A framework for using SACADA to enhance the qualitative and quantitative basis of HRA

Katrina M. Groth*

Systems Risk and Reliability Analysis Lab, Center for Risk and Reliability, Department of Mechanical Engineering, University of Maryland, College Park, MD

Abstract: The purpose of this research is to explore ways to use Bayesian methods with data from the US Nuclear Regulatory Commission's (NRC) Scenario, Authoring, Characterization, and Debriefing Application (SACADA) system. SACADA is a database designed to enable collection of nuclear power plant (NPP) control room simulator and crew training data to improve both operator training and human reliability analysis (HRA). This paper presents a framework to use SACADA data and causal modeling to enhance the qualitative and quantitative aspects of HRA. The framework is a multi-faceted approach involving causal models as well as multiple sources and types of data. Elements of the framework include a comprehensive set of performance influencing factors, crew failure modes, Bayesian Network causal models, Bayesian parameter updating, and temporal modeling. This paper also outlines a path forward for developing the framework to enhancing the technical basis of HRA and enabling streamlined use of SACADA data as the volume and variety of data increases.

Keywords: HRA data, Bayesian updating, SACADA, Bayesian Networks

1. INTRODUCTION

A critical challenge for the field of human reliability analysis (HRA) is the need for traceable, data-informed models that provide a defensible basis for risk-informed decision making. Currently the U.S. Nuclear Regulatory Commission (NRC) is pursing data collection through the recently developed Scenario, Authoring, Characterization, and Debriefing Application (SACADA) framework and database [1]. The SACADA database is one of several international data collection activities focused on collecting human performance data from nuclear power plant (NPP) control room simulator scenarios. It also offers a common basis and structure for HRA data collection, analysis, and exchange. SACADA is actively being populated with data and, in parallel, the initial data is being analyzed to provide insight into the development and use of SACADA [2].

As SACADA and similar databases become more mature, an important research question emerges: *how can this data be used to improve HRA?* As a first step toward answering that question, the U.S. NRC asked three teams to develop methods for using SACADA data to quantify the probability of a human failure event associated with a given performance context. In HRA, this quantity is called the human error probability (HEP), which is a conditional probability with the conditioning factors representing the context of performance in terms of performance influencing factors (PIFs), also called performance shaping factors (PSFs). This paper is one outcome of the analysis being conducted by the University of Maryland. The methods developed by the two other teams are also presented in papers at this conference [3], [4].

This paper defines a framework for using SACADA data to improve both the qualitative and quantitative basis of HRA using causal Bayesian Networks and Bayesian parameter updating. The SACADA data is described in Section 2 of this paper. Section 3 describes the approach to development of the framework. The proposed framework is described in Section 4.

* kgroth@umd.edu

2. DESCRIPTION OF SACADA DATA

2.1. Database structure

The SACADA data taxonomy is described in detail in [1] and the current state of development is also described in another paper in this conference [2]. SACADA collects data about multiple facets of crew and system (machine) performance in NPP control rooms. SACADA provides a detailed structure for collecting information about the characteristics of a simulator scenario, the plant conditions, a task-level breakdown of activities involved in responding to conditions in those scenarios, crew roles, and the performance outcome (aggregated at the task-level) for multiple crews that have performed the scenario. SACADA provides a common set of elements for capturing the context of the scenario (which is generally the same across all crews), and a set of performance factors that are used for debriefing crews when performance is considered less-than-satisfactory.†

Each scenario starts with a plant initial condition and contains one or more plant malfunctions that are pre-programmed to occur in the simulator during the exercise. These scenarios are designed before the crews are run on the scenario, and multiple crews run each scenario.

Each scenario is decomposed into a series of tasks or *training objective elements (TOE),* each of which represents one activity that the crews must complete to respond to the specified plant condition or malfunction. The TOE is the basic data unit for SACADA. In general, there are several TOEs involved in the response to each malfunction in a scenario. Each TOE is associated with one of five macrocognitive functions: monitoring/detecting; diagnosis; response planning/decision making; manipulation/execution; and communication/coordination.‡ Example TOEs include things like "ensure the charging cooling pump 1A is in service," "announce transition to procedure [number]" and "monitor [system X] and identify increasing trend in pressure." From an HRA perspective, these TOEs resemble tasks or sub-tasks. Many HRA methods are designed to be used at the scenario or event level, and thus would include multiple TOEs.

For each TOE, the data collection team characterizes the context of the scenario using situational factors (SFs), which are similar to the HRA concepts of PIFs. There are approximately 29 SFs in SACADA, although only a subset are used for each TOE depending on which type of macrocognitive function is associated with the TOE. Some of these factors are rated as one of two states (present/absent), some have up to four states, and some represent a summation of multiple constituent factors rated on the two point (present/absent) scale, as described in [1]. The SFs document the context of each TOE, and they do not change depending on which crew is running the scenario. The outcome of the crew performance for each TOE is ranked on a four-point scale (ordered from best to worst performance, where "SAT" is an abbreviation of the word satisfactory): SAT+, SAT, SATΔ, UNSAT.

For crews that receive a score of SATΔ or UNSAT, a second worksheet is completed to capture the causes of the degraded performance. The worksheet contains approximately 21 performance factors (PFs) that are also similar to PIFs. As with SFs, only a subset of the PFs are used depending on which macrocognitive functions are involved. Several of the PF factors are rated as being in one of two states (present/absent), some have up to six states, and some represent a summation of multiple factors rated on the two point (present/absent) scale as described in [1]. These PFs are used to describe the reasons for errors (or near misses) for the crew on a specific TOE.

2.2. Current data

† Satisfactory performance is defined from a training perspective. Use of this data for HRA purposes requires some additional considerations, which are described later in this paper.

‡ In SACADA, this is defined as "external communication," meaning communication beyond the crew or team. It is worth noting that this may be a narrower interpretation of the macrocognitive concept of "communication," which also involves the concept of teamwork and within-team communication.

This section provides summary information about the current SACADA data set provided for descriptive purposes. It is important to caution against misinterpretation of the summary information provided in this section. Because of the causal nature of the underlying factors, the author cautions that **no statistics of conclusions about HRA or human performance should be drawn from these descriptive data, because they are aggregated across contexts and because not all combinations of contexts are represented in the current data.** Some contexts may be over- or under-represented in the training data when compared to the contexts experienced in real operational events.

As of July 2017, SACADA contains data from 86 simulator scenarios with results coming from both published international simulator experiments and non-published operator training activities. Within these 86 scenarios, there are 329 malfunctions and a total of 2,155 TOEs. Each scenario was performed by several crews, and on average each TOE was performed by 12 crews. In total, the current SACADA database contains 26,153 crew-TOEs (this number represents the sum-product of the 2,155 TOEs and the variable number of crews that attempted each TOE). Of the 2,155 TOEs, 149 TOEs had one or more crews with a rating of UNSAT and 219 TOEs with a rating of SATΔ. Of the 26,153 crew-TOE data points, there were 209 scores of UNSAT and 261 scores of SATΔ.

3. APPROACH TO DEVELOPMENT OF THE FRAMEWORK

After initial analysis of the SACADA data and review of existing HRA needs, the next step was to consider the desirable factors of the framework for using this data to enhance HRA. The approach to development of the proposed method was based on the desirable characteristics of advanced HRA methods outlined by many HRA studies (e.g., [5]). In particular, it was decided that during the first stage of this work, the method should meet the following criteria:

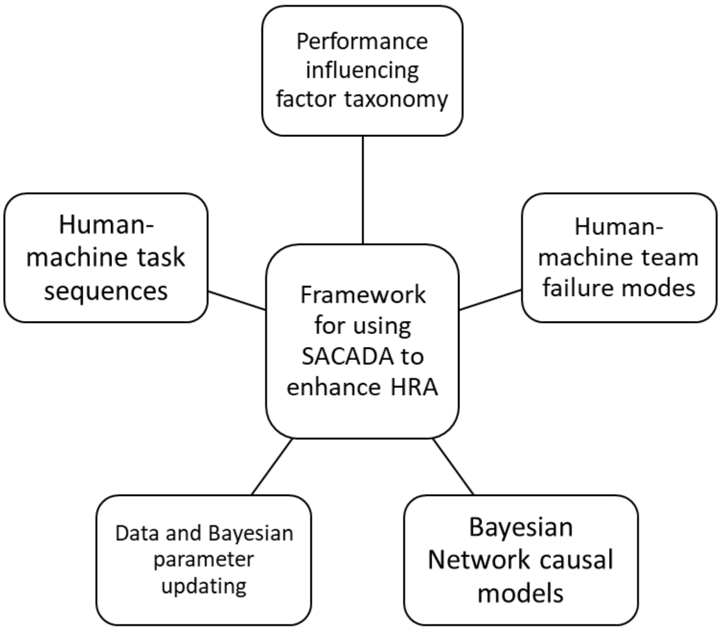
1. The proposed method should be based on a causal model of human-machine performance.[§] That model should be rooted in both cognitive science and systems science.
2. The structure of the model should provide explicit representation of the causal factors that affect human-machine performance.
3. The method should support the quantitative and qualitative aspects of HRA as a part of probabilistic risk assessment (PRA), including quantification of HEPs.
4. The method should provide a framework that is both data-informed and model-informed.
5. The method should be flexible enough to accommodate changes in SACADA structure as SACADA is developed further.
6. The method should be able to incorporate additional data, models, and information to address contexts or factors that are not represented in SACADA.
7. The framework should be capable of providing quantitative insights that can be used to help the data collection teams improve human performance (e.g., via training).

4. PROPOSED FRAMEWORK

The main elements that are included in the proposed framework are illustrated in Figure 1. These elements are a comprehensive set of PIFs, human failure modes, Bayesian Network causal models, Bayesian parameter updating, and temporal evolutional of performance. Each of these elements is described in more detail in this section.

[§] I use the term “human-machine performance” to emphasize that there is an important role of both machines and humans in the concept of human reliability.

Figure 1: Main elements of the proposed framework for using SACADA to enhance HRA



4.1 A comprehensive set of PIFs

The first element in the proposed framework is a comprehensive taxonomy of PIFs [6]. The taxonomy provides a consistent vocabulary and structure for combining data and information from multiple sources and at multiple levels of detail in a transparent and repeatable way. The Groth and Mosleh taxonomy also provides non-overlapping, orthogonal set of causal factors, meaning that each PIF is defined uniquely. The term “orthogonal” is used to indicate that, while the factors are uniquely defined, they may not be independent in a statistical or causal sense.

4.2 Human-machine team failure modes

This aspect of the framework involves having a defined set of failure modes for the human-machine team. These capture the ways that the human-machine teams could fail in responding to a condition or malfunction. The term “human-machine failure mode” is designed to reflect that the HRA concept of a human failure event (HFE) involves a contribution from both the human response and the machine response. In the proposed framework, each of these failure modes would be represented in a causal model as described in Section 4.3.

The definition of these human-machine team failures modes could be achieved via multiple approaches. One option is to use the failure modes or failure-mode-identification approaches defined by the existing HRA methods, which acknowledge both a cognitive and a machine element (e.g., methods such as IDA **[7]**, IDAC **[8]** IDHEAS **[9]**, or PHOENIX **[10]** and HRA research activities **[11]–[15]**). Another option is to use first-principles reliability techniques (e.g. by doing a “human-machine” FMEA or HAZOP with consideration of macrocognitive functions and machine functions). Another approach would be to define one failure mode for each of the five macrocognitive functions used in SACADA.

4.3 Bayesian Network (BN) causal models

In the proposed framework, BN causal models are used to capture the detailed causal pathways and interdependencies among PIFs. In this framework, BNs are used because of their ability to model cause-and-effect relationships and the use of probabilistic inference. Furthermore, BNs provide the ability to reason about any variable in the model, enabling quantitative insights relevant to improving performance.

This approach will allow all relevant PIFs (both observable and unobservable) to be used in model development. In addition, it enables the explicit inclusion of data collection elements within the model structure, and these data collection elements could enable using data from different versions of SACADA.

These causal relationships would be initially developed from cognitive science and systems science, and could eventually be developed or validated by using SACADA data. The BN model will be developed by creating an explicit map between PIFs (first element described above) and the human-machine team failure modes (second element described above). This could be accomplished by using the causal mapping approach developed and illustrated in the work of Zwirglmaier, Straub, and Groth **[16]**. The starting point for development of this causal mapping has been developed as part of NRC's work on defining a cognitive basis for HRA **[17], [18]** and in the IDAC model **[19]**. The size of the resulting BN models could be reduced (e.g., to facilitate quantification) by using node reduction algorithms **[16]**. A second approach would be to follow the approach of **[20]** and use factor analysis clustering, or structure learning algorithms techniques directly on the SACADA data.

4.4 Data and Bayesian parameter updating

The quantitative parameters of the BN models would be populated using multiple sources of data. Prior information on the relationship between the PIFs and the human-machine failure modes could be defined by using an existing HRA method (e.g., as illustrated in **[21]** using SPAR-H **[22]**) or other HRA databases (e.g., **[23]–[27]**). Prior information on the PIF prior probabilities could be assembled from a variety of published sources in cognitive science and HRA, or via formal expert elicitation with HRA experts. The SACADA data would be used as information to Bayesian update multiple parameters within the model, using a Bayesian updating approach **[28]**.

There are several reasons to include this aspect in the framework. First, it allows the causal model to be populated with appropriate information from multiple sources, including data. It also enables the model to include variables and information which are unlikely to be represented in the data (e.g., control room environment). This also provides a way to address the reality that some factors will be over- (or under-) weighted in the training contexts (e.g., high task complexity).

4.5 Human-machine task sequences

The final aspect of the framework involves modeling the sequential aspects of human-machine activities associated with the response to a malfunction. This addresses the need to treat an HFE as the outcome of a process involving several sequential activities or tasks involving different macrocognitive functions. This notion that human failure involves a series of activities and that there is dependency between HFEs has been acknowledged in even the earliest HRA methods **[29]**. It is considered in depth in simulation-based methods which explicitly model sequences of activities or subtasks involved in PRA event (e.g., **[7], [8], [30], [31]**). This sequential and semi-temporal aspect of performance needs to be modeled explicitly in order to reflect the fact that that failure is a process, not a single event.

SACADA provides the first opportunity to use data inform this process. This opportunity arises as a result of several coupled aspects of the data. The first aspect is the detailed consideration of TOEs at the level of macrocognitive tasks (rather than at the higher event level used in many HRA methods). The second is the alignment of these TOEs with responses to specific malfunctions (more akin to the event level in HRA), and the third is that the PIFs are collected at the TOE level. This provides new potential to transform the treatment of sequential and temporal aspects of the PIFs and the activities that comprise an HRA event. A mechanism for modeling this dependency would be through the use of Dynamic Bayesian Networks, which were first introduced in **[12], [32], [33]** and further expanded within the HUNTER framework **[30], [31], [34]**.

5. CONCLUSION

The SACADA database provides a unique opportunity to enhance the foundations of HRA. This work provides a new framework for enhancing the foundations of HRA by using SACADA data together with scientific information and new modeling approaches. The next steps of this work involve further developing each aspect of the proposed framework. The framework proposed in this work will enable a path toward an HRA vision that is both model-based and data-informed, enhancing the technical basis of HRA and enabling streamlined use of SACADA data as the volume and variety of SACADA data increases.

Acknowledgements

This research was partially supported by the U.S. Nuclear Regulatory Commission (NRC). The content is solely the responsibility of the authors and does not represent the official views of the NRC.

References

[1] Y. J. Chang *et al.*, “The SACADA database for human reliability and human performance,” *Reliability Engineering & System Safety*, vol. 125, no. 0, pp. 117–133, 2014.
[2] Y. H. J. Chang and C. Franklin, “SACADA Data for HEP Estimates,” presented at the Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 14), Los Angeles, CA, 2018.
[3] P. Nelson and R. Grantom, “Methodology for Determination of Human Error Probabilities from Simulator Sourced Data,” presented at the Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 14), Los Angeles, CA, 2018.
[4] M. A. Azarm, I. S. Kim, C. Marks, and F. Azarm, “Analyses methods and pilot applications of SACADA Database,” presented at the Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 14), Los Angeles, CA, 2018.
[5] A. Mosleh and Y. H. Chang, “Model-based human reliability analysis: prospects and requirements,” *Reliability Engineering and System Safety*, vol. 83, no. 2, pp. 241–253, Feb. 2004.
[6] K. M. Groth and A. Mosleh, “A data-informed PIF hierarchy for model-based Human Reliability Analysis,” *Reliability Engineering and System Safety*, vol. 108, pp. 154–174, Dec. 2012.
[7] S. H. Shen, C. Smidts, and A. Mosleh, “IDA: a cognitive model of nuclear power plant operator response during abnormal conditions,” PhD Thesis, University of Maryland at College Park, 1994.
[8] Y. H. J. Chang and A. Mosleh, “Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents: Part 1 - Overview of the IDAC Model,” *Reliability Engineering and System Safety*, vol. 92, no. 8, pp. 997–1013, Aug. 2007.
[9] J. Xing, G. Parry, M. Presley, J. Forester, S. Hendrickson, and V. Dang, “An Integrated Human Event Analysis System (IDHEAS) for Nuclear Power Plant Internal Events At-Power Application,” NUREG-2199, Volume 1, May 2017.
[10] N. J. Ekanem, A. Mosleh, and S.-H. Shen, “Phoenix – A model-based Human Reliability Analysis methodology: Qualitative Analysis Procedure,” *Reliability Engineering & System Safety*, vol. 145, pp. 301–315, Jan. 2016.
[11] J. Oxstrand, D. L. Kelly, S.-H. Shen, A. Mosleh, and K. M. Groth, “A Model-Based Approach to HRA: Qualitative Analysis Methodology,” in *Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 11)*, Helsinki, Finland, 2012.
[12] A. Mosleh, S.-H. Shen, D. L. Kelly, J. H. Oxstrand, and K. Groth, “A Model-Based Human Reliability Analysis Methodology,” in *Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 11)*, Helsinki, Finland, 2012.
[13] K. M. Groth, S.-H. Shen, J. Oxstrand, A. Mosleh, and D. Kelly, “A Model-Based Approach to HRA: Example Application and Quantitative Analysis,” in *Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 11)*, Helsinki, Finland, 2012.

[14] J. Park, Y. Kim, and W. Jung, “A framework to estimate HEPs from the full-scope simulations of NPPs: unsafe act definition, identification, and quantification,” Korea Atomic Energy Research Institute (KAERI), Daejeon, Korea, KAERI/TR-6401/2016, Jun. 2016.
[15] Y. Kim, J. Park, and W. Jung, “A classification scheme of erroneous behaviors for human error probability estimations based on simulator data,” *Reliability Engineering & System Safety*, vol. 163, pp. 1–13, 2017.
[16] K. Zwirglmaier, D. Straub, and K. M. Groth, “Capturing cognitive causal paths in human reliability analysis with Bayesian network models,” *Reliability Engineering & System Safety*, vol. 158, pp. 117–129, Feb. 2017.
[17] A. M. Whaley *et al.*, “Building a Psychological Foundation for Human Reliability Analysis,” US Nuclear Regulatory Commission, Washington DC, NUREG-2114, Aug. 2012.
[18] A. M. Whaley *et al.*, “Cognitive Basis for Human Reliability Analysis,” US Nuclear Regulatory Commission, Washington DC, NUREG-2114, Jan. 2016.
[19] Y. H. J. Chang and A. Mosleh, “Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents. Part 4: IDAC causal model of operator problem-solving response,” *Reliability Engineering and System Safety*, vol. 92, no. 8, pp. 1061–1075, 2007.
[20] K. M. Groth and A. Mosleh, “Deriving causal Bayesian networks from human reliability analysis data: A methodology and example model,” *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 226, no. 4, pp. 361–379, Aug. 2012.
[21] K. M. Groth and L. P. Swiler, “Bridging the gap between HRA research and HRA practice: A Bayesian Network version of SPAR-H,” *Reliability Engineering and System Safety*, vol. 115, pp. 33–42, Jul. 2013.
[22] D. Gertman, H. Blackman, J. Marble, J. Byers, and C. Smith, “The SPAR-H Human Reliability Analysis Method,” US Nuclear Regulatory Commission, Washington DC, NUREG/CR-6883, 2005.
[23] J. Park and W. Jung, “A database for human performance under simulated emergencies of nuclear power plants,” *Nuclear Engineering and Technology*, vol. 37, no. 5, p. 491, 2005.
[24] J. Park and W. Jung, “OPERA–a human performance database under simulated emergencies of nuclear power plants,” *Reliability Engineering and System Safety*, vol. 92, no. 4, pp. 503–519, 2007.
[25] A. B. Skjerve and A. Bye, Eds., *Simulator-based Human Factors Studies Across 25 Years: The History of the Halden Man-Machine Laboratory*. London: Springer, 2010.
[26] S. Y. Choi and W. Jung, “Qualitative human event analysis with simulator data by using HuRAM+ and HERA,” in *Proceedings of the European Society for Reliability Annual Meeting (ESREL 2013)*, Amsterdam, 2013.
[27] B. Hallbert *et al.*, “Human Events Repository Analysis (HERA) System Overview,” US Nuclear Regulatory Commission, Washington DC, NUREG/CR-6903, Vol. 1, 2006.
[28] K. M. Groth, C. L. Smith, and L. P. Swiler, “A Bayesian method for using simulator data to enhance human error probabilities assigned by existing HRA methods,” *Reliability Engineering & System Safety*, vol. 128, pp. 32–40, 2014.
[29] A. D. Swain and H. E. Guttmann, “Handbook of Human Reliability Analysis with emphasis on nuclear power plant applications,” US Nuclear Regulatory Commission, Washington DC, NUREG/CR-1278, 1983.
[30] R. Boring *et al.*, “Human Unimodel for Nuclear Technology to Enhance Reliability (HUNTER): A Framework for Computation-based Human Reliability Analysis,” in *Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 13)*, Seoul, Korea, 2016.
[31] R. Boring, D. Mandelli, J. Joe, C. Smith, and K. Groth, “A Research Roadmap for Computation-Based Human Reliability Analysis,” Idaho National Laboratory, Idaho Falls, ID, INL/EXT-15-36051, Jul. 2015.
[32] K. M. Groth, “A Data-Informed Model of Performance Shaping Factors for Use in Human Reliability Analysis,” PhD Thesis, University of Maryland, College Park, MD, 2009.

[33] N. J. Ekanem and A. Mosleh, “Human failure event dependency modeling and quantification: A Bayesian network approach,” in *Proceedings of the European Society for Reliability Annual Meeting (ESREL 2013)*, Amsterdam, 2013.
[34] R. Boring *et al.*, “Integration of Human Reliability Analysis Models into the Simulation-Based Framework for the Risk-Informed Safety Margin Characterization Toolkit,” Idaho National Laboratory, Idaho Falls, ID, INL/EXT-16-39015, Jun. 2016.