OPERATIONAL AND LABORATORY VERIFICATION TESTING OF A HEURISTIC ON-LINE WATER MONITORING SYSTEM FOR SECURITY

DAN KROLL

HACH HOMELAND SECURITY TECHNOLOGIES, 5600 Lindbergh Drive
LOVELAND, COLORADO 80539 USA
DKROLL@hach.com

KARL KING

HACH HOMELAND SECURITY TECHNOLOGIES, 5600 Lindbergh Drive
LOVELAND, COLORADO 80539 USA
KKING@hach.com

Drinking water systems are vulnerable to attack. While homeland security initiatives commonly focus on aerial CW/BW attacks, they have tended to ignore the far more inexpensive and easy to orchestrate attacks on drinking water distribution networks. An approach that utilizes off-the-shelf broad-spectrum analytical instruments coupled with advanced interpretive algorithms to provide detection-response networks for water is described. This report summarizes the development of interpretive algorithms applied to drinking water instrumentation and the implications for immediately engineering and deploying distribution based detection-protection systems. Data obtained from in house testing along with testing at ECBC was used to produce fingerprint response data from traditional as well as non-military threat agents. Data obtained from a Battelle/EPA ETV study addresses issues such as long-term deployment and ability to detect and characterize contaminants. Loop testing at ECBC and Battelle demonstrates extrapolation of beaker testing fingerprints to flowing systems. Real world deployment data is used to demonstrate recognition and classification of actual events and heuristic capabilities of the system along with its potential role in enhancing water quality above and beyond its obvious security aspects. The system is shown to be a practical measure to detect and characterize backflow events involving both chemicals and bioagents.

Keywords: Water Distribution; Monitoring; On-Line; Algorithm, Multi-parameter

1. Our Water Systems are Vulnerable

The water that we use in our homes and businesses is supplied to us through a vast network of interconnected pipes known as the distribution system. The vulnerability of this system to attack is not a subject that most people dwell upon. Initially after the attacks of 9/11, government experts tasked with ensuring the safety of our water supply declared the system to be secure. EPA administrator Christie Whitman stated on

10/18/2001, "*People are worried that a small amount of some chemical or biological agent –a few drops for instance- could result in significant threats to the health of large numbers of people. I want to assure people –that scenario can't happen. It would take large amounts to threaten the safety of a city water system. We believe it would be very difficult for anyone to introduce the quantities needed to contaminate an entire system.*"[1]

Unfortunately these sorts of statements apply only to large reservoirs and water sources and not to the distribution system. Various industry experts soon pointed out the error of declaring our water supply to be safe. Researchers from the US Air Force, Army Corps of Engineers and Hach Homeland Security Technologies (HST) have calculated that an attack on drinking water distribution systems can be mounted fro between \$0.05 and \$5.00 per death, using rudimentary techniques, and amass casualties in the thousands over a period of hours via a method known as a backflow attack.[2,3,4,5]

Wherever water can be drawn from the distribution system, it would be possible to use a pump to inject material back into the system. Once the contaminant is present in the pipes, the normal movement of water in the system acts to disseminate the contaminant throughout the network. The introduction point can be anywhere in the system. Due to the vast expanse of the distribution network, physically securing the system against such attack becomes impossible. That leaves monitoring as the only viable method to protect against such events.

2. Distribution System Monitoring

Monitoring in the distribution system is a difficult proposition. The shear number and diversity of potential threat agents that could be utilized in an attack against the system makes monitoring for them on an individual basis an effort that is doomed to failure from the start. What is needed is a broad-spectrum analyzer that can respond to any possible threat and even unknown or unanticipated events.

The difficult mission of detecting such a wide variety of potential threats is not the only challenge confronting a monitoring system for the distribution system. The environment that any such sensor would be exposed to is extremely harsh. Extreme variability in water conditions is routinely encountered in the pipes. Much of the existing water supply infrastructure is also aging and in poor condition. This results in conditions of corrosion and scaling that may cause the fouling of sensors that are not robust enough to operate under such conditions on an extended deployment timeframe. Biofilms may also form on exposed surfaces leading to sensor failure over time. What is needed is an extremely rugged sensor that is capable of withstanding long-term deployment and has the ability to respond to all types of threat agents.

3. The Hach HST Approach

Rather than attempting to develop individual sensors to detect contaminants or classes of contaminants, the Hach HST approach was to utilize a sensor suite of commonly available off-the-shelf water quality monitors such as pH, electrolytic conductivity, turbidity, chlorine residual and total organic carbon (TOC) linked together in an intelligent network. The logic behind this is that these are tried and true technologies that have been extensively deployed in the water supply industry for a number of years and have proven to be stable in such situations. One of the difficulties encountered when designing such a device is that the normal fluctuations in these parameters found within the water can be quite pronounced.

The problem then becomes, can we differentiate between the changes that are seen as a result of the introduction of a contaminant and those that are a result of everyday system perturbation? The secret to success, in a situation such as this, is to have a robust and workable baseline estimator. Extracting the deviation signals in the presence of noise is absolutely necessary for good sensitivity. Several methods of baseline estimation were investigated. Finally, a proprietary, patent pending, non-classical method was derived and found to be effective.

In the system as it is designed, signals from 5 separate orthogonal measurements of water quality (pH, Conductivity, Turbidity, Chlorine Residual, TOC) are processed from a 5-paramater measure into a single scalar trigger signal in an event monitor computer system that contains the algorithms. The signal then goes through the crucial proprietary baseline estimator. A deviation of the signal from the established baseline is the derived. Then a gain matrix is applied that weights the various parameters based on experimental data for a wade variety of possible threat agents. The magnitude of the deviation signal is then compared to a preset threshold level. If the signal exceeds the threshold, the trigger is activated.

The deviation vector that is derived from the trigger algorithm is then used for further classification of the cause of the trigger. The direction of the deviation vector relates to the agents characteristics. Seeing that this is the case, laboratory agent data can be used to build a threat agent library of deviation vectors. A deviation vector from the monitor can be compared to agent vectors in the threat agent library to see if there is a match within a tolerance. This system can be used to classify what caused the trigger event. This system can also be very useful in developing a heuristic system for classifying normal operational events that may be significant enough in magnitude to activate the trigger. When such an event occurs the profile for the vector causing it is stored in a plant library that is named and categorized by the system operator. When the event trigger is set off the library search begins.

The agent library is given priority and is searched first. If a match is made, the agent is classified. If no match is found, the plant library is then searched and, the event is identified if it matches one of the vectors in the plant library. If no match is found, the event is classified as an unknown and can be named if an investigation determines its cause. This is very significant because no profile for a given event need be present in the libraries for the system to trigger. This gives the system the unique ability to trigger on unknown threats. Also, the existence of the plant library with its heuristic ability to learn plant events results in a substantial and rapid decrease in unknown alarms over time. The developed system has been subjected to strenuous testing in both laboratory and field scenarios as detailed in the remainder of this paper and has been found to be an effective tool for surveillance of the distribution system.

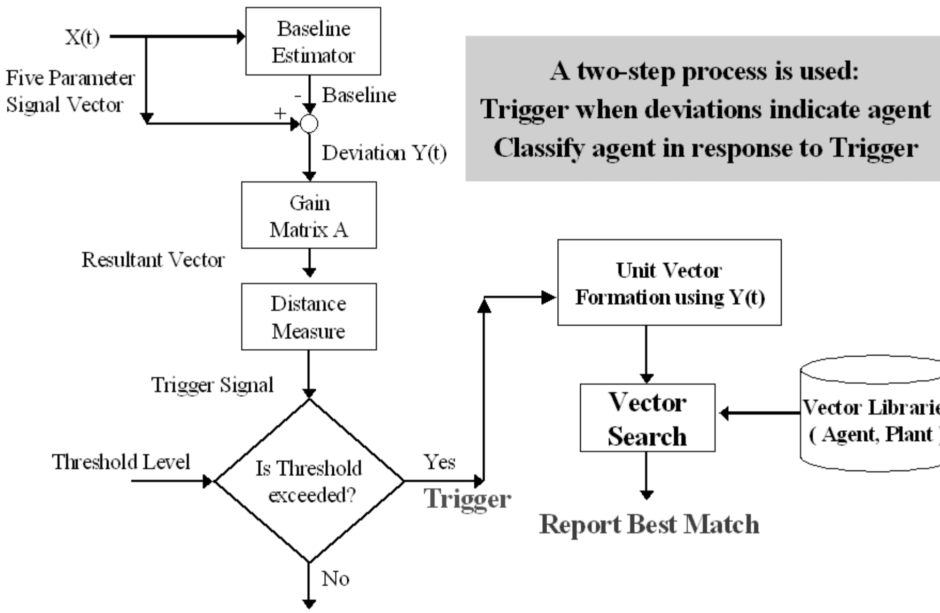


Fig 2. The use of intelligent algorithms with standard bulk parameter monitoring equipment allows for a robust system that is capable of triggering on and classifying a wide diversity of threat agents including unknown events.

4. Battelle EPA ETV Verification Testing

In the fall of 2004, the developed technology was submitted for testing to the EPA Environmental Technology Verification (ETV) testing program run by Battelle. The ETV Program develops testing protocols and verifies the performance of innovative technologies that have the potential to improve protection of human health and the environment. ETV was created to accelerate the entrance of new environmental technologies into the domestic and international marketplace. The category being tested was Multi-parameter Water Monitors for Distribution. The technologies verified needed to function on line and be capable of monitoring chlorine and at least one other parameter. The tested technologies were evaluated for accuracy of instrument readings versus reference methods, ability to maintain integrity during long term deployment, system to system variability, ability to respond to contaminants, and classification of unknown contaminants (The Hach system was the only one evaluated for ability to classify contaminants).

4.1 Accuracy[6]

The on-line instruments were evaluated on an individual basis versus standard laboratory methods for the given parameters.

Table 1. Accuracy

Evaluation Parameter		Cl	Turb	Cond.	pH	TOC
Stage 1 Accuracy	**Units 1 & 2, range of %D (median)**	-47.4 to 4.5 (-3.9)	-53.9 to –1.3 (-34.1)	-15.5 to 8.1 (2.2)	-6.6 to 3.1 (0.9)	-64.7 to 147.5 (-14.8)

Overall accuracy of all of the parameters measured was very good and correlation to laboratory method was strong. Small absolute values in turbidity and TOC lead to large %D. Due to sampling errors in some cases on-line instrument readings may be more reliable than the reference method e.g. chlorine.

4.2 Inter-unit Reproduceability[6]

Two separate but identical sets of instrumentation were deployed for the duration of the study. Measurements versus reference interments were compared through out the course of the study.

Table 2. Reproducibility

Parameter		Cl	Turb.	Cond.	pH	TOC
Inter-unit Reproducibility	Slope (intercept)	0.98 (0.03)	0.97 (0.005)	0.92 (4.19)	1.06 (-0.40)	0.97 (0.31)
	r^2	0.994	0.881	0.961	0.919	0.991
	p-value	0.779	0.884	0.006	0.517	0.374

With the exception of conductivity both units generated similar results.

4.3 Contaminant Injection[6]

During this phase of the test the instruments were installed in a recirculating pipe loop. Various contaminants were injected to determine if they altered the baseline response pattern of the instruments. Similar injections were preformed after the long-term deployment test to ascertain if there had been any degradation of instrument response.
Total chlorine and TOC were dramatically affected by injections of nicotine, *E. coli*, and Aldicarb; and turbidity, pH, and conductivity were affected by some or all of the injections, but not as consistently as total chlorine and TOC. Aldicarb altered pH during testing after extended deployment but not before. Agreement in both cases with reference readings indicates that the instruments were functioning properly and, the difference was in the injection preparation.

Table 3. Response to Contaminant Injection

Parameter			Cl	Turb	Cond	pH	TOC
Initial response to injected contaminants	**Nicotine**	**Ref**	-	(a)	NC	NC	+
		Hach	-	+	NC	NC	+
	Arsenic Trioxide	**Ref**	-	(a)	+	+	NC
		Hach	-	+	+	+	NC
	Aldicarb	**Ref**	-	(a)	NC	NC	+
		Hach	-	+	NC	NC	+
Response after extended deployment	**E. coli**	**Ref**	-	+[(b)]	+	-	+
		Hach	-	+	+	-	+
	Aldicarb	**Ref**	-	+[(b)]	NC	-	+
		Hach	-	+	NC	-	+

(a) Relatively large uncertainty in the reference measurements mad it difficult to detect a significant change

(b) Magnitude of change different between duplicate injections

+/- Parameter measurement increased/decreased upon injection

NC No change in response to contaminant injection

4.4 Long-term Deployment[6]

During this phase of the testing, the systems were operated continuously for 52 days with only normal maintenance, such as reagent replenishment, being performed. During the course of the test, instruments were regularly compared to reference instruments. See Table 4. At the end of the 52 days a second response to contaminant injection procedure was performed. See table 3. The results of the extended deployment study indicate that the system can be effectively deployed for long periods of time with only routine maintenance. Relative large %D in turbidity and TOC measurements are artifacts of the low total values for these parameters encountered during the testing procedure and do not indicate a problem with these sensors.

Table 4. Post-Extended Deployment Results

		Unit 1		Unit 2	
Parameter	**Reference Average (SD)**	**Average (SD)**	**%D**	**Average (SD)**	**%D**
Free chlorine	1.03 (0.03)	0.98 (0.02)	-4.9	0.98 (0.02)	-4.9
Turbidity	0.17 (0.02)	0.16 (0.03)	-5.9	0.15 (0.04)	-11.8
Temperature	22.66 (0.16)	22.61 (0.03)	-0.2	23.70 (0.06)	4.6
Conductivity	356 (1)	380 (1)	6.7	357 (1)	0.3
pH	8.59 (0.01)	8.40 (0.01)	-2.2	8.61 (0.00)	0.2
TOC	0.88 (0.01)	0.70 (0.01)	-20.5	0.91 (0.01)	3.4

4.5 Inter-Unit Reproduceablity[6]

Two Hach units were compared, using data collected from reference samples throughout the verification test to determine whether similar results were generated. See table 5.

Table 5. Inter-Unit Reproducibility Evaluation.

Parameter	**Slope**	**Intercept**	**r^2**	**t-test p-value**
Free chlorine	0.98	0.03	0.994	0.779
Turbidity (outlier removed)	0.97	0.005	0.881	0.884
Temperature	0.72	7.68	0.758	5.5 × 10-6
Conductivity	0.92	4.19	0.961	0.006
pH	1.06	-0.40	0.919	0.517
TOC	0.97	0.31	0.991	0.374

Shading indicates a significant difference between the two units.

For free chlorine, pH, TOC, and turbidity, the linear regression had coefficients of determination greater than 0.91 and slopes within 6% of unity, indicating similar and repeatable results. The t-test p-values confirmed the sensors were generating statistically similar results. The conductivity meters had a linear regression coefficient of determination of 0.961 and a slope of 0.92, indicating that the data were highly correlated with one another. The t-test generated p-values significantly less than 0.05, which indicated that the results from the two conductivity sensors were significantly different. This difference was driven by the small amount of variability in the conductivity measurements; therefore, the small difference between the means of the two units was statistically significant. It is important to note that the offsets in the measured parameters do not affect the performance of the algorithm because the baseline is removed and the classification is performed based only on deviations from baseline.

4.6 Contaminant Classification[6]

During the final stage of the verification test thirteen contaminants (See Table 6) were injected at a concentration of approximately 15 mg/L, in duplicate, into a 1500 foot straight line pipe and allowed to flow past the monitoring sensors. Every contaminant injection resulted in the system exceeding the trigger threshold and producing a corresponding agent alarm. Each minute-by-minute search of the agent library can result in more than one agent being identified. For both Hach Units, the agent alarms occurred as few as eight times and as many as 79 times during the 20-minute injection periods. No agent alarms occurred outside of the 20-minute injection periods. If the system recognized deviations from the baseline, the agent library identified and recorded these deviations as "unknown" event. Due to the dynamic nature of the leading and trailing edges of the injected contaminant, it is possible that an injection event generated alarms other than the known injected contaminant. Table 6 shows all contaminant injections classified according to the fraction of agent alarms attributable to the correctly classified injected contaminant. The data is depicted concisely by classification rates divided into five levels: Level 5 – greater than 70%, Level 4 – between 31% and 69%, Level 3 – between 1% and 30%, Level 2 – injected contaminant not identified but other contaminants were identified, and Level 1 – no injections detected.

Table 6. Contaminant Classification Results

	Injection 1		Injection 2	
Contaminant	**Unit 1**	**Unit 2**	**Unit 1**	**Unit 2**
Aldicarb	4	4	4	4
Arsenic trioxide	2	2	2	2
Colchicine	4	4	4	4
Dicamba	4	5	5	5
Dichlorvos	4	3	3	2
E. coli	3	2	4	2
Ferricyanide	5	5	5	5
Fluoroacetate	5	5	4	4
Glyphosate	4	3	2	2
Lead nitrate	5	5	5	5
Mercuric chloride	4	4	4	4
Methanol	4	4	4	3
Nicotine	2	2	2	2

From the tests conducted on Hach Version 1 System, weak results were obtained for Methanol and Dichlorvos, while poor results were obtained for Glyphosate, Nicotine, Arsenic Trioxide and *E. coli*. The data from the tests on VERSION 1 of the HHST technology at the EPA center were recorded at the time of the tests. The Event Monitor Trigger System also acts as a data logger and provides a copy of the sensor signals recorded during the tests. This situation afforded us the ability to analyze failures detected in the VERSION 1 tests, improve and upgrade software, and then replay new versions of the technology to test for efficacy. A variety of causes were found to affect the test results of VERSION 1.

Because of a misunderstanding, the original Version 1 threat agent library included Round-UP Herbicide (a form of glyphosate), while the ETV protocol used pure glyphosate. When pure Glyphosate was added to the agent library in Version 3, the system correctly classified the agent.

HHST had previously tested nicotine (in house, and at ECBC) with good results. However, the data from the ETV test facility revealed that the excessive mixing method employed prior to injection had caused the nicotine base to react with the carbon dioxide in the air, changing the chemical nature of the contaminant. The Agent Library was improved by adding a signature for reacted nicotine, and Version 3 shows the positive test results. These two examples demonstrate the sensitivity of the system, and how the comprehensive data structure of its Agent Library derives its classification accuracy.

In addition while first developing the Agent Library, HHST employed bench-scale chlorine analyzers that contained EDTA (a metal sequestering agent) as a reagent component, whereas the EMTS sensor panel includes chlorine analyzers that do not use this substance. The EPA/ETV tests revealed this flaw, and VERSION 3 includes upgraded library signatures. Signatures for some other agents were examined for tabular errors and those were corrected as needed. This second set of test results could not be included in the ETV report, as any re-testing was not a part of the original test protocols. Following analysis and upgrades, two succeeding algorithm versions were produced; the test results from VERSION 3 is summarized in Table7.

Table 7. Classification Results from Algorithm version 3

	Injection 1		**Injection 2**	
Contaminant	**Unit 1**	**Unit 2**	**Unit 1**	**Unit 2**
Aldicarb	4	4	4	4
Arsenic trioxide	4	5	4	5
Colchicine	4	4	4	4
Dicamba	4	3	5	5
Dichlorvos	2	2	3	3
E. coli	4	3	4	3
Ferricyanide	5	5	5	5
Fluoroacetate	5	5	5	5
Glyphosate	4	4	4	4
Lead nitrate	5	5	5	5
Mercuric chloride	5	5	5	5
Methanol	4	4	4	3
Nicotine	4	4	4	4

5. ECBC Testing

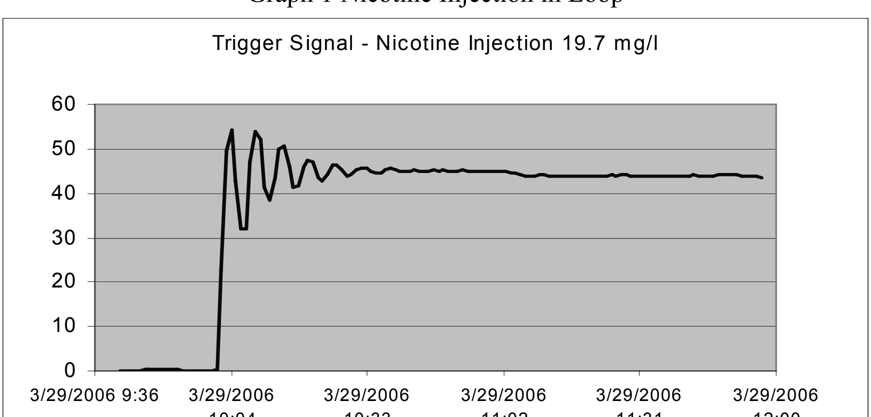
The purpose of this effort was to challenge water distribution systems and sensors, with agent simulants and real threat agents, in order to characterize the response of the distribution system and Early Warning System to agents. Agent concentrations and water solutions were varied to allow for the development and demonstration of distribution methodologies and performance data acquisition. In addition, this work evaluates the effectiveness of Hach Homeland Security Technologies real-time detection technology and provides important information necessary for the U.S. Army to perfect its theories of operation and response mechanisms. The scope of the work performed during these tests was two fold. The first part of the test was to perform beaker studies on agents that are not available for use in the Hach Laboratories in Colorado such as VX, Sarin, Soman, Ricin and Anthrax etc. The second part of the testing protocol called for verification of signatures in a flowing loop to validate the transfer of the beaker signature data to real world scenarios.

It is known that in a real attack on a water distribution system the concentrations of the agents would vary throughout the distribution system. The concentrations tested were to either infectious amounts, $ID_{50,}$ for replicating agents, or LD_{50} amounts for chemical agents. Each agent was tested at three dose values, as defined in a test plan matrix. Two types of disinfectant are commonly used in water distribution systems: free chlorine, and monochloramine. It is necessary to test agents in both types of media to have information representative of each type. It is also clear that variable amounts of chlorine would be in the distribution system water, so tests used different solutions: 0.2 ppm Free Available Chlorine (FAC), 1 ppm FAC, and 2 ppm Monochloramine, as these are the limits of typical system concentration. Monochloramine is usually less variable in concentration and can be tested at typical values. Real-time loop tests were run on nicotine, ricin, BA, and methanol.

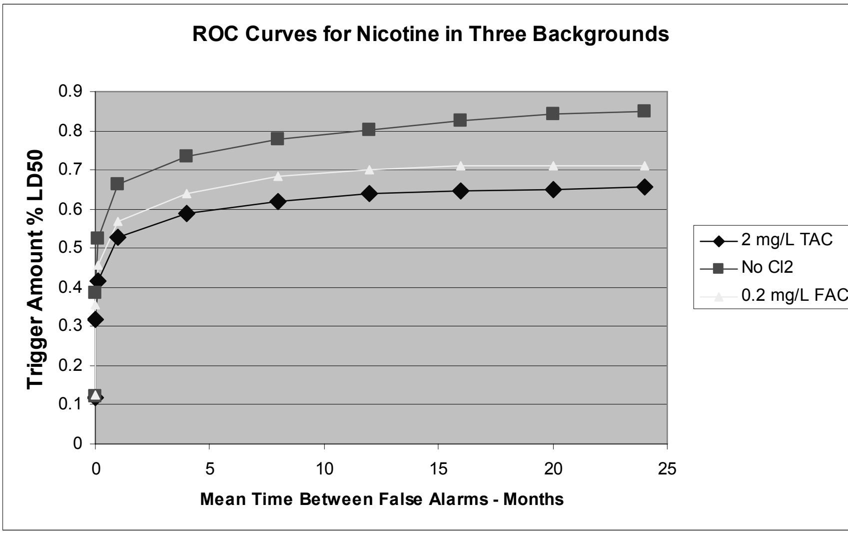
All fingerprints were successfully developed and ROC curves were generated for all agents tested. It was also found that the fingerprints developed from the lab work could be successfully transferred to a flowing system by successfully triggering and classifying agents in the flow loops. Because of security concerns and confidentiality, only selected nicotine test data is provided in graphs 1 and 2.

The test concentration of nicotine was 19.7 mg/L. In a single test run, nicotine was recognized by the system, with detection angles ranging from 0.94 degrees (essentially a perfect match) to 9.84 degrees (a weak match). The fluctuating trigger signal is due to the mixing dynamics of a recirculating loop.

Graph 1 Nicotine Injection in Loop



Graph 2 ROC Curves for Nicotine in Three Backgrounds

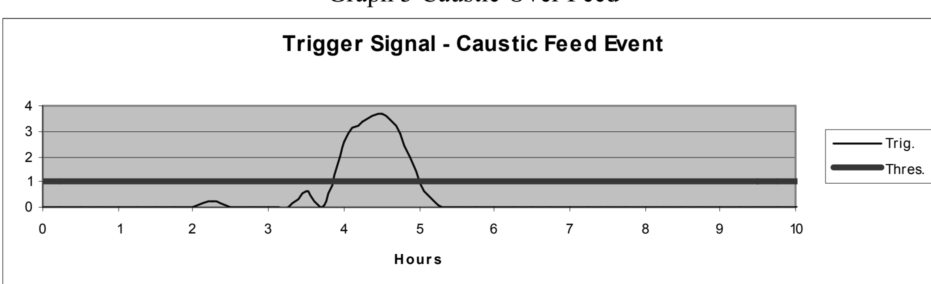


Threshold trigger levels range from 0.36 to 1.26 for all curves

6. Real World Testing

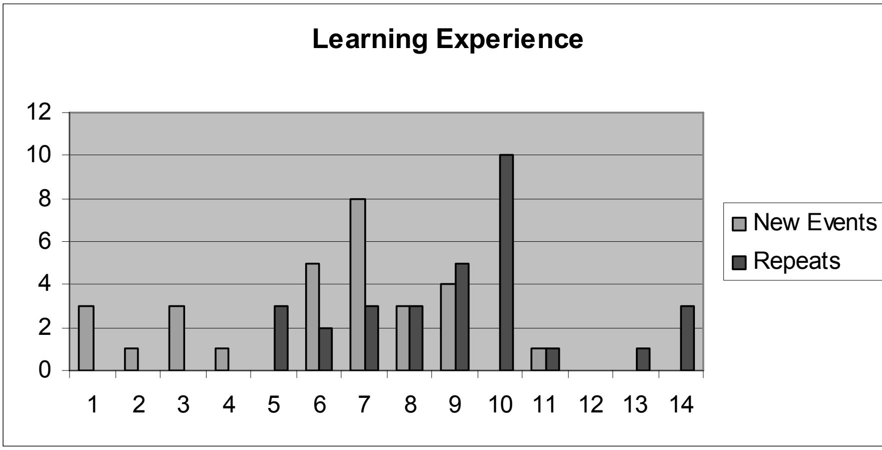
The described system has been deployed in a variety of real world venues across the United States to determine theories of deployment and response and to verify robustness of the trigger and learning ability. To date over 120,000 hours of real world data has been collected. Several actual incidents have been recorded and learned. An example is the caustic over feed event depicted in Graph 3. Graph 4 shows a very noisy real world situation and demonstrates how quickly the system can learn commonly occurring unknown events to help reduce their occurrence.

Graph 3 Caustic Over Feed



In this situation, the water utility plant used caustic feed to control the water pH. Misdelivery of a more concentrated form of the caustic resulted in the feed of excess caustic. The pH and conductivity of the water deviated enough to trigger a Plant Event for the system to “learn” and identify its reoccurrence.

Graph 4. Learning Experience Rate at a Field Site



The data in this case represent a real world deployment situation that had very noisy water quality. In this scenario there were 26 unique trigger events in the first 11 days of operations. All were fingerprinted and learned by the system. 11 of the events were repeated. This demonstrates that common events are rapidly learned by the system resulting in a rapid decrease in unknown alarms.

7. Conclusion

The designed and tested system makes use of an integrated array of robust common water quality monitoring sensors coupled with interpretive algorithms to recognize and classify significant water quality deviations. Extensive in house and 3rd party verification testing as well as extensive deployment at field sites has demonstrated the suites ability to fill the analytical gap that currently exists for distribution network monitoring and serve the purpose of an early warning system in the water distribution network. Hopefully the systems unique ability to learn will result in not only increased safety from terror related events but will morph into an operational tool that will find everyday use in improving water quality operations and ensure a better quality drinking water product to consumers.

References

1. C. Whitman. Environmental News: Whitman allays fears for water security; Possibility of successful contamination is small. EPA Press Release. Thursday, October 18th (2001).
2. D. Kroll, Mass casualties on a budget. Confidential paper. Hach HST (2003).
3. Army Corps of Engineers, Calculations on threat agents, requirements, and logistics for mounting a successful backflow attack. (2003)
4. T. P. Allman, Drinking water distribution system modeling for predicting the impact and detection of intentional contamination, Masters Thesis, Colorado State University, Dept of Engineering, Fort Collins Colorado, Summer (2003).
5. M. Ginsberg, V. Hoch and D. Kroll. Terrorism and the security of water distribution systems. *Sword and Ploughshares*. January. (2005)