Robust Combinatorial Optimization Formulations for Sensor Placement in Water Networks

William Hart
Sandia Labs

Collaborators at Sandia Labs, CU Denver, UC Berkley, U New Mexico, ASU, EPA/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.

Overview

- What type of robust optimization techniques can be used with general MILP formulations?
- Driver application: sensor placement in water distribution networks
- Key issue: how can we bound risk while maximizing performance?

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

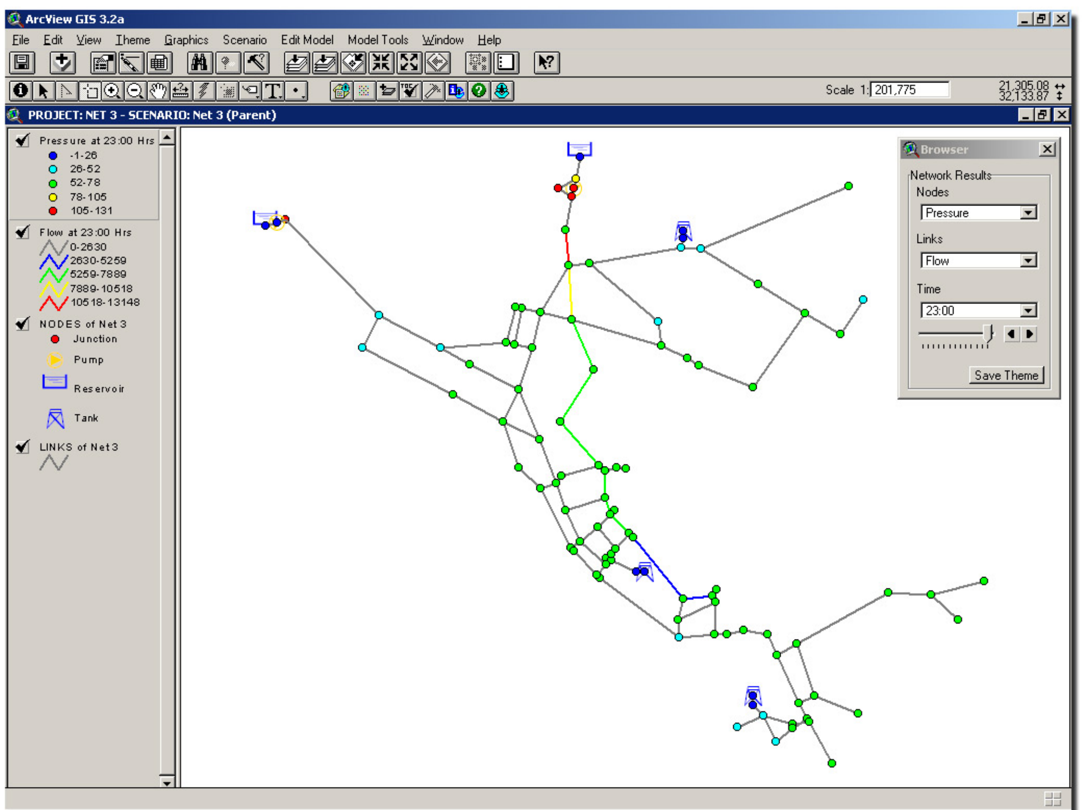
What is a water distribution network?

<u>Drinking Water</u>

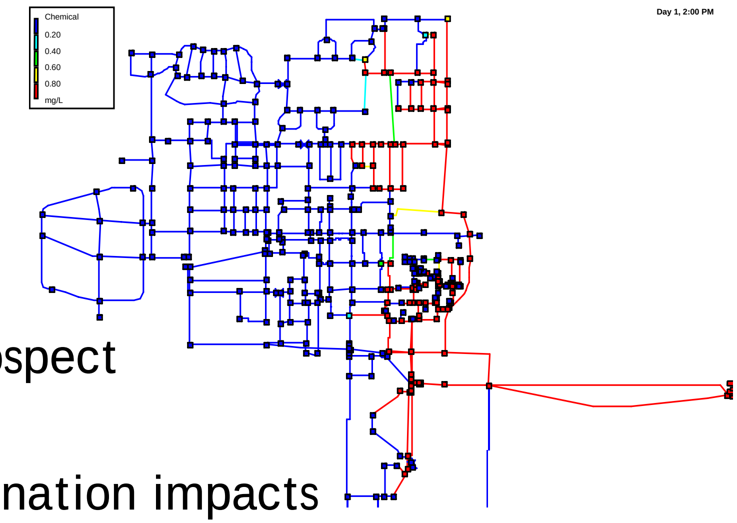
- Water source
- Treatment facilities
- Transmission systems
- Distribution systems

<u>Wastewater</u>

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



A Motivating Threat Scenario



Contaminant Injection

Risk: moderate-high

- Technically difficult to accomplish
- Potential terrorists fascinated by this prospect

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…

What data do we need for sensor placement?

- Population consumption
  - Location and time
  - Individual characteristics: health, age
- Attack 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!

Integer Programming for Sensor Placement

IPs can be used to model sensor placement for water security

- Berry et al (2003, 2004); Watson et al (2004)

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

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

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

IP model:

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

Two RO Approaches

Interval Uncertainties

- Data lies within a specified interval about an estimated value
- Formulate models that find the best solution for all possible uncertain values within the interval
- Consider side-constraints that exploit additional knowledge about the uncertainties

Attack Location Uncertainties

- Contamination impact varies with attack location
- Formulate models that minimize or constrain the risk of a *catastrophic* attack
- Consider different notions of risk: worst-case, VaR, CVaR

Note: Our focus is on developing robust MILP models

Interval Uncertainties

Idea: consider uncertainties as simple errors on given numerical estimates

$$\min_{x \in X} \max_{c \in U(c_L, c_U)} c^T x$$

where

$$X \text{ is the set of feasible integer points}$$

$$U(c_L, c_U) = \{c : c_L \le c \le c_U\}$$

Note: This RO problem can be trivially reformulated as an IP model!

Bounded Interval Uncertainties

Example: population estimates

- The total population is probably well-known
- The estimated population at any given site may have considerable error

Modeling interval uncertainties as a sum-restricted ball

Let δ be the population estimate and let $\hat{\delta}$ be the true population.

Let Δ be the total population.

We assume that $\Delta = \sum_i \delta_i = \sum_i \hat{\delta}_i$.

We assume that $\delta_L \leq \delta \leq \delta_U$.

Note: this model of uncertainty works for various model parameters

- attack probabilities, consumption demands, etc.

Optimization with Bounded Uncertainties

Idea: minimize the worst case value with bounded uncertainty

$$\min_{x \in X} \max_{c \in B(\hat{c}, \varepsilon)} c^T x$$

where

$$B(\hat{c}, \varepsilon) = \left\{ c : (1-\varepsilon)\hat{c} \le c \le (1+\varepsilon)\hat{c}, \sum_k c_k = \sum_k \hat{c}_k \right\}$$

We consider proportional uncertainty intervals to simplify our presentation…

There are two cases that reflect different applications/models

1. Linearly weighted uncertainty
2. Bilinear uncertainty

Linearly Weighted Uncertainty

Example: minimizing extent of contamination (EC)

- Objective: $\sum_{a \in A} \sum_{i \in L} \alpha_a w_{ai} x_{ai}$
- w_{ai} is the pipe length contaminated, which is well-known
- α_a is uncertain

RO formulation with bounded interval uncertainty

$$\min_{x \in X} \boxed{\max_{\alpha \in B(\hat{\alpha}, \varepsilon)} \sum_{a \in A} \sum_{i \in L} \alpha_a w_{ai} x_{ai}}$$

Idea: reformulate this RO formulation using the dual of this subproblem

Sandia National Laboratories

Solving Linearly Weighted RO Problems

The optimal solution for EC can be obtained by solving the following IP:

(LWRO)

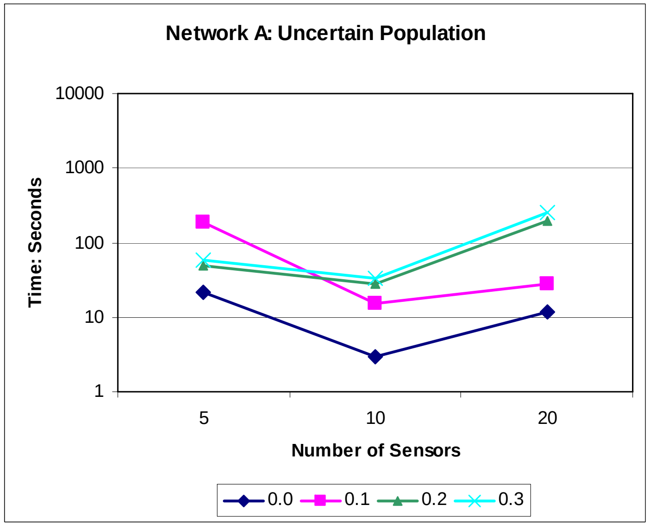
$$
\begin{aligned}
\min \quad & \sum_{a\in A}\sum_{i\in L}\hat{\alpha}_a w_{ai}x_{ai} + \varepsilon\hat{\alpha}(\gamma-\mu) \\
\text{s.t.} \quad & \pi + \mu_a - \gamma_a = \sum_i w_{ai}x_{ai} \\
& \mu,\gamma \geq 0 \\
& x \in X
\end{aligned}
$$

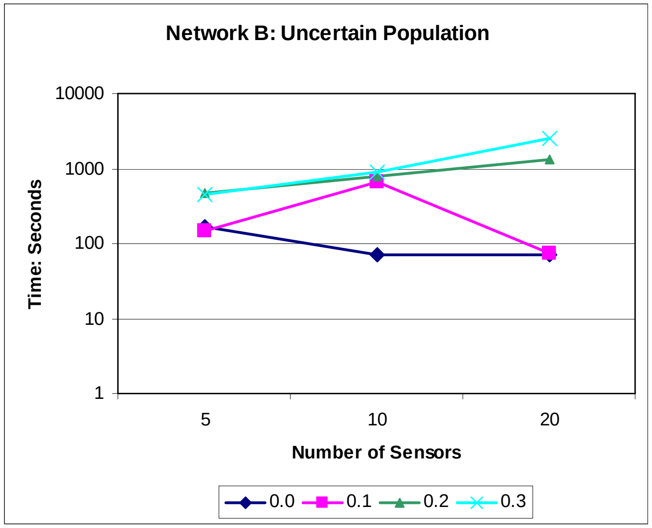
Interval uncertainty can be addressed by solving a related MILP!

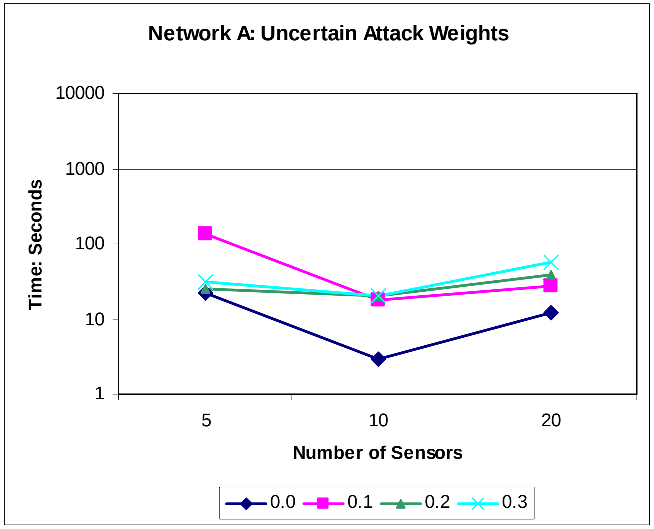
Preliminary Experiments with LWRO

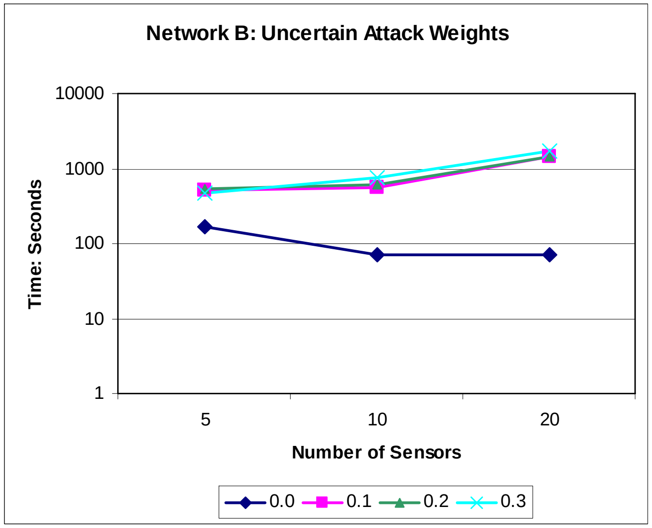
Network A: 97 junctions, 234 pipes

Network B: 470 junctions, 1198 pipes









A Special Case: Unweighted Uncertainties

Example: minimizing the number of failed detections (NF)

- Objective: $\sum_{a \in A} \sum_{i \in L} \alpha_a x_{ai}$
- α_a is uncertain

Theorem: The optimal solution to the problem

$$\min_{x \in X} \max_{\alpha \in B(\hat{\alpha}, \varepsilon)} \sum_{a \in A} \sum_{i \in L} \alpha_a x_{ai}$$

is the optimal solution to the problem

$$\min_{x \in X} \sum_{a \in A} \sum_{i \in L} \alpha_a x_{ai}$$

for all $\varepsilon > 0$.

Conclusion: proportional interval uncertainty does not impact this formulation!

Bilinearly Weighted Uncertainty

Example: minimizing the population exposed (PE)

- Objective: $\sum_{a \in A} \sum_{i \in L} \alpha_a w_{ai} x_{ai}$
- w_{ai} is the population exposed to contaminant, which is uncertain
- α_a is uncertain

RO formulation with bounded interval uncertainty

$$\min_{x \in X} \boxed{\max_{\substack{\alpha \in B(\hat{\alpha}, \varepsilon) \\ w \in B(\hat{w}, \varepsilon)}} \sum_{a \in A} \sum_{i \in L} \alpha_a w_{ai} x_{ai}}$$

This is NOT a MILP.

Note: even solving this bilinear maximization problem is hard!

Solving Bilinearly Weighted RO Problems

1. Consider the impact of uncertainties for α_a and w_a independently (using the linearly weighted technique)

2. Sample one of the uncertainties and solve a problem of the form

$$\min_{x \in X} \max_{\alpha \in \Xi} \max_{w \in B(\hat{w},\varepsilon)} \sum_{a \in A} \sum_{i \in L} \alpha_a w_{ai} x_{ai}$$

   This can be reformulated as a MILP.

3. Build a custom MINLP solver
   - Linearize the subproblem problem using standard techniques (McCormick's linear relaxations)
   - Reformulate as a dual to provide an overall lower bound
   - Branch on choice variables and uncertainty regions

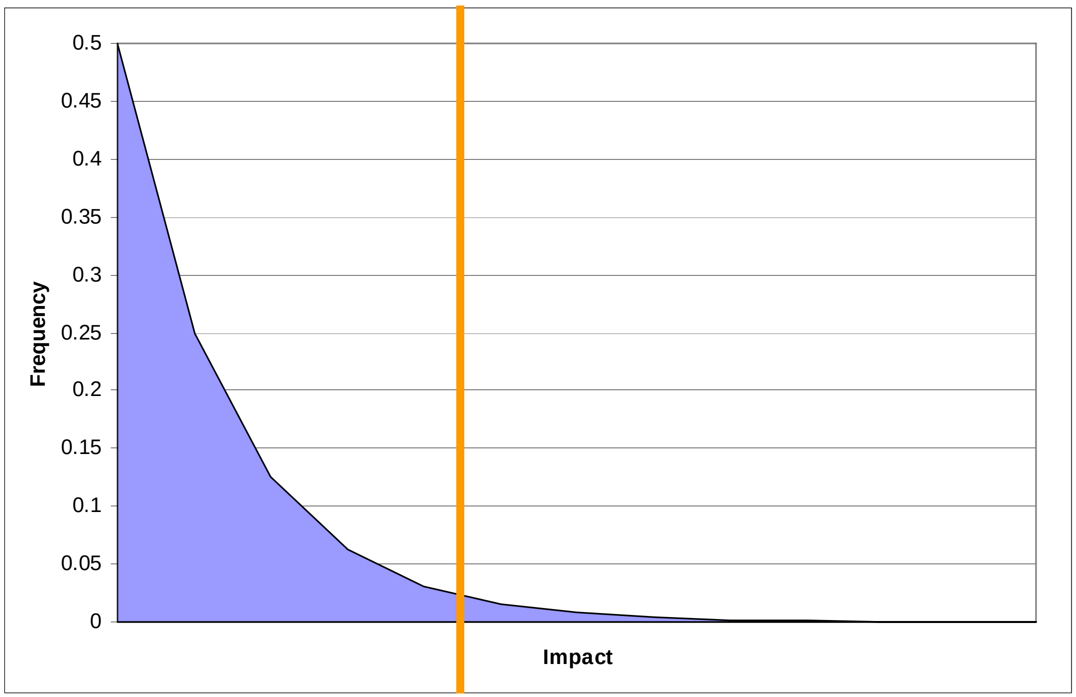
Solving Bilinearly Weighted Problems (cont'd)

4. Heuristically search *X* and evaluate solutions using the bilinear subproblem.
   - May not be computationally feasible
   - It is NP-hard to solve this subproblem
   - Can compute approximation values (using McCormick bounds)

Attack Location Uncertainties

Idea: consider the risk of a catastrophic attack

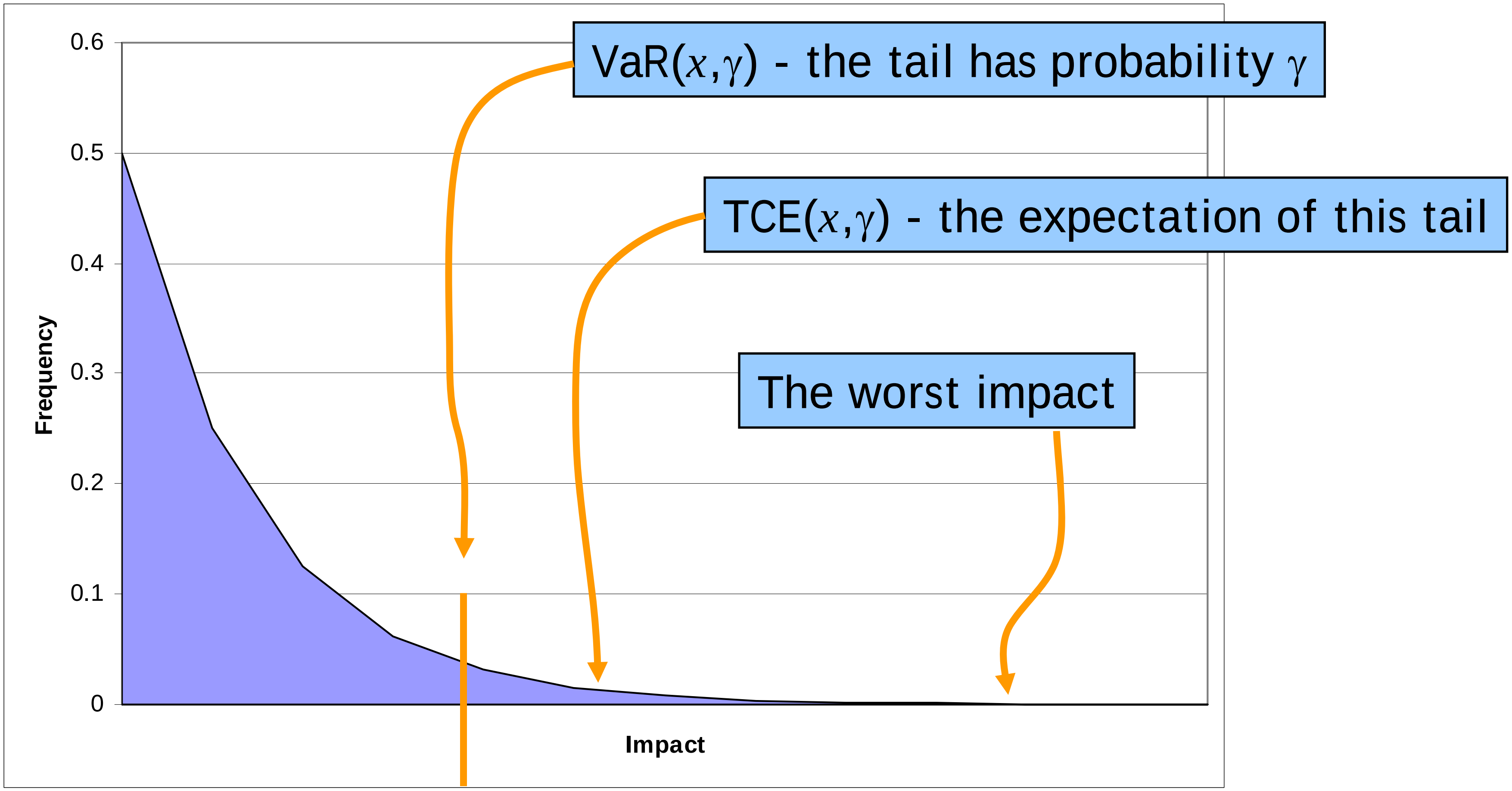
Goal: limit the risk that the worst impacts have a big effect



Risk Measures of Interest

- Worst-Case
  - The value of the worst impact
- Value-at-Risk (VaR)
  - VaR(x,γ) is the value of the $1-\gamma$ quantile of the impact distribution
- Conditional Value at Risk
  - Tail Conditional Expectation (TCE) is the average loss in worst 100γ percent of the impact distribution
  - CVaR(x,γ) is an approximation to TCE

Examples of Risk Measures



Sandia National Laboratories

Defining CVaR

$$\text{Let } \Lambda_{\mathrm{i}} = \begin{cases} 1 & \text{if } \sum_i w_{ai} x_{ai} \geq \mathrm{VaR}(x,\gamma) \\ 0 & \textit{otherwise} \end{cases}$$

$$\text{Let } y_i = \max\left\{0, \sum_i w_{ai} x_{ai} - \mathrm{VaR}(x,\gamma)\right\}$$

$$\begin{aligned} \mathrm{TCE}(x,\gamma) &= \frac{\sum_a \alpha_a \left(\sum_i w_{ai} x_{ai}\right) \Lambda_a}{\sum_a \alpha_a \Lambda_{\mathrm{a}}} \\ &= \mathrm{VaR}(x,\gamma) + \frac{\sum_a \alpha_a \left(\sum_i w_{ai} x_{ai} - \mathrm{VaR}(x,\gamma)\right) \Lambda_{\mathrm{a}}}{\sum_a \alpha_a \Lambda_{\mathrm{a}}} \\ &= \mathrm{VaR}(x,\gamma) + \frac{\sum_a \alpha_a y_a}{\sum_a \alpha_a \Lambda_a} \end{aligned}$$

Now $\sum_a \alpha_a \Lambda_a \approx \gamma$, so we can approximate $\mathrm{TCE}(x,\gamma)$ with

$$\mathrm{CVaR}(x,\gamma) = \mathrm{VaR}(x,\gamma) + \frac{\sum_a \alpha_a y_a}{\gamma}$$

Risk-Constrained MILPs

Goal: Minimize expected impact while constraining the risk of a catastrophic attack

Worst-case MILPs are easy to formulate

VaR-constrained formulations are messy

- Can formulate with a simple quadratic constraint
- Can use binary indicator variables in a MILP formulation

CVaR-constrained formulation…

CVaR-Constrained Formulation

Let Ω be a bound on CVaR

The following formulation bounds CVaR

$$\begin{array}{ll} \min & \sum_a \sum_i \alpha_a w_{ai} x_{ai} \\ & v + \frac{1}{\gamma} \sum_a \alpha_a y_a \leq \Omega \\ & y_a \geq \sum_i w_{ai} x_{ai} - v \\ & y_a \geq 0 \\ & x \in X \end{array}$$

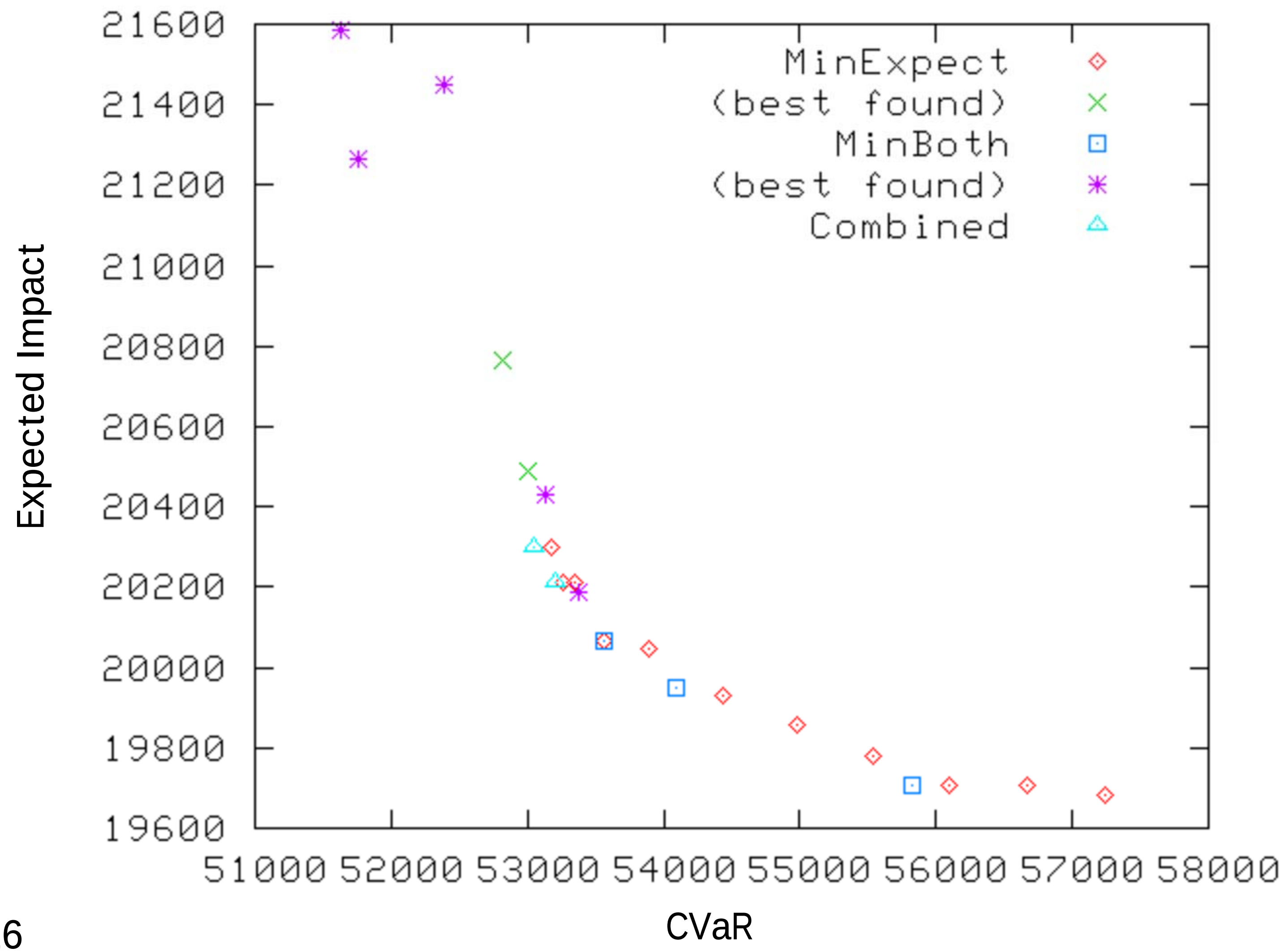
Note: we can include CVaR in both the objective as well!

Note: v is VaR when the CVaR constraint is binding

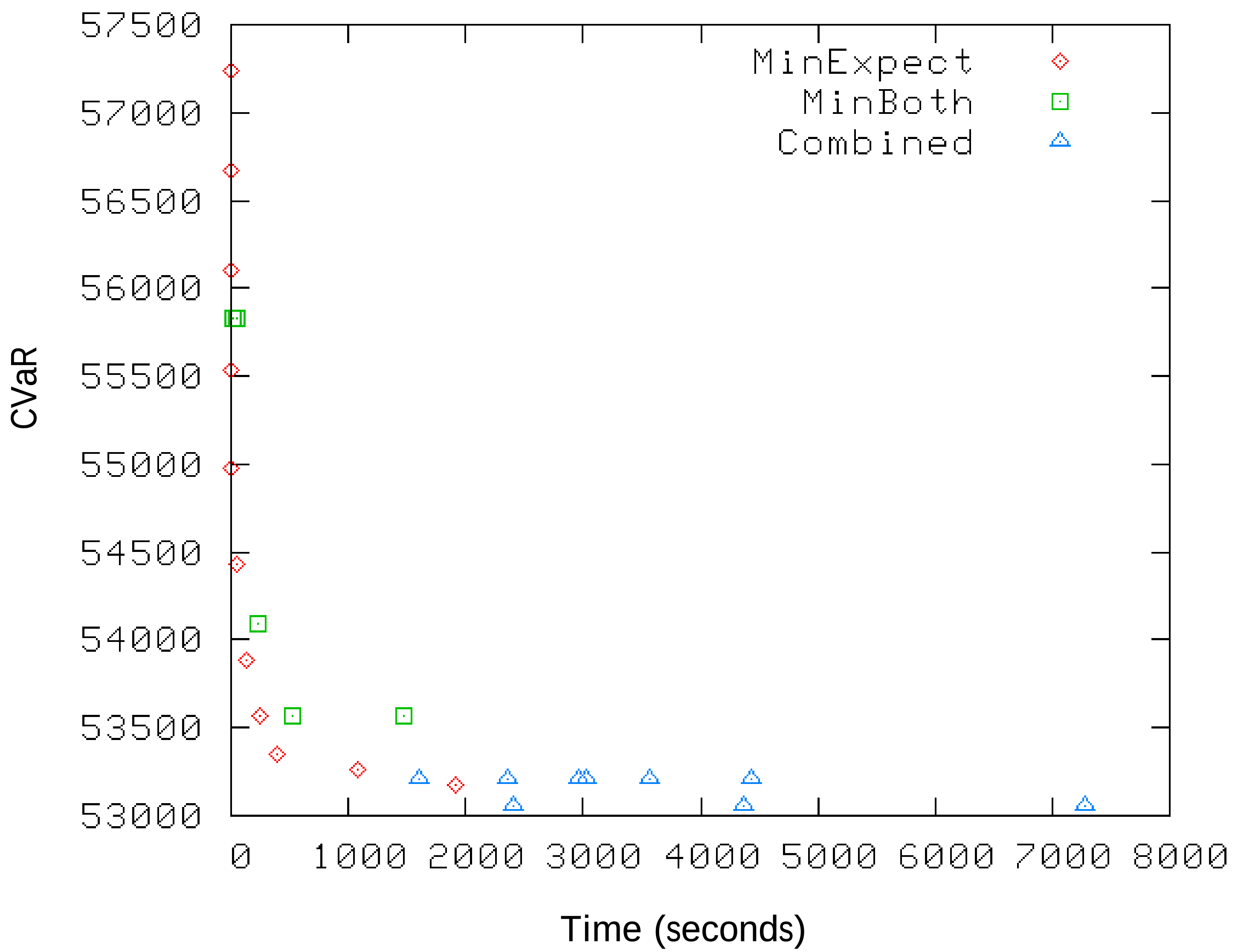
Sandia National Laboratories

Experimental Experience with CVaR MILPs

Considered Performance/CVaR trade-off for Network B



Runtime vs. Optimal CVaR



Empirical Observations

1. There appears to be a clear spike in the runtime
   - In many trials we were not able to find (near-)optimal solns
   - CPLEX had difficulty finding good incumbent solutions
   - We weren't able to solve a minimum CVaR model on this network

2. This formulation consistently found dominated solutions
   - Need an additional bias to guarantee pareto optimal solutions
   - Found undominated solutions with weighted risk objective

Sandia National Laboratories

Ongoing Work

- Developing methods to more efficiently enumerate pareto front
  - Considering bi-objective MILP methods
- Working on reduced fidelity models that aggregate sensor locations
  - Fast solutions with guaranteed approximation bounds
  - May lead to branch-and-price solver
- Analyzing the structure of VaR/CVaR risk measures in MILP formulations
  - E.g. tighter-constrained MILPs seem to take longer. Why?

Acknowledgements

- Robert Carr (SNL)
- Harvey Greenberg (SNL)
- Goren Konjevod (ASU)
- Eric Lauer (UNM)
- Henry Lin (Berkeley)
- Tod Morrison (CU Denver)
- Cindy Phillips (SNL)