Privacy Concerns in Upcoming Residential and Commercial Demand-Response Systems

Mikhail A. Lisovich and Stephen B. Wicker
School of Electrical and Computer Engineering
Cornell University, Ithaca, New York 14853–3801
Email: mal86@cornell.edu, wicker@ece.cornell.edu

***Abstract*—We explore the privacy concerns arising from the collection of power consumption data in current and future demand-response systems. We claim that in a lax regulatory environment, the detailed household consumption data gathered by advanced metering (AM) projects can and will be repurposed by interested parties to reveal personally identifying information such as an individual's activities, preferences, and even beliefs. To develop this claim, we begin with an overview of demand-response technologies and their deployment trends, mentioning both the parties interested in the data and their motivations. We proceed to formalize the notion of privacy and list the types of personal information which can be estimated with current and upcoming monitoring technologies. To support our list, we conduct a small-scale monitoring experiment on a private residence. Our results show that personal information can be estimated with a high degree of accuracy, even with moderately sophisticated hardware and algorithms. We discuss the implications of our results for future demand-response projects. Our paper concludes with guidelines for data-handling policies which ensure the protection of privacy.**

***Index Terms*—NG-SCADA, Protection, Privacy.**

I. Introduction

The next decades will see a transformation of our nation's power distribution systems. Next generation Supervisory Control and Data Acquisition (NG-SCADA) architectures will precipitate an exponential increase in both the data and control available to consumers and utilities. Utilities are increasingly adopting automated metering, advanced demand response architectures, microgirds, and other systems which will provide cost savings and flexibility in power generation, increase grid reliability, and create new modes of consumer-utility interaction.

This transformation is already well underway. Recent years have seen several pilot microgrid projects [1], as well increased deployment of Advanced Metering Infrastructure (AMI) systems by major utilities across the US and in other countries. AMI systems in particular have been deployed on a large scale by utilities in California [2], Ontario [3], and elsewhere. According to a 2006 Federal Energy Regulatory Comission [4] staff report, six percent of meters installed in the US are 'smart' meters supporting some advanced metering project, and the number steadily continues to increase.

Next generation SCADA projects will provide many advantages to both the utilities and the consumer. For the power companies, automated metering will reduce collection costs, while the ability to capture detailed usage information will allow for large-scale load research. The results of this research will allow utilities to improve generation planning, rate development, demand side management and distribution planning [CITE] . In addition, load research will provide essential information for projecting the effects of various demand side management programs and novel pricing structures. For the consumer, the projects will result in more information, more control over power use, and the ability to actively participate in power generation. However, increased availability of data, along with emerging use cases, will inevitably create privacy and security issues.

This paper is part of a larger effort by the TRUST Center[1] to explore the confluence of sensor networking, power distribution, privacy, and security issues that will emerge from a substantial increase in power system monitoring at the consumer level. In this paper, we choose to focus on the privacy risks arising from the collection of power, gas, and water consumption data in current and future demand-response systems. Our main claim is that in a lax regulatory environment, the detailed household consumption data gathered by advanced metering projects can and will be repurposed by interested parties to reveal personally identifying information about the programs' participants.

Let us elaborate. Although current projects implement measures to safeguard individuals' privacy and confidentiality, we believe that there exist strong motivations for entities involved in advertising, law enforcement, and even criminal enterprises to collect and repurpose power consumption data. These entities may collaborate with utilities, pressure them, or simply steal the desired data.

Consumption data in the hands of these entities raises serious ethical concerns - without proper safeguards, these data may be used to commit fraud, initiate unsolicited and invasive advertising, and in the case of law enforcement, to conduct warrantless searches that may infringe on individuals' Fourth Amendment rights (see Section III-A for an example). Because of these concerns, there is a need for industry-wide discussion on the privacy aspect of data collection, as well as on developing data-handling policies which will allow the technology to evolve while safeguarding privacy.

The rest of this paper is concerned with systematically developing and substantiating our claim. In Section II, we

[1]TRUST is a multi-university collaboration focused on privacy & security and consisting of research engineers, social scientists, and students.

familiarize the reader with the current state of advanced metering technology and projections for its evolution. We also mention some of the parties interested in the data and their motivations for obtaining and repurposing it. In Section IV, we aim to formalize these parties' impact on individual privacy by discussing a 'privacy metric' (encompasses the ways that privacy can be infringed). In Section V, we prove that repurposing is feasible from a technical standpoint by conducting a small-scale monitoring experiment on a private residence. Our results show that personal information can be estimated with a high degree of accuracy, even with moderately sophisticated hardware and algorithms. In Section VI, we discuss how our experimental methods can be extended to large scales (finishing the claim). Finally, having made apparent the need for discussion, in Section VII we provide general guidelines for proper data handling and chart directions for future work to be undertaken by our TRUST collaborators . Section VIII concludes.

II. Technology Overview

To familiarize the reader with the technical aspects of the issue, we begin with a brief overview of demand response technologies. We focus primarily on Advanced Metering (AM) and Nonintrusive Load Monitoring (NILM) systems. In each case, want to highlight the types of available raw data, as well as access points by which it can be collected by authorized and/or unauthorized parties. For a more complete overview of AMI and NILM, we refer the reader to [5] and [6], respectively.

A. Advanced Metering

In a typical Advanced Metering setup, the customer is equipped with solid state electronic meters that collect time-based data at daily, hourly or sub-hourly intervals. The types of available devices differ from project to project, but may include electricity, gas, and water meters. These meters have the ability to transmit the collected data through commonly available fixed networks such as Broadband over Power Line (BPL), Power Line Communications (PLC), and public networks (e.g., landline, cellular, paging). The meter data are received by the AMI host system and sent to the Meter Data Management System (MDMS) that manages data storage and analysis to provide the information in useful form to the utility [5]. The typical building blocks of an AMI system are shown in Figure 1.

A typical AMI system outputs hourly or sub-hourly interval data on power consumption and may also take daily data on gas and water consumption. The meter reads, dated and time-stamped, are collected from all devices and recorded either by an intermediate node or by the central processing entity. Data access at intermediate points - [Cite Cardell] The data is required to be reasonably complete and accurate. In [10], the specifications are that $>$ 98% of all meter reads make it to the intermediate node, and that the readings have a precision of at least 10 Watt-hours (0.01 kWh).

As mentioned in the introduction, AMI systems have already been deployed on large scales (refer to [4] for detailed

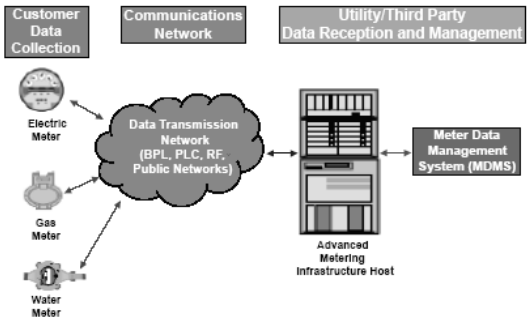


Figure taken directly from [5]

Fig. 1. AMI Building Blocks

deployment information). The technology's significant market presence makes is (worthwhile to consider?)

B. Non-Intrusive Load Monitoring

A NILM system goes a step further, processing power consumption datat determine the operating schedules of individual electrical loads. This is typically done by disaggregating the collected data stream into individual load signatures and matching each signature with signatures stored in a device database. As with AMI, data is usually sent to an intermediate node to be processed into useful forms.

These systems are used for a wide variety of purposes, including load research, evaluating impact of rate structure changes, implementing incentive programs for particular appliance usage patterns, and handling high-bill complaints [7]. However, they are important to us because appliance usage information can easily be used to extract behavior information (and thus, NILM systems will be important in substantiating our main claim. See next section). For now, we'd like to mention that deployment has so far been confined to pilot pilot, though numerous utilities, including Oklahoma Gas & Electric, China Light and Power, EGAT (Thailand), Buckeye Power (Ohio) and Commonwealth Edison (Chicago) have deployed systems involving up to several hundred sensors [Phone Conv With Bill Rush] .

Current NILM systems require data with a second/sub-second resolution for proper operation. Because of this, processing is usually done locally, at the electricity meter. However, there are no technical constraints preventing NILM algorithms from running remotely, and useful results may be obtained even with data from an AMI system.

III. Players, Use Cases and Motivations

Utilities typically have policies which provide a certain amount of protection for utility records and personal information. For example the California Energy Commission requires written consent for the release of personal data related to billing, credit, and power usage [8]. Utility records may be released in certain circumstances if customer not identified, though exceptions are made for law enforcement. Note that data security and data handling practices are promulgated from utility to third party through contract and audit, so any Meter

Data Management System operators are bound by the same data-handling rules as the utilities.

Given these policies, there exist agencies, organizations and individuals who have natural motives to use power consumption data for purposes other than load research and demand response. These interested parties fall into two categories, those likely to obtain some/all of the information in the current regulatory environment, and those likely to seek it though illegal means. In the former case, the utility may engage in partnership in a for-profit venture or be required to cooperate by the federal government. In the latter case, the expected proliferation of access points may facilitate unauthorized access. We will now list and describe some of these entities, citing precedent where appropriate.

A. Law Enforcement Agencies

Law enforcement agencies have easy access to public utility records, and in some cases routinely use them to seek out drug producers. KXAN Austin recently reported that the Austin Police Department has an agreement that allows it to access Austin Energy power usage records without a search warrant[9]. Investigators have used their access to screen consumers for possible drug production, relying on the fact the heat lamps and watering systems used to grow marijuana indoors can vastly increase an average energy bill. In their response to complaints by the American Civil Liberties Union, police and utility representatives claim that such techniques comply with all state and federal investigative laws. While this claim is disputed and the Austin incident is an isolated case (many utilities require a subpoena for releasing records), the case sets a precedent for increasingly sophisticated future use of consumption data for law enforcement purposes.

There is legal precedent to consider such use an invasion of privacy. In the case of Kyllo v. U.S. (2001) [10], the Supreme Court considered the use of thermal imaging devices to identify the use of heat lamps in a private residence. The court ruled that öbtaining by sense-enhancing technology any information regarding the interior of the home that could not otherwise have been obtained without physical intrusion into a constitutionally protected areaẗonstituted a search and therefore violated the defendant's 4th Amendment rights.

B. Employers

One parameter that can easily be estimated from power usage data is Presence - whether or not person(s) are present within a residence (see Sections III, IV). An employer concerned with productivity or false sick-day claims might use presence information to monitor its employees. A 2006 article in The Denver Post [11] details the use of GPS technology embedded in phones to track employees during the work day. In the article, the director of the Electronic Privacy Information Center expresses concern that the technology may be used for off-work tracking, emphasizing the fact that no clear-cut privacy legislation exists to protect workers from potential abuse.

C. Government

Power consumption may be used as a backup tracking technique to track suspected criminals, terrorists (necessary?)

D. Marketing Partners

Behavoir and appliance usage information may potentially be used for directed advertisements. For example, some NILM systems are powerful enough identify specific appliance brands, and may even identify malfunctioning appliances [CITE]. A marketing company partnering with a utility may use this data to send customers targeted advertisements for repair/upgrade. While not as invasive as the above use cases, targeted advertising of this sort may meet with consumers disapproval and must be considered.

E. Other Utilities

Other Power Companies or subsidiaries may be interested in the data for their own load research and service development purposes.

F. Criminals

Unauthorized Access [8] Current state of security - site Judith Cardell's talk on 'power substation security'

IV. Formalizing Privacy

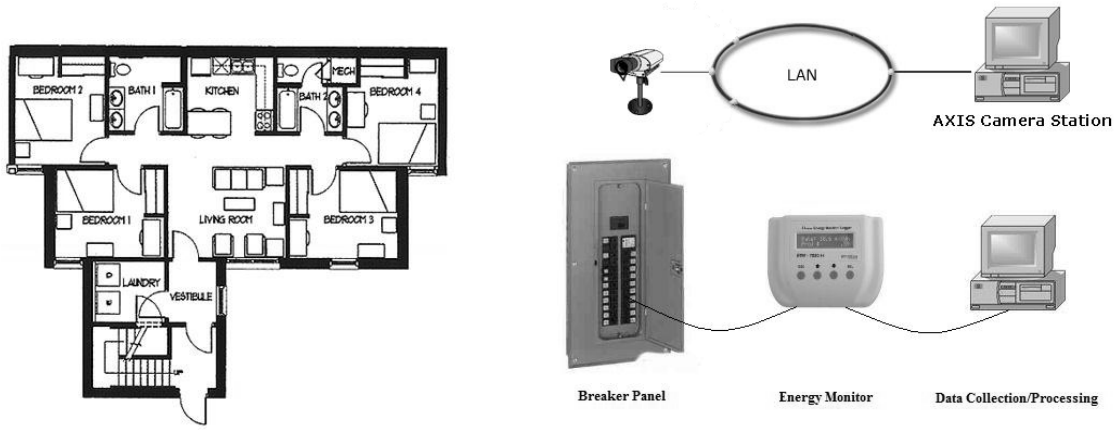
The previous section showed by way of examples that the evolution of monitoring technology creates real threats to individual privacy. However, it's not apparent just how these threats can be quantified, especially as a function of available data. There is a need for a 'privacy metric', which takes associates the degree of data availability (accuracy of readings, time resolution, types of readings, etc) with potential privacy risks, providing a robust and reliable indicator of overall privacy.

In this section, we briefly show how to approach the construction of such a metric. Although the actual construction is the subject of future work, the insights we gain while thinking about it can be applied to our 'proof of concept' demonstration.

To construct a privacy metric, we need to better understand the nature of the information which can be extracted from available sensor data. Thus, we will start by suggesting a formal framework for extrapolating activity.

A. Extrapolating Activity

Extrapolating activity may be thought of in two stages - during the first 'intermediate' stage, NILM in combination with data from other sensors is used to extract appliance usage, track an individual's position, and match particular individuals to particular observed events. During the second stage, the intermediate data is combined with contextual data (such as the number/age/sex of individuals in the residence, tax and income records, models of typical human behavior). Together, these data are used to identify activities, behaviors, preferences, beliefs, and so on. The two stages are not cleanly separated



(a) (b)

Fig. 2. Experimental Setup: (a) shows the floorplan of the residence; (b) shows the camera and electrical data gathering setups

- raw data may be used directly to estimate a parameter of interest, and determination of some intermediate parameters may rely on contextual information. However, many parameters in the second stage rely on the same intermediate data (ex: sleeping habits and eating habits may both be extrapolated from tracking data.)

Note that the nature of the sensors will necessarily lead to 'sample impoverishment' - the data collected will almost certainly be insufficient for accurate tracking and event assignment. For example, if several individuals arrive at the house at once, one can't assign the event 'living room light turns on' to a particular individual with any degree of certainty. Also, a person moving through a residence without triggering any appliances or temperature/humidity sensors is invisible to the system. This limitation will have to be taken into account when defining second-stage parameters .

There is a clear upper limit for first stage - at most, the gathered information will reveal everything that's happening in the house, yielding precise information about all movements, activities, and even the condition of appliances (though it may not be possible to achieve this limit with current or future in-home sensing systems). However, it's more difficult to define an absolute performance metric for the second stage - the number of specific preferences and beliefs that can be estimated is virtually limitless. In order to develop a comprehensive 'privacy metric', one needs to carefully define a list of 'important' parameters, basing importance both on how fundamental a parameter is (how many other parameters may be derived from it) and on home/business owners' expectations of privacy. Expectations of privacy, in turn, are partially based on previous abuse incidents (such as the one in Section III-A). The list of second stage parameters may be hierarchical, with more specific parameters being used to evaluate more general ones. Once an appropriate list is defined and 'importance' values assigned, it is possible to determine the sufficiency of available data based on requirements of current and future NILM, tracking, and other relevant algorithms.

The list of important second-stage parameters form the evaluation criteria. Algorithms for estimating the parameters, along with the corresponding data requirements, provide a method for evaluating the sufficiency of the available data. Together, these provide a metric for how much information may potentially be disclosed by a particular monitoring system. Developing a comprehensive privacy metric is the subject future work for the TRUST Center.

V. EXPERIMENT

Although it is known that first-stage parameters such as appliance usage may be accurately estimated (see performance chart in [7]), to our knowledge no one has tried to extrapolate activity from power consumption data. In this paper we want to prove that activity extrapolation is feasible, thus lending credibility to our thesis and providing an experimental precedent which our collaborators can cite in future efforts. To do this, we conduct a small-scale monitoring experiment on a private residence.

A. Experimental Setup

We conducted our experiment in a typical student residence (Figure 2a). For data gathering, we used the Brultech EML energy usage monitor. Figure 2b shows the data gathering setup. The energy monitor was attached to the residence's breaker panel and sent real-time power usage information to a workstation responsible for data collection. The station recorded power usage at intervals of 1 or 15 second(s) and with a resolution of 1 Watt. The same workstation then ran the NILM and behavior extraction algorithms . To evaluate the system's performance, we placed a network of cameras around the residence. We elected to use the Axis 206 network camera (position shown in Figure 2a), which we connected to a workstation using an Ethernet switch. The workstation ran the AXIS Camera Station software and recorded motion events for later processing. The camera control setup is shown in Figure 2b.

B. Experimental Protocols

The experiment was run semi-continuously over a period of two weeks. This time frame allowed us to obtain repeat data for pattern matching while accounting for time constraints.

The power and camera data collection software was shut down on a semi-daily basis for archiving, maintenance, and manual video data processing.

Electrical data was collected from the house breaker panel and stored as .txt file recording input in 1 or 15 second intervals and with 1 Watt resolution. The data is kept in its raw form for the duration of the experiment and analyzed after its conclusion.

Camera data was collected by the Axis Camera Station software and stored in mpeg format at a resolution of 320x240 at 4 fps. At regular intervals, video data was manually analyzed and processed into activity logs. Upon completion of the logs, the original video data was deleted. Activity logs had the following format:

Date/Time Subject Activity

Here, the subject could be any of the house's three residents or a guest. While residents were identified by name, guests were identified only as *Guest_x*. Possible activities included turning any of the household appliances on or off (ex: *kitchen_lamp_1_on*), entering or leaving the residence, sleeping, preparing meals, taking a bath, or having a party. Note that because the cameras were not put in individual rooms, the resulting activity logs were not fully complete. However, this arrangement respected the residents' privacy and lead to more natural behavior, while the collected data were sufficient to estimate parameters of interest (see Section V-D for the parameters).

Finally, experiment's participants interacted with the system by going about their daily routines. No specific action, other than notifying their guests about the experiment, was required of of the participants.

C. Privacy Protections

The experiment involved potentially serious intrusions into the participants' private lives. Therefore, when designing the experiment we took steps to maximize the participants' comfort, minimize potential for embarrassment, and protect their confidentiality.

First, each participant was given a consent form explaining the experiment, detailing their rights, urging them to ask questions, and highlighting the completely voluntary nature of their participation (participants were free to withdraw from the experiment at any time without penalty). They were also given contact information which they could use to reach us if they had any questions or concerns.

Secondly, video logs were processed by one of the household's residents, which eased the participants' anxiety at being videotaped.

Thirdly, all electrical and video data was kept secure and confidential. Collected data was stored in a password-protected folder, able to be accessed only by individuals directly involved in the project. Also, all publicly available results were stripped of any potentially identifying information.

Finally, we made sure that the project complied with the Cornell Human Subjects Testing guidelines. It has been reviewed and approved by the Cornell Institutional Review Board. The approval request form, consent form, and Experimental Setup & Protocol documents are available from the authors upon request.

D. Parameters to be Estimated

We chose several parameters which we beleived were both revealing and possible to estimate using our data gathering equipment and processing algorithms. They were:

1) Presence/Absence - whether or not someone was present at the house
2) Number of Individuals - if presence we detected, we estimated the number of individuals present.
3) Appliance Use - microwave, stove, water heater, TV, misc appliances etc.
4) Sleep/wake cycle - when, on average, each individual woke up and went to sleep.
5) Miscellaneous Events: Breakfast, Dinner, Shower (if water heater present), Party, etc.

More formally, we begin by combining all data into a single timeline. For each parameter, we partition this timeline into segments, with each segment assigned some value. For most parameters, the value is binary, indicating whether a person present or absent, asleep or awake, etc. The sole exception is the 'Number of Individuals parameter', which is assigned a set value from the partition $\{0, 1, 2, 3, > 3\}$. For a specific parameter, the i^{th} 'on' event is defined by T_i^{on} and T_i^{off}. An example partition for the 'Presence/Absence event is shown below: [DIAGRAM]

E. Behavior Extraction Algorithms

NILM, provided by Prof Baranski, [12],[13]
Behavior extraction algorithms developed by us

F. Performance Metrics and Evaluation

Once energy use data was gathered and processed with NILM / behavior extraction algorithms, we wished to compare the results against reference results obtained from the camera data. To do this, we used the following metrics:

Failure to Detect/ Misdetection:

Define T_{thresh}.

If a camera event occurs, but a corresponding NILM event does not occur within T_{thresh} seconds, declare a *Failure to Detect*.

If a NILM event occurs, but a corresponding camera event does not occur within T_{thresh} seconds, declare a *Misdetection*.

For each successful detection, compute square error between boundary intervals and/or lengths. Average SE over all successful detections.

G. Results

Graphs/interpretations.

VI. Discussion

Concerns for future systems: List of additional parameters that can be estimated with combined power/water/gas consumption

Look at California Energy Commission - raw data is there to run NILM/data extraction algorithms. Potential applications

VII. Guidelines

Paper meant to fit in with work done by the Berkeley School of Law. We draw attention to the issue and make a case for it, while future efforts by the TRUST team & others find solutions

Determining solutions that guard privacy of information is beyond the stated scope of this work. However, we believe the solutions will be in the form of policy, which will establish:

1) Hard prohibitions against relaying certain types of data, [8]
2) Protocols which do most of the data processing at stations located inside the residence or business.
3) Strong user control over information leaving the residence. This will allow research, demand response, etc.. to be done strictly with consent from the consumer.

NOTE: This is where we'll summarize Dr. Mulligan's recent work.

VIII. Conclusion

The conclusion

Acknowledgment

The authors would like to thank Dr. Michael Baranski, who lent us the use of his NILM algorithms. His assistance improved the effectiveness of our behavior extraction tools and helped us make a more persuasive case experimentally.

The authors would also like to sincerely thank Devashree Trivedi, who provided a helpful presence and equally helpful input during every stage of the project, and who single-handedly ran data gathering during the experimental stage.

Finally, the authors would like to thank Derieree Mulligan, Judith Cardell, and others who were very helpful throughout the project duration.

References

[1] "Certs microgrid test bed," Website. [Online]. Available: http://certs.aeptechlab.com/

[2] C. P. U. Commission, "Proceedings on demand response and advanced metering." [Online]. Available: http://www.cpuc.ca.gov/PUC/hottopics/1Energy/R0206001.htm

[3] O. M. of Energy, "Functional specification for an advanced metering infrastructure," jul 2006. [Online]. Available: http://www.energy.gov.on.ca/english/pdf/electricity/smartmeters/Functional_Specification_for_Advanced_Metering_Infrastructure.pdf

[4] "Assessment of demand response and advanced metering," Staff Report, aug 2006. [Online]. Available: http://ferc.gov/legal/staff-reports/demand-response.pdf

[5] E. P. R. Institute, "Advanced metering infrastructure (ami)." feb 2007. [Online]. Available: http://www.ferc.gov/EventCalendar/Files/20070423091846-EPRI\%20-\%20Advanced\%20Metering.pdf

[6] C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, N. Les, and P. Armstrong, "Power signature analysis," *IEEE Power and Energy Magazine*, vol. 1, no. 2, pp. 1540–7977, Mar. 2003.

[7] "Single point end–use energy disaggregation (speed) marketing brochure," 2001. [Online]. Available: http://www.enetics.com/downloads/SPEED\%20Brochure.pdf

[8] Mulligan, Deirdre K. and Lerner, Jack I. and Jones, Erin and King, Jen and Sislin, Catlin and Wilson, Bethelwel and Hall, Joseph, "Privacy and the law in demand response energy systems," Samuelson Law, Technology and Public Policy Clinic, 2006. [Online]. Available: http://www.truststc.org/pubs/36/Jones_PrivacyAndLawInDemandResponse..pdf

[9] K. A. News, "High utility bills may lead police to your door," nov 2007. [Online]. Available: http://www.kxan.com/global/story.asp?s=7322955

[10] S. C. of the United States, "Kyllo v. united states," Opinion of the Supreme Court, Jun. 2001. [Online]. Available: http://www.law.cornell.edu/supct/html/99-8508.ZS.html

[11] T. McGhee, "Gps technology tracks employees," The Denver Post, dec 2006. [Online]. Available: http://certs.aeptechlab.com/

[12] M. Baranski and V. Jurgen, "Detecting patterns of appliances from total load data using a dynamic programming approach," in *Proc. IEEE Fourth International Conference on Data Mining, (ICDM '04)*, Nov. 2004, pp. 327–330.

[13] ——, "Genetic algorithm for pattern detection in nialm systems," in *Proc. IEEE International Conference on Systems, Man and Cybernetics*, vol. 4, oct 2004, pp. 3462–3468.