., penetration testing [Bis07], [Lin75]) are appropriate?

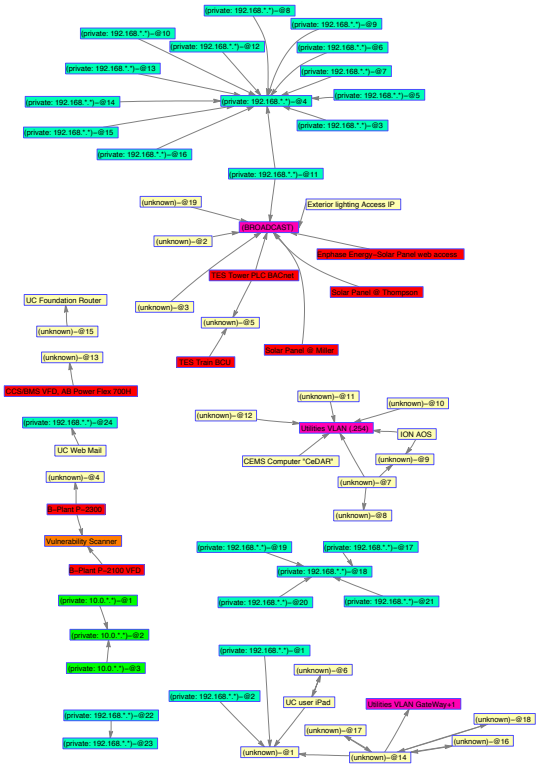

Fig. 4. One snapshot of communication traffic map observed inside a CPS SCADA network.

While examining these questions on the data that we collected, our discussions with the operators confirmed that there is general confusion even on the operational side in terms of what traffic patterns should be deemed typical and what should not. In Fig. 4 we show the equipment active during the period we sniffed traffic. The figure shows arrows connecting nodes the node sent a message addressed to another device. Fig. 5 shows the packet transmission of different protocols during the sniffing experiment. The top is for PLC data or supervisory command transmission, using the Dropbox Sync

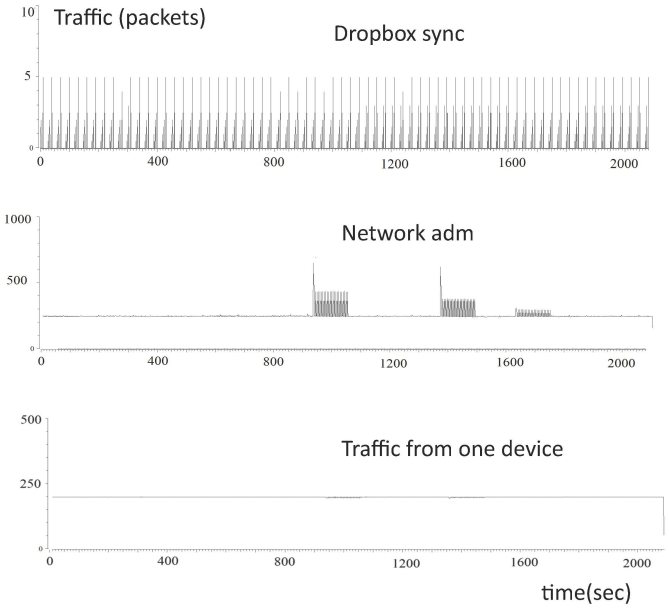

Fig. 5. The traffic sent from one physical device

LAN protocol.[1] The intermittent traffic is periodic with a maximum transmission rate of 5 packets/sec. The figure in the middle shows the traffic of all other protocols, for connection setup or network administration, such as Ethernet/IP implicit packets and multi-cast advertising packets to announce a device's presence. This traffic occupied most of the network resources, with a steady rate about 200 packets/sec most of the time plus three clearly distinguishable spiking period, whose maximum transmission rate reached over 500 packets/sec. Our further study discovered that the 200 packets/sec traffic mainly came from the PLC of a specific device (named "Boiler4") in the plant, which was recognized by the manager as the newest and more *wired* PLC among the ones available in the facility. This PLC was using ENIP/UDP to set up connections with other devices and after the efforts failed, it began to broadcast over 31 MB in that period. Another notable observation was the presence of traffic from a network vulnerability scanner. It is possible that this interaction would activate processes that are not designed for a close network such as the SCADA network.

3) Measurement and Monitoring Sensors: The archival electrical data that can be retrieved from the SCADA database server is recorded in a very coarse resolution, every 15 minutes. The electrical data points for an AC network are collected from electrical voltage/current transmitters or switches installed at transformers supplying the physical devices at the plant, which include the voltage V(Volt) and current magnitude I (Ampere), the real and reactive power P (kilo-Watt) and Q (kilo-Var), frequency f(Hz), power factor pf(the cosine of the phase angle difference θ (deg) between the voltage and current phasors), etc. The voltage, current, and real/reactive powers can be measured at three separate phases

[1]Dropbox: What is LAN sync? https://www.dropbox.com/help/137/en

(a,b,c) and the mean, maximum, and minimum values recorded over each observing periods. Due to the insufficiency of electric sensors, aggregate measurements are taken instead of measuring individual devices at some subsystems.

A sample from a week of electrical measurements from the electrical meter in one sub-plant is shown in Figure 6. We estimated the probability density function (PDF) and the cumulative distribution function (CDF) of the data. Fig. 7 shows a sample empirical PDF/CDF curves of voltage, current, real/reactive power, power factor (in percentage), lagging phase angle, and AC frequency, computed based on the month worth measurements taken from one chilling device in our examined plant. It can be seen that the voltage (a little bit above the rated voltage level 4160V) and frequency (around 60 Hz) are quite stable and roughly follow a Gaussian distribution with a very small variance. This is because the electric power grid usually provides strict voltage and frequency regulations in order to maintain the power quality and grid stability. Other data types (i.e., the current, real/reactive power, the power factor and the lagging phase angle) are clearly multimodal, due to the ON/OFF nature of the sources. There are two distinct sets of states, which respectively correspond to the low-power and high-power states of the grid.

For a device operated in an alternative-current (AC) grid, given a near-constant frequency as we observed, the consumed real/reactive power and the power factor can be written as:

$$\begin{aligned} &P = VI\cos(\theta), \quad Q = VI\sin(\theta), \\ &pf = \cos(\theta) \end{aligned} \tag{1}$$

Therefore the real/reactive powers, the current, the power factor and the lagging phase angle are inherently correlated together. One interesting phenomenon is that during the low-power (low-current as well) period, the device's electrical meters tend to read abnormally large lagging phase angle therefore very low power factor. The system operator attributed this to two likely reasons: the meter errors or some operating defects of the device. Since this phenomenon is common among many electrical meter measurement data of various devices in our examined plant, we believe it is more likely caused by the meter errors. However, if the devices tend to have extremely low power factors at their low-power level, this may introduce potential harm to the grid quality and result in utility penalties to the plant owner, since poor lagging power factors may cause extra voltage drops, additional transmission loss even flow overload in the grid; in fact, the utility usually requires their customers to keep an ideal power factor about $0.95 \sim 1.00$, otherwise some onsite compensation has to implement in order to improve the factor.

We examined the electrical meter errors by comparing directly measured real/reactive power values with computed values, based on other data types, using the equations (1). We compute the real and reactive power as:

$$\begin{aligned} &\hat{\theta} = \cos^{-1}(pf), \\ &\hat{P} = VI\cos(\hat{\theta}), \quad \hat{Q} = VI\sin(\hat{\theta}). \end{aligned} \tag{2}$$

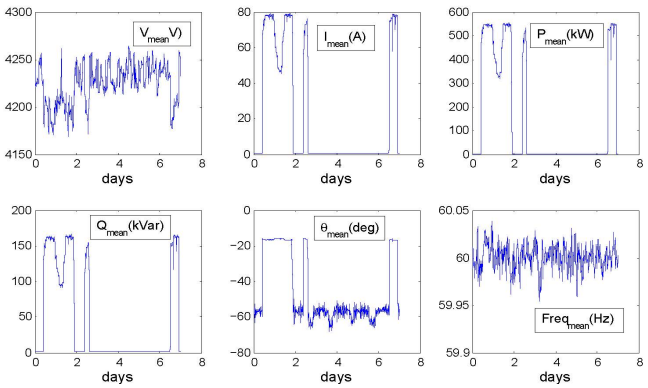

Fig. 6. Electrical measurements taken from the facility.

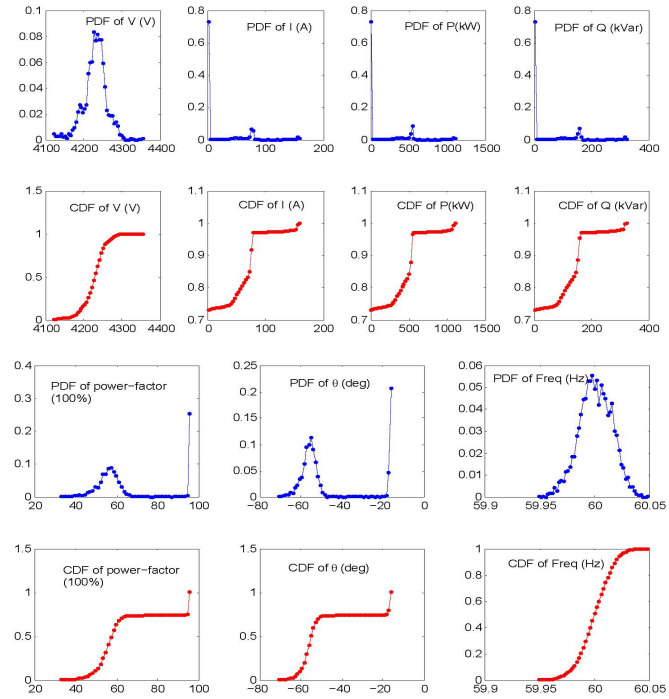

Fig. 7. The empirical probability density function (above) and cumulative distribution function (below) of some electrical facility measurements.

Then the discrepancy between the measured and computed real/reactive power are

$$\text{err}_P = P - \hat{P}, \ \text{err}_Q = Q - \hat{Q}. \tag{3}$$

Fig. 8 shows the discrepancy derived from the same measurement data of Fig. 7. It can be seen that the discrepancy relates with the two distinct power-level states: during the high-power-level period (normal-loaded) the discrepancy is trivial; during the low-power-level period (light-loaded), the discrepancy is large and can be as high as 100% of the measured value. Fig. 9 presents the mutual-correlation image-map to show the correlation strength between the variables from a specific set of data points, drawn according the Pearson coefficients, which

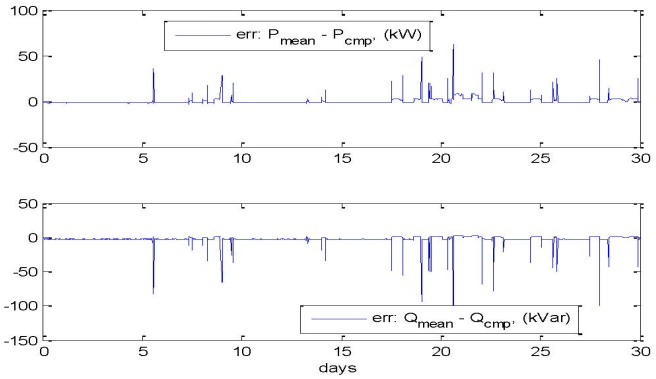

Fig. 8. The error in real (above) and reactive (below) power.

is defined as :

$$\rho(X,Y) = \frac{E[(X - E_X)(Y - E_Y)]}{\sqrt{E[(X - E_X)^2]E[(Y - E_Y)^2]}}, \qquad (4)$$

where $E(\cdot)$ represents the expectation (or the mean-value) of a statistical variable. From the correlation map, one can see that the whole set of data points can be clearly divided into three groups. The first group comprise the current I, the real/reactive power P and Q, which are strongly correlated to with a correlation coefficient close to 1.0 (as shown as dark red in the image map) ; the second group contains the power factor pf and the discrepancy of real power err_P, which maintain a moderately strong correlated with the first-group variables and among themselves and correlation coefficients around $0.6 \sim 0.8$ (shown as orange and red in the image map); while the third group include the discrepancy of reactive power err_Q, the frequency f and the voltage V which have very weak correlation among themselves and also a very weak correlation with the variables from the other two groups and the correlation coefficients mostly close to be zero (shown as yellow or green in the map), however, with two exceptions here which will be discussed in the following:

The first exception is the voltage V exhibiting a moderately strong negative correlation with the first group variables ($\rho = -0.8 \sim -0.6$). This is no surprise because they have been related together by (1). The correlation coefficients are negative because a large power consumption (i.e., real/reactive power and the current as well) usually causes more voltage drops along the transmission line.

The second exception is the errors in the real and reactive power values, i.e., err_P and err_Q, which have moderately strong mutual negative correlation. This is possibly a feature of the electrical meter erroneous measurement pattern.

From our analysis we believe that the empirical PDF/CDFs and the correlation statistics of measurement data can be derived and utilized to define hypothesis tests for anomaly detection. Particularly the distinct statistical characteristics related with the two power-level states could be used to design a specification-based detection and distinguish whether the anomaly is caused by meter errors or malicious attacks.

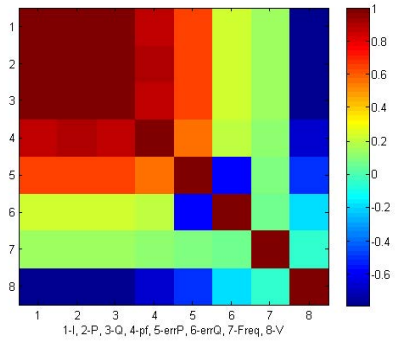

Fig. 9. The correlation map of the measurement data points.

III. Conclusions and Future Efforts

Accessing real data to compile an inventory like this can be challenging. The primary conclusions that we can draw from this study is that the networking of several devices has given the operators a greater sense of safety and flexibility, but that, in turn, the activity within the network is broadly misunderstood. The picture we derived confirms the concern that there is no form of support in these network deployments for contingency planning or analysis of events that originates in the cyber network and that attests the lack of testing models for these networked infrastructures in the industry suppliers that provide this equipment. The heterogeneity of devices connected to the network is particularly daunting, and daunting are also the challenges of networked control. Future research includes creating models and simulations that can shed some light on the first tasks that we outlined in our project objectives and represent our initial steps towards a computer model that can replace and inform these case studies.

Acknowledgments

This research was supported in part by the Director, Office of Computational and Technology Research, Division of Mathematical, Information, and Computational Sciences of the U.S. Department of Energy, under contract number DE-AC02-05CH11231. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect those of any of the sponsors of this work.

References

[Bau10] Todd Baumeister. Literature review on smart grid cyber security. Technical Report Technical Report CSDL-10-10, , Department of Information and Computer Sciences, University of Hawaii, 2010.

[BEP+08] Matt Bishop, Sophie Engle, Sean Peisert, Sean Whalen, and Carrie Gates. We Have Met the Enemy and He is Us. In *Proceedings of the 2008 New Security Paradigms Workshop (NSPW)*, Lake Tahoe, CA, September 22–25, 2008.

[Bis07] Matt Bishop. About Penetration Testing. *IEEE Security and Privacy Magazine*, pages 84–87, Nov/Dec 2007.

[KHLF10] Himanshu Khurana, Mark Hadley, Ning Lu, and Deborah A. Frincke. Smart-Grid Security Issues. *IEEE Security & Privacy*, 8(1):81–85, 2010.

[Lev11] N.G. Leveson. *Engineering a Safer World: Systems Thinking Applied to Safety*. MIT Press, 2011.

[Lin75] Richard R. Linde. Operating System Penetration. In *Proceedings of the AFIPS National Computer Conference*, pages 361–368, May 19–22 1975.

[SS75] Jerome H. Saltzer and Michael D. Schroeder. The Protection of Information in Computer Systems. *Proceedings of the IEEE*, 63(9):1278–1308, 1975.

[Sym11] Symantec. W32.Stuxnet Dossier (v.1.4). http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/w32_stuxnet_dossier.pdf, February 2011.