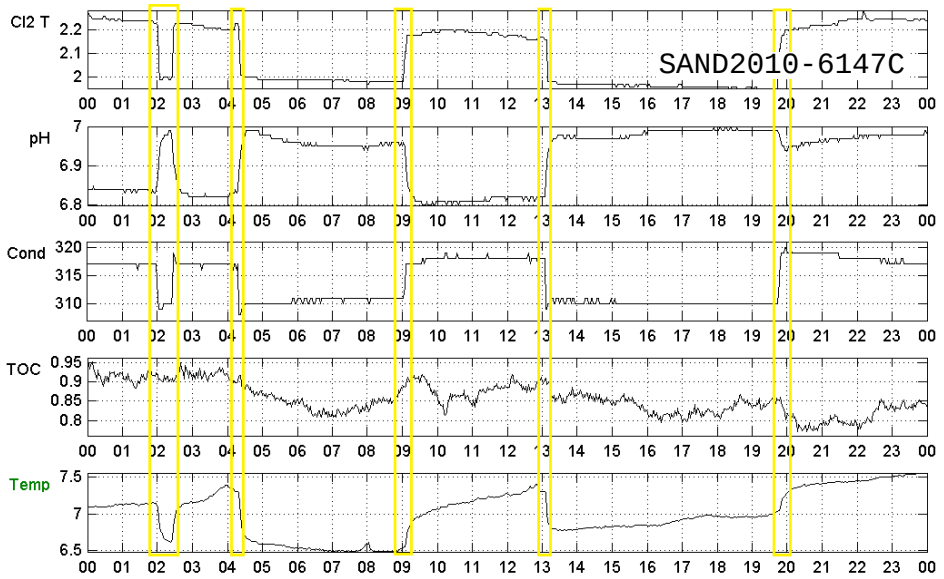


Combining Water Quality & Operational Data for Improved Event Detection

WDSA 2010, Tucson AZ

September 13, 2010

[1] David B. Hart, [1] Sean A. McKenna, [2] Terra Haxton, [2] Regan Murray

[1] Sandia National Laboratories, Albuquerque NM

[2] National Homeland Security Research Center, US EPA, Cincinnati OH

CANARY was developed through an InterAgency Agreement between the U. S. Environmental Protection Agency and Sandia National Laboratories.

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

A Department of Energy National Laboratory

Introduction

- False alarms have a significant negative effect on the usefulness of security systems:
  - Investigating false alarms takes time and resources
  - Too many false alarms decreases human sensitivity to real alarms – the "boy who cried wolf" syndrome
- Changing algorithm parameters, such as event thresholds and levels, reduces false alarms, but at a cost to sensitivity
- How can we decrease the false alarm rate without losing sensitivity?
  - Start using operational data to eliminate false alarms within the algorithms themselves

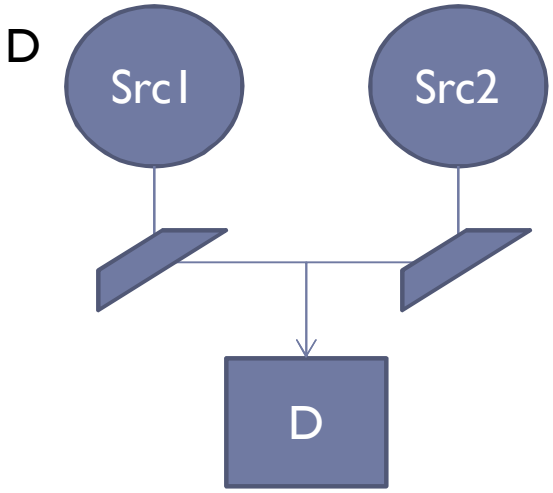
Outline

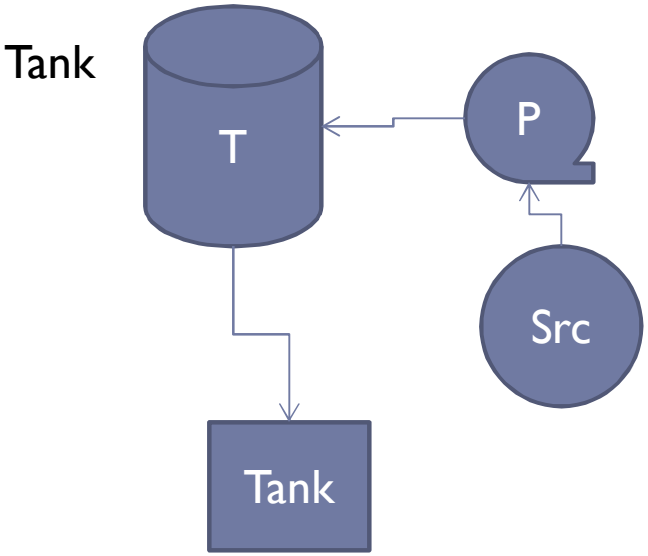
- Examples of Problem Data
- Method 1: Direct Operations Signal Use
- Method 2: Indirect Operations Signal Use
- Method 3: Dynamic Operations Signal Use
- Conclusions and Results

Sandia National Laboratories

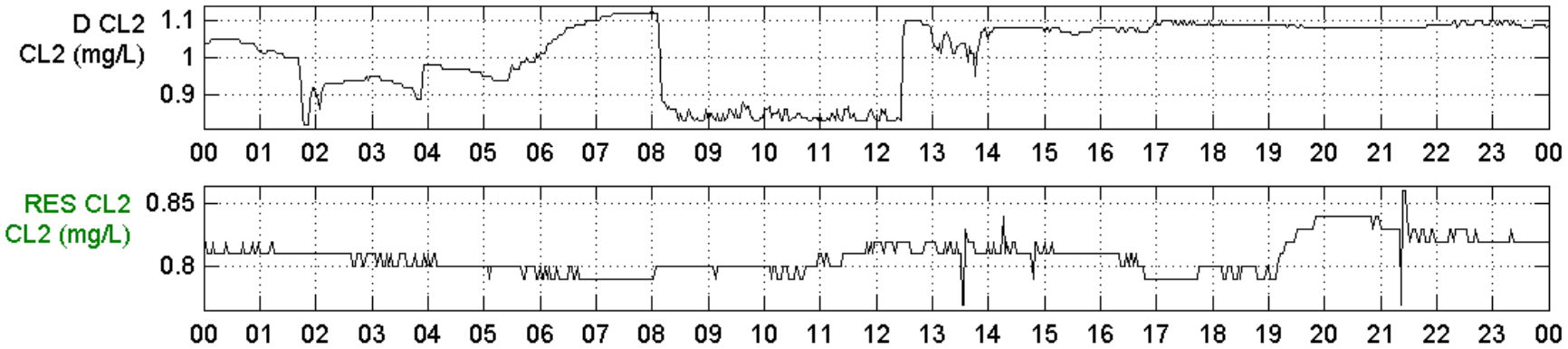
Problem Data Examples

- Two Systems – "D" and "TANK"
  - Data is from "Battle of Event Detection Systems"[1]
- "D" serves as an example of in-system variation
- "Tank" serves as an example of at-source variation



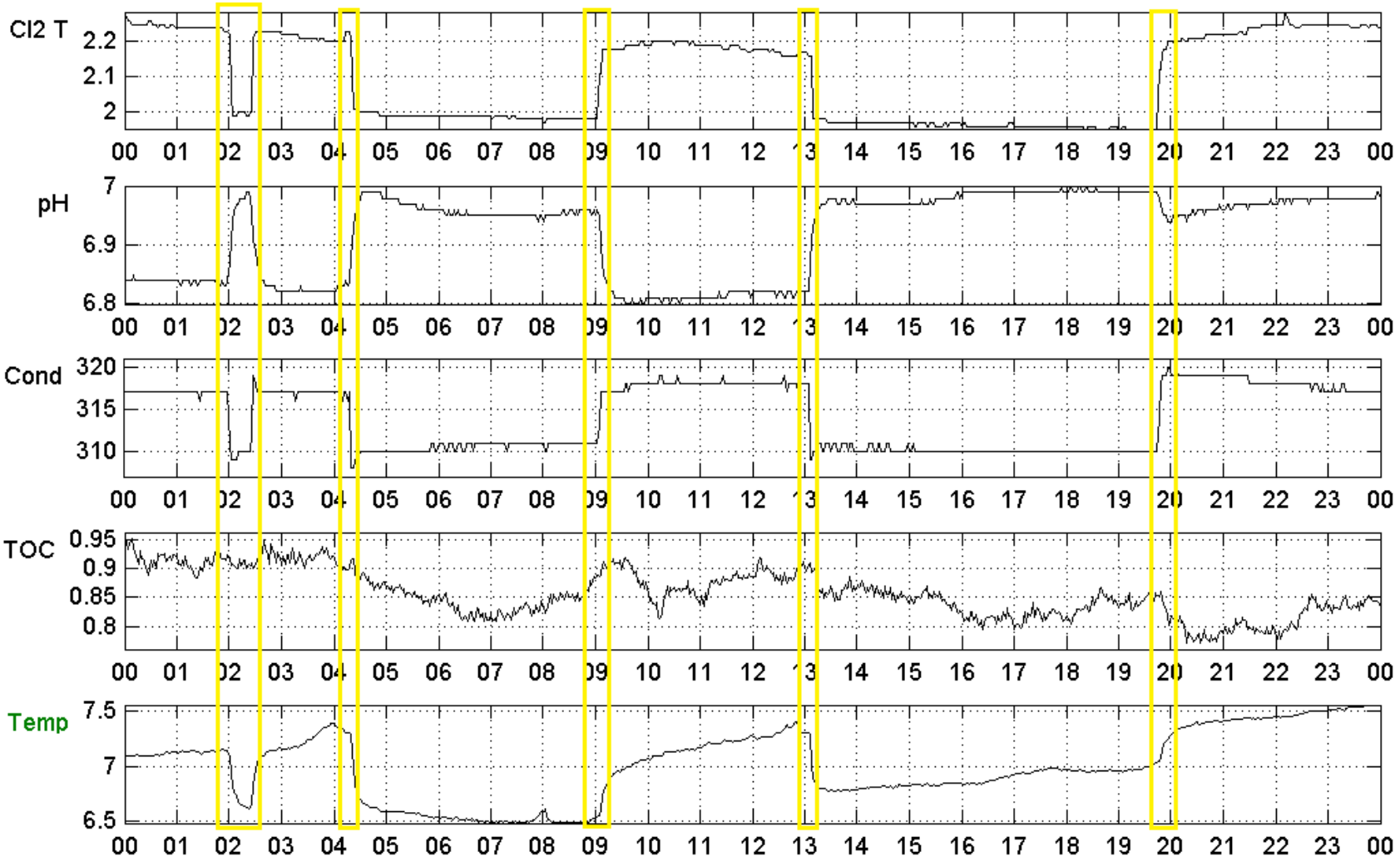


Problem Data Examples: “D”



Sandia National Laboratories

Problem Data Examples: “Tank”



Direct Operations Signal Use

- This method uses a direct signal, such as a pump status, to "recalibrate" the algorithm. In CANARY, this means that alarms at a station are turned off for a few time steps after the operations data changes.
- Advantages:
  - Extremely simple to add to any algorithm – serves the essentially the same function as a "calibration" signal
- Disadvantages:
  - Alarms are disabled for a short period of time
  - Possible to spoof such a signal if SCADA system is compromised
  - Requires a very good system model to calculate lag times unless station and operations equipment are co-located

Sandia National Laboratories

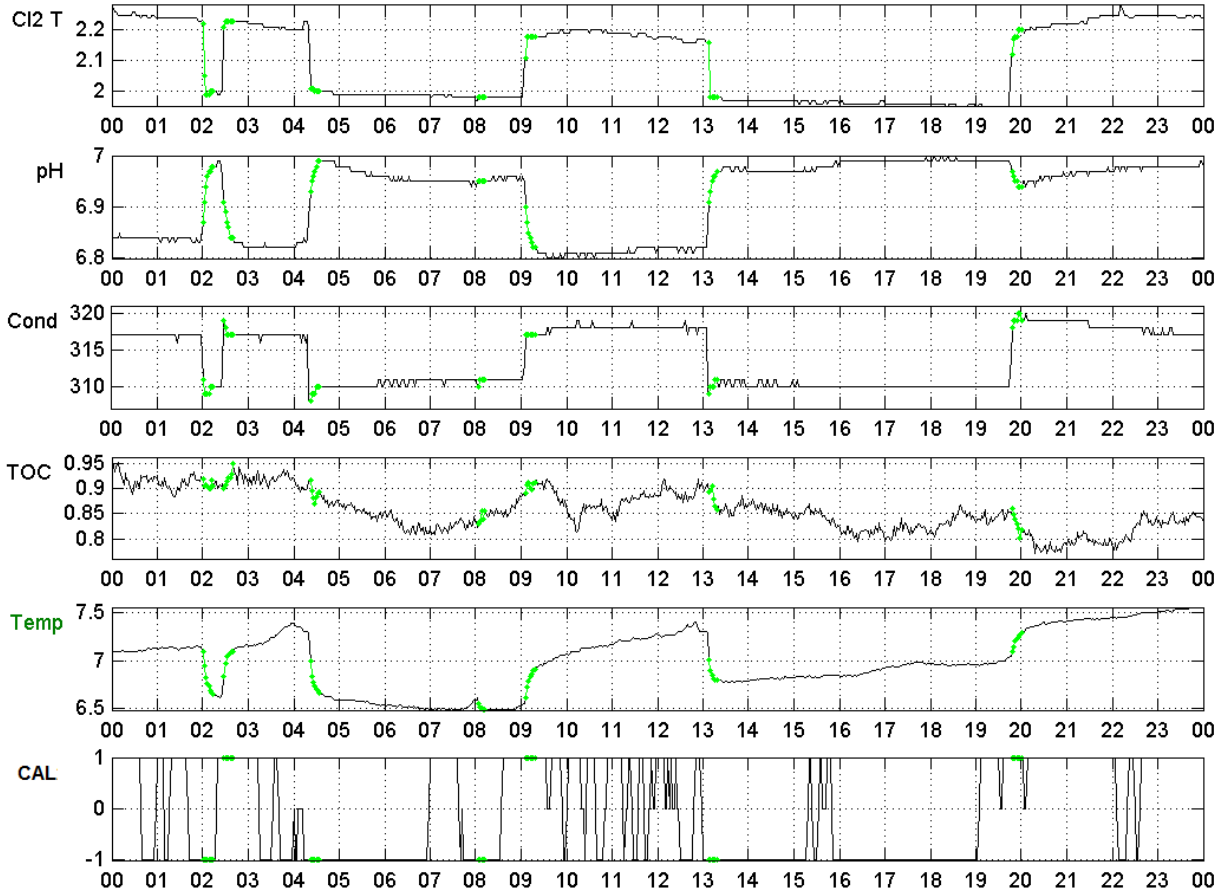
Indirect Operations Signal Use

- Same basic "recalibration" method as direct use, but does not require discrete signals
  - Transform language added to CANARY allows for derivatives, trends, and changes in non-discrete signals (such as pressure or tank levels) to be calculated on the fly as a "composite signal"
  - Transform language can be applied to any SCADA signal or previously created composite signal
- Advantages:
  - Much harder to "spoof" a derived signal's data
  - Can use looser lag times combined with some other trigger (such as temperature)
  - Possibilities of combined signals to analyze are endless
- Disadvantages:
  - Ideally, composite signals would be defined/calculated within the SCADA or data management system, not in the event detection software
  - Possibilities are endless (and therefore require time and system knowledge to create the most useful composite signals)

Sandia National Laboratories

Indirect Operations Signal Use

An example of composite signals using the "Tank" location



$$dir[k] = \sum_{i=1}^{10} T[k] - T[k-i]$$

$$S[k] = \begin{cases} 1 & dir[k] > 0 \\ 0 & dir[k] = 0 \\ -1 & dir[k] < 0 \end{cases}$$

$$CAL[k] \equiv S[k] \neq S[k-2]$$

$$CAL2[k] \equiv CAL[k] \cap (|Cond[k] - Cond[k-2]| > 5)$$

Green highlights are places where the full CAL2[k] composite signal evaluates to "TRUE"

Dynamic Operations Signal Use

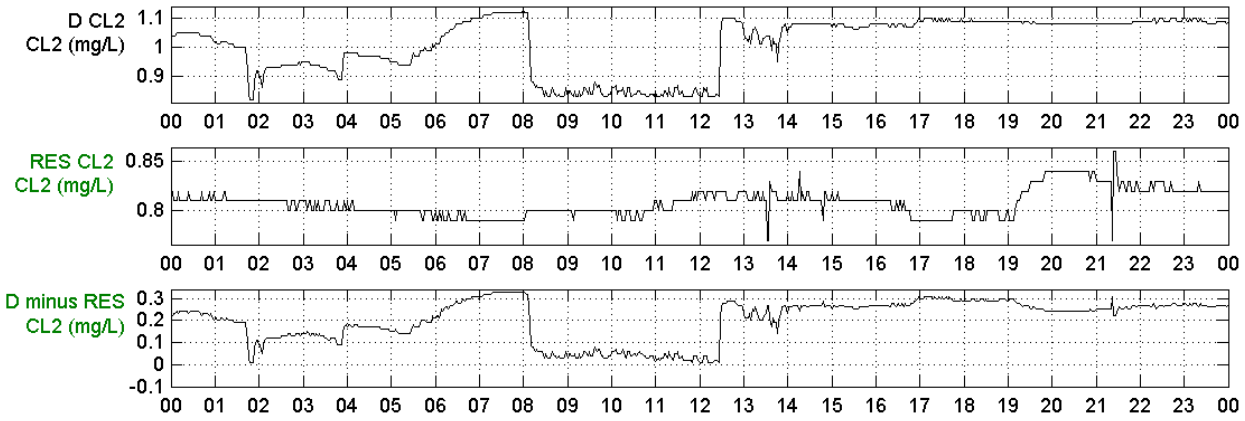
- This method uses the composite signals transform language, but applies it differently than the previous example
- For an in-network monitoring station, changes are as likely to be due to changing demand patterns and flow directions as from operational changes
- Rather than try to model these flow changes, the composite signals are used to create a dynamic set of set-points, based on both the upstream sources
- Each upstream source, combined with an estimate of decay or change in the given parameters, is used to create a set-point range

Sandia National Laboratories

Dynamic Operations Signal Use

- Once the dynamic set point ranges are created, as long as the monitoring point data is within one of these ranges, crated from current upstream monitoring data, then the change from one band to another is ignored

$$Cl_{Dyn}[k] = Cl_{D}[k] - Cl_{RES}[k]$$



Sandia National Laboratories

Dynamic Operations Signal Use

- Advantages:
  - Incredibly flexible method to use set-point methods with variable set-point bounds
  - Uses up-stream data for comparison, but without the need for super-accurate flow lags or as detailed system analysis as previous methods
- Disadvantages:
  - Requires method to be part of consideration in where to place sensor stations
    - Talking with source of "D" data led us to realize the "D_RES" was slightly downstream, rather than upstream, of "D" – while results are promising, a better, more complete data set is really necessary for good analysis.

Sandia National Laboratories

Results on the "TANK" site

- The previous methods were applied to the "TANK" data set to find the reduction in false alarm rates
- The composite calibration signal reduced the false alarms at TANK from 1 alarm per day to 2 per week over a 4-month data set
- Four different simulated event strengths were then added to then check sensitivity. The results are tabulated below.

$$Y_i[k] = X_i[k] + pf_i\sigma_i$$

p value for Equation	0.25	0.5	1.0	1.5
TANK Sharp Events	=14/25	=22/25	=23/25	=24/25
TANK Smooth Events	=8/25	=13/25	=15/25	=15/25

Conclusions

- We believe that incorporating operations data into event detection systems and algorithms is essential to reducing false alarms without adversely effecting sensitivity to real events
- The methods presented show promise in helping reduce false alarms, but all require some tradeoffs or planning in the placement of sensor stations
- Additional methods, such as pattern matching and clustering may help reduce false alarms even more, but will likely require operations data to help distinguish between "good" and "bad" patterns

Sandia National Laboratories

Acknowledgements

- The authors would like to thank Katie Umberg, from the US EPA Office of Water, and Srinivas Panguluri from the Shaw Group, Inc., for their aid in providing data and algorithm testing opportunities, as well as the many software testers who have helped us debug and develop algorithms.

- The U.S. Environmental Protection Agency through its Office of Research and Development funded and collaborated in the research described here under an Interagency Agreement with the Department of Energy's Sandia National Laboratories. It has been reviewed by the Agency and approved for publication but does not necessarily reflect the Agency's views. No official endorsement should be inferred. EPA does not endorse the purchase or sale of any commercial products or services.

Sandia National Laboratories