SAND2007-6889C

Limited-Memory Techniques for Sensor Placement in Water Distribution Networks

William Hart
Jonathan Berry
Erik Boman
Cynthia Phillips
Lee Ann Riesen
Jean-Paul Watson

wehart@sandia.gov
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.

Water Security

National Water Security Goals

- Protect long-term availability of national water resources
- Protect the operation of water utility distribution systems
- Protect water resources and infrastructure from improper use

Universal Vulnerabilities in Water Systems

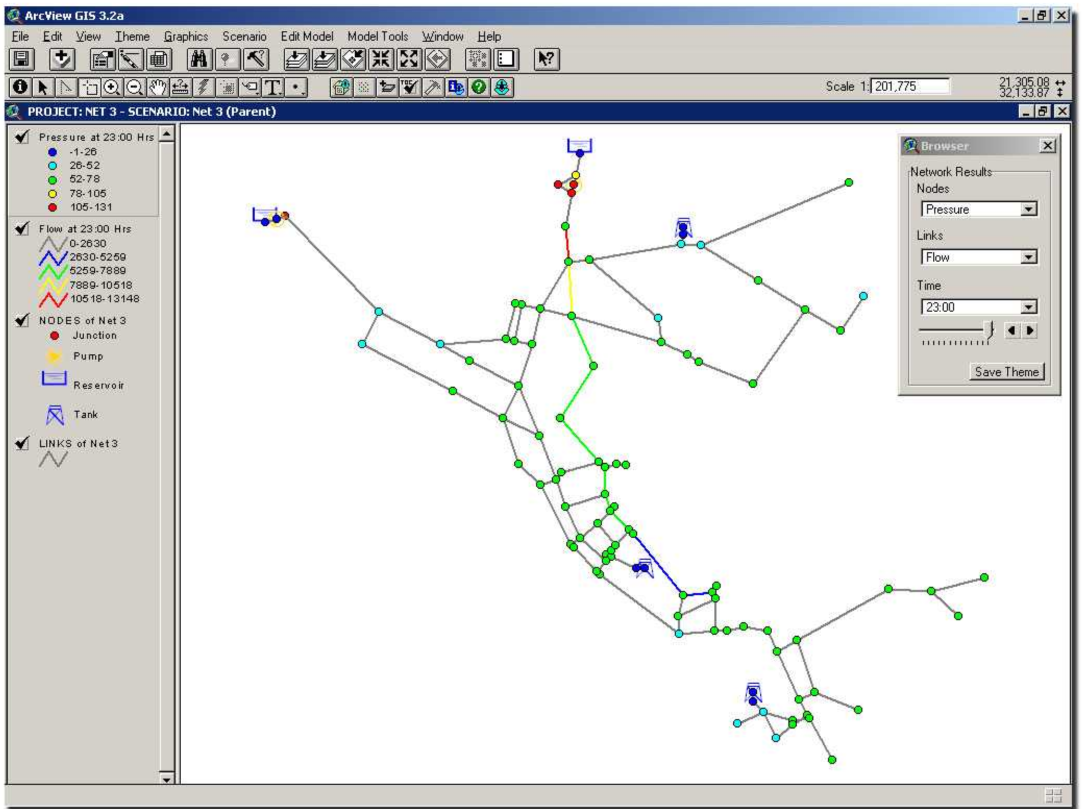
- Plant access
- Source Water
- Water storage
- Water distribution

Sandia National Laboratories

What is a water distribution network?

Drinking Water
- Water source
- Treatment facilities
- Transmission systems
- Distribution systems

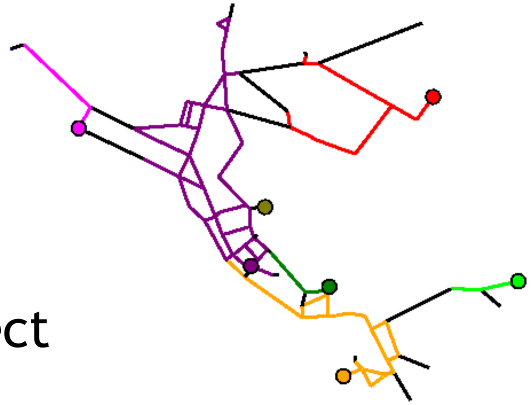
Wastewater
- Wastewater source
- Collection system
- Treatment facility
- Receiving water body



Sandia National Laboratories

A Motivating Threat Scenario

Contaminant Injection



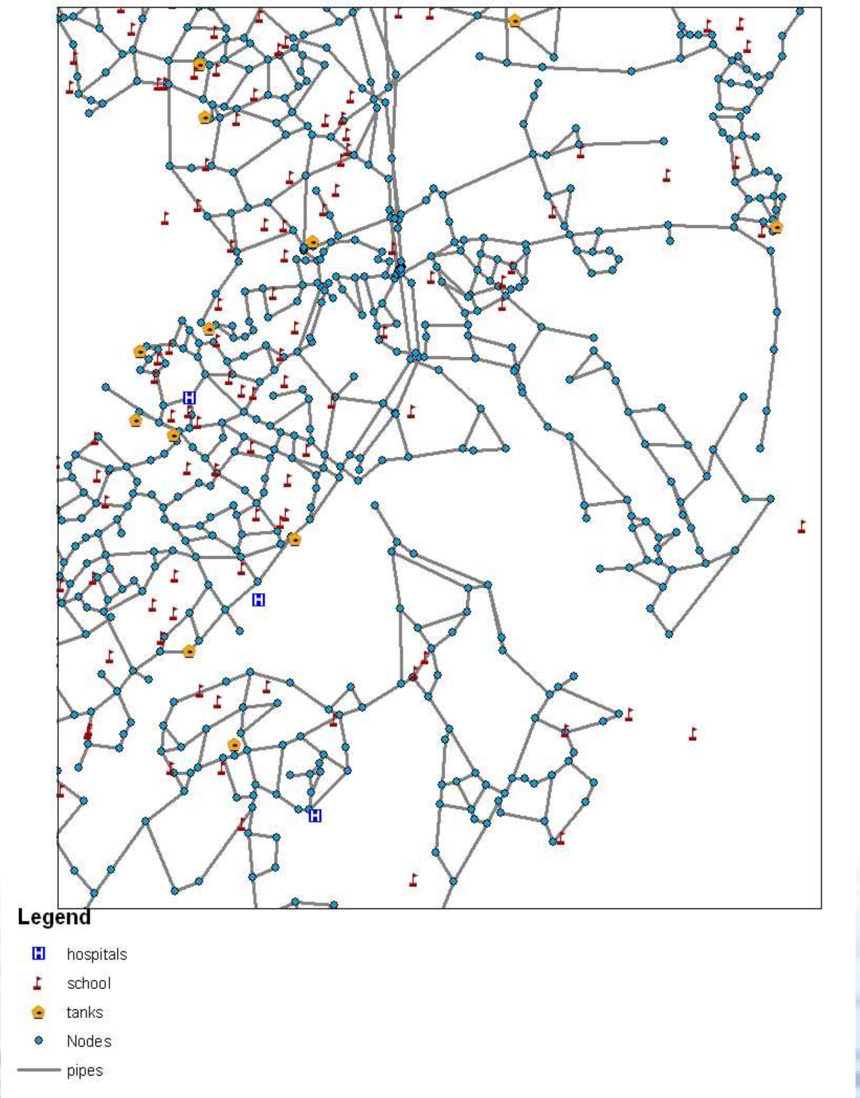
- Risk: moderate-high
  - Technically difficult to accomplish
  - Potential terrorists fascinated by this prospect
  - Potential economic/health impacts are high
- Impact: public health impacts, network contamination impacts
- Mitigation:
  - Use of detection equipment
- Response:
  - Coordination with public health institutions
  - Proactive identification of contaminant source
  - Decontamination procedures

Designing a Contaminant Warning System (CWS)

Technical Goal: placement of sensors for the CWS within a budget

Possible objectives:

- Minimize response time
- Minimize health impacts
- Minimize extent of contamination
- Minimize volume of water that enters the water network
- Minimize number of failed detections
- Minimize cost
- Minimize political risk…



Sandia National Laboratories

Modeling Contamination Incidents

- Given: Set of incidents = (location, time) pairs
- Simulate the evolution of a contaminant plume
- For each event determine
  - Where/when event can be observed
  - Contamination impact prior to that observation
- Sensors
  - Detect without errors
  - A detection raises a general alarm
  - Response delay can be factored into the contamination impact

Contaminant Transport Modeling

Water movement (direction, velocity in each pipe) determined by

- Demand (consumption)
- Pumps
- Gravity
- Valves
- Sources/tanks

Current (most trusted) simulator

- EPANET code computes hydraulic equations to determine flows
- Discrete-event simulation for contaminant movement

Note: can only model transport of water soluble contaminants!

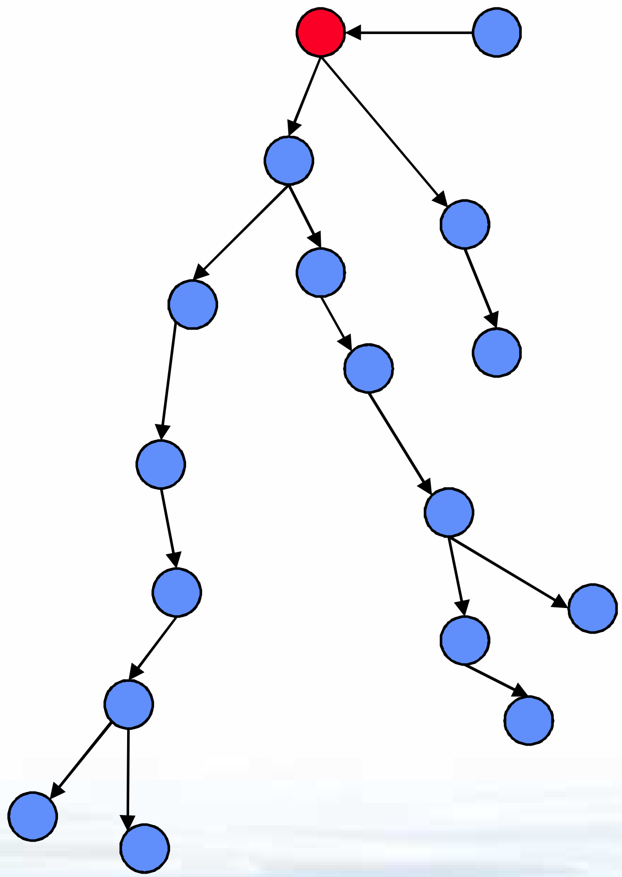
Witnessing an Incident

Consider an incident at the red node, with current flow directions as indicated.

When a contaminant reaches a node, a sensor at that node can "witness" the incident

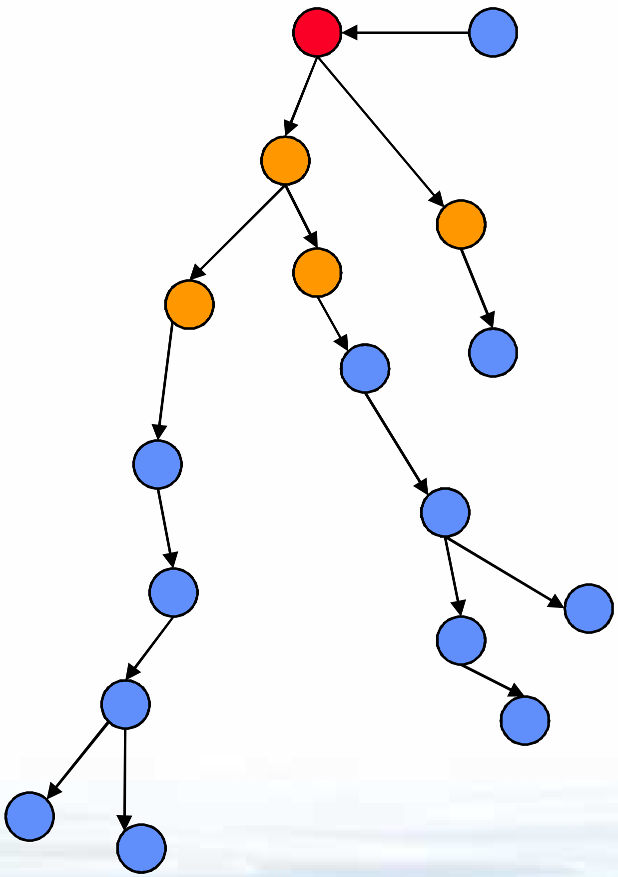
The "possible witness list" is shown (pictorially) below.

Possible Witnesses:

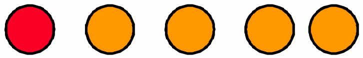


Witnesses of an Incident

At the first EPANET reporting step, the orange nodes have experienced contamination.

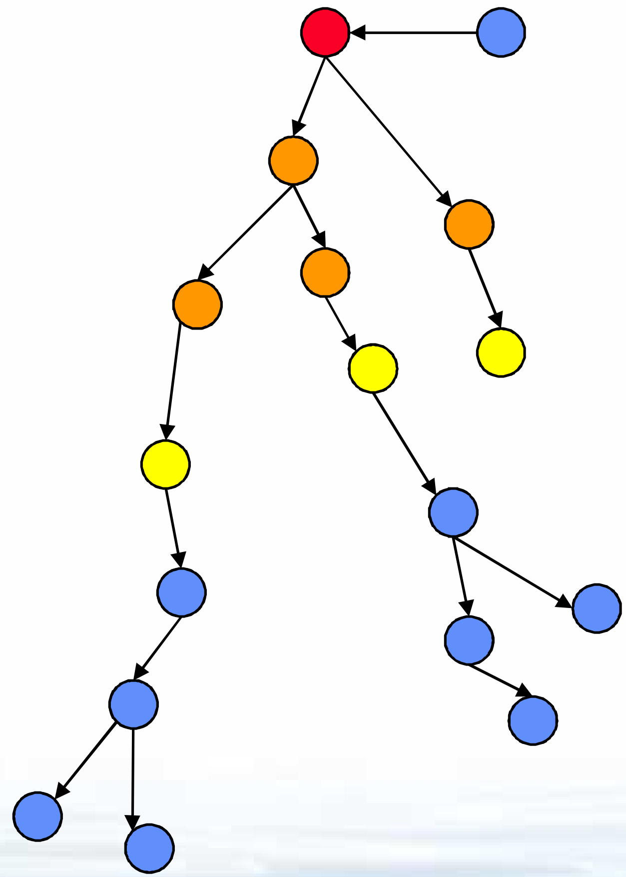


Possible Witnesses:

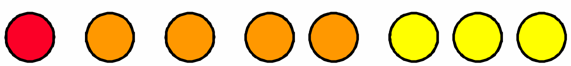


Witnessing an Incident

At the second EPANET reporting step, the yellow nodes have experienced contamination.



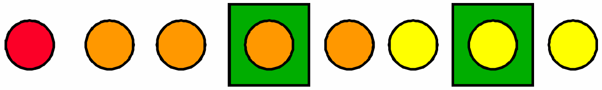
Possible Witnesses:



Witnessing an Incident

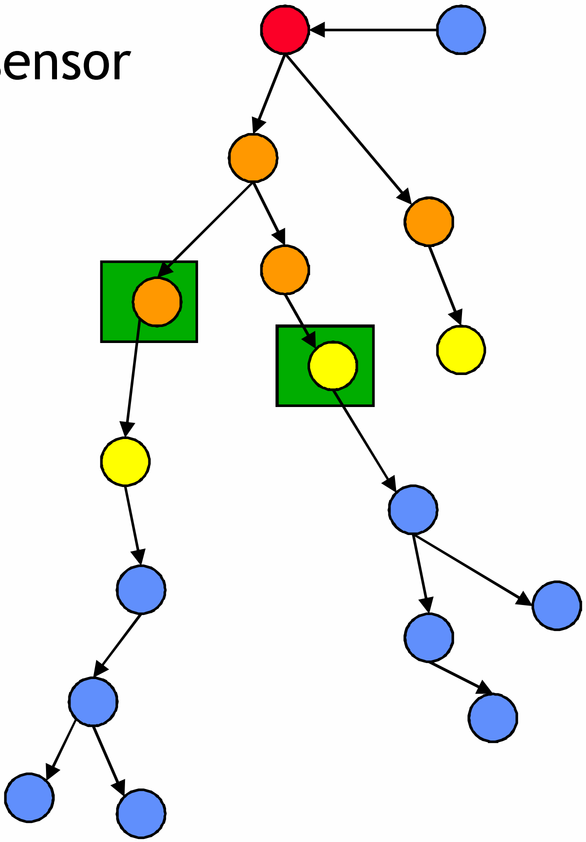
Simulator gives ordered list of nodes where a sensor could witness contamination

Witnesses:



This example has two (green) sensors.

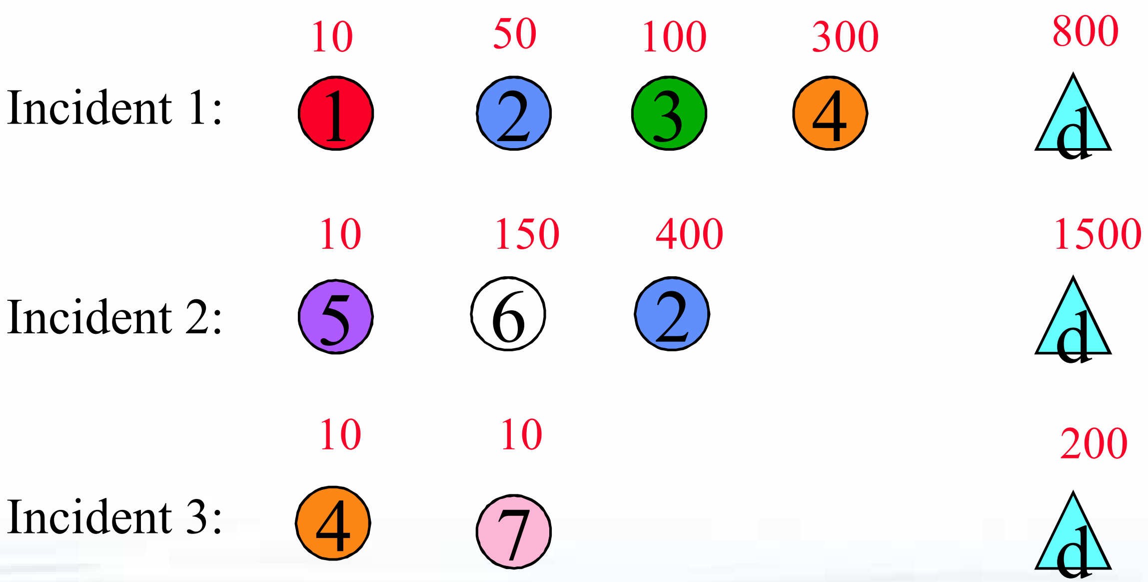
Perfect sensor model: first sensor in list detects the incident.



Evaluating a Sensor Placement

- Impact in red

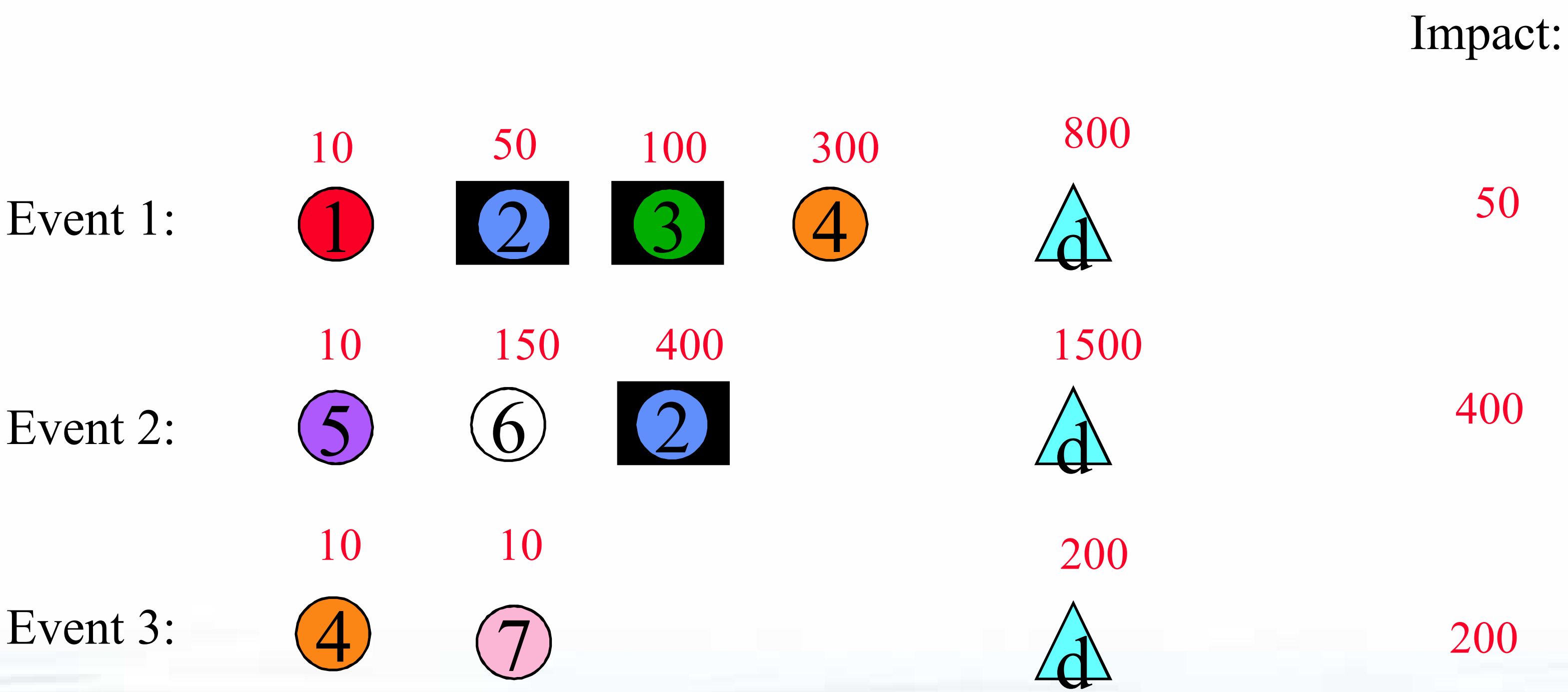
= dummy node (represents failure to detect)



Evaluating a Sensor Placement

- Impact in red

d = dummy node (represents failure to detect)



Choose sensors 2 and 3 (black)

Integer Programming Formulation

IPs can be used to model sensor placement for water security

– Berry et al (2003, 2006); Watson et al (2004)

Objective: $\sum_{a \in A} \sum_{i \in L} \alpha_a w_{ai} x_{ai}$

- α – contamination likelihood
- *w* – contamination impact
- *x* – witness variable
- *s* – sensor placement variable

$$
\begin{aligned}
&\text{minimize} \sum_{a \in A} \sum_{i \in L} \alpha_a w_{ai} x_{ai} \\
&s.t. \\
&\sum_{i \in L} x_{ai} = 1 && \forall a \in A \\
&x_{ai} \le s_i && \forall a \in A, i \in L \\
&\sum_{i \in L} s_i \le S_{\max} \\
&s_i \in \{0,1\}
\end{aligned}
$$

IP model:

- Can capture different objectives/networks
- Can be solved with COTS software
  – We need a 64-bit workstation to solve large instances

Sandia National Laboratories

Scalability Challenge

- Full-size problems have
  - 10,000+ nodes
  - 100's of interesting times of day
  - Multiple seasons
  - Weekends/weekdays
  - Special events
  - Multiple contaminant types
- Lots of witness variables
  - trivial upper bound: (# events) x (# nodes)
- Space can be an issue
  - 64-bit workstations
- Linear programming relaxation can be difficult to solve

Planners want to develop CWS designs on their PCs!

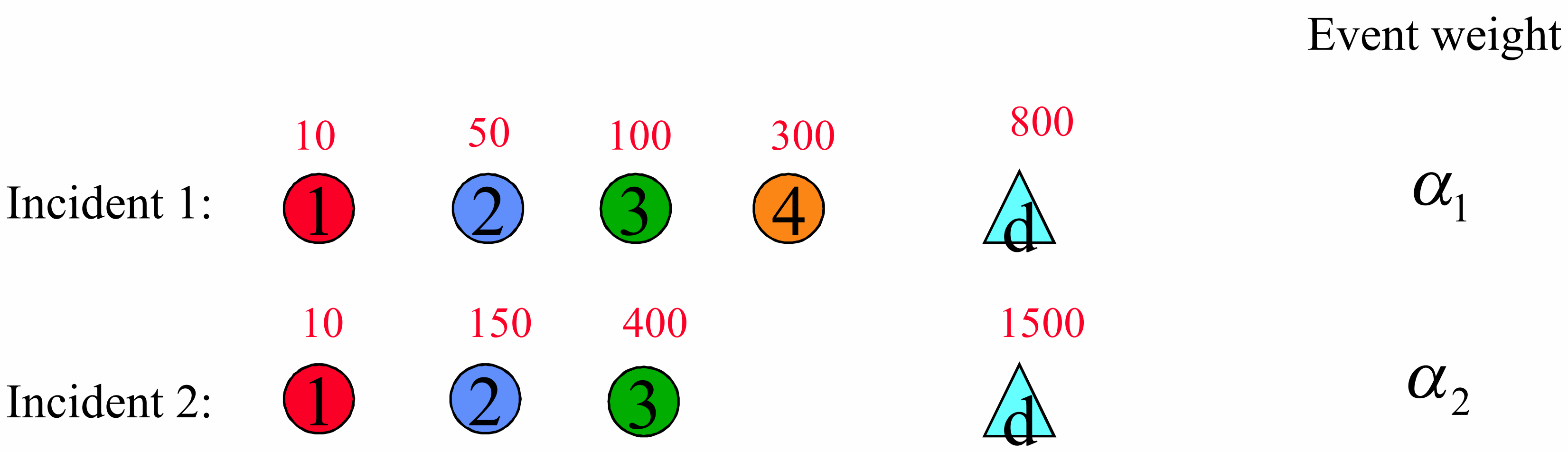
Limited-Memory Strategies for SP

- Aggregated Integer Programming Formulations
  - Incident aggregation
  - Witness aggregation

- Heuristic solvers
  - GRASP Heuristic
  - Lagrangian Heuristic

Incident Aggregation

- Sorted witness list of Incident 2 is a prefix of Event 1’s list ➔ merge to a single event with no error



Sandia National Laboratories

Incident Aggregation

- New incident has the same objective impact as two previous ones

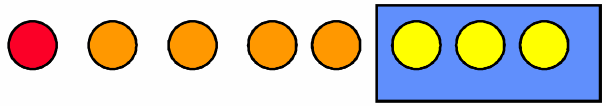
	1	2	3	4	d	Incident weight
Incident 1:	10	50	100	300	800	α_1
Incident 2:	10	150	400	1500	1500	α_2
Aggregated:	10	$\frac{50\alpha_1 + 150\alpha_2}{\alpha_1 + \alpha_2}$	$\frac{100\alpha_1 + 400\alpha_2}{\alpha_1 + \alpha_2}$	$\frac{300\alpha_1 + 1500\alpha_2}{\alpha_1 + \alpha_2}$	$\frac{800\alpha_1 + 1500\alpha_2}{\alpha_1 + \alpha_2}$	

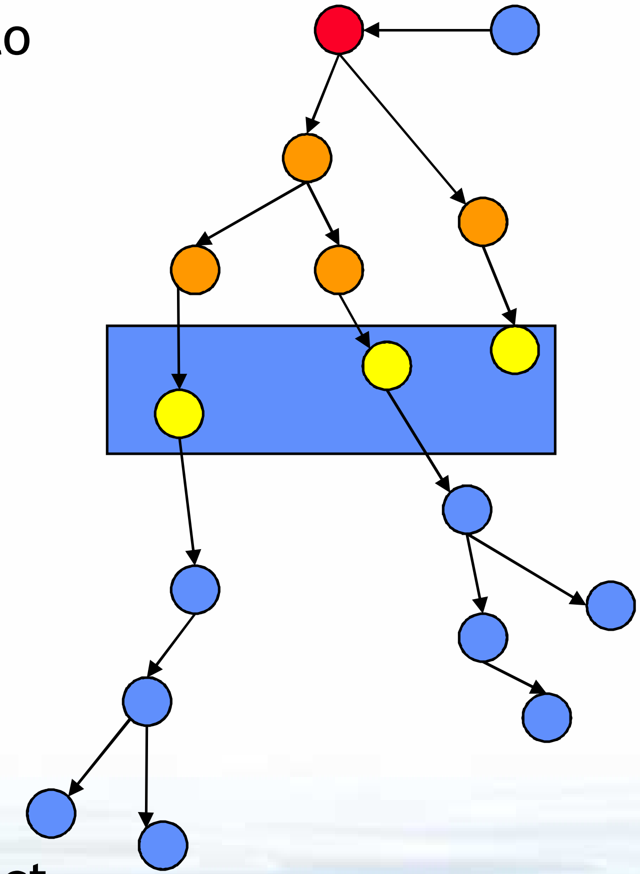
Aggregated Weight: $\alpha_1 + \alpha_2$

Problem Size Reduction: Witness Aggregation

Witness aggregation: Use one variable to represent a group of witnesses with approximately equal impact. Ensure there is one sensor in the group.

Witnesses:

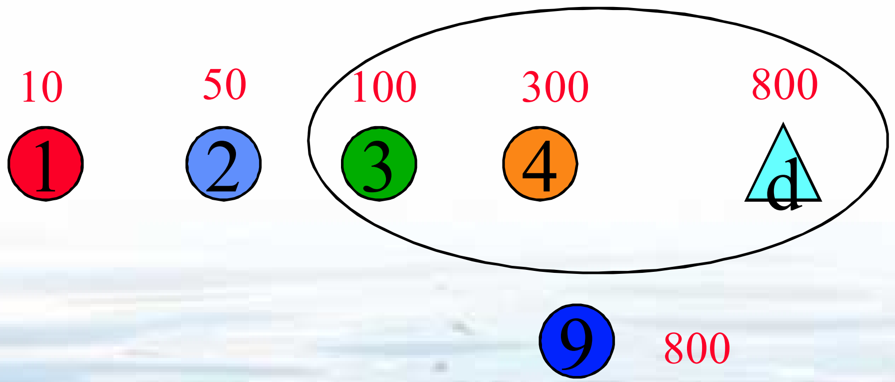




Typically see 10x reduction even if
Only combine witnesses with same impact

Witness Aggregation

- Group witnesses that are contiguous in the list for an event.
- Starting with dummy, group elements such that worst and best impacts satisfy a "closeness" measure.
- Closeness is parameterized between 0 and 1
  - One extreme only aggregate equal impacts (no error)
  - Other extreme aggregates all witnesses into one (total error)
- Option to distinguish any detection from failure
- Weight of "supernode" is largest of any member node



Witness Aggregation Methods

- PMR (Percent of Maximum Range)
  - R is largest difference between best witness and worst (dummy), taken over all events.
  - Group witnesses if difference in impact $\leq \rho R$, for $0 \leq \rho \leq 1$

- Ratio aggregation
  - Two witnesses can merge if ratio of high/low impact $\leq \frac{1}{r}$, $0 \leq \mathrm{r} \leq 1$

Aggregation summary

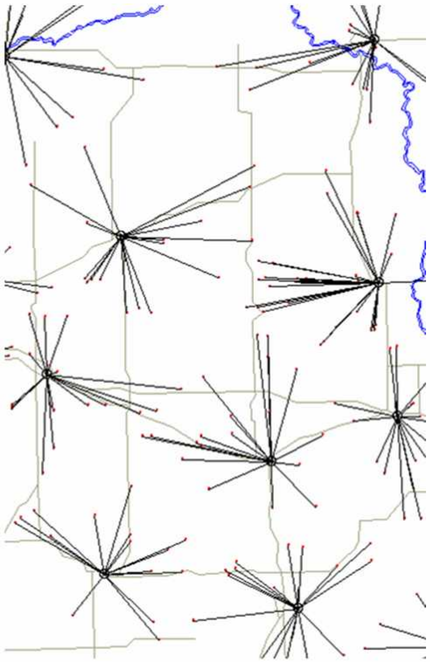
- Scenario aggregation
  - No error
  - Requires prefix (e.g. differ in contaminant only)
- PMR witness aggregation
  - Guaranteed size reduction
  - Unbounded error
- Ratio witness aggregation
  - Guaranteed error bound
  - No guarantee on size reduction

Impact: using both scenario aggregation and ratio witness aggregation (ratio 0.5) ran 100 times faster than using no aggregation and gave less than 1% error.

GRASP Heuristic

Grasp: Multistart local search

- Neighborhood swaps sensor location with non-location
- Can rapidly solve problems with 10,000's of junctions (SNL-3 in 154 seconds)
- Heuristic solutions are often optimal
- Uses sparse matrix representation, but still requires superlinear space.



Sandia National Laboratories

Lagrangian Relaxation for p-Median Problems

- What is the biggest challenge in solving this formulation well?:

$\alpha_j w_{ij}$ (→ c_{ij})

$$\min \sum_{i,j \in A} c_{ij} x_{ij}$$

$$\text{s.t.} \quad \sum_{i \in L} x_{ij} = 1, \qquad j \in A$$

$$x_{ij} \le y_i \qquad i \in V, j \in A$$

$$\sum_{i \in L} y_i = p$$

$$y_i \in \{0,1\} \qquad i \in V$$

$$x_{ij} \ge 0 \qquad i \in V, j \in A$$

"Every event must be witnessed by a sensor."

Lagrangian Relaxation for p-Median Problems

- Solution strategy: lift those tough constraints out of the constraint matrix into the objective. (e.g., Avella, Sassano, Vasil'ev, 2003)

$$\min \sum_{i,j \in A} c_{ij} x_{ij} +$$

s.t.

$$x_{ij} \leq y_i \qquad i \in L, j \in A$$

$$\sum_{i \in L} y_i = p$$

$$x_{ij} \geq 0 \qquad i \in L, j \in A$$

$$y_i \in \{0,1\} \qquad i \in V$$

Good news: remaining problem easy to solve!

Bad news: some of the original constraints might not be satisfied.

Solving a Relaxed Problem

New problem:

$$\min\left(\sum_{i\in L, j\in A}\left(\mathrm{c}_{\mathrm{ij}}-\lambda_j\right)x_{ij}\right)+\sum_{j\in A}\lambda_j$$

subject to:

$$x_{ij} \le y_i \quad \text{for } i \in L, j \in A$$

$$0 \le x_{ij} \le 1, \quad y \in \{0,1\}$$

$$\text{For fixed } \lambda_{j,} \text{ let } \rho(i) = \sum_{j\in A}\max\left(0, c_{ij}-\lambda_j\right)$$

Set $y_i = 1$ for the p locations with lowest values of $\rho(i)$.
Set $x_{ij} = 1$ if y_i=1 and c_{ij}-λ_j < 0, x_{ij}=0 otherwise.

- Gives valid lower bound on the best p-median cost
- Linear space, O(W + pn) time for n locations, W potential witnesses.

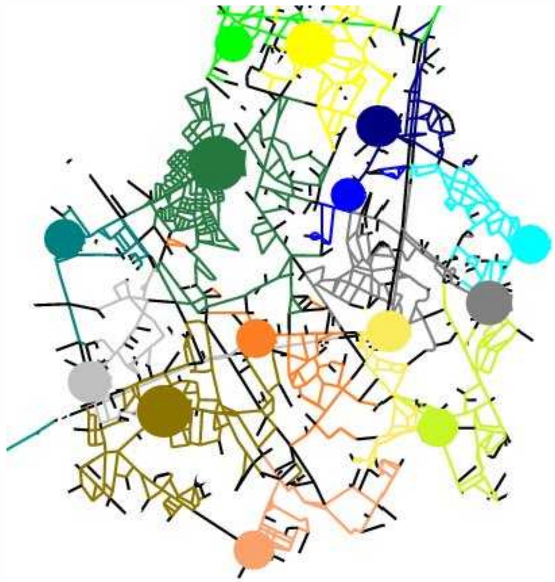
Simple computational example

- Network with about 3358 nodes, 1621 events, 5 sensors
- Memory: Lagrangian: 45028kb, Heuristic: 154424kb
- Memory increases rapidly, by 13,000 events, Heuristic ~1Gb

Objective	Solver	Gap	Time
Pop. Exposed	Lagrangian	.008	84.3s
Pop. Exposed	Heuristic	0	33.8
Extent Contam.	Lagrangian	0	73.1s
Extent Contam.	Heuristic	0	33.2
Mass Consumed	Lagrangian	.049	85.4s
Mass Consumed	Heuristic	0	41.7
Vol. Consumed	Lagrangian	.641	104.7s
Vol. Consumed	Heuristic	0	44

Impact Summary

TEVA-SPOT: software developed by EPA’s TEVA Program to identify the best placement of sensors to improve water security



Features:

- Scalable solvers for large-scale problems
  - 10,000s junctions and pipes
- Solvers that can optimize many different objectives
- Flexible specification of performance constraints
- Fast solvers
- Methods for rigorously evaluating solver performance

Impact:

- Designed sensor placements for EPA’s Water Security Initiative (WSI)
- Used to find prototypical designs for 9 partner utilities participating in the EPA’s TEVA Program
- TEVA-SPOT is the recommended method for future WSI utility participants

Outstanding Challenges

- Modeling realistic sensors
  - Imperfect sensor models
  - Direct optimization with real sensor data
  - Distributed detection models

- Addressing data uncertainties
  - Population water consumption
  - Demand flows

- Extensibility to multiple “scenarios”
  - Winter vs Spring vs …

Ideas for Related Work

- Integrate static sensor placement with manual sampling
  - We've done some preliminary work in this area
  - These problems have more of a scheduling nature, and they are harder to solve
- Apply SP model to related domains
  - Can apply directly to detect airborne contaminants
  - NOTE: network flows do not impact the formulation, except for how the impact data is generated!
- Develop SP models for other objectives
  - Example: place sensors to facilitate source location
  - Early work in this led to our collaboration with the EPA…!

THE END

Modeling Assumptions

- Sensors are perfect
- Sensors raise a general alarm
  - Can model a response delay
- Fixed set of demand patterns for “typical” day
  - Seasonal variations
  - Special events
  - Weekday/weekend
- Accurate extended-period simulation of the water distribution network

What data do we need for sensor placement?

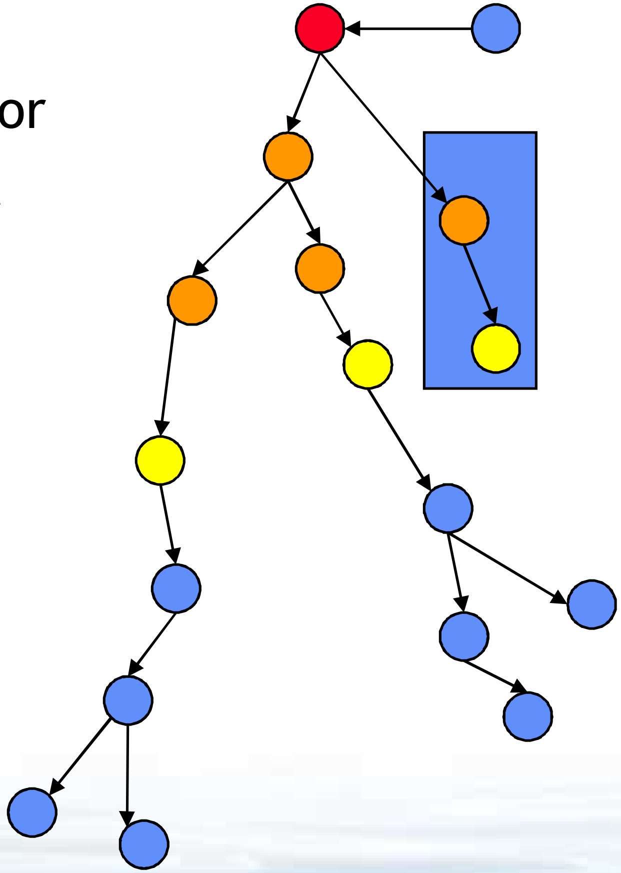
- Population consumption
  - Location and time
  - Individual characteristics: health, age
- Contamination risks
  - Location and time
  - Contaminant type
  - Duration of impact
- Network model
  - Physical topology
  - Demand characteristics through time
  - Variability in demands

Note: there are major uncertainties in many of this data!

Aggregation Techniques: Skeletonization

In skeletonization, geographically related locations can be combined.

- Commercial codes: H2OMAP, Skelebrator
- Merging pipes, hydraulically equivalent
- Dropping pipes alters hydraultics/flow
- Merging nodes alters demand/impact
- ? Mapping back to all-pipes model?



Witnesses: