Probabilistic Gas Leak Rate Estimation Using Submodular Function Maximization With Routing Constraints

Kalvik Jakkala and Srinivas Akella

IEEE Robotics and Automation Letters (RA-L), April 2022

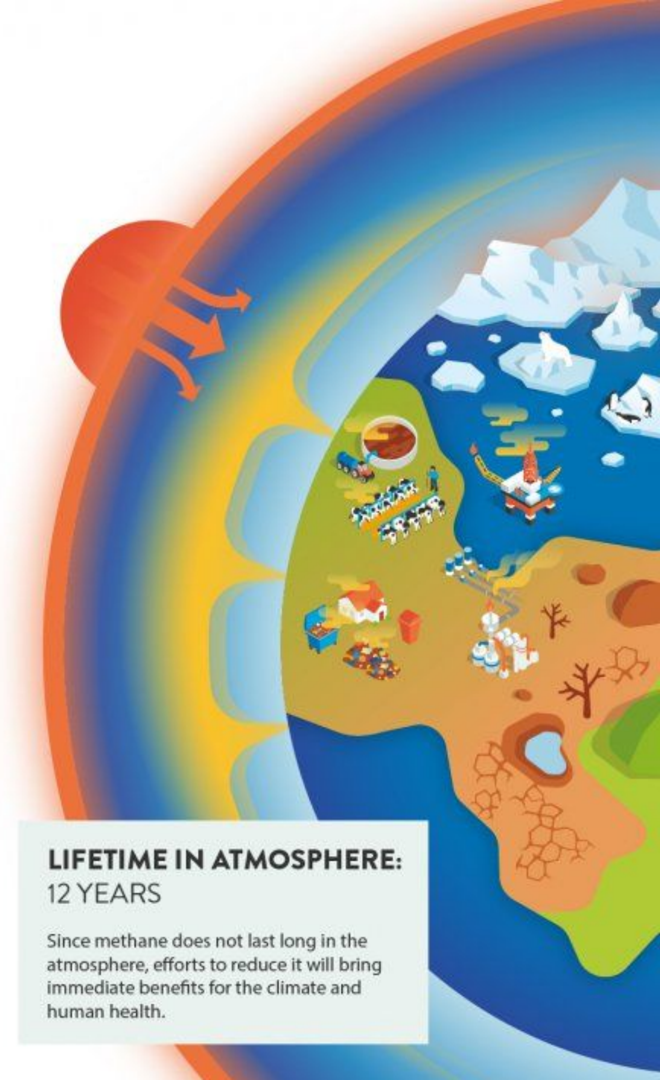
To be presented at the IEEE International Conference on Robotics and Automation (ICRA) 2022

Outline

- Problem
- Assumptions
- Prior approaches
- Method
  - Leak rate estimation
  - Informative Path Planning (IPP)
- Limitations
- Conclusion

Problem

- Monitor fugitive gas leaks
  - Harmful greenhouse gases
- Non-trivial problem
  - Environmental factors
  - Simultaneous leaks
  - Road network constraint
  - Informative Path Planning (IPP) problem
  - Distance budget



https://www.ccacoalition.org/en/slcps/methane (figure source)

Assumptions

- Flat terrain
- Non-overlapping gas plumes
- Constant environmental factors
- Known candidate leak locations
- Threshold leak rates to find source

Permian basin oil fields

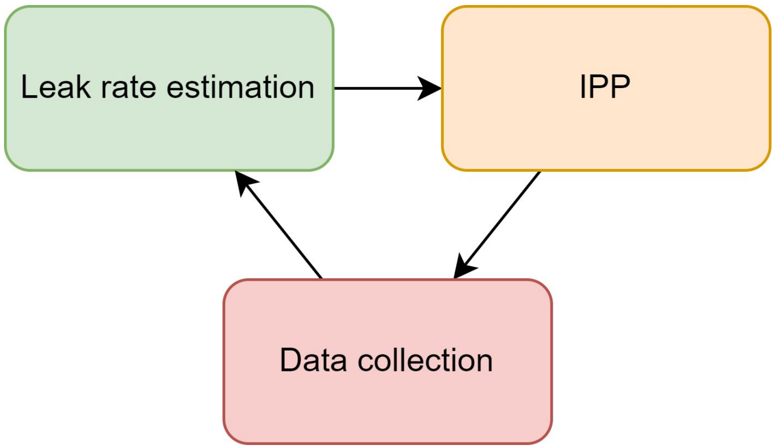
https://fortune.com/2018/05/25/photograph-permian-basin-oil/ (figure source)

Outline

- Problem
- Assumptions
- Prior approaches
- Method
  - Leak rate estimation
  - Informative Path Planning (IPP)
- Limitations
- Conclusion

Prior approaches

- Satellites [Pandey et al. 2019]
- Fixed sensor networks [Allen 2020]
- Neural Networks [Travis et al. 2020]
- Bayesian approach [Albertson et al. 2016]
    - Fix prior over leak rate
    - Mutual information for IPP
    - Collect data
    - Calculate posterior of leak rates
    - Repeat until convergence



Prior approaches

- Bayesian approach
  - GEV type II prior
  - Likelihood

$$p(m|s) = \frac{1}{\sigma_e\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{m - C(s,x,y,z)}{\sigma_e}\right)^2\right]$$

  - Gas dispersion model

$$C(s,x,y,z) = \frac{s}{U}\left(\frac{\bar{A}}{\overline{z}(x)} exp\left[-\left(\frac{Bz}{\overline{z}(x)}\right)^2\right]\right) \times \left(\frac{1}{\sqrt{2\pi}\sigma_y} exp\left[-\frac{1}{2}\left(\frac{y}{\sigma_y}\right)^2\right]\right)$$

Prior approaches

- Bayesian approach
  - Posterior (intractable to an analytical solution)

$$p(s|m) = \frac{p(m|s)p(s)}{\int_{s_1 \epsilon S} p(m|s_1)p(s_1)ds_1}$$

Outline

- Problem
- Assumptions
- Prior approaches
- Method
  - Leak rate estimation
  - Informative Path Planning (IPP)
- Limitations
- Conclusion

Method (Gaussian assumption)

- Gaussian prior

$$p(s) \sim \mathcal{N}(\mu_s, \sigma_s^2)$$

- Likelihood (gas dispersion model factorization)

$$C(s,x,y,z) = s \underbrace{\frac{1}{U}\left(\frac{\bar{A}}{\overline{z}(x)} \exp\left[-\left(\frac{Bz}{\overline{z}(x)}\right)^2\right]\right)\left(\frac{1}{\sqrt{2\pi}\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{y}{\sigma_y}\right)^2\right]\right)}_{constant,\ independent\ of\ function\ input\ 's'} = s\mathcal{A}(x,y,z)$$

$$\begin{aligned} p(m|s) &= \frac{1}{\sigma_e\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{m - C(s,x,y,z)}{\sigma_e}\right)^2\right] \\ &= \frac{1}{\sigma_e\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{m - s\mathcal{A}(x,y,z)}{\sigma_e}\right)^2\right] \end{aligned}$$

$$p(m|s) \sim \mathcal{N}(s\mathcal{A}, \sigma_e^2)$$

Method (Gaussian assumption)

- Analytical posterior
    - Linear in the number of data samples

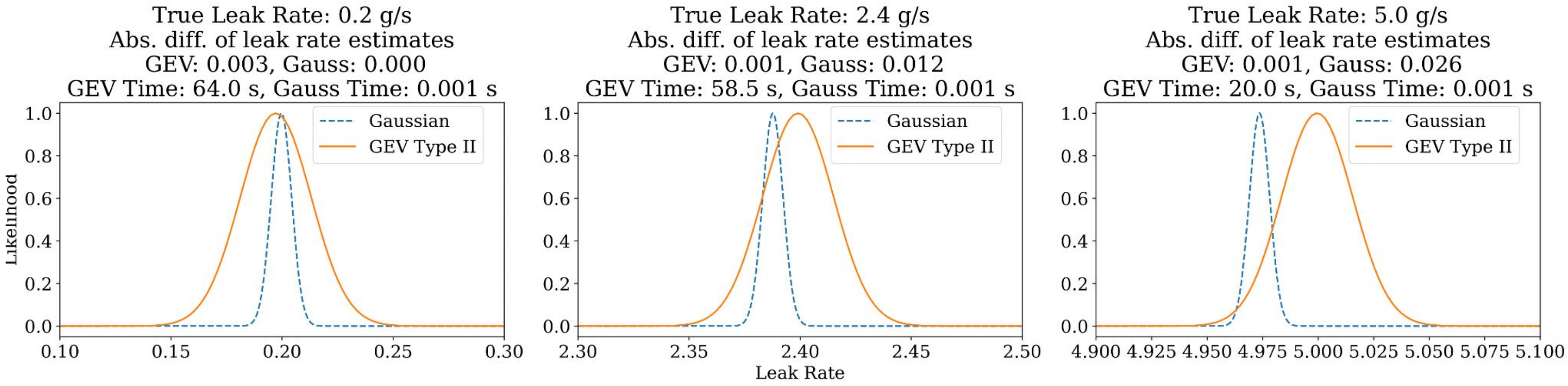
$$p(s|M) \propto \mathcal{N}\left(\frac{M^T\mathbf{A}\sigma_s^2 + \mu_s\sigma_e^2}{\mathbf{A}^T\mathbf{A}\sigma_s^2 + \sigma_e^2}, \frac{\sigma_e^2\sigma_s^2}{\mathbf{A}^T\mathbf{A}\sigma_s^2 + \sigma_e^2}\right)$$

M set of gas concentration data samples collected from the field

$\mathbf{A}$ vector containing the output of the factorized gas dispersion model with unit leak rate

Simulated experimental results

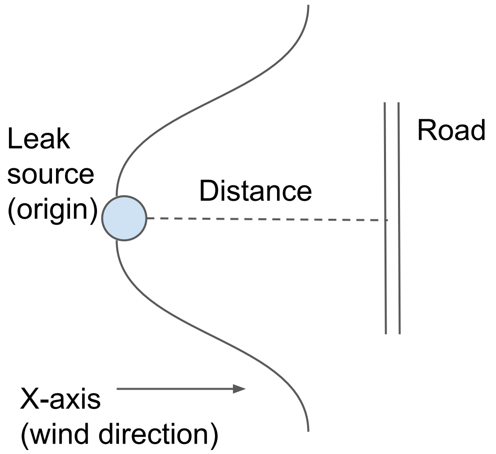
- Gaussian prior convergence results
  - Converges to true leak rate
  - Five orders of magnitude faster



Simulated experimental results

- Predicts correct path information order
  - Five orders of magnitude faster

Distance to oil well along X-axis (meters)	Compute Time (secs)	
	GEV Type II prior	Gaussian prior
0.2	22.42391	0.00017
0.8	16.90775	0.00016
1.4	30.65850	0.00017
2.4	41.51963	0.00021
3.0	42.78649	0.00017
5.0	41.21629	0.00020



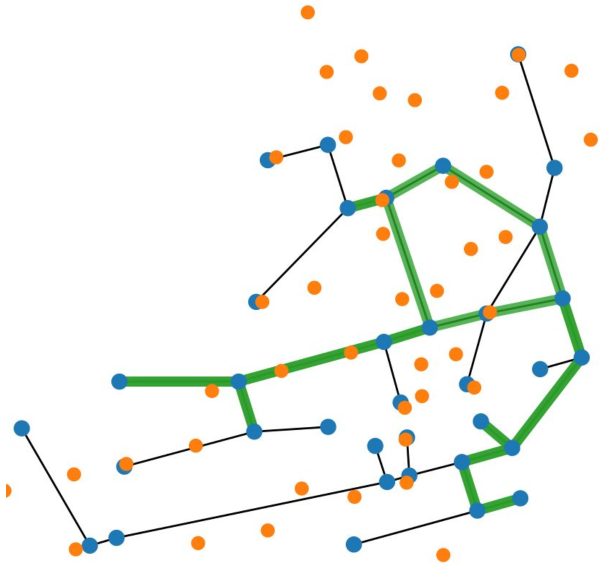
Outline

- Problem
- Assumptions
- Prior approaches
- Method
  - Leak rate estimation
  - Informative Path Planning (IPP)
- Limitations
- Conclusion

Prior approaches (IPP)

- IPP literature
  - Focuses on scalar fields (e.g., gas concentration)
  - Not directly applicable
  - We plan paths only for estimating leak rates

- Prior approaches considered brute force search
  - Expected entropy reduction (EER)
  - Does not scale to large graphs
  - Exponential number of possible paths
  - Paths might not approach any leak sources



Example road network

Method (IPP)

- EER measures mutual information

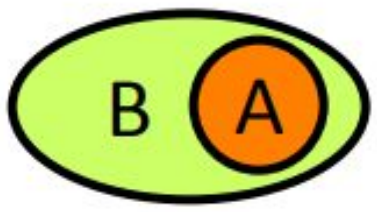
$$\varphi[S;M] = H(S) - \int_{m \epsilon M} H(S|m)p(m)dm$$

  - Submodular function (diminishing returns)

$$f(X \cup \{u\}) - f(X) \geq f(Y \cup \{u\}) - f(Y)$$
$$\forall X \subseteq Y \subset T \text{ and } u \in T \backslash Y$$

Method (IPP)

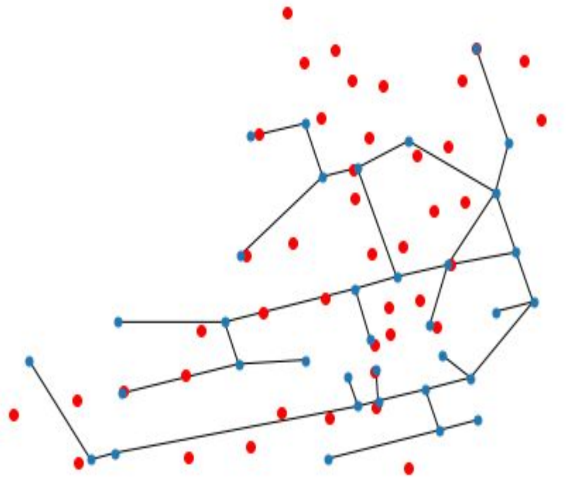
- EER measures mutual information
  - Submodular function (diminishing returns)



http://luthuli.cs.uiuc.edu/~daf/courses/Opt-2019/BilevelSubmodularMaterial/submodularity-slides.pdf (figure source)

Method (IPP)

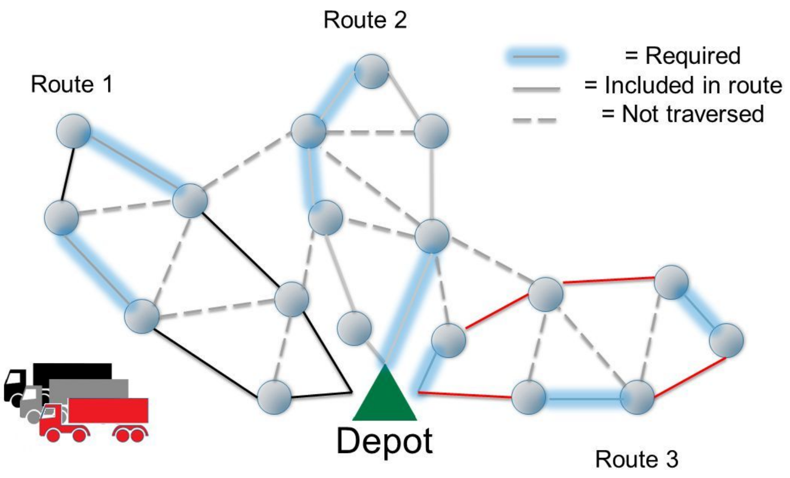
- Generalized Cost-benefit (GCB) algorithm [Zhang and Vorobeychik 2016]
  - Maximizes submodular functions (EER)
  - Imposes *node* routing constraints
  - Runtime approximation guarantee
  - Computationally expensive (faster than brute force)



Example road network

Method (IPP)

- Arc routing
  - Data collection along edges (i.e., continuous sensing)
  - Not a complete graph
  - Required set of edges to visit
  - Might need additional edges to form a path



https://slideplayer.com/slide/7663745/ (figure source)

Method (IPP)

- Arc routing variant
  - Replace TSP with ARP routing solver
  - Can be used in road networks
  - Use deterministic ARP solver (faster convergence)
  - Substantial speedup from Gaussian assumption (EER computation)

Algorithm 1: The modified Generalized Cost-benefit Algorithm (MGCB). $\hat{c}$ is the routing function: TSP when W is the node set V of the graph G, and ARP when W is the edge set E. φ is the submodular cost function (EER), and $W' \backslash x^*$ is the set W' without the element x^*.

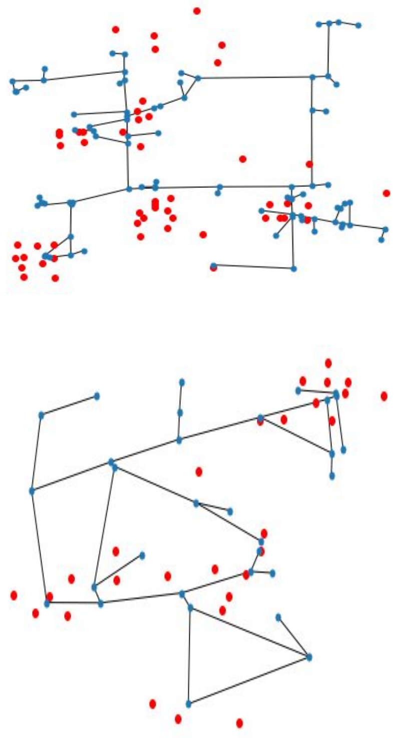
Data: $b > 0, W$
Result: Walk $S \subset W$

$A := \arg\max\{\varphi(x) \mid x \in W, \hat{c}(x) \leq b\}$
$Z := \emptyset$
$W' := W$
while $W' \neq \emptyset$ **do**
    for $x \in W'$ **do**
        $\Delta_\varphi^x := \varphi(Z \cup x) - \varphi(Z)$
        $\Delta_c^x := \hat{c}(Z \cup x) - \hat{c}(Z)$
    end
    $Y := \{x \mid x \in W', \hat{c}\,(Z \cup x) \leq b\}$
    if $|Y| == 0$ **then**
        break
    end
    $x^* = \arg\max\{\Delta_\varphi^x / \Delta_c^x \mid x \in Y\}$
    $Z := Z \cup x^*$
    $W' := W' \backslash x^*$
end
return $\arg\max_{S \in \{A, Z\}} f(S)$

Simulated experimental results

- Dataset
  - Oil well clusters from the Permian basin (U.S.A.)
  - K-means clustering
  - Extracted the road networks
  - 134 total graph

Statistic	Min	Max	Mean
Number of oil wells	11	145	35.50
Number of vertices	15	420	76.28
Number of edges	16	484	81.07
Avg. connectivity [31]	1.0	1.10	1.04



Example graphs

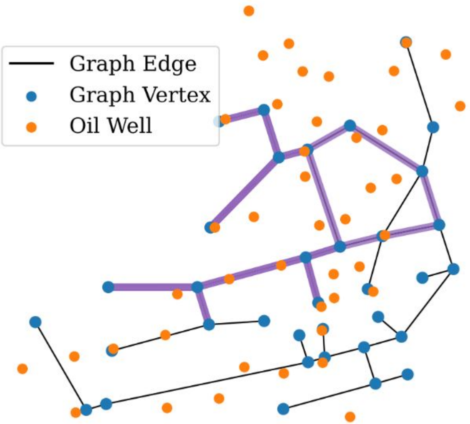
Simulated experimental results

- GCB is faster than path iteration (brute force)
- MGCB at least one order of magnitude faster than GCB
- Converges to the same solution as vanilla GCB

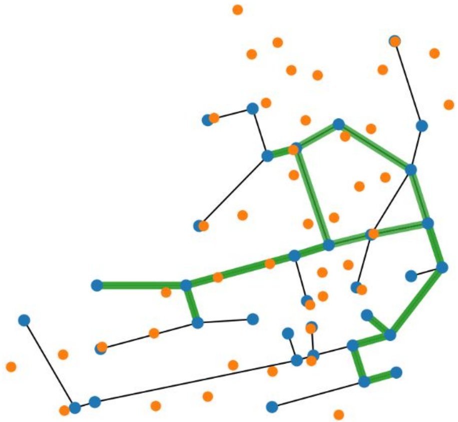
Method	Mean MSE	Std. Dev. of MSE	Mean Time (secs)	Std. Dev. of Time
Path Iter	1.6132E-01	1.2989E+00	1203.81	4.18
GCB TSP	2.0583E-04	1.4948E-03	1209.58	22.64
GCB RPP	2.5680E-05	1.3238E-04	1211.50	17.45
MGCB TSP	2.0583E-04	1.4948E-03	129.41	305.48
MGCB RPP	**2.5680E-05**	1.3238E-04	450.02	512.24

Simulated experimental results

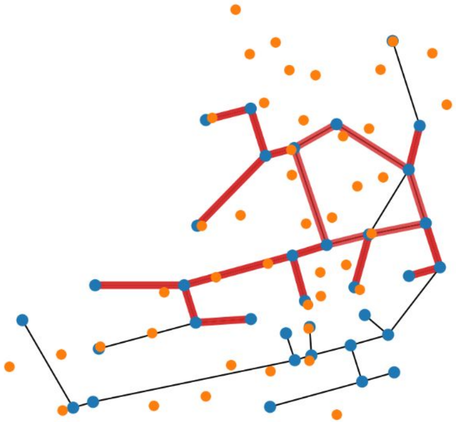
- One of the graphs from our dataset along with the solution from each solver



(a) Path iteration (1.80E-5 MSE)



(b) Modified GCB with TSP (2.20E-5 MSE)



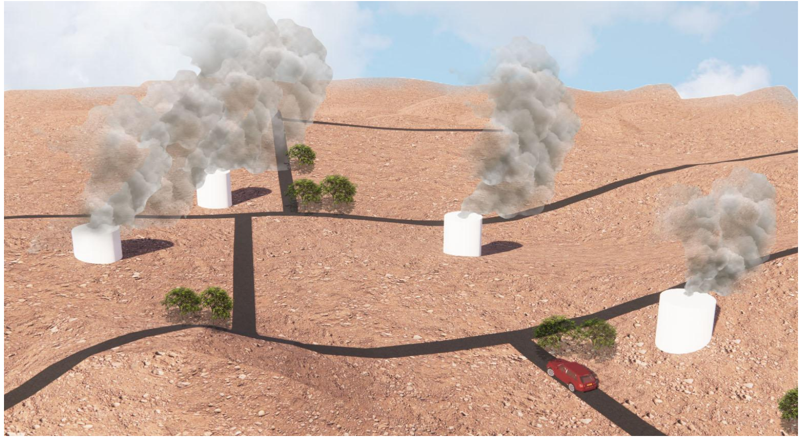
(c) Modified GCB with RPP (1.08E-5 MSE)

Outline

- Problem
- Assumptions
- Prior approaches
- Method
  - Leak rate estimation
  - Informative Path Planning (IPP)
- Limitations
- Conclusion

Limitations

- Overlapping gas plumes
  - Leaks from sources in proximity
- Source localization
  - Leaks from pipelines and storage tanks
- Dynamic environmental conditions
  - Changing wind direction, speed, and temperature



Conclusion

- Computationally efficient leak rate estimation
  - Gaussian prior assumption
  - Gas dispersion model factorization
  - Five orders of magnitude faster
- Informative path planning
  - Substantially faster EER computation
  - Leveraged submodularity with the GCB algorithm
  - Generalized GCB to arc routing constraints
  - Improved GCB algorithm’s compute cost
  - At least one magnitude faster than vanilla GCB

Takeaway

- Gaussian assumption
  - Can your phenomena be modeled with a Gaussian?
  - Forest fire smoke dispersion
  - Water pollutants
- Submodularity
  - Is your objective function submodular?
- There are other variants of IPP
  - Does our IPP variant apply to your problem?
- How would you address our limitations
  - Overlapping gas plumes
  - Source localization
  - Dynamic environmental conditions

Acknowledgments and Resources

- Funding: This work was supported in part by NSF Award IIP-1919233

- Paper: https://ieeexplore.ieee.org/document/9706242
- Code: https://github.com/UNCCharlotte-CS-Robotics/Gas-Leak-Estimation

We thank Saurav Agarwal for thoughtful discussions and making his RPP code available in the Line Coverage Library. We also thank Eben Thoma and Sayantan Datta for helpful comments and Geethika Jakkala for creating Fig. 1 (slide 25).