EXAMPLE USE OF THE SACADA DATA TO INFORM HRA

Y. James Chang[1], Carmen Franklin[1], Laurence Criscione[1]

[1] *U.S. Nuclear Regulatory Commission: Washington, DC, 20555-001, USA Code, and Email Address*

The U.S. Nuclear Regulatory Commission (NRC) sponsors the development of the Scenario Authoring, Characterization, and Debriefing Application (SACADA) system to collect the licensed operator performance information in simulator exercise of nuclear power plants (NPPs). This includes simulator training, exams, and experiments. The objective is to provide human performance indications with statistical basis to inform the human error probability (HEP) estimates in human reliability analysis (HRA). The SACADA system includes software designed for NPPs' training department to use in their routine operator simulator training to collect performance information. The operator simulator training is operated in a time constrained environment, therefore, the SACADA taxonomy is concise to collect key human performance information within a short time. In addition, the collected data can be used for HRA and improving operator simulator training. This enables a long term data collection to be performed by plant staff.

Since 2012, the SACADA system has been implemented in a US nuclear power station's operator simulator training program and a research institute's operator simulator experiment program to collect the operator simulator performance data. The data are entered by the operator trainers, the experiment designers, and the crew. These sources provide high quality data entry to increase the SACADA data population. This paper demonstrates the use of the context similarity data analysis approach to analyze the SACADA data to inform the development of NRC's Integrated Human Event Analysis Methodology for General HRA (IDHEAS-G)[1] by using the 7042 data points available in 2014.

I. INTRODUCTION

The Scenario Authoring, Characterization, and Debriefing Application (SACADA) system was developed by a multi-disciplined team sponsored by the U.S. Nuclear Regulatory Commission (NRC) to collect nuclear power plant operator performance information in simulator exercises. The developers included NRC staff, nuclear plant operator trainers, and experts in the domains of human reliability analysis (HRA), cognitive engineering, and probabilistic risk assessment. SACADA's objective is to provide data-based operator performance information and statistical indications to inform HRA in task analysis and human error probability (HEP) estimates. Since 2012, the SACADA system has been implemented in a US nuclear power station and a research institute to collect their operator performance data in simulator exercises. A few nuclear power stations are piloting the SACADA system to decide whether to routinely use the system in their operator training programs. This paper uses the data in the SACADA database as of April 2015 to demonstrate the use of SACADA data for HRA.

The following is a brief description of SACADA data structure and taxonomy. The detailed SACADA taxonomy can be found in Chang, et al[2]. A simulator training scenario starts with a plant initial condition that could be steady state or a 'normal' transient (e.g., increasing or decreasing reactor power). The initial condition also includes latent failures of equipment or instrumentation. After the simulation starts, the operator trainers actuate a set of plant malfunctions, usually one at a time, for operators to respond to. Each malfunction requires operators to perform a number of activities to successfully handle the malfunction. Principally this includes detecting the occurrence of the abnormal condition, understanding the issue, planning for responses and performing actions to manipulate the system. Understanding the issue in nuclear plant operation generally is represented by entering the correct procedure. The procedures provides the response plan. One the correct procedure is entered, the operator's responsibility is to correctly follow and execute the procedures. For certain actions that need to be immediately performed, the operators are trained to recognize the symptoms and to carry out the required actions immediately. A malfunction example is the emergency diesel generators (EDGs) loss of cooling water during operation. If the cooling is not restored in time or the EDGs are not tripped in time, the EDGs would be damaged by overheat. Give this malfunction, the operators need to promptly detect the loss of cooling water. This could be represented by alarms, flow meter, or the corresponding valve status indications, etc. After noticing the initial abnormal symptoms, the operators would check the related indications to identify and confirm the issue. Based on the situation, the corresponding actions are expected to be performed to handle this malfunction. In SACADA, the key operator responses are explicitly identified. They typically represent the training objective element (TOE) in operator simulator training.

The crew are evaluated based on their performance carrying out the TOEs. General guidance is available for evaluating the crew performance. Some TOEs' success criteria are explicitly specified. For example, the success criterion of the TOE "manually tripping the EDGs" (in a loss of EDG cooling event) is to manually trip the EDGs before the automatic trip.

Every TOE contains context information and performance results. The performance results include performance disposition and performance details. The context represents the performance challenges. Each reactor unit typically has five to six operating crews working in shifts to operate the reactor. All crews take the same simulator training (i.e., using same scenarios). Therefore, for the same scenario, the TOEs' context is identical for all crews. In SACADA, the context is characterized based on macrocognitive functions including detecting (information), understanding (the situation), deciding (choosing response strategies), acting (performing the response strategy), and external-communication. This human-centered taxonomy provides a convenient means to group the TOEs of the same context, but may be different in the specifics of the tasks, to provide human-centered system-neutral performance indications.

Different crews may respond to the same TOE differently, therefore, they show different performance. The performance disposition is performed to indicate the overall performance. This generally includes excellent (SAT+), satisfactory (SAT), satisfactory with deficiencies (SATΔ), and unsatisfactory (UNSAT). If a TOE's performance is rated as SATΔ or UNSAT, additional performance detail information is collected. This includes error types, error specifics, error causes, team error recovery, and overall effects on the scenario.

The TOEs are the basic data analysis unit in SACADA. A data point is a TOE performed by a crew. The 7042 data points mentioned in the abstract are the total crew-TOEs used in this paper's analysis.

II. DATA OF THE ANALYSIS

As of April, 2015, there were 7042 TOEs in the SACADA database that have the required information for data analysis. Each TOE is a data point. The data sources include nuclear power plant (NPP) operator simulator training and operator simulator experiments (two difference sources). Table 1 shows an overview of the data. The data show that among the 7042 TOEs, their cognitive types are relatively evenly distributed. The UNSAT ratios of the cognitive functions show similar distributions. In the SACADA taxonomy, each cognitive type has a set of performance influencing factors (PIFs) to characterize the context. Because human performance is sensitive to the context, a set of explicit PIFs characterize the context and facilitate the analysis of context effects on human performance.

The purpose of this paper is to demonstrate the use of SACADA data for HRA. The statistical values shown in this paper are only for illustration purpose. They are not the results that can be directly applied to calculate HEPs in HRA. In addition, a validation of data quality has not been performed. This may affect the values shown in table 1.

Table 1 Data overview

Cognition Type	# of TOE	# of UNSAT	UNSAT Ratio (%)
Detecting or Monitoring	1367	10	0.73
Understand or Diagnosis	1001	16	1.6
Deciding or Response Planning	1990	17	0.86
Action or Manipulation	1838	33	1.8
External Communication	846	13	1.5
TOTAL	**7042**	**89**	**--**

II. DATA ANALYSIS

II.A. UNSAT Examples

This section provides four UNSAT examples to illustrate the human performance classified as UNSAT. The first example is a large loss of reactor coolant event and subsequent loss of the emergency recirculation. While performing the main emergency operating procedures (EOPs), i.e. "loss of reactor or secondary coolant" and "loss of emergency recirculation" EOPs, the crew needs to monitor the critical safety function tree (CSFT) for appropriate procedure transition. A TOE is to transfer to the "response to inadequate core cooling" EOP when the core exit temperature exceeds a pre-set temperature. The crew performance of this TOE was classified as UNSAT because the crew took too long to transfer to the EOP after the temperature threshold was reached. The supervisor noticed the issue but with delay. After the supervisor noticed the issue, the crew entered the correct EOP. The performance deficiency type (or error mode) was classified as poor teamwork, specifically in coordination. The contributing factor (or PIF) was performing multiple tasks. The crew was implementing multiple procedures at the time the core exit temperature reached the procedure transfer threshold. The performance deficiency was recovered with delay by the supervisor. Overall, the delay in error recovery did not impact plant control.

The above description of the performance deficiency (or human error) shows the types of information SACADA collected for the TOEs with performance dispositions of SATΔ or UNSAT. The information includes error mode (i.e., teamwork), error details (i.e., coordination), error causes (i.e., performing multiple tasks), error recovery (i.e., recovered with delay by supervision), and impact on the scenario (i.e., no impact on plant control). In addition, the remediation action information (to prevent the same deficiency from re-occurring) is also collected. The remediation information is not discussed in this paper.

The second example is a steam generator tube rupture (SGTR) event. After the ruptured steam generator was isolated and the reactor coolant system (RCS) was cooled down and depressurized, a TOE was to terminate the safety injection and to control the RCS pressure and makeup flow to maintain the RCS pressure stable at the SG pressure. This TOE was classified as UNSAT. The performance deficiency type was classified as teamwork in coordination. The personal attitude was identified as the main contributor to the poor coordination. The performance deficiency was not recovered in this case. In this TOE, the crew did not comment on what exactly happened, e.g., the safety injection was not terminated or the RCS pressure was not controlled appropriately.

The third example is a main steam line break (MSLB) event. In this scenario, prior to the MSLB malfunction, the crew responded to malfunctions of a loss of instrumentation air and a loss of instrumentation. Immediately after the MSLB, the crew should, by memory, perform the post reactor trip immediate actions. This TOE was classified as UNSAT because the crew did not verify that the 480V load center breakers closed during performance of the immediate actions. This deficiency was classified as an indicator monitoring issue. The main contributor was attention distraction because of responding to the malfunctions prior to the MSLB event. The mistake was recovered with delay by the crew's questioning attitude. This performance deficiency did not impact plant control.

The fourth example is in a large loss of coolant accident. An important reactor operator's action is to secure all pumps taking suction from the Refuel Water Storage Tank (RWST) when the RWST's level is below a set point. In this TOE, the reactor operator did not secure all pumps aligned to the RWST prior to the water level reaching the set point. This was because the reactor operator thought the RWST water level would stop decreasing after swapping the low head safety injection's suction from the RWST to the sump. The secure all pump instruction is written as a caution within the procedure, therefore, the operator is expected to memorize the caution when performing the procedure. In this TOE, the deficiency was not recovered and resulted in the required actions not being performed.

As indicated in the above four examples, a TOE's performance disposition is evaluated based on the crew performance from the training requirement, not from HRA. The success criteria are more stringent in training requirements than in HRA. For example, the above examples 1 and 3 have delayed error recovery and the plant control is not affected. However, they were classified as UNSAT. Therefore, the UNSAT in training does not equal to the human error in HRA, which typically means that the human performance deficiency resulted in loss of plant function. In addition, the TOEs include operator actions not modeled in PRAs. The SACADA TOE coverage is much broader than the human actions modeled in PRAs. For example, timely declaration of an emergency action level (EAL) is in SACADA but is not modeled in PRAs.

II.B. Data-Based Quantitative Performance Indications

II.B.1. Context-Similarity Based Quantitative Performance Indication

When developing the SACADA taxonomy, the developers envisioned using the context-similarity approach to use SACADA data to inform HRA. Human performance of a task could be significantly different if performed in different contexts. In the NRC's Integrated Human Event Analysis System – General methodology (IDHEAS-G), the analysis of a human failure event (HFE) is to identify the tasks critical to the success of the HFE. The HFE's HEP is sum of these critical tasks' HEPs. A critical task's HEP is calculated by identifying its key macrocognitive functions and specifying the context of each macrocognitive function. The context is represented by a set of performance influencing factors (PIFs). The critical tasks HEP is calculated based on the macrocognitive functions and the PIFs. In SACADA, every TOE includes the dominant macrocognitive function and the PIFs' statuses. The information is entered by operator trainers or the researchers who design the simulation scenarios. The information can be used to inform the HEPs of the IDHEAS-G's critical tasks that, in turn, inform the HFE's HEPs.

II.B.2. Performance Influencing Factors for Deciding Macrocognitive Function

Within the 7042 data points (TOEs) used for this analysis, 1990 of them are classified as the decision making (or response planning) macrocognitive function. This is classified by the operator trainers or researchers who conduct the simulator experiments. Example decisions are transferring to a new procedure or how the actions should be performed, e.g., deciding the reactor coolant system cooldown rate. This section focuses on the deciding macrocognitive function TOEs.

SACADA uses a few PIFs to characterize the context of performing a deciding macrocognitive function. These PIFs include only the PIFs that affect the deciding macrocognitive function (cognition-specific PIFs) and the PIFs that affect all macrocognitive functions (overarching PIFs). The deciding cognition-specific PIFs include decision basis, decision familiarity, and decision uncertainty. The overarching PIFs include workload, time criticality, extent of communication, and a number of miscellaneous PIFs. Note that these PIFs do not intend to be a complete list of PIFs for making decision for all applications, instead, these PIFs are identified specifically for operator simulator exercises. This enables a concise taxonomy to be practically implemented in nuclear plants' operator simulator training program that has high time constraints for data collection. For example, environmental factors are not expected to be present in simulator exercises, therefore, environmental factors are not included in the SACADA taxonomy. The PIFs affecting the deciding macrocognitive function are discussed below:

Decision Basis: Three decision basis options as described below:

- Procedure: The decision is driven by procedures or other guidance.
- Skill: Skill-driven decision; without procedure, operator can make decisions from memory. Examples are immediate operator actions after a reactor trip.
- Knowledge: No procedure applicable; crew relies on engineering or technical knowledge and operating experience.

Decision familiarity: In crew simulator exercises, procedures are generally available to the situation. However, the procedures may not be perfectly match to the scenario. Therefore, the procedure cannot be literally followed. Instead, the operator has to, based on the scenario, interpret the procedure instruction as necessary to implement the procedure as intended. The decision familiarity includes the following three options:

- Standard: Crew has previously trained on this challenge.
- Anomaly: Standard training must be adapted to fit an anomalous situation (e.g., the procedures do not cover the circumstances).
- Novel: This involves a change to the way the challenge is addressed, such as a new procedure, scenario, or role.

Decision uncertainty: This refers to the clearness of the criteria for making the decision. Four options are available:

- Clear: No uncertainty or competing goals – clear decision criteria.
- Uncertain: Lack of information or ambiguous decision criteria.
- Competing Priorities: Multiple competing goals, foreseeable severe consequences.
- Conflicting Guidance: Policies, practices and procedure have conflicting guidance.

The overarching PIFs, miscellaneous PIFs, and their available status are shown in Table 2. The overarching PIFs have multiple optional statuses. The miscellaneous PIFs' statuses are either present or not present.

Table 2 SACADA's overarching PIFs and miscellaneous PIFs.

Workload: o Normal: All crew members have peer check and backup. o Concurrent Demand: one crew member has own task with no backup; all others have normal peer check and backup. o Multiple Concurrent Demands: Overloaded, no peer check. Everyone has their own task with no backup.
Time Criticality: o Expansive Time Available o Normal Time Available o Barely Adequate Time Available: e.g., high tempo, time pressured tasks.
Extent of Communication: o Normal: Standard level of three-way communication within control room, with occasional onsite communication. o Extensive Onsite: High level of close communication with on-site operators (e.g., to coordinate fire response). o Extensive within Control Room: High level of close communication within control room (e.g., to coordinate actions between board operators):
Miscellaneous: o Non-Standard: Anomalous conditions forcing the operator to account for previous discoveries/incidents/failures. o Noisy Background: Loud background noise makes communication challenging. o Coordination: Requires close coordination with on-site personnel. o Communicator Unavailable: Designated communicator is needed but is not available. o Multiple Demands: Multiple competing demands on attention/distractions. o Memory: Demand on memory.

II.B.3. Context Specific Performance Indications

Among the 1990 decision TOEs, 1282 TOEs have decision basis classified as procedure based. Among the 1282 TOEs, 959 TOEs have decision familiarity classified as standard. Among the 959 TOEs, 861 TOEs have decision uncertainty classified as clear. This process continues to identify the number of the TOEs that met the additional requirements including the workload status is concurrent demand, the time criticality is normal, the extent of communication is either extensive within the main control room (MCR) or extensive onsite, and no presence of miscellaneous PIFs. To that end, 201 TOEs are identified with the specific PIF combinations. The 201 TOEs include 16 different tasks in nine different scenarios. The same tasks are performed by a number of crews. None of the 201 TOEs had a performance that was dispositioned as UNSAT. Therefore, for the specific context, SACADA data shows zero UNSAT out of 201 response opportunities (0/201). This is an example of SACADA showing the combination information of context, number of response opportunities, and number of UNSAT responses to provide context-specific performance indications.

II.B.4. Performance Influencing Factors Effects on Human Performance

A PIF's effect on performance can be evaluated by applying the same analysis process described in section 3.2.3 with changing the PIF to a different status. It is similar to performing the one-factor-at-a-time sensitivity analysis. For example, the 201 TOEs identified in the previous paragraph have the same combination of PIFs' statuses. To evaluate the effect of changing the decision basis from procedure-based to knowledge-based, the analysts would perform the same process but change the status of the PIF of interest to a different status. In this demo, only the decision basis' status is changed from procedure-based to knowledge-based. All the other PIFs' statuses remain the same. This identifies 115 TOEs. There is one UNSAT response in the 115 TOEs. Therefore, the SACADA data indicates that changing the decision basis from procedure-based to knowledge-based of the specific context (i.e., decision macrocognitive function, decision familiarity is standard, decision uncertainty is clear, workload is concurrent demand, etc.) the human performance changed from zero UNSAT responses in 201 response opportunities (0/201) to one UNSAT response in 115 response opportunities (1/115). This not only demonstrates a data-based indication of PIF's effects on human performance but also a context-specific, instead of general, PIF effects. This could be important because a PIF's effects on human performance could vary from context to context. Current HRA methods' implementation of PIFs' effects on human performance is constant. In HRA methods, PIFs' effects on human performance do not change when the other PIFs' statuses are different. SACADA data could provide indications of the context's effects on PIFs' effects on human performance. This could lead to identification of the cliff effects (sharp changes in human performance) of certain PIF combinations. The information can be useful to inform the error forcing context mentioned in the A Technique for Human Event Analysis (ATHEANA) method[3].

II.B.5. Overall Difficulty's Effects on Human Performance

The human performance cliff effects occur when the overall situation approaches human limitations. Using a two dimensional diagram (X-Y plot) with the X-axis representing the overall difficulty of performing a task, and the Y-axis representing the likelihood of failure, psychological evidence suggests that the change of failure likelihood can be divided into three regions on the X-axis. In the low difficulty region (the first region), the failure likelihood increases slowly with an increase of the difficulty level. When the difficulty level reaches a threshold the failure likelihood increases rapidly. This threshold represents the boundary between the first and second regions. In region two, a small change in task difficulty would significantly increase failure likelihood. In other words, in the second region failure likelihood is very sensitive to the change of the overall task difficulty. This pattern continues until the task difficulty increase to reach to another threshold (between the second and third regions). In the third region, an increase in task difficulty only results in small increase in failure likelihood. Psychological experiments show the threshold between the first and second region occur when there are a few negative PIFs.

Analysis of the 7042 SACADA data points did not reveal the same failure likelihood pattern as described in psychological literature. The analysis is performed by surrogating the level of task difficulty with the number of the negative PIFs. More negative PIFs imply that the overall difficulty is higher. The number of negative PIFs, the number of UNSAT responses, the number of data point (TOEs) and the UNSAT ratios are show in Table 3 for the TOEs of understanding macrocognitive function in the SACADA database. No negative PIFs indicates that the understanding or diagnosis task is straightforward. Figure 1 is a plot of the Table 3 data. In Figure 1, the X-axis is the number of negative PIFs which is used as a surrogate of the overall difficulty of performing the diagnosis TOEs. The Y-axis is the UNSAT ratio which is an indication of failure probability. The red line shows the X-Y plot that, in theory, should show the '∫' shape relation between the difficulty of performing the TOEs and the TOEs' failure probabilities. As shown in Figure 1, the UNSAT ratio slightly increases when the number of negative PIFs increased from zero to one. When the number of negative PIF increases from one to two, the UNSAT ratio increases significantly. By theory, when the number of negative PIFs increases, the UNSAT ratio should increase. However, the SACADA data do not show that trend. Instead, when the number of negative PIFs is three, four or five, their UNSAT ratio is much lower than when the number of negative PIF sis two. The crews performed well in the situations where the number of negative PIFs is four and five. There are no UNSAT responses in these two situations. These two situations represent the most challenging context to crew performance among the TOEs of the understanding macrocognitive function of the analyzed data. This is a surprise to the '∫' shape hypothesis.

Analysis of the number of negative PIFs' effects on the UNSAT ratio is also performed for other macrocognitive functions. The results do not show the '∫' shape performance curve. There are many potential reasons to explain this surprise such as insufficient data points, the PIFs' effects on performance are different from PIF to PIF, the operators being more cautious in performing difficult tasks (thus resulting in lower UNSAT ratios), and the quality of context characterization, etc. It would require more effort to analyze the data and more data may be needed to reach meaningful conclusions. However, the purpose of this demonstration is not to show that SACADA results match psychological hypotheses, but to show how the SACADA data can be used to inform HRA.

Table 3 The classification of the understanding TOEs based on number of negative factors.

# of Negative PIFs	# of UNSAT	Data Points	UNSAT %
0	1	377	0.27
1	2	362	0.55
2	12	132	9.09
3	1	86	1.16
4	0	29	0.00
5	0	15	0.00

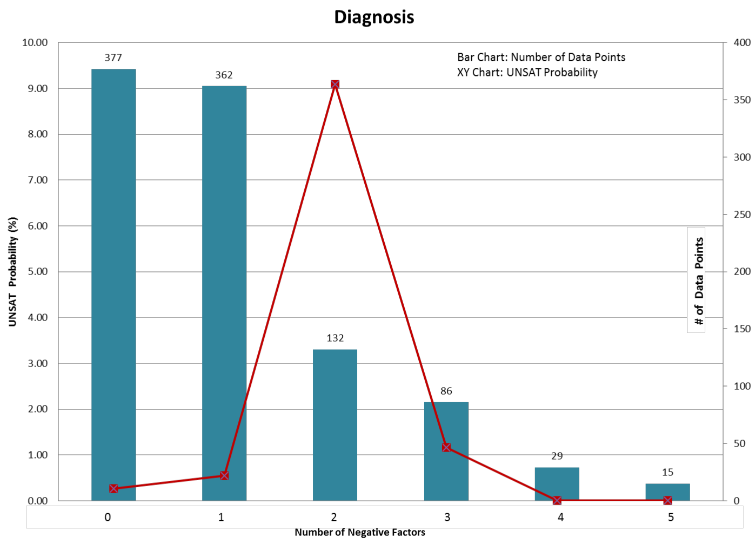

Fig. 1 TOEs of the understanding macrocognitive functions and UNSAT ratio in the SACADA database of this analysis.

III. CONCLUSIONS

This paper uses the limited data in the SACADA database as of April 2015 to demonstrate ways that SACADA data can be used to inform HRA methods, qualitatively and quantitatively. It is also feasible to analyze the UNSAT TOEs to identify the underneath relation between error mode and causes (i.e., PIFs). Analysis of the 89 UNSAT responses is not discussed in this paper. The demonstration has, at least, two significant outcomes: (1) It demonstrates that the SACADA data structure and taxonomy provide a convenient means to use the context-similarity approach to inform HRA methods; (2) It provides context-dependent PIF effects on human performance. This enables having data to perform in-depth studies of PIF effects on human performance. It opens potential to identify the generic forms of error forcing contexts. SACADA has demonstrated that it can be used in nuclear power plant operator simulator training to collect crew performance data. This opens the potential for having a large number of data points to perform data-based or data-driven human performance indications of the nuclear plant control MCR crew. As of May 2016, there are more than 10,000 data points in the SACADA database. The NRC is in the process of analyzing the existing data to gain insights on the meaning of the data and to identify the areas of improvement of the SACADA system. The NRC desires to collaborate with nuclear power plants and research organizations to use the SACADA system to collect and share the operator simulator exercise data to improve HRA and operator simulator training.

REFERENCES

1. Xing, J. and Chang, Y.J. The general methodology of an integrated human event analysis system (IDHEAS), a working draft, in-progress report, ML#16074A389, U.S. Nuclear Regulatory Commission
2. Chang, Y.J., Bley, D., Criscione, L., Kirwan, B. Mosleh, A., Madary, T., Nowell, R., Richards, R., Roth, E.M., Sieben, S., Zoulis, A. “The SACADA Database for Human Reliability and Human Performance” the journal of Reliability Engineering and System Safety, 125(2014) 117-133.
3. U.S. Nuclear Regulatory Commission, “Technical Basis and Implementation Guidelines for A Technique for Human Event Analysis (ATHEANA),” NUREG-1624, Rev.1, (2000).