SANDIA REPORT

SAND2012-1922

Unlimited Release

Printed April 2012

Processing Large Sensor Data Sets for Safeguards: The Knowledge Generation System

Maikael A. Thomas, Robert F. Matthews, Heidi A. Smartt

Prepared by

Sandia National Laboratories

Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from

U.S. Department of Energy

Office of Scientific and Technical Information

P.O. Box 62

Oak Ridge, TN 37831

Telephone: (865)576-8401

Facsimile: (865)576-5728

E-Mail: reports@adonis.osti.gov

Online ordering: http://www.doe.gov/bridge

Available to the public from

U.S. Department of Commerce

National Technical Information Service

5285 Port Royal Rd

Springfield, VA 22161

Telephone: (800)553-6847

Facsimile: (703)605-6900

E-Mail: orders@ntis.fedworld.gov

Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2012-1922
Unlimited Release
Printed April 2012

Processing Large Sensor Data Sets for Safeguards: The Knowledge Generation System

Maikel A. Thomas, Robert F. Matthews, Heidi A. Smartt
International Safeguards and Technical Systems Department
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87158-1371

ABSTRACT

Modern nuclear facilities, such as reprocessing plants, present inspectors with significant challenges due in part to the sheer amount of equipment that must be safeguarded. The Sandia-developed and patented Knowledge Generation system was designed to automatically analyze large amounts of safeguards data to identify anomalous events of interest by comparing sensor readings with those expected from a process of interest and operator declarations. This paper describes a demonstration of the Knowledge Generation system using simulated accountability tank sensor data to represent part of a reprocessing plant. The demonstration indicated that Knowledge Generation has the potential to address several problems critical to the future of safeguards. It could be extended to facilitate remote inspections and trigger random inspections. Knowledge Generation could analyze data to establish trust hierarchies, to facilitate safeguards use of operator-owned sensors.

This page intentionally left blank

Processing Large Sensor Data Sets for Safeguards:

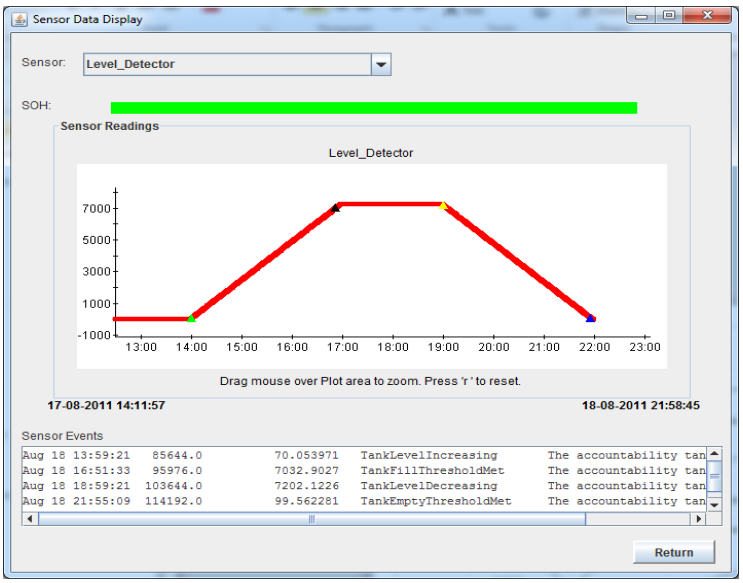
The Knowledge Generation System

Executive Summary

Modern nuclear facilities, such as reprocessing plants, present inspectors with significant challenges due in part to the sheer amount of equipment that must be safeguarded. Near-real-time accountancy (NRTA) approaches only compound this complexity due to the number of measurements and accountancy calculations performed. With continually growing amounts of sensor data, it is becoming increasingly difficult to analyze these large datasets to extract meaningful patterns and higher order systemic trends to ensure timely detection of diversions.

The Sandia-developed and patented Knowledge Generation system was designed to automatically analyze large amounts of safeguards data to identify anomalous events of interest by comparing sensor readings with those expected from a process of interest and operator declarations. The user defines significant thresholds and events with respect to each sensor, and the process is modeled as a state machine, with sensor events governing how and when states change. As a result, Knowledge Generation automatically compares the operator's declarations against the modeled process and sensor data, and reports out-of-order events, events missing from the declaration, or declared events that did not occur.

Recently the Knowledge Generation codebase was reassembled and demonstrated using simulated input accountability tank sensor data representing part of a reprocessing plant from the Separations and Safeguards Performance Model. The figure below is a screenshot of the tank level in a typical fill-mix-empty cycle. Knowledge Generation was able to successfully identify anomalous events (not conforming to the process as designed) and discrepancies between actual operations and operator declarations.



An important next step in development of Knowledge Generation as a safeguards tool is to model and process additional sensors from a large reprocessing plant. The Separations and Safeguards Performance Model would be employed to assist in developing the state machine model and to provide simulated sensor data.

Knowledge Generation has the potential to address several problems critical to the future of safeguards. It could be extended to be a near-real-time tool to facilitate remote inspections and, via remote monitoring, flag anomalies that could trigger random inspections. Further, Knowledge Generation could analyze data to establish trust hierarchies, in which the outputs of authenticated sensors could be used to authenticate data from operator-owned sensors. For example, authenticated accountability tank sensor data could be analyzed for patterns to establish a level of trust of the operator sensors in the surge tank.

Introduction

Safeguards Challenges in Reprocessing

Modern reprocessing plants present challenges to safeguards because of their size and complexity. For example, an early consideration of the requirements for safeguards at the Rokkasho Reprocessing Plant (RRP) in Japan estimated 45 full-time inspectors would be needed to implement safeguards according to IAEA criteria (OTA 1995, p. 126). A similar number would be required at THORP in the UK, and twice that number at La Hague, France. Given that there are a total of 200 full-time personnel in the IAEA inspectorate, traditional approaches to inspection are not sustainable.

A second challenge is presented by the large material inventories contained in these plants. At any time, the chemical processing area at RRP could contain up to 800 kg of Pu (Ehinger 2004). Accurate accountancy is crucial, because diversion of even a small fraction of the inventory could constitute a significant quantity (e.g. 8 kg for Pu).

The combination of the need for accurate accountancy and the limited availability of human inspectors has led to a search for unconventional safeguards approaches. To improve the timeliness of accountancy information, near-real-time accountancy (NRTA) techniques collect authenticated, safeguards sensor data and compute material balances on a daily or weekly basis, as opposed to current timeliness goals of one to three months, depending on the form of the nuclear materials present. Statistical methods applied to the larger inventory data sets generated by NRTA can improve the sensitivity of accountancy to small diversions (OTA 1999, p115).

In addition to NRTA, use of the plants' own process monitoring data has been proposed for safeguards purposes. Such data as flow rates, temperatures, etc. taken at a number of points in the process could provide enhanced visibility of the process to locate diversions more closely and allow more timely intervention. As the Office of Technology Assessments reported, "[s]afeguards experts point out that when various statistical tests are applied to a sequence of process control data and to measurements taken at various points in the plant, and these measurements are combined with a thorough understanding of the plant's designed operating conditions, sensitivity to diversion detection improves over the case in which only annual material balance measurements are used." (OTA 1999, p121)

Both NRTA and process monitoring have the side effect of multiplying the amount of data collected by the safeguards system. This is a problem because a sensor reading that may be of significance for safeguards is likely swamped in the massive amounts of other sensor readings

collected by the system. In addition, use of process monitoring in conjunction with NRTA may contribute to build-up of error over the multiple sensors, which could cause the total error to exceed the significant quantity (Cipiti 2011). Alternative, automated methods for processing the sensor data are needed to handle and extract knowledge from large data sets.

Objectives of this Study

This study was proposed to investigate the applicability of the Sandia-developed Knowledge Generation system to the problems of analyzing large safeguards data sets. Four deliverables were proposed:

1. Description of the principles of operation of the baseline Knowledge Generation system
2. Discussion of issues affecting application of the Knowledge Generation system to continuous processes, such as those encountered in fuel reprocessing plants
3. Design goals for an upgraded Knowledge Generation system with applications to continuous processes
4. Preliminary discussion of a developmental testing program for the upgraded Knowledge Generation system.

1. Knowledge Generation Principles of Operation

Knowledge Generation (Brabson 1999) is a process data analysis system originally developed for safeguards applications. Its inputs include models of the process and the sensors that monitor it, sensor data from process operations, and declarations of activities from the process operator. The output of the Knowledge Generation system includes a list of events, the times at which they occurred, and discrepancies, if any, between -the process's actual operation and the declarations. These discrepancies could be indicative of safeguards violations, such as diversions of safeguarded material; deficiencies in the operator's process model, possibly due to incorrect design information; or malfunctions in process equipment. Figure 1 illustrates the Knowledge Generation concept.

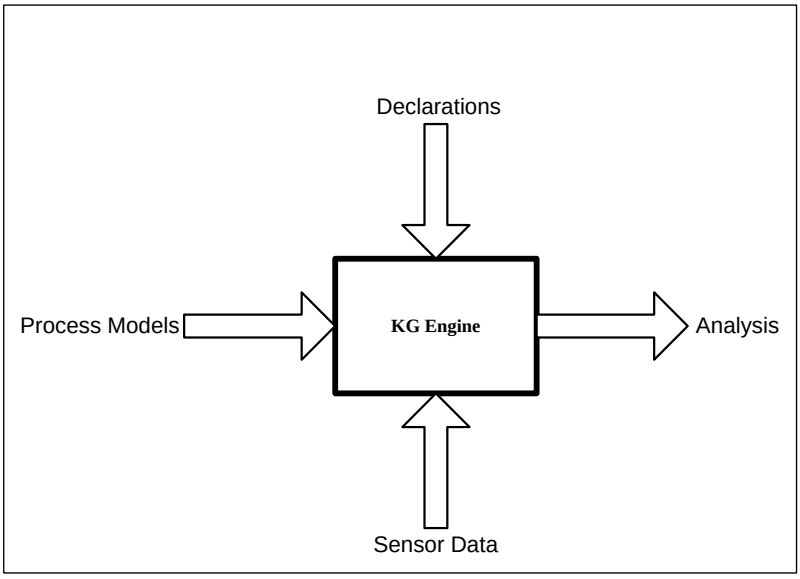


Figure 1. Knowledge Generation Concept

The Knowledge Generation system has two major components. The Event Generator accepts sensor data from the process and recognizes significant sensor events as they occur. It takes as input the possible sensor events and the raw sensor data associated with them. As a simple example, if the door sensor associated with the entrance door to a facility changes from closed to open, the Event Generator issues a door-open event. The output is a list of all transitions recognized in the sensor data along with the associated event timestamp.

Lists of sensor data, transition rules, and operator declarations are the input to the Analysis Engine sub-system, which models a process as a state machine. The state-machine model input to Knowledge Generation describes a process as a series of states with transition rules that govern how the state changes when events occur. The Analysis Engine automatically compares the operator's declarations against the actual operation of the process, recognizing out-of-order events, events missing from the declaration, or declared events that did not occur. The results are presented to the inspector or reviewer for further investigation.

The inherent ability to compare actual operations to the operator's declarations makes Knowledge Generation a potentially valuable tool in safeguards applications. Demonstrations of Knowledge Generation previous to this project, however, were on processes with relatively simple relationships between events and process state. For example, in one application, Knowledge Generation was used to monitor the movement of fuel assemblies in the fresh fuel and spent fuel areas at a reactor facility (Damico 2002). Multiple motion sensors in the facility had discrete outputs: "on" when they detect movement, "off" when no motion is taking place. By monitoring the relative order and timing of the sensor signals, Knowledge Generation could deduce the direction of motion.

Applying Knowledge Generation to a reprocessing plant would require a more sophisticated ability to recognize events. For example, to recognize that a tank is filling, Knowledge

Generation requires the ability to handle continuous inputs (the fluid level in the tank) and to compare successive level measurements over time to determine whether and how fast the tank is filling.

Applying Knowledge Generation to reprocessing plants safeguards is facilitated by the system's ability to compare actual facility operation against the operator's declarations. To do this, the Knowledge Generation system requires the following inputs:

- A high-fidelity, state-machine model of the plant's processes. Building such a model may be a daunting task for a large and complex system like a modern reprocessing plant[1], but it would be a product of the design information verification (DIV) effort before plant commissioning. Like the declared design information, the model would have to be updated as the plant and its processes evolve over time.
- Sensor data. The process data inputs to the model could come from both authenticated sensors installed as part of the inspection regime and from the operator's process monitoring system. The most likely source for sensor data would be the raw sensor database.
- Operator declarations. The process of converting declarations to a form Knowledge Generation can use is fairly straightforward. Ease of use was a design principle for the Knowledge Generation human interface.

Given this information, Knowledge Generation produces a continuous display of the status of the process along with notifications of departures from the declared process.

2. Issues Affecting Application of Knowledge Generation to Continuous Processes

The Knowledge Generation system was demonstrated initially at Sandia's Integration, Test, and Evaluation Laboratory (ITEL) in a material handling process (Brabson 2000). The demonstration simulated retrieval of material from storage, loading into a vehicle, transporting the material, and unloading it at its destination. Monitoring the process involved tracking items by location and status (moving, in storage, etc.). Sensors monitoring the process included door-open sensors, break beams, and motion detectors. All had discrete outputs (door open/door closed, motion detected/not detected, etc.).

Applying Knowledge Generation to reprocessing plants and processes requires monitoring processes where the material to be safeguarded is not in countable items but in solution. In reprocessing, material monitoring relies on flow rate, pressure, density, and other continuous variables.

To explore the applicability of the existing Knowledge Generation system to reprocessing processes a demonstration was conducted using simulated reprocessing sensor data. The two objectives for the demonstration were

[1] In Knowledge Generation, the state machine model is input as a Java program. Thus, the analyst creating the model must be or have access to a programmer who can develop the program according to Knowledge Generation specifications.

- Assemble the Knowledge Generation code, and recompile it for current hardware and architecture.
  The code was developed more than 10 years ago and was intended to run on a different hardware and network configuration than is available for this demonstration. In 2002, the system was rewritten in a current, portable, object-oriented language (Java) and development tools were available for reconstituting the system. Minor modifications were required to convert the system from its original client-server configuration to allow it to run as a stand-alone application.
- Demonstrate the use of Knowledge Generation with continuous sensor data as would be prevalent in a reprocessing plant application.
  While the original Knowledge Generation system was developed with the ability to accept continuous (analog) sensor data, its early demonstrations used discrete sensors. The demonstration was designed to use the analog input capability to monitor a reprocessing scenario. Since actual sensor data from a reprocessing plant was not available, the data used in the demonstration was provided by the Separations and Safeguards Performance Model (Cipiti 2010).

Demonstration Scenario

Due to the size and complexity of a reprocessing plant we opted to model an accountability tank. The accountability tank is a critical step in the reprocessing pipeline and generally has authenticated sensors for safeguards purposes. The accountability tank is located at the output of the front end of a reprocessing plant, where the spent fuel assemblies are chopped and dissolved and the hulls are removed. Clarified solution from the dissolver flows into the accountability tank until it is full. When the tank is full, the input is shut off and the solution is mixed to ensure uniform composition. After two hours of mixing, inspectors take samples from the tank for destructive analysis to determine the fissile material content of the solution. Then the tank is emptied and the solution sent on to the separation process for extraction of uranium and plutonium.

The accountability tank state-machine model is shown in Figure 2. In the diagram, the circles represent all of the possible states of the accountability tank—Empty, Fill, Drain, and Mix. The arcs between the states show the possible transitions between them. The transitions are labeled with the sensor readings corresponding to them. For example, the transition between Empty and Fill is labeled “Level/time,” meaning that a certain change of the level of solution in the tank over time is associated with the transition between the Empty and Fill states.

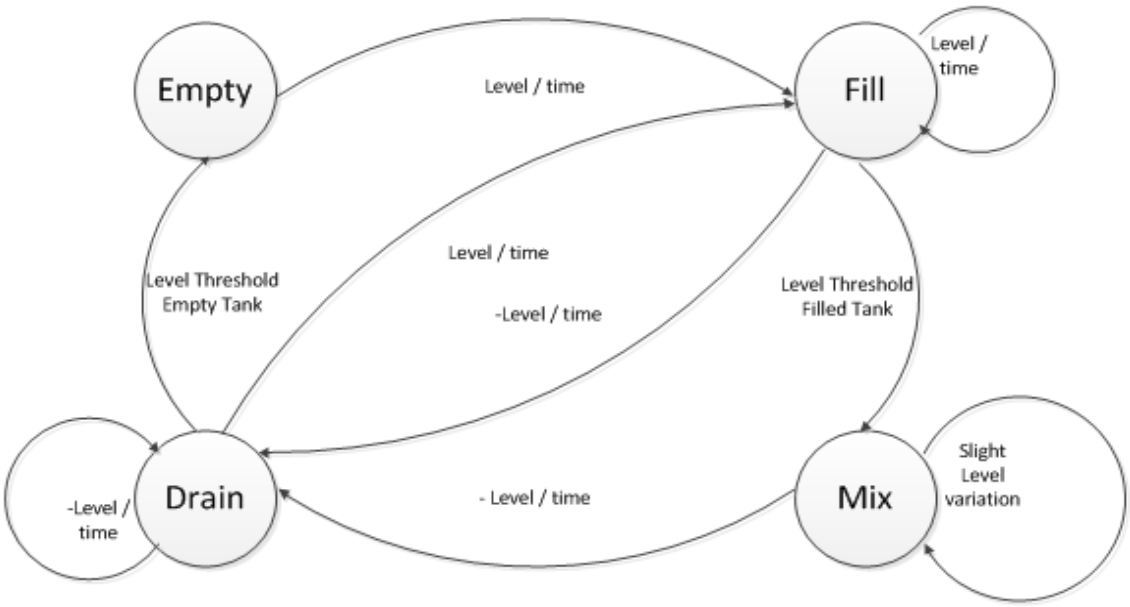


Figure 2. Accountability Tank State Machine Model

Set-Up

The first step in setting up Knowledge Generation to analyze a process is to define the sensors in the process using the Event Generator user interface tool shown in Figure 3. The dialog asks for the sensor type (discrete or analog), the possible sensor states, and rules for transition to the next process state. In the figure, the sensor is defined as an analog tank level sensor for the accountability tank. The transition rule states that if while the tank is filling the solution level rises above the defined threshold, the event TankFillThresholdMet has occurred. In this case, the next state is Mix (Figure 2). Note that a sensor can be associated with multiple states as long as each sensor and sensor rule is associated with a single state transition.

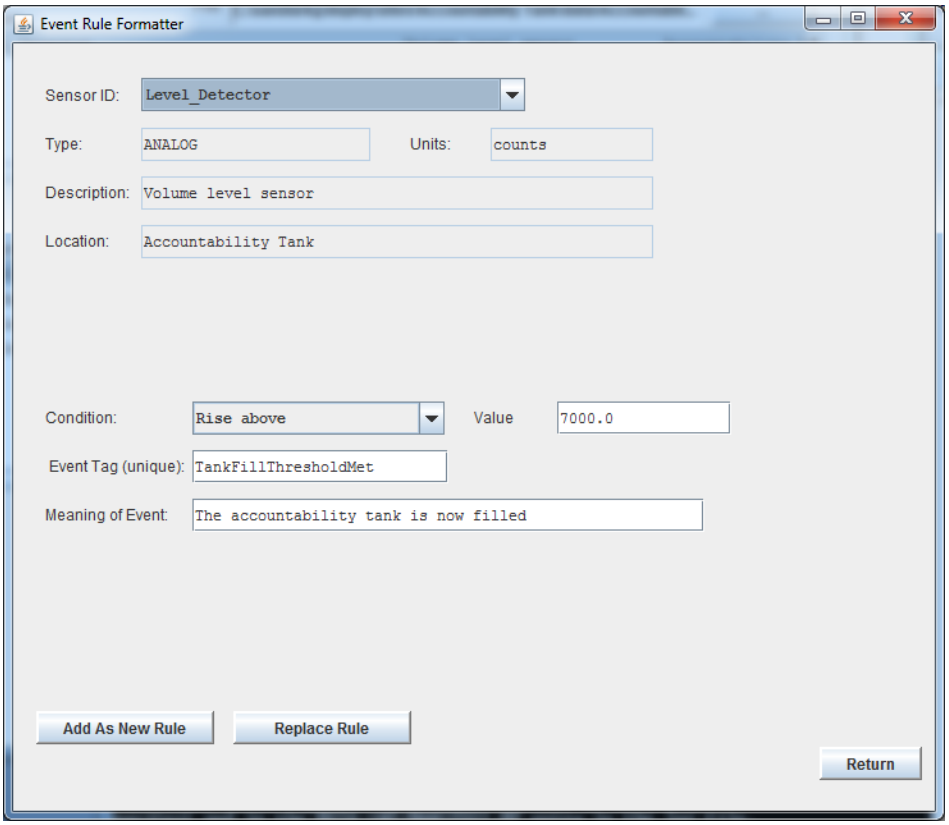


Figure 3. Sensor Definition

Running the Model

Once all the sensors and rules have been entered, raw sensor data is entered from a file, and the sensor rules are applied to identify events. Figure 4 shows how events are identified in the sensor data, denoted by color-coded triangles on the graph superimposed on top of the raw sensor data. The text box below the graph shows the events that were identified, with timestamp and sensor value that triggered the event, in a complete Fill-Mix-Drain-Empty cycle of the accountability tank.

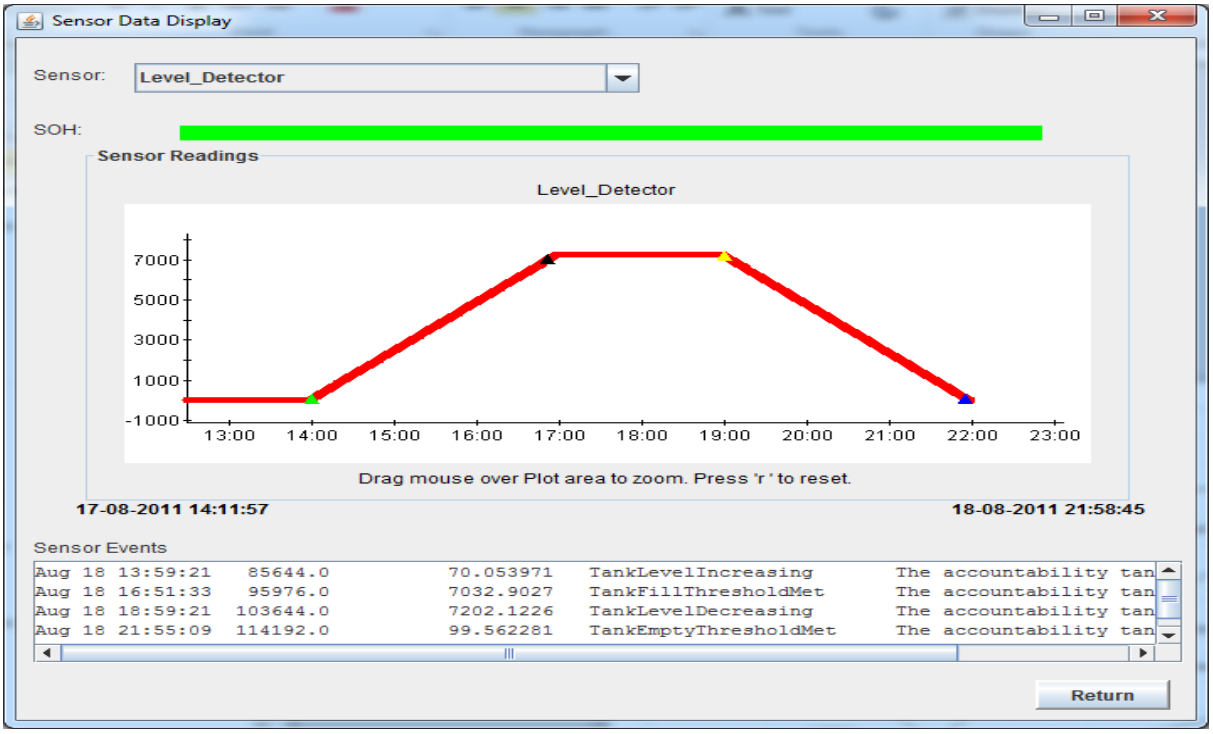


Figure 4. Accountability Tank Events

Knowledge Generation issues an error message if a sensor input does not match the conditions of the state with which it is associated. For example, in Figure 5 the sensor shows a decrease in tank level during the Fill state. Because such an anomaly could indicate an attempt at diversion, Knowledge Generation marks it with an error message.

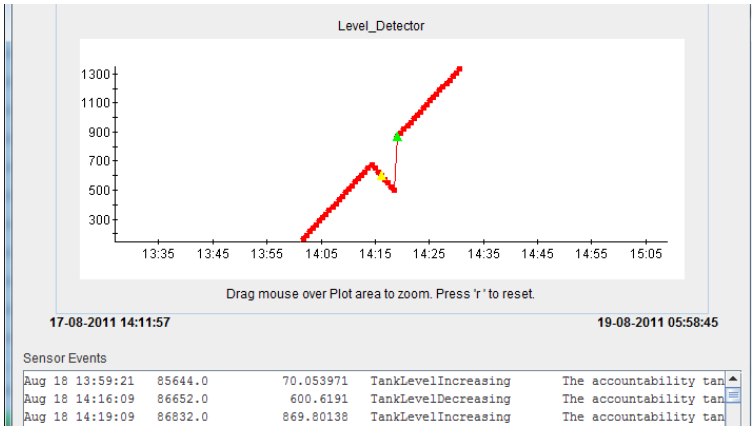


Figure 5. Unexpected Sensor Reading During Event

The operator's declarations consist of a list of events the operator declares to have happened. While Knowledge Generation runs the process, it compares the events identified in the sensor inputs against the declarations. It displays a report like that in Figure 6 with the results of the comparison.

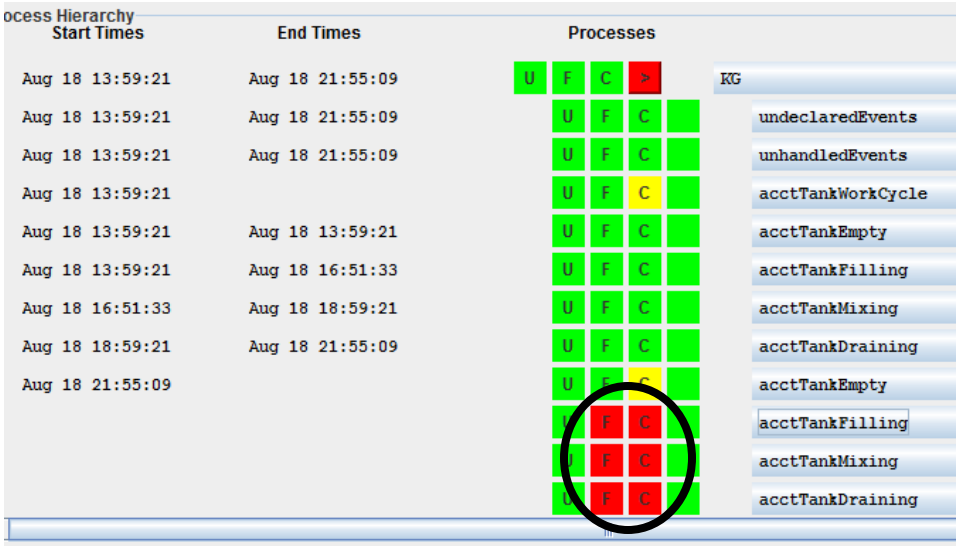


Figure 6. Events List Showing Missing Events

In Figure 6, Knowledge Generation reports that the sensor data shows two successive tank filling-mixing-draining cycles. The operator's declaration included only one cycle, so it marks the events in the second cycle in red. Clicking on one of the events displays a message to the effect that the event occurred but was not declared. Knowledge Generation also identifies events that were declared but did not occur.

Demonstration Conclusions

The objective of the demonstration was to answer two main questions:

1. Can Knowledge Generation software be migrated successfully to modern computing hardware and system architecture?
2. Can Knowledge Generation handle and produce meaningful results from data associated with the current generation of reprocessing plants?

Both questions were answered in the affirmative, at least to the level of detail we were able to explore in this study. The Knowledge Generation codebase was usable in a modern computing system with some modifications to accommodate differing system architectures between when the system was developed 10 years ago and the present day.

Knowledge Generation is able to accept data from continuous-variable sensors like level gauges and flow meters. It is capable of using such data to identify events that are meaningful in terms of the processes, and it can recognize anomalies when the sensor outputs do not match those expected for the process. It is also able to recognize discrepancies between operator declarations and the process as it was actually executed—either events that occurred but were not declared or declared events that did not occur.

3. Applications and design goals for an upgraded Knowledge Generation system

A potentially valuable application of Knowledge Generation could be as a near-real-time tool that could facilitate remote inspections or flag anomalies that could trigger random inspections. Remote inspection is envisioned as a way to reduce the resources required to ensure compliance with non-proliferation obligations. Knowledge Generation could aid inspectors and local and regional authorities in handling the masses of data gathered by both authenticated sensors and operator's process control system sensors in support of remote inspection. (Zendal 2010).

Knowledge Generation could help facilitate the use of process monitoring sensors in safeguards applications by establishing trust hierarchies (Damico 2011). In such hierarchies, the outputs of authenticated sensors could be compared to readings of other sensors in the system to which they are related. For example, an authenticated level sensor in the accountability tank could be used to authenticate signals from the sensor in the surge tank. Efforts at tampering would appear to Knowledge Generation as discrepancies and be reported as such.

The primary design goal for a Knowledge Generation system is ensuring its capacity to handle large data sets and complex processes. The accountability tank process used in this study to exercise Knowledge Generation's capabilities is not representative of an entire reprocessing plant. Therefore, future work should include efforts to ensure that Knowledge Generation can handle the amount and types of sensor data from an entire plant's operation.

Some consideration should also be given to evaluating and revising the user interface, given the amount of data that is required to set up the state-machine model and enter the declarations. The current Knowledge Generation system requires the state-machine model to be entered in text form, using a programming language. While this method is flexible and powerful, it can be burdensome, especially for large models. Future development of Knowledge Generation should include alternate entry methods, including graphical techniques like that used to develop the Separations and Safeguards Performance Model.

4. Preliminary discussion of future testing

Testing to verify the capability of Knowledge Generation to process large data sets could employ models like the Separations and Safeguards Performance Model. The model simulates operations of a large reprocessing plant and its sensors. It allows accumulation of data simulating months of plant operations in a short time, allowing Knowledge Generation to be exercised on very large data sets. The model can simulate diversions of material at a variety of points in the process, allowing the sensitivity of Knowledge Generation to diversions to be tested. It can also simulate tampering with sensors, to test the ability of the trust hierarchy concept to detect such tampering. Using the model, it should be possible to quickly verify the capacity of Knowledge Generation to handle sensor data from large, modern reprocessing plants.

References

Brabson, J. 1999. Knowledge Generation system requirements document.

Brabson, John and Sharon M. DeLand. 2000. *Knowledge Generation*, SAND2000-2728C. Albuquerque, NM: Sandia National Laboratories.

Cipiti, Benjamin and Owen Zinaman. 2010. *Separations and safeguards model integration*, SAND2010-5962. Albuquerque, NM: Sandia National Laboratories.

Cipiti, Benjamin. 2011. Separations and safeguards performance modeling for advanced reprocessing facility design, *Journal of Nuclear Materials Management*, XXXIX, no. 2 (Winter).

Damico, Joe. 2002. Final report for JNC international fellow Joe Damico.

Damico, J., et al. 2011. Information assurance for international safeguards process monitoring applications, *Proceedings of 52nd INMM Annual Meeting*. Deerfield, IL: Institute of Nuclear Materials Management.

Ehinger, M., et al. 2004. Solution monitoring applications for the Rokkasho Reprocessing Plant, *Proceedings of the 7th International Conference on Facilities Operations-Safeguards Interface*. La Grange Park, Il: American Nuclear Society, Inc.

U.S. Congress, Office of Technology Assessment (OTA). 1995. *Nuclear Safeguards and the International Atomic Energy Agency*, OTA-ISS-615. Washington, DC: U.S. Government Printing Office.

Zendal, M., et al. 2010. Enhancing and optimizing safeguards implementation by remote safeguards inspections. IAEA-CN-184/266. Vienna: International Atomic Energy Agency.

Distribution

1	MS0889	Technical Library, 9536 (electronic copy)
1	MS1371	Dianna S. Blair, 6832 (electronic copy)

Sandia National Laboratories