SAND2017-8580PE

Sensor Placement for Municipal Water Networks

Jon Berry (Sandia National Laboratories)
Michael Davis (Argonne National Laboratory)
Kate Klise (Sandia National Laboratories)
William Hart (Sandia National Laboratories)
Robert Janke(US EPA)
Regan Murray (US EPA)
Cynthia A. Phillips (Sandia National Laboratories)
Jean-Paul Watson (Sandia National Laboratories)

Sponsored by the US Environmental Protection Agency (EPA) National Homeland Security Research Center (NHSRC)

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.

The Sensor Placement Problem

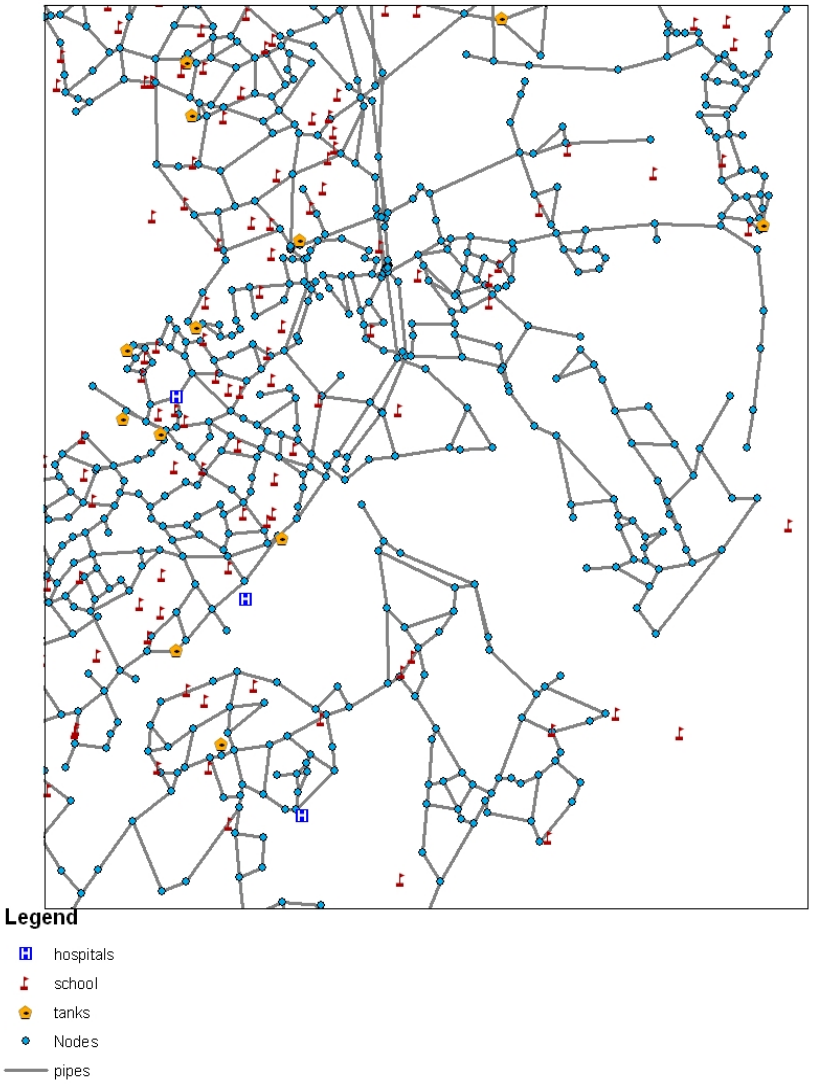
Issue: Contamination released in a municipal water network

Goal: develop early warning system

- Protect human populations
- Limit network remediation costs

Place sensors on

- Utility-owned infrastructure
- Schools
- hospitals



CCR Center for Computing Research

Sandia National Laboratories

Accidental contamination is harmful too

- Outbreak of Cryptosporidium
  - Milwaukee, WI, 1993
- Undetected equipment failure
- 403,000 (est) infected
- > 4400 hospitalized, 69 deaths
- $31.7M in medical costs
- $64.6M lost productivity).
- $90M new water-treatment system.

https://public.health.oregon.gov/LaboratoryServices/imageLibrary/Pages/crypto2.aspx

MacKenzie, et. al. 1994. A massive outbreak in Milwaukee of cryptosporidium infection transmitted through the public water supply. New England J. Medicine 331 161–167.

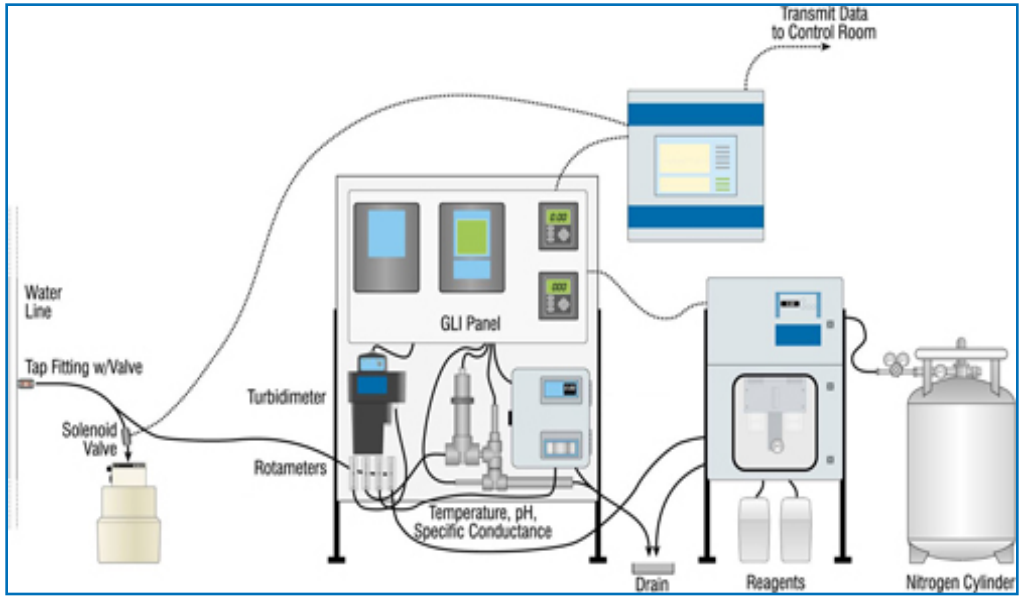
Corso, P. S. et. al. 2003. Cost of illness in the 1993 waterborne Cryptosporidium outbreak, Milwaukee, Wisconsin. Emerging Infectious Diseases 9 426–431.

CCR Center for Computing Research

Sandia National Laboratories

Modeling Assumptions - Sensors

- Sensors are expensive: Cost to buy and install
- Limited number of sensors (sensor budget)
  - Initially assume they are perfect
- Sensors raise a general alarm
  - Wireless communication (e.g. Internet)
  - Can model a response delay

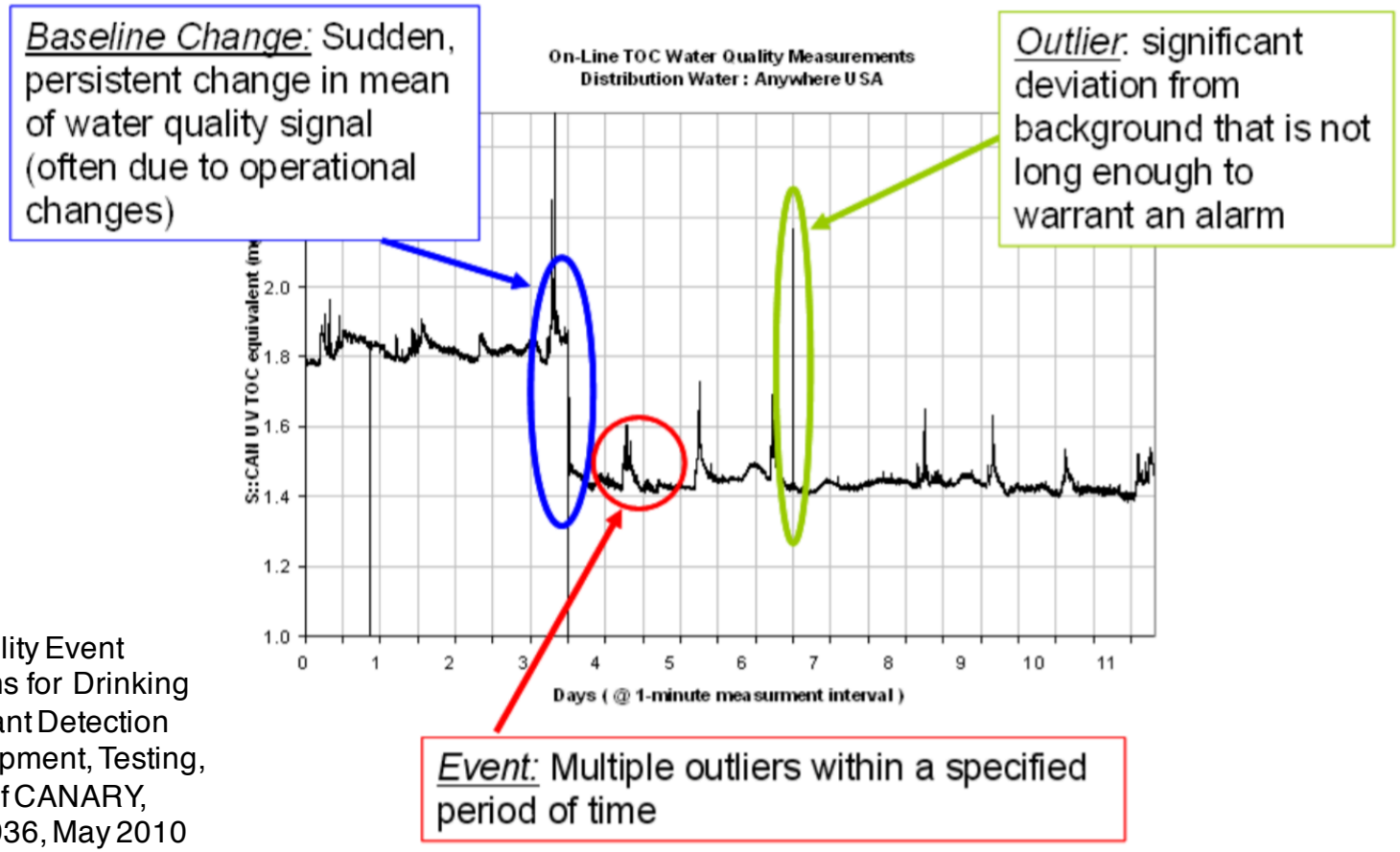


CCR Center for Computing Research

Sandia National Laboratories

Just General Water Quality Sensors

- Measure things like pH, chlorine, electrical conductivity, oxygen-reduction potential, total organic carbon …



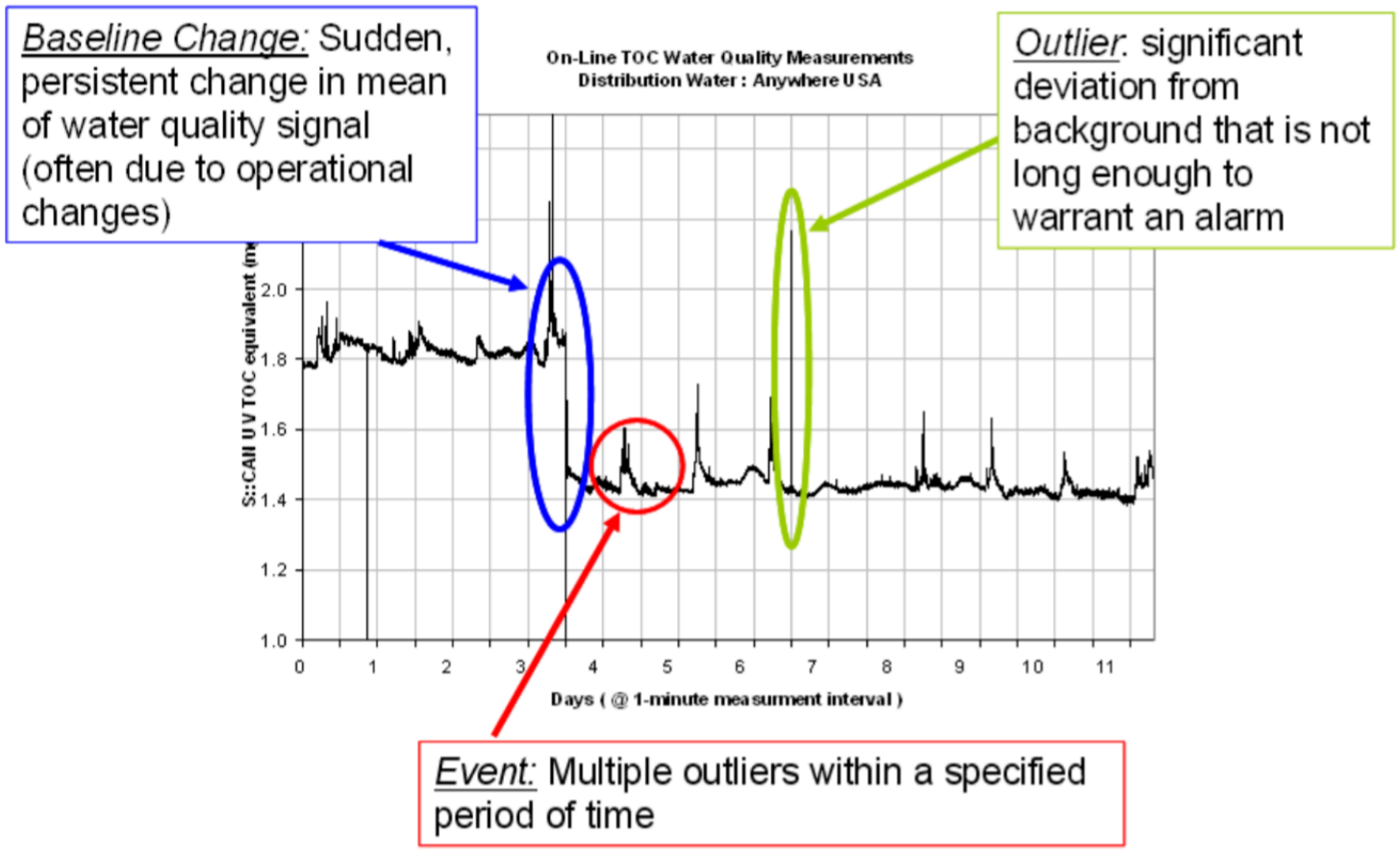
From: Water Quality Event Detection Systems for Drinking Water Contaminant Detection Systems. Development, Testing, and Application of CANARY, EPA/600/R-010/036, May 2010

CCR Center for Computing Research

Sandia National Laboratories

CANARY Event Detection System

- Learns ”normal” value
- Shifting time window



CCR Center for Computing Research

Sandia National Laboratories

Contaminant Transport Modeling

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

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

Contamination plume depends on

- Location of injection
- What is injected
- How long it is injected

Contaminant Transport Modeling

- Demand for water drives water movement
- Assume we know sets of demand patterns for “typical” day
  - Seasonal variations
  - Special events
  - Weekday/weekend

Consider many scenarios at once:
Stochastic program

Current (most trusted) simulator

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

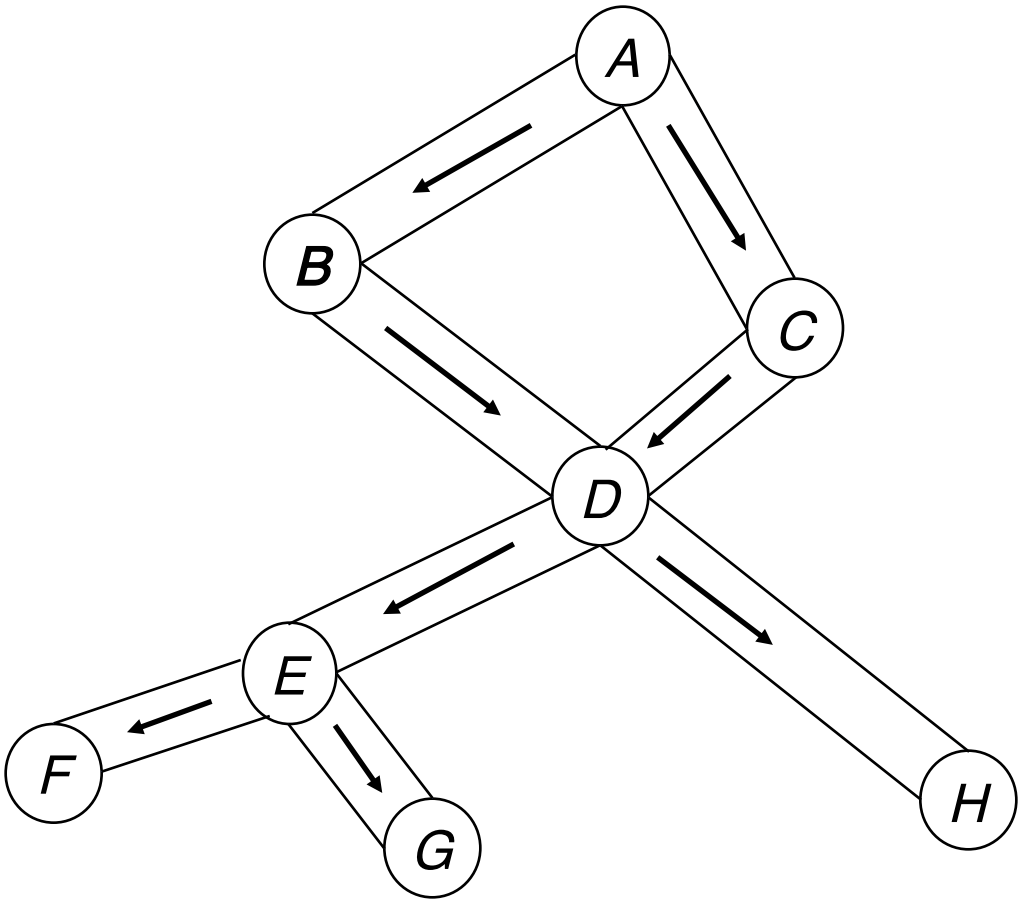
Modeling Events

- Given: Set of events = (location, time) pairs
- Simulate the evolution of a contaminant plume
- For each event determine
  - Where/when event can be observed
  - Amount of damage prior to that observation

- Measures of damage/impact:
  - Population exposed
  - # deaths
  - Volume of contaminant release
  - Total pipe length contaminated
  - Time to detection
  - # failed detections

Simulating contamination transport and Impact

- Total impact: 0



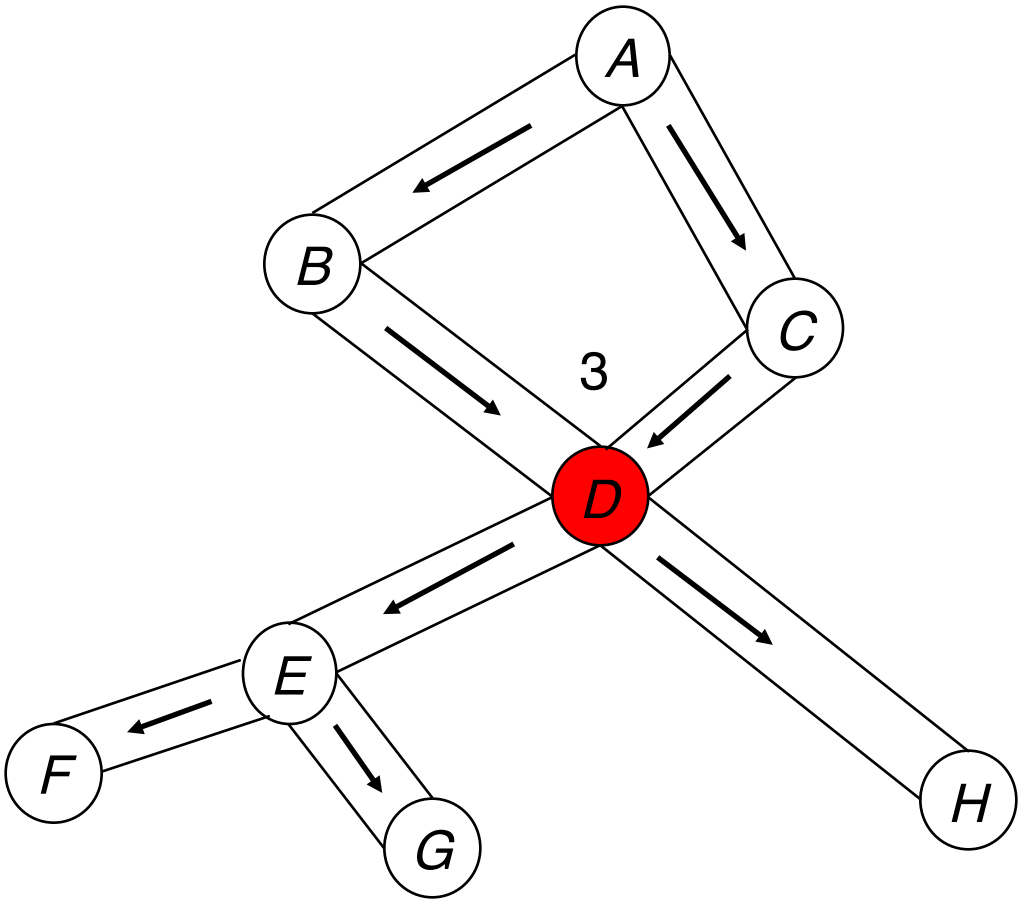
- Nodes can represent neighborhoods (granularity)

CCR Center for Computing Research

Sandia National Laboratories

Simulating contamination transport and Impact

- Total impact: 3



Injection happens over time:

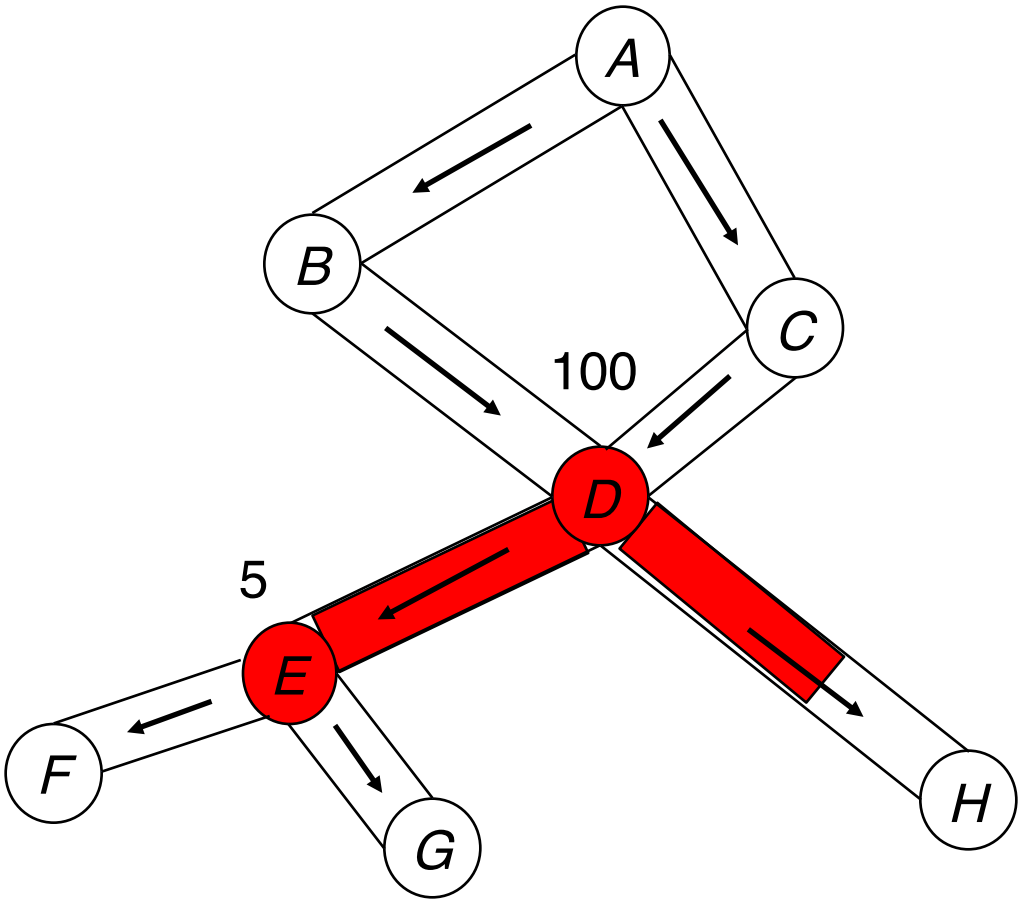
- Contaminant
- Strength
- Volume/rate
- Duration

CCR Center for Computing Research

Sandia National Laboratories

Simulating contamination transport and Impact

- Total impact: 105



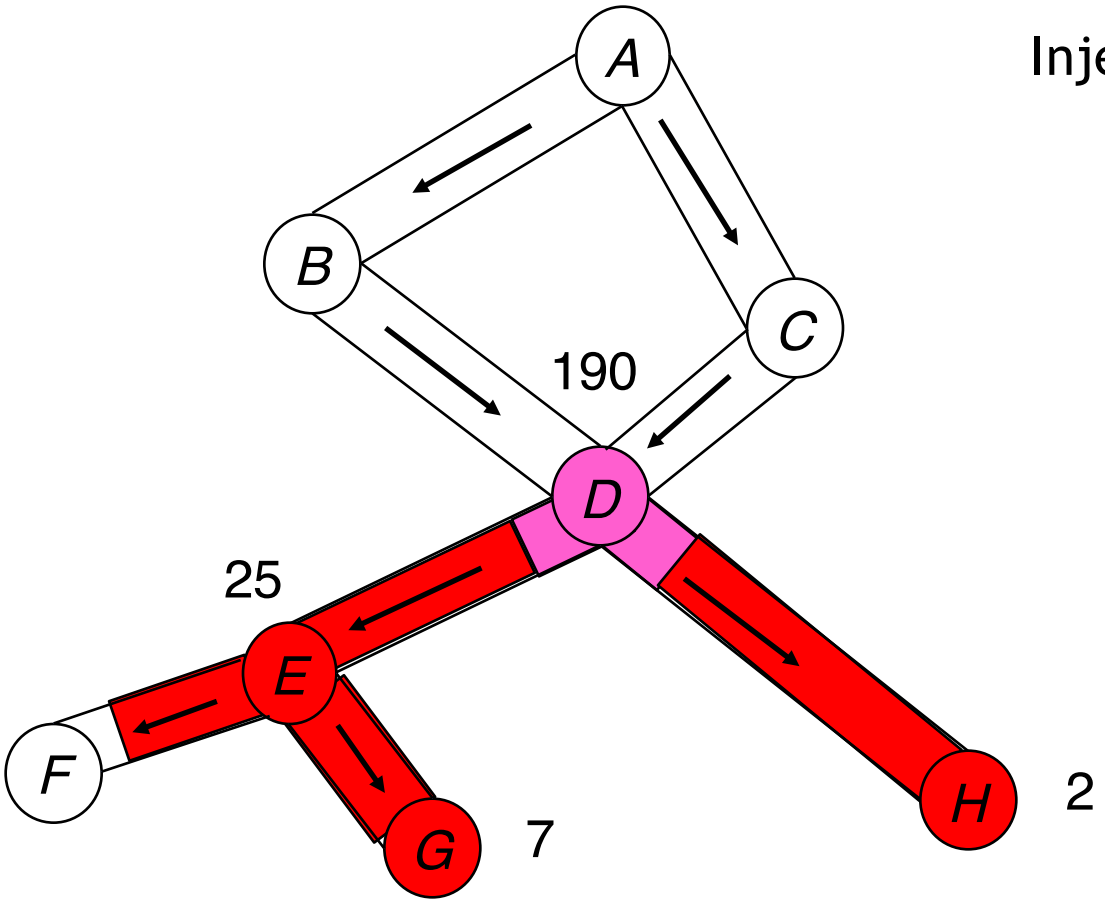
- Speed varies by pipe
- Speed and direction can change over time

CCR Center for Computing Research

Sandia National Laboratories

Simulating contamination transport and Impact

- Total impact: 224



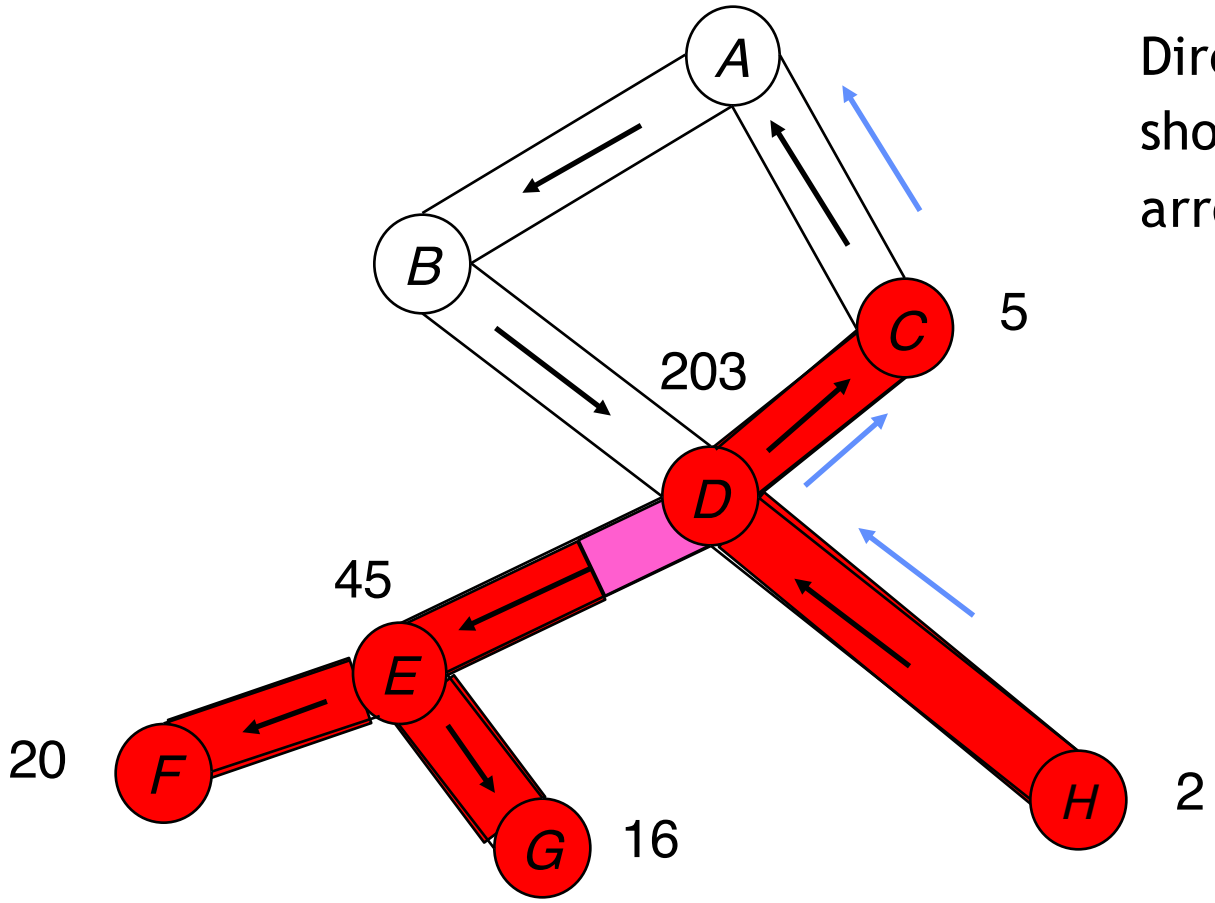
Injection has stopped

CCR Center for Computing Research

Sandia National Laboratories

Simulating contamination transport and Impact

- Total impact: 291



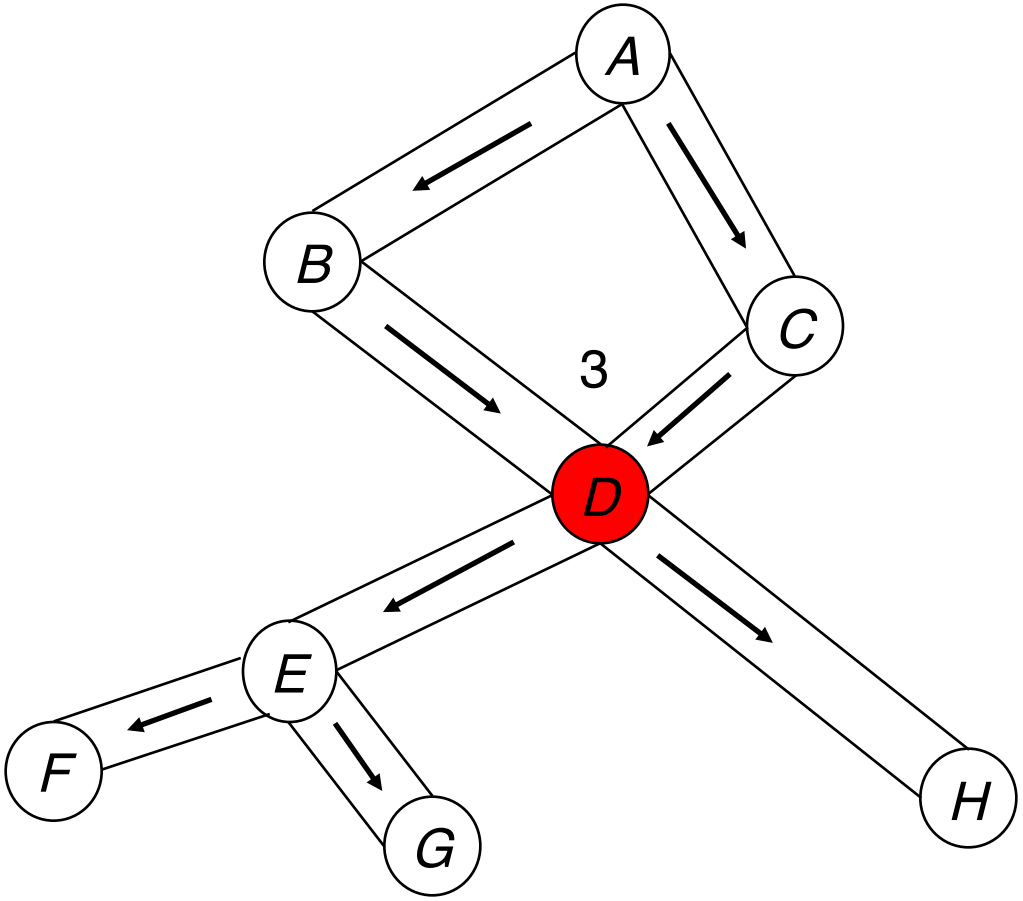
Direction reversal on shown with blue arrows

CCR Center for Computing Research

Sandia National Laboratories

Impact at Sensor Detection, No Delay

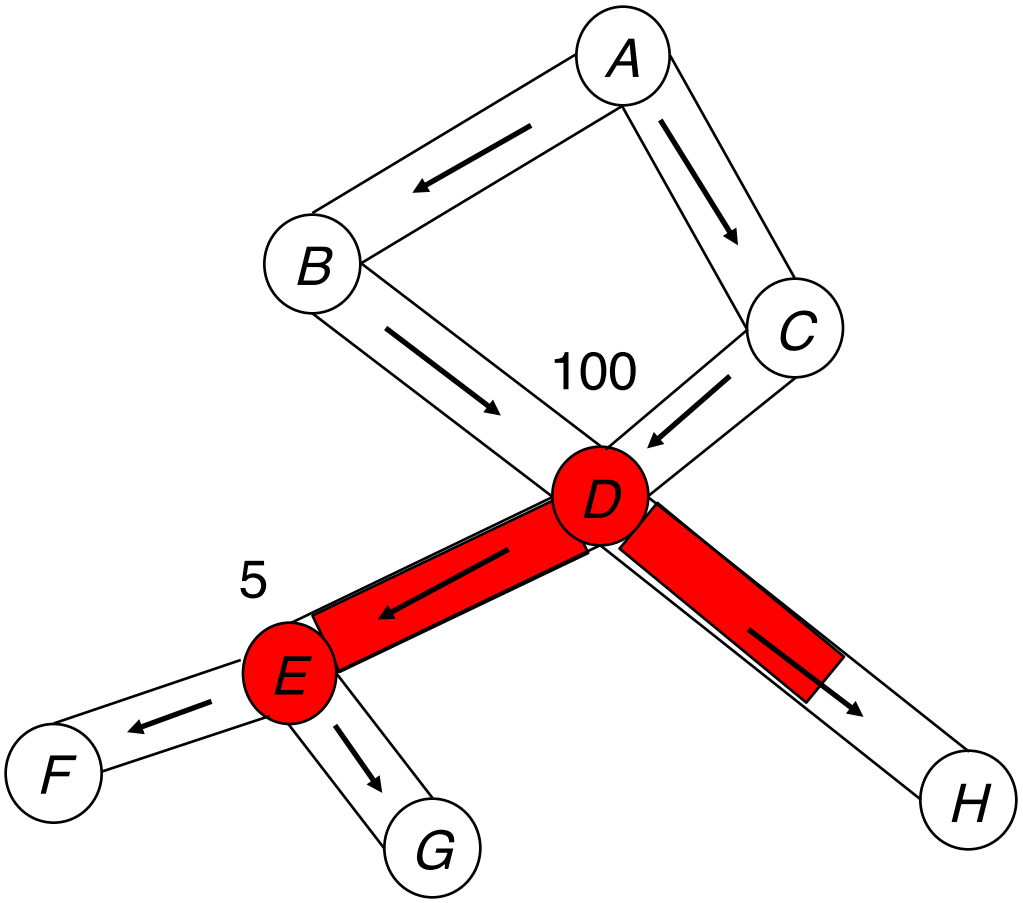
- Total impact: 3



Sensor at:	impact
D	3

Impact at Sensor Detection, No Delay

- Total impact: 105



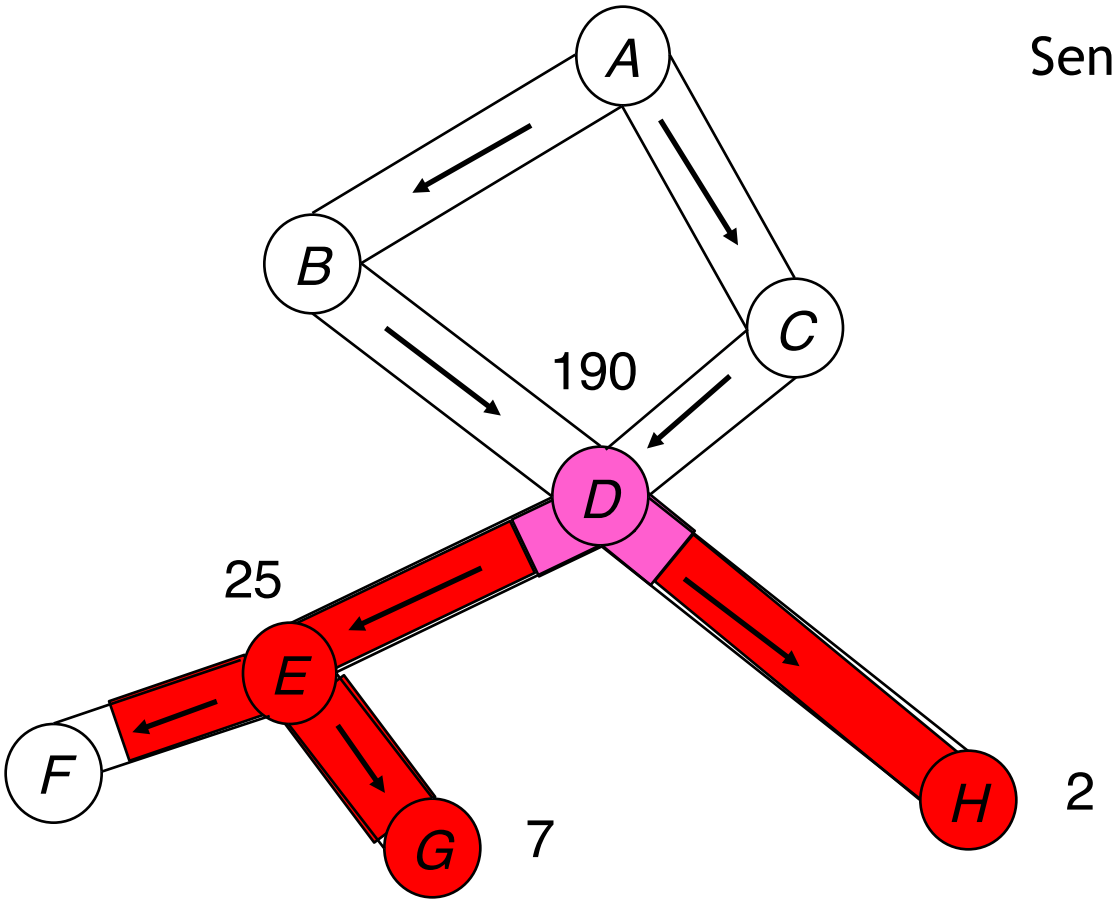
Sensor at:	impact
D	3
E	105

CCR Center for Computing Research

Sandia National Laboratories

Impact at Sensor Detection, No Delay

- Total impact: 224



Sensor at:	impact
D	3
E	105
G	224
H	224

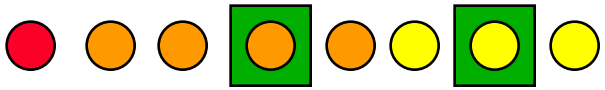
CCR Center for Computing Research

Sandia National Laboratories

Witnessing an Event

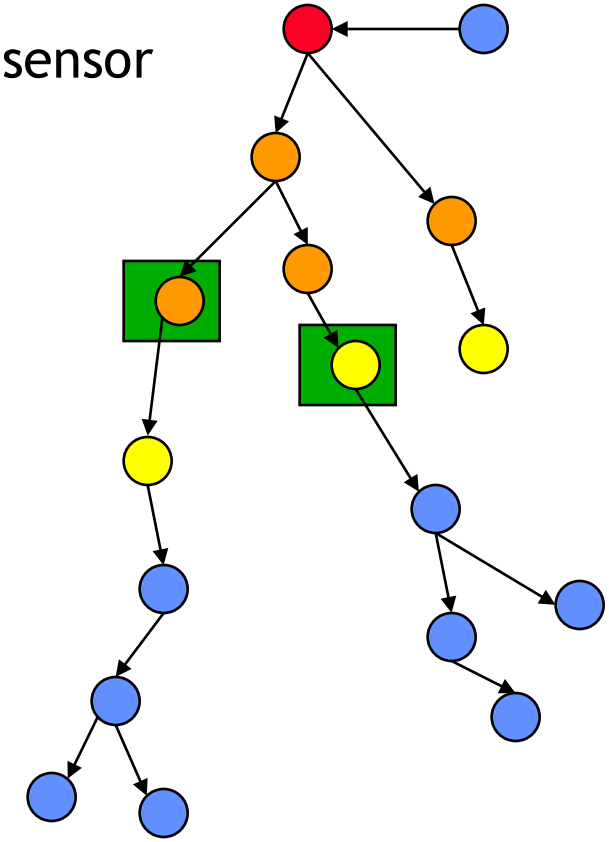
Simulator gives ordered list of nodes where a sensor could witness contamination

Witnesses:



This example has two (green) sensors.

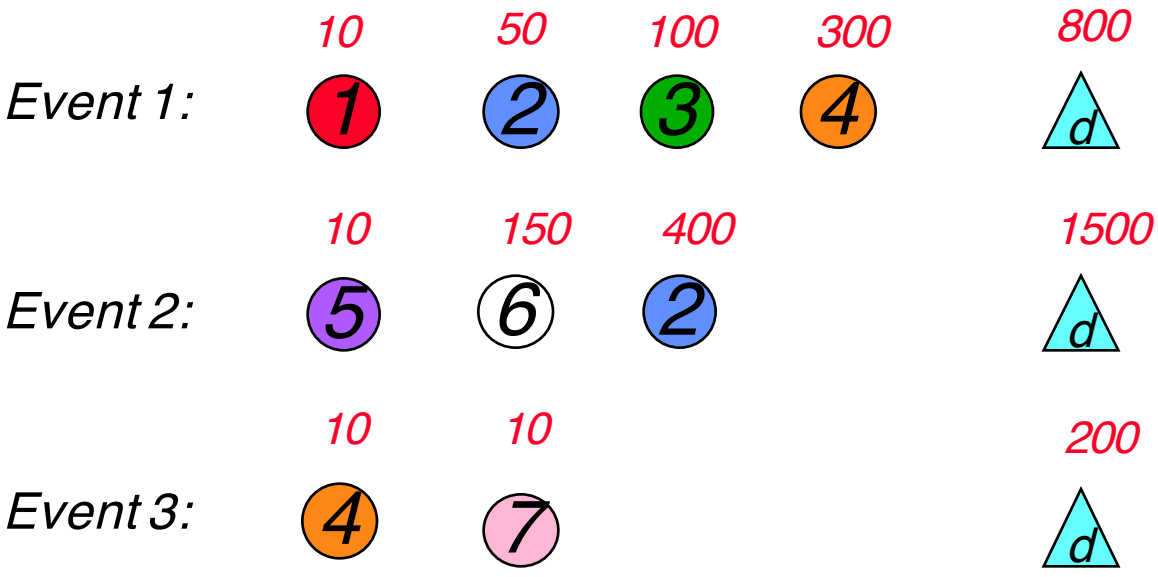
Perfect sensor model: first sensor in list detects the event.



Evaluating a Sensor Placement

- Impact in red

d = dummy node (represents failure to detect)



CCR Center for Computing Research

Sandia National Laboratories

Evaluating a Sensor Placement

- Impact in red

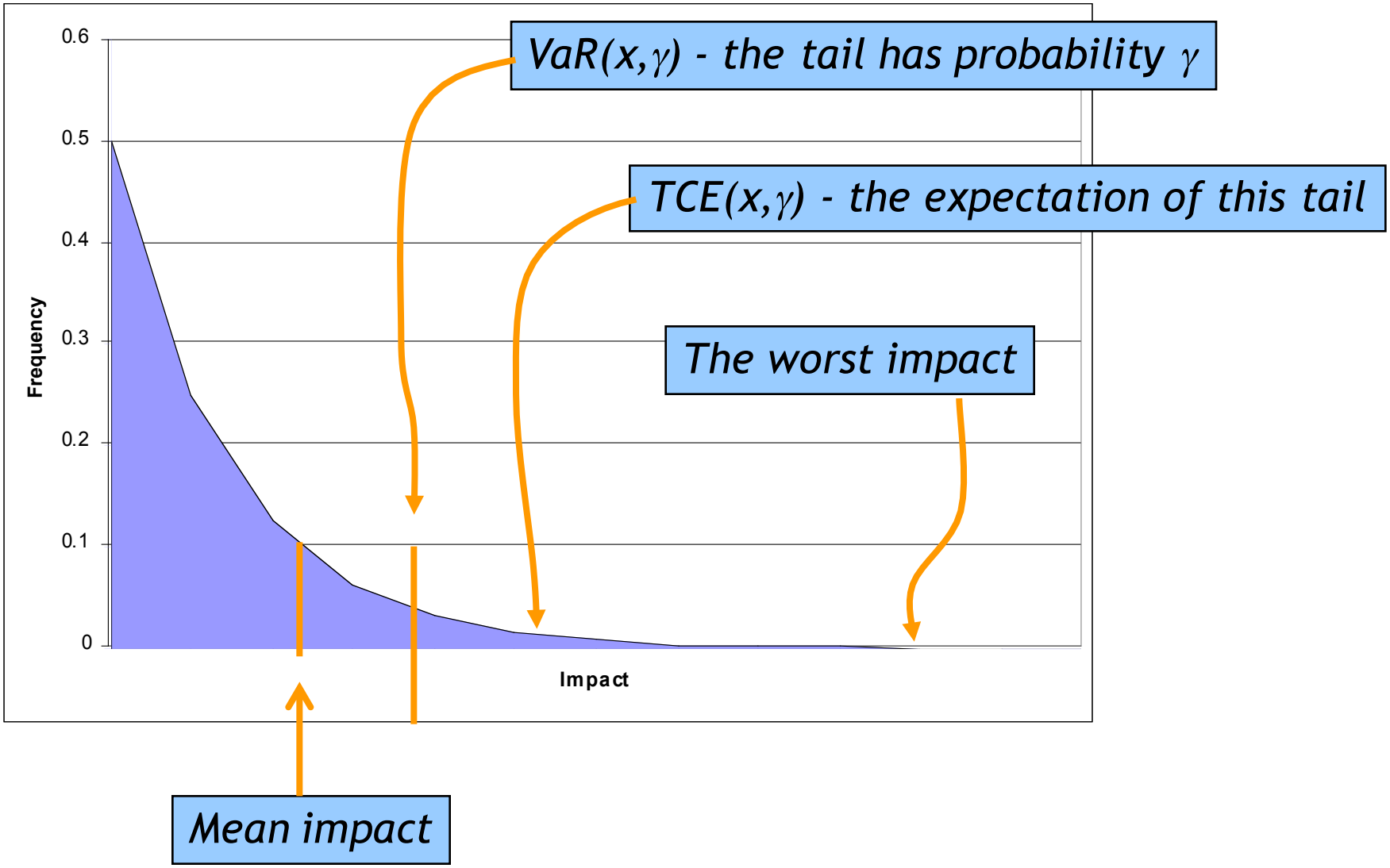
d = dummy node (represents failure to detect)

Choose sensors 2 and 3 (black)

CCR Center for Computing Research

Sandia National Laboratories

Examples of Risk Measures



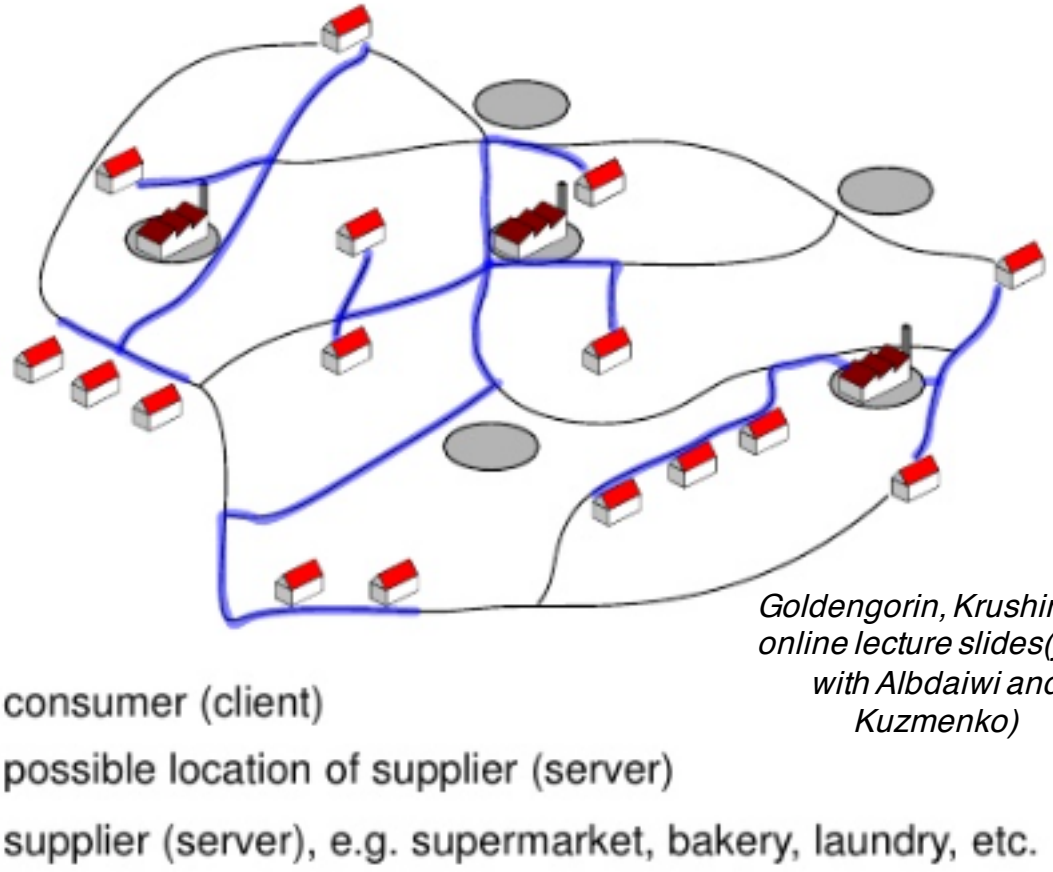
CCR Center for Computing Research

Sandia National Laboratories

Facility Location: p-median

Sensor placement (average impact) as a p-median problem:

- Sensors = Facilities
- Network locations = potential facility locations
- Events = Customers to be "served" (witnessed)
- "Distance= impact.



Goldengorin, Krushinsky online lecture slides(joint with Albdaiwi and Kuzmenko)

- Pick *p* locations and assign each customer to an open location to minimize the total distance.

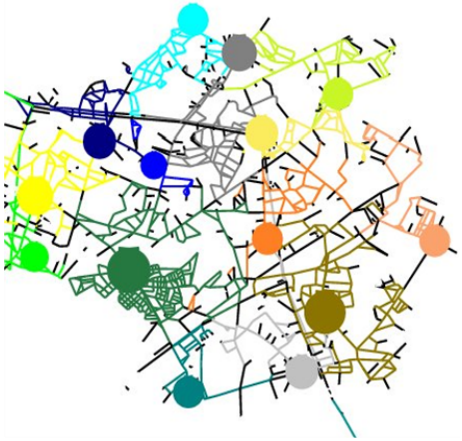
CCR Center for Computing Research

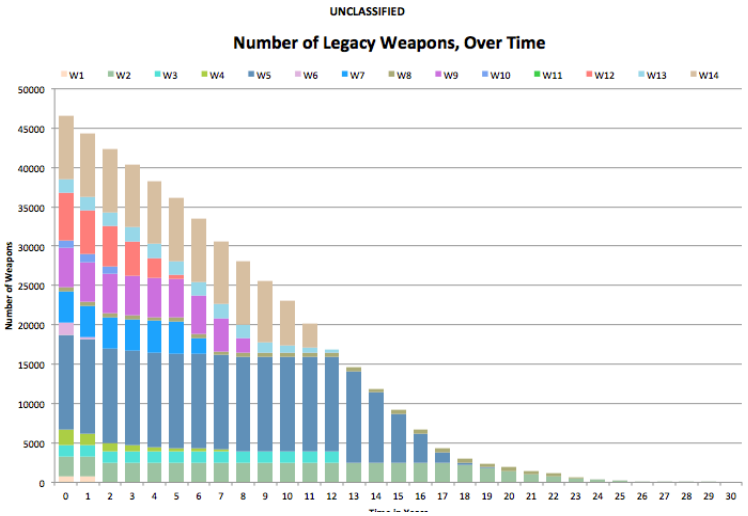
Sandia National Laboratories

Integer Programming

- An expressive, practical type of optimization
- Optimize a linear objective subject to linear constraints
- Variables can take integer values
  - 0 or 1 corresponds to yes/no

- Solve (optimally or within bounds) with intelligent enumeration with commercial or free solvers.





CCR Center for Computing Research

Sandia National Laboratories

Integer Programming

For example:

$$y_i = \begin{cases} 1 & \text{if we place a sensor at location } i \in \mathcal{L}, \\ 0 & \text{Otherwise} \end{cases}$$

$$y_1 + y_2 + \ldots + y_n \leq p$$

enforces sensor budget limit

Other constraints:

- Pick a witness for each scenario
- Allow a location to witness only if it has a sensor

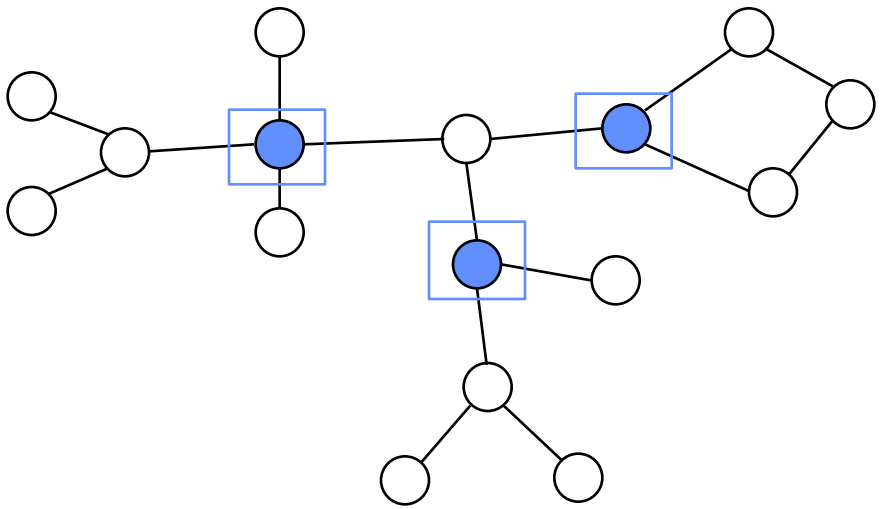
Objective is total impact based on chosen witnesses

CCR
Center for Computing Research

Sandia National Laboratories

Heuristic sensor placement

- Pick starting locations 1 by 1 randomly with greedy weighting
- Hill climbing
  - Move one sensor to a sensor-free location that gives the best reduction in impact
  - Until no move improves
  - Simplified form of GRASP = Greedy Randomized Local Search Procedure



Real Networks for Experiments

- 6 Networks
- Heuristic runtime depends mostly on # nodes

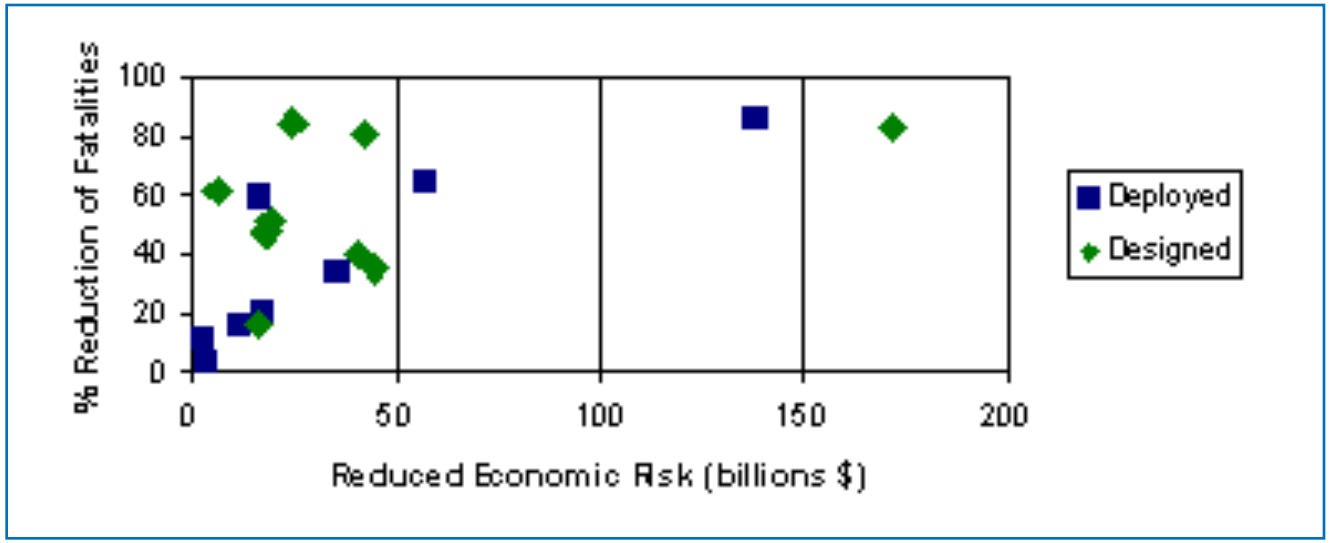
Name	# nodes	# contamination events	#impacts (varies)	impact file size
Net2Morph	3358	1621	1.2M	22M
NetI	6809	6671	4.7M	82M
BWSN	12527	10552	8.2M	156M
NetE	13634	8679	57M	1G
NetB	42698	28675	36M	744M
NetN	48164	9162	38M	772M

minimize mean circa 2005

- System was a bit overloaded, so memory failure may not be fair, but large problems typically do fail for memory.
- 25 sensors, minimizing population exposed

Network Name	heuristic value	heuristic runtime	IP value (opt)	IP runtime
Net2Morph	205.1808	24.44s	205.1808	116s
NetI	33.3841	179s	33.3841	3000s
BWSN	64.0051	531.32s	memory failure	N/A
NetE	41.1494	545.13	memory failure	N/A
NetB	216.3516	5407.97	memory failure	N/A
NetN	969.3954	2235.29	memory failure	N/A

Application to 9 Large Water Utilities



From Edelman competition

- Median **reduction in fatalities** is 48% (4–87%)
- Median **reduction in economic costs** due to lost lives is $19 billion ($3–172B)
- Median **reduction in decontamination and recovery costs** of $29 million ($5–340M)

Mystery

GRASP (greedy heuristic) dominates integer programming for minimizing mean impact

- Almost always gets the optimal solution
- 10x faster
- Lower space

There is no structural reason

It's easy to forget that there is no theory to support optimal results

Imperfect Sensors

- Sensor at location *i* detects with fixed probability p_i
  - Assume independence (well spaced geographically)
- In practice, base on water quality zones

1. Detects contamination with probability p_1
2. Detects contamination with probability $(1 - p_1)\, p_2$
3. Detects contamination with probability $(1 - p_1)\, (1 - p_2)\, p_3$

Assuming there is a sensor in each location

Imperfect Sensors

- Sensor at location *i* detects with fixed probability p_i
  - Assume independence (well spaced geographically)
- In practice, base on water quality zones

10	*50*	*100*	*300*	*800*	*Impact*
.1	*.3*	*.25*	*.5*	*1*	*Raw success probability p_i*
.1	*.27*	*.16*	*.24*	*.23*	*Witness probability if All 4 locations have sensors*

- Witness an event if all sensors that see it first fail, and you succeed
- This becomes a nonlinear optimization

CCR Center for Computing Research

Sandia National Laboratories

One-Imperfect Witness Approximation

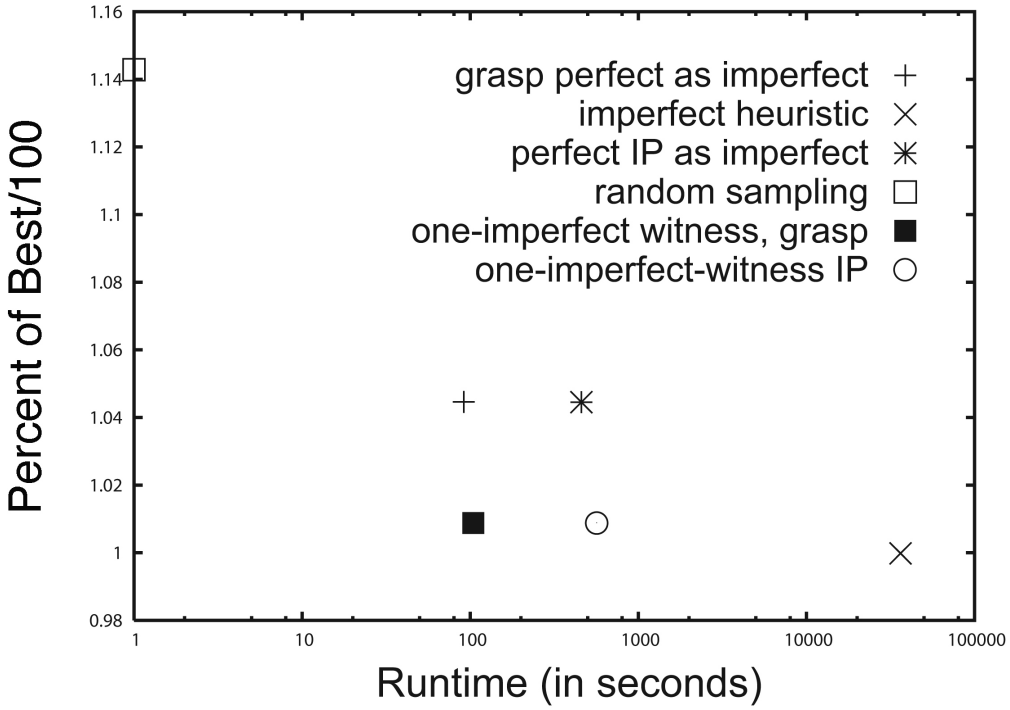
- Sensor a location *i* detects with fixed probability p_i
- Only consider the best sensor for each event
  - No "back up"
- Adjusted impact: $d'_{ai} \rightarrow p_i d_{ai} + (1 - p_i) D_a$,
  where D_a = dummy impact for event *a*

1	2	3	4	d	
10	50	100	300	800	Impact
.1	.3	.25	.5	1	Raw success probability p_i
721	575	625	550	800	One-imperfect-witness impact d′

Some methods for solving w/imperfect sensors

- Ignore imperfection
- Exact linear integer program based on zones
- Nonlinear solver (fractional)
- Local search with imperfect-sensor objective
- Random Sampling
- One-imperfect witness

11,575 nodes
9705 events
40 sensors



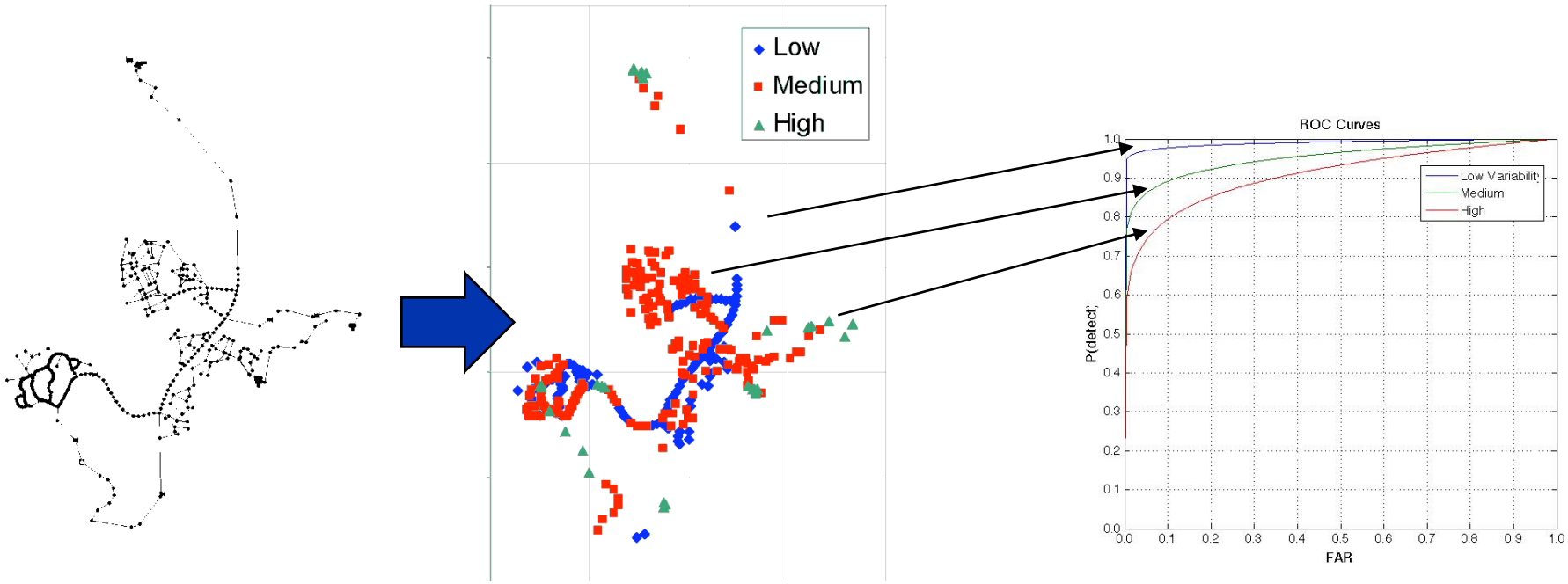
CCR Center for Computing Research

Sandia National Laboratories

A Discrete Tuning Model for Sensors

Water quality variability depends on the sensor location

- Coarse sensor classes
- Characterize junction variability as high/medium/low
- Allow for three sensor tuning levels for each variability type
- Determines both failure probabilities and conditional delay

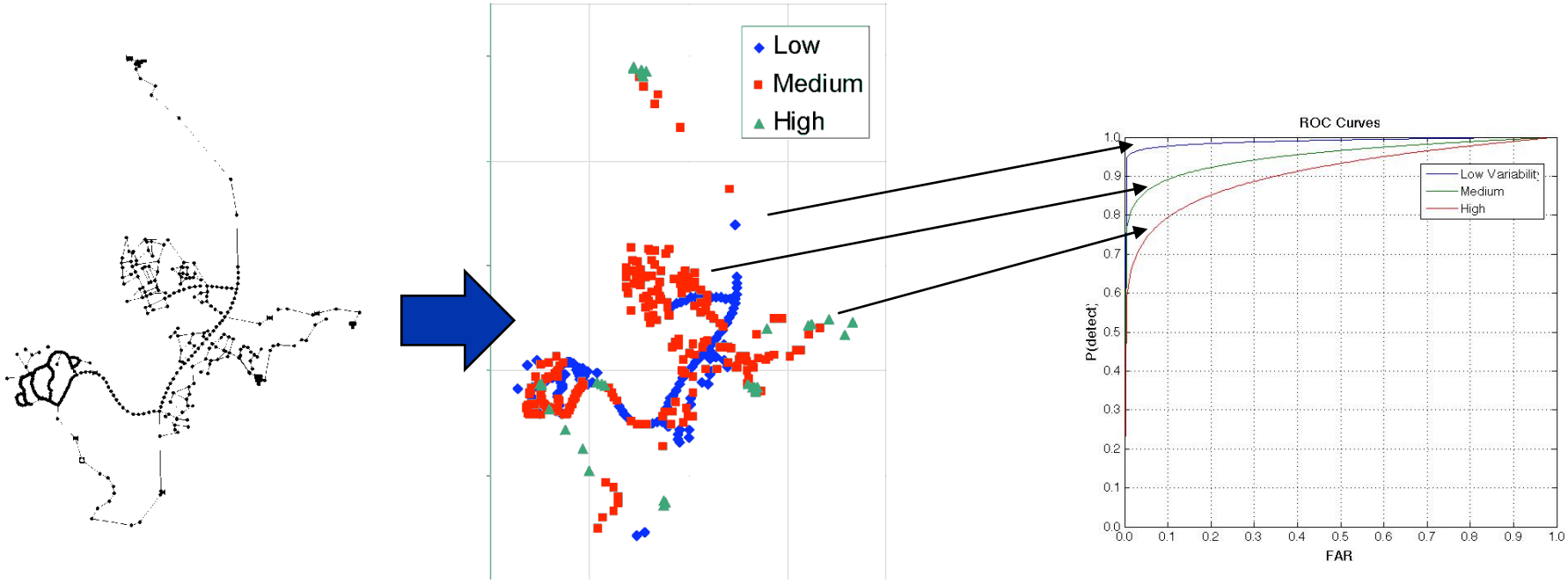


CCR Center for Computing Research

Sandia National Laboratories

A Discrete Tuning Model for Sensors

- Can modify the one-imperfect witness formulation to allow tuning
- Solve with an integer program to enforce false positive limit
- Results: When the false alarm tolerance is too low (say, 0.5/week) we cannot usually use even 10 or 20 sensors



CCR Center for Computing Research

Sandia National Laboratories

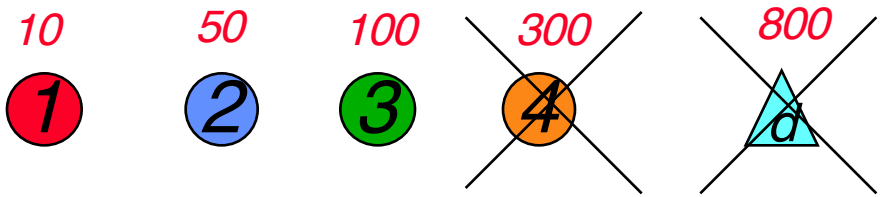
New Objective: Minimize Worst-Case Impact

Can use GRASP with a different objective

But now there's a better wa:

Find minimum number of sensors to achieve worst case impact W

- For each scenario, remove all witnesses with impact > W

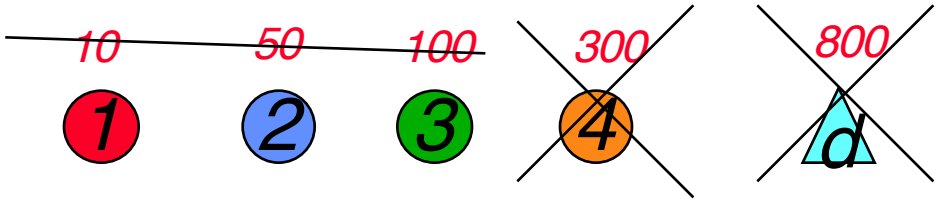


  - If none left, W infeasible for any # sensors

Minimize Worst Case

Find minimum number of sensors to achieve worst case impact W

- For each scenario, remove all witnesses with impact > W



  - If none left, W infeasible for any # sensors
- Ignore impact value on remaining locations, just a set

Set Cover:

- A set of sensor locations covers a scenario if at least one feasible location selected
- Find the smallest covering set using IP

Results - minimizing worst case

Name	Heuristic Value	Heuristic Runtime	Binary search value (opt)	Binary Search Runtime	% Error
$Net2Morph_b$	861	1131s	852	61s	1
$Net2Morph_w$	1166	778s	1059	62s	10
Net1 (best)	897	16,072s	890	355s	0.8
Net1 (worst)	570	5731s	518	330s	12
BWSN (best)	1037	33,040s	1037	844s	0
BWSN (worst)	1092	26,339s	980	816s	11
NetE	1070	47,792s	1025	2508s	4.4
NetB	8472	666,622s	8320	8600s	1.8
NetN	7286	358,697s	6851	4186s	6.3

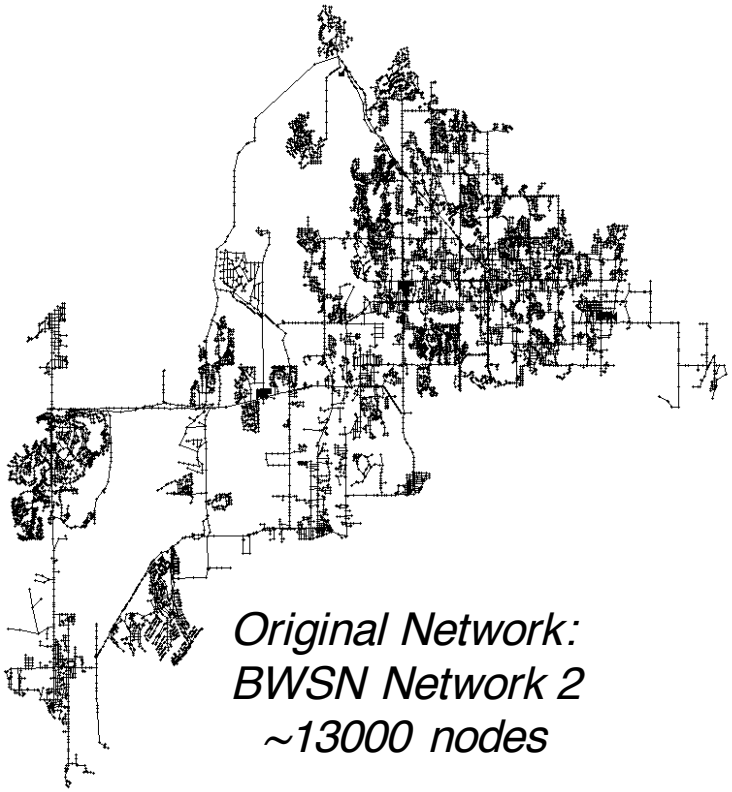
Best(b)/worst(w): out of various impact objectives (toxicities)

Major Issue: Water Utilities Resources

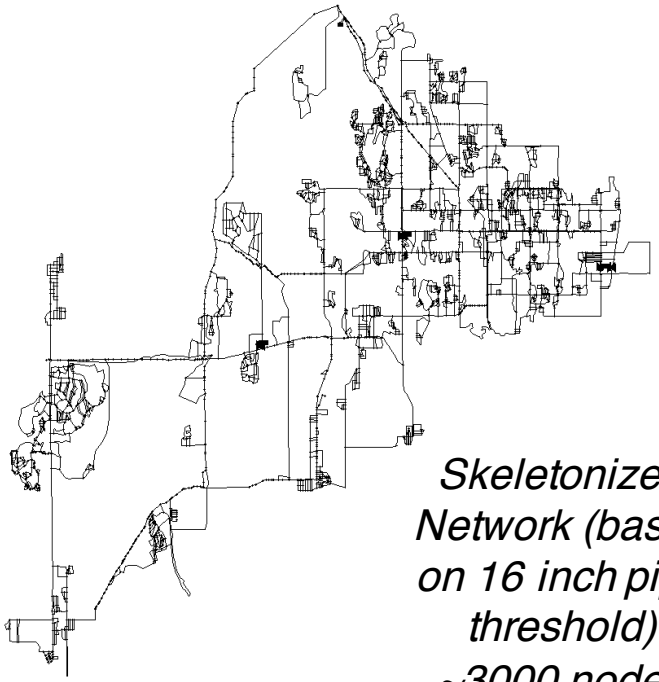
- Very small computing resources
  - Run in small space (before the cloud)

- One method: Lagrangian Solver
  - Move event coverage to a penalty term
  - Linear space
  - Gives fractional solution -> round
  - Excellent lower bounds (95-99%)

Skeletonized Networks



Original Network: BWSN Network 2 ~13000 nodes



Skeletonized Network (based on 16 inch pipe threshold) ~3000 nodes

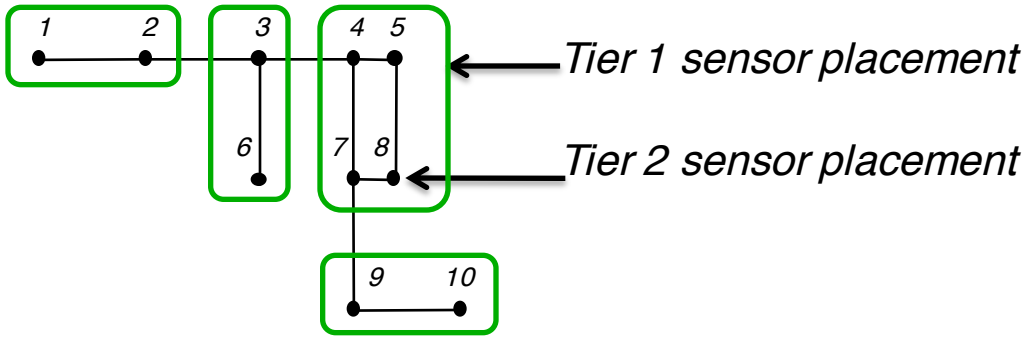
- Skeletonization reduces space
  - Doesn't match hydraulics
  - Picks supernodes instead of real nodes

CCR Center for Computing Research

Sandia National Laboratories

Space Issues

- Two-tiered process
- Use hydraulics from full model



Method and Solver		Memory Footprint (GB)	Mean Impact	MIRE
Two-tiered sensor placement	GRASP/ GRASP	5.3	80.1	5.0%
	Lagrangian /GRASP	2.4	83.5	9.4%

- Network2, 10 sensors, high toxicity
- Skeletonization only had relative error of 200%

CCR Center for Computing Research

Sandia National Laboratories

Some Final Thoughts

- Model requires only a list of witnesses and impacts for each event
- Model is stable as simulator (EPANET) improves

- Same basic model works in other settings
  - Detecting airborne contaminants
  - Placing gas detectors in chemical plants
  - Blog watching

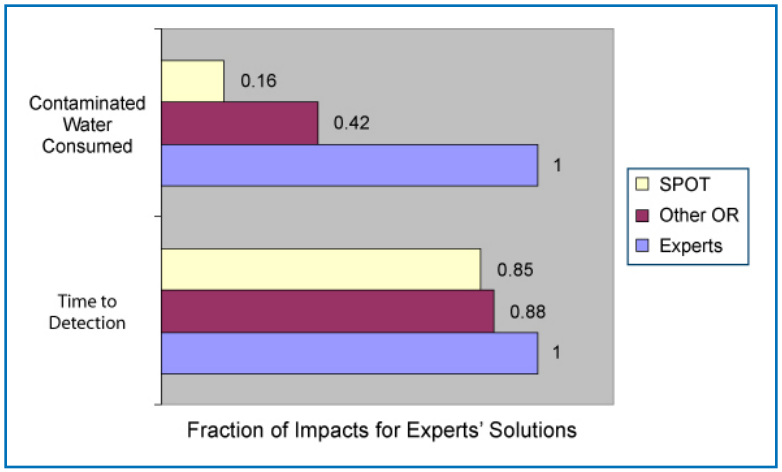
- Codes are openly available
  - TEVA-SPOT: Threat-Ensemble Vulnerability Analysis-Sensor Placement Toolkit, now part of Water Security Toolkit (WST)
  - CANARY

- More recent work has been on response to contamination

More Final Thoughts

- Battle of the Water Sensor Networks
  - Showed that (for p-median model), codes do better than experts
  - No declared winner. Multiple interpretations
  - Lesson: design such things carefully: GRAPH500, benchmarks

- The ASCE/EWRI/water community is very welcoming of new ideas
  - Our entre to this area was designed by/for computer scientists
  - It took years to deprecate it



CCR Center for Computing Research

Sandia National Laboratories