Robust Optimization with Applications in Networking

ANRL Seminar

Siddartha Devic

The University of Texas at Dallas

April 23, 2019

Outline

1. Introduction
   - Example & Motivation [1]
   - Mathematical Formulation
2. Application in Network Flows [4]
   - Min-Cost Flow Problem
   - Γ-Robust Models
3. RO for VM Consolidation
   - Problem Overview
   - Problem Formulation
   - Results

What is Robust Optimization?

Traffic Flows Single link in a communication network, and two traffic flows that can use this link.[1]

What is Robust Optimization?

Traffic Flows Single link in a communication network, and two traffic flows that can use this link.[1]

Solving LP Each (x, y) historical pair corresponds to some amount of traffic through $x_1 = x$ and $x_2 = y$.

What is Robust Optimization?

Traffic Flows Single link in a communication network, and two traffic flows that can use this link.[1]

Solving LP Each (x, y) historical pair corresponds to some amount of traffic through $x_1 = x$ and $x_2 = y$.
Say, average $\hat{x}_1 = 477$, average $\hat{x}_2 = 637$:

What is Robust Optimization?

Traffic Flows Single link in a communication network, and two traffic flows that can use this link.[1]

Solving LP Each (x, y) historical pair corresponds to some amount of traffic through $x_1 = x$ and $x_2 = y$.
Say, average $\hat{x}_1 = 477$, average $\hat{x}_2 = 637$:

$$\begin{aligned} \textbf{maximize } & x_1 + x_2 \\ s.t.\ 477x_1 + 637x_2 & \leq 1024 \\ 0 \leq x_1, x_2, & \leq 1 \end{aligned}$$

What is Robust Optimization?

Traffic Flows Single link in a communication network, and two traffic flows that can use this link.[1]

Solving LP Each (x, y) historical pair corresponds to some amount of traffic through $x_1 = x$ and $x_2 = y$.
Say, average $\hat{x}_1 = 477$, average $\hat{x}_2 = 637$:

$$\begin{aligned} \textbf{maximize } & x_1 + x_2 \\ s.t. \;\; & 477x_1 + 637x_2 \leq 1024 \\ & 0 \leq x_1, x_2, \leq 1 \end{aligned}$$

Feasible solutions $(0, 1) \times (0, 1)$

What is Robust Optimization?

How good is a solution, say $x_1, x_2 = (1, 0.86)$?

What is Robust Optimization?

How good is a solution, say $x_1, x_2 = (1, 0.86)$?

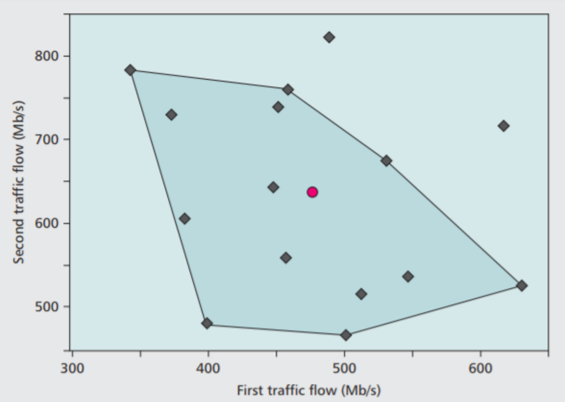


Figure: Historical sample points for aforementioned LP. Red point is average.

LP's Are Not Enough

- In our previous example, with the average constraint, 8/15 historical points overload the link

LP's Are Not Enough

- In our previous example, with the average constraint, 8/15 historical points overload the link
- We would like to do better, and have an efficient solver with some "robustness"

LP's Are Not Enough

- In our previous example, with the average constraint, 8/15 historical points overload the link
- We would like to do better, and have an efficient solver with some "robustness"
- The problem is we have uncertain coefficients in the LP.

$$\begin{aligned} \textbf{maximize } & x_1 + x_2 \\ s.t. \;\; q * x_1 + p * x_2 &\leq 1024 \\ 0 \leq x_1, x_2, &\leq 1 \end{aligned}$$

Find q, p s.t. less than some 15% of flows from sampled data overload the link.

LP's Are Not Enough

- In our previous example, with the average constraint, 8/15 historical points overload the link
- We would like to do better, and have an efficient solver with some "robustness"
- The problem is we have uncertain coefficients in the LP.

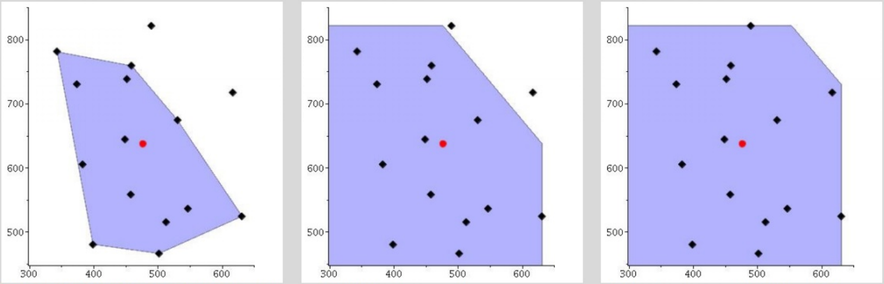
$$\begin{aligned} \textbf{maximize } & x_1 + x_2 \\ s.t. \;\; q * x_1 + p * x_2 & \leq 1024 \\ 0 \leq x_1, x_2, & \leq 1 \end{aligned}$$

Find q, p s.t. less than some 15% of flows from sampled data overload the link.

- "Robustly" solving the LP w.r.t. sampled data.

Robustly Solving LP's

Instead of fixing the coefficients, consider them to be drawn from some "uncertainty set".



(a) convex hull of 13 out of 15 historical data points

(b) Γ-robustness with averages as nominal values

(c) Γ-robustness with 67% quintiles as nominal values

Figure: Some feasible uncertainty sets

Robustly Solving LP's

- Consider our previous LP

$$\begin{aligned} \textbf{maximize } & x_1 + x_2 \\ s.t. \;\; q * x_1 + p * x_2 & \leq 1024 \\ 0 \leq x_1, x_2, & \leq 1 \end{aligned}$$

Robustly Solving LP's

- Consider our previous LP

$$\begin{aligned} \textbf{maximize } & x_1 + x_2 \\ s.t.\ \ q * x_1 + p * x_2 & \leq 1024 \\ 0 \leq x_1, x_2, & \leq 1 \end{aligned}$$

Add the constraint that $(q, p) \in \mathcal{D}$, where $\mathcal{D}$ is a feasible convex uncertainty set.

Robustly Solving LP's

- Consider our previous LP

$$\begin{aligned} \textbf{maximize}\ & x_1 + x_2 \\ s.t.\ \ q * x_1 + p * x_2 &\leq 1024 \\ 0 \leq x_1, x_2, &\leq 1 \end{aligned}$$

  Add the constraint that $(q, p) \in \mathcal{D}$, where $\mathcal{D}$ is a feasible convex uncertainty set.
- New objective is two-fold:
  1. Satisfy constraints everywhere within uncertainty set
  2. Maximize sum amongst all robust feasible solutions

Robustly Solving LP's

- Consider our previous LP

$$\begin{aligned} \textbf{maximize } & x_1 + x_2 \\ s.t. \;\; q * x_1 + p * x_2 & \leq 1024 \\ 0 \leq x_1, x_2, & \leq 1 \end{aligned}$$

  Add the constraint that $(q, p) \in \mathcal{D}$, where $\mathcal{D}$ is a feasible convex uncertainty set.

- New objective is two-fold:
  1. Satisfy constraints everywhere within uncertainty set
  2. Maximize sum amongst all robust feasible solutions
- Motivated by the fact that "worst case" optimization is too restrictive.

Formal Description (for completeness)

Uncertain LP A collection $\{\min_x\{c^T x \ : \ Ax \leq b\} \ : \ (c, A, B) \in \mathcal{U}\}$

Formal Description (for completeness)

Uncertain LP A collection $\{\min_x\{c^T x \;:\; Ax \leq b\} \;:\; (c, A, B) \in \mathcal{U}\}$

Robust Counterpart $\min_x\{\sup_{(c,A,B)\in\mathcal{U}} c^T x \;:\; Ax \leq b \; \forall(c, A, B) \in \mathcal{U}\}$

Formal Description (for completeness)

Uncertain LP A collection $\{\min_x\{c^T x \ : \ Ax \leq b\} \ : \ (c, A, B) \in \mathcal{U}\}$

Robust Counterpart $\min_x\{\sup_{(c,A,B)\in\mathcal{U}} c^T x \ : \ Ax \leq b \ \forall(c, A, B) \in \mathcal{U}\}$

Notes

- Supremum comes from guaranteed value of the original objective, which is upper bound for a minimization problem. [2]

Formal Description (for completeness)

Uncertain LP A collection $\{\min_x \{c^T x \; : \; Ax \leq b\} \; : \; (c, A, B) \in \mathcal{U}\}$

Robust Counterpart $\min_x \{\sup_{(c,A,B) \in \mathcal{U}} c^T x \; : \; Ax \leq b \; \forall (c, A, B) \in \mathcal{U}\}$

Notes

- Supremum comes from guaranteed value of the original objective, which is upper bound for a minimization problem. [2]
- Objective can be uncertain

Formal Description (for completeness)

Uncertain LP A collection $\{\min_x\{c^T x \ : \ Ax \leq b\} \ : \ (c, A, B) \in \mathcal{U}\}$

Robust Counterpart $\min_x\{\sup_{(c,A,B)\in\mathcal{U}} c^T x \ : \ Ax \leq b \ \forall(c, A, B) \in \mathcal{U}\}$

Notes

- Supremum comes from guaranteed value of the original objective, which is upper bound for a minimization problem. [2]
- Objective can be uncertain
- Solution is only important in $\mathcal{U}$
- Cannot violate constraints set by $\mathcal{U}$ given input satisfying in the feasible uncertainty set

Network Flows

[Gast et al.], min-cost flow

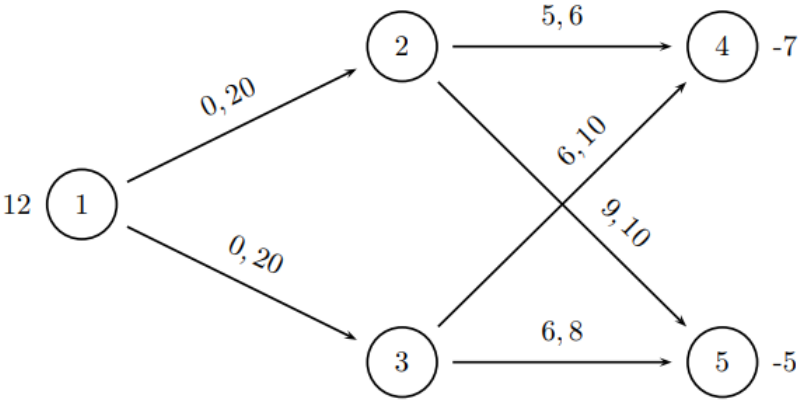


Figure: Min-cost flow problem. Arcs are labeled with (cost, capacities) and $b < 0$ is demand, $b > 0$ is supply.

Min-Cost Flow

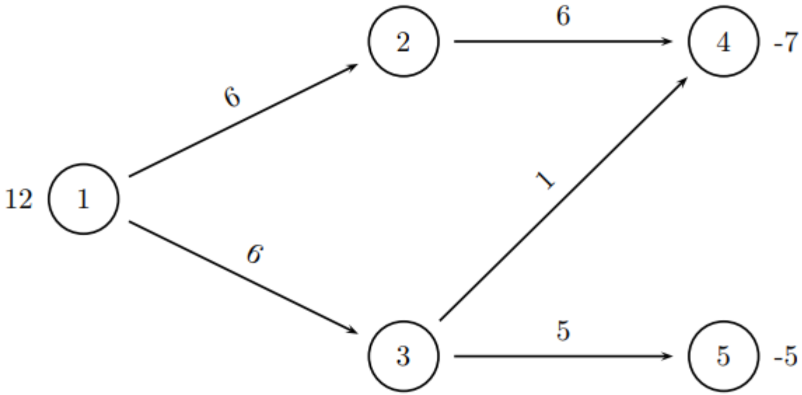


Figure: Solution to flow problem on previous slide.

Normal LP Formulation

$$
\begin{aligned}
\textbf{minimize} \quad & \sum_{(i,j)\in E} c_{ij} f_{ij} \\
\text{s.t.} \quad & \sum_{\{k|(j,k)\in E\}} f_{jk} - \sum_{\{i|(i,j)\in E\}} f_{ij} = b_j, \quad \forall j \in V \\
& 0 \leq f_{ij} \leq u_{ij}, \quad \forall (i,j) \in E
\end{aligned}
$$

Uncertain Min-Cost Flow

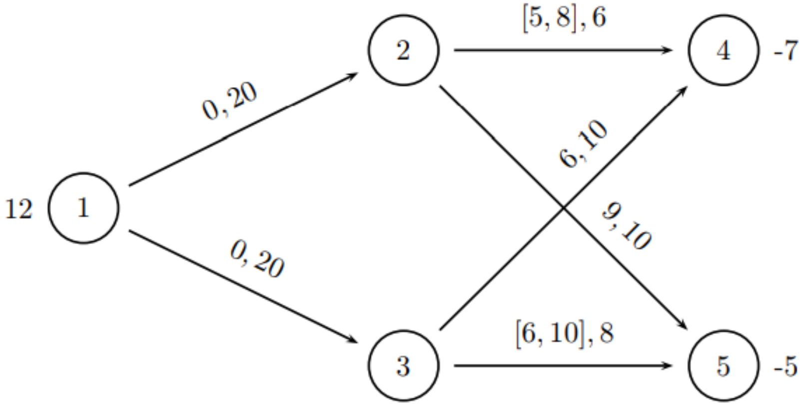


Figure: Uncertainty in edge costs.

Robust LP Formulation (Bertismas and Sim [3])

$$
\begin{aligned}
\min \sum_{(i,j)\in\mathcal{A}} & \tilde{c}_{ij} x_{ij} \\
\text{s.t.} \sum_{\{j:(i,j)\in\mathcal{A}\}} x_{ij} - & \sum_{\{j:(j,i)\in\mathcal{A}\}} x_{ji} = b_i && \forall i \in \mathcal{N} \\
0 \leq x_{ij} & \leq u_{ij} && \forall (i,j) \in \mathcal{A}
\end{aligned}
$$

with underlying graph $G = (\mathcal{N}, \mathcal{A})$, positive costs $\tilde{c}_{ij}$ where $\tilde{c}_{ij}$ takes values in $[c_{ij}, c_{ij} + d_{ij}]\,, c_{ij}, d_{ij} \geq 0, (i,j) \in \mathcal{A}$, and d_{ij} is uncertainty in cost edges.

Γ-Robustness

Let's give some examples for Γ-robustness before formulating it explicitly.
The uncertain network flow problem:

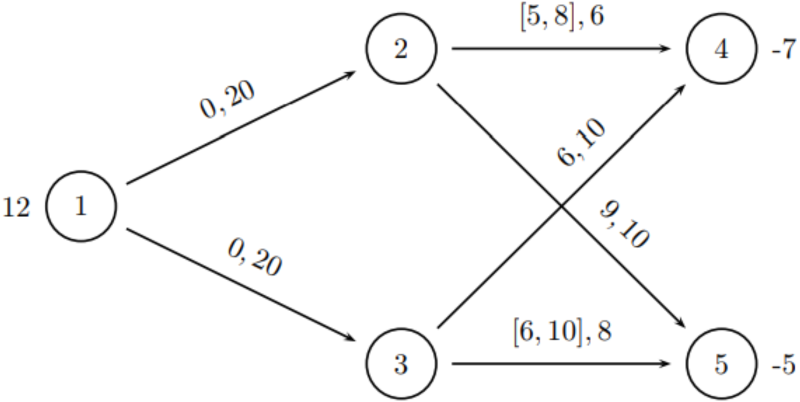


Figure: Min-cost flow problem. Arcs are labeled with (cost, capacities) and $b < 0$ is demand, $b > 0$ is supply.

Γ-Robustness

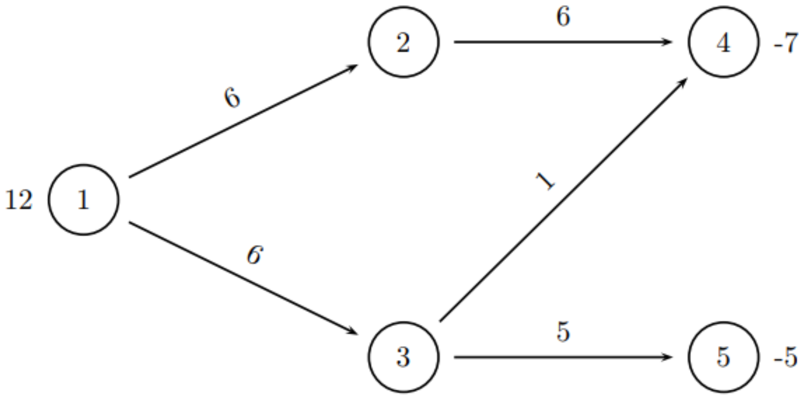


Figure: $\Gamma = 0$, best case cost 66, worst case cost 104

Γ-Robustness

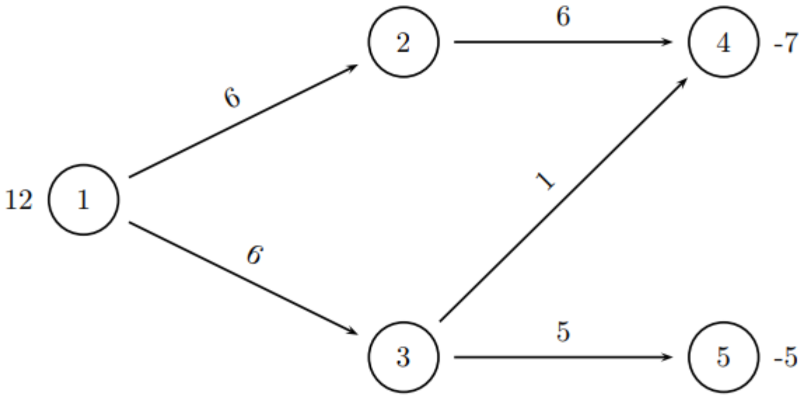


Figure: Γ=0, best case cost 66, worst case cost 104

This is identical to the original LP's solution!

Γ-Robustness

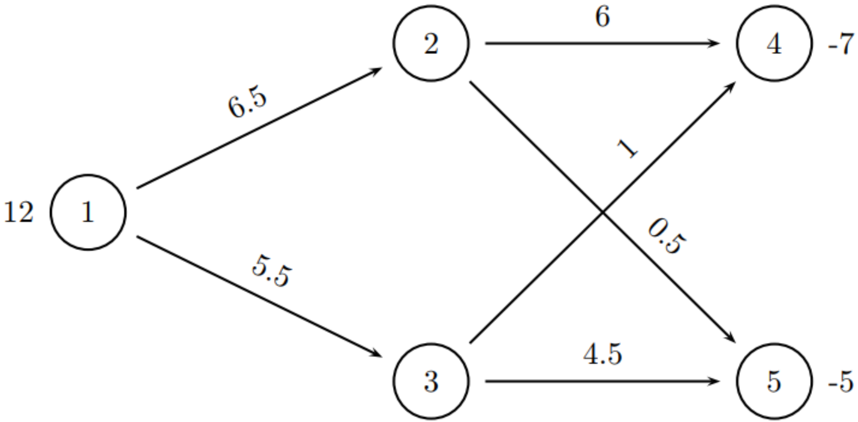


Figure: $\Gamma = 1$, medium-conservative robustness with best case cost 67.5, worst case cost 103.5

Γ-Robustness

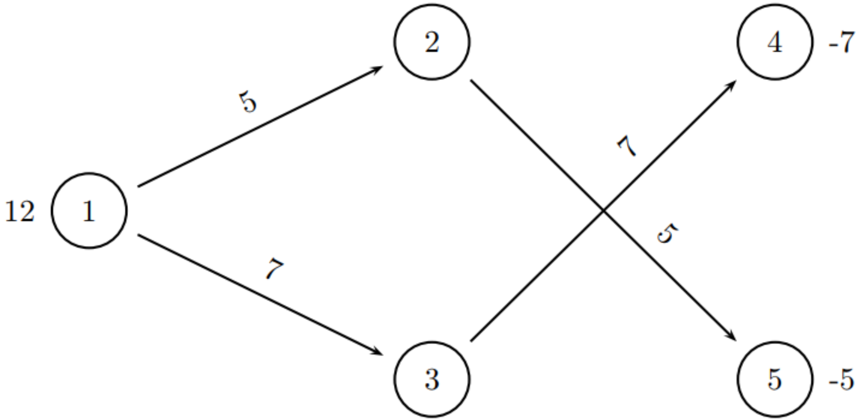


Figure: $\Gamma = 2$, best case cost 87, worst case cost 87

Formulation of Γ-Robust Min-Cost Flow

$$\min \quad c^T x + \max_{\{S | S \subseteq \mathcal{A}, |S| \leq \Gamma\}} \sum_{(i,j) \in S} d_{ij} x_{ij}$$

Formulation of Γ-Robust Min-Cost Flow

$$\begin{aligned} \min \quad & c^T x + \max_{\{S \mid S \subseteq \mathcal{A}, |S| \leq \Gamma\}} \sum_{(i,j) \in S} d_{ij} x_{ij} & \\ \text{s.t.} \quad & \sum_{\{j:(i,j)\in\mathcal{A}\}} x_{ij} - \sum_{\{j:(j,i)\in\mathcal{A}\}} x_{ji} = b_i & \forall i \in \mathcal{N} \\ & 0 \leq x_{ij} \leq u_{ij} & \forall (i,j) \in \mathcal{A} \end{aligned}$$

and $\Gamma \in [0, |\mathcal{A}|]$ controls robustness.

Formulation of Γ-Robust Min-Cost Flow

$$
\begin{aligned}
\min \quad & c^T x + \max_{\{S \mid S \subseteq \mathcal{A}, |S| \leq \Gamma\}} \sum_{(i,j) \in S} d_{ij} x_{ij} & \\
\text{s.t.} \quad & \sum_{\{j:(i,j) \in \mathcal{A}\}} x_{ij} - \sum_{\{j:(j,i) \in \mathcal{A}\}} x_{ji} = b_i & \forall i \in \mathcal{N} \\
& 0 \leq x_{ij} \leq u_{ij} & \forall (i,j) \in \mathcal{A}
\end{aligned}
$$

and $\Gamma \in [0, |\mathcal{A}|]$ controls robustness.

- We only consider Γ-Robustness, but [4] has this same example with at least five other types of robustness formulations, all fairly recent.

Case Study: RO for VM Consolidation [5]

- Increasing need for more and more physical servers $\rightarrow$ downtime during non-peak hours
- Resource inefficient if VMs spread across multiple physical servers
  - e.g. 42% of operational cost of Amazon datacenter is power consumption + cooling
- Drives VM consolidation problem

VM Consolidation Problem

- Reliant on various factors like VM load, time of day, etc.

VM Consolidation Problem

- Reliant on various factors like VM load, time of day, etc.
- Many issues related to predicting these things
  - Hard to predict VM workload in real world scenarios (Dynamic workload due to cloud infrastructure)

VM Consolidation Problem

- Reliant on various factors like VM load, time of day, etc.
- Many issues related to predicting these things
  - Hard to predict VM workload in real world scenarios (Dynamic workload due to cloud infrastructure)
- Furthermore, presence of uncertain data at optimization time may lead to solutions that are useless in practice (non-robust).
  - small deviations in input data values may lead to situations where found optimal solution is even not feasible anymore.

Robust Optimization Assumptions

- Probability distribution of uncertain data is not known beforehand

Robust Optimization Assumptions

- Probability distribution of uncertain data is not known beforehand
  - rather, uncertain data lies within some uncertainty set

Robust Optimization Assumptions

- Probability distribution of uncertain data is not known beforehand
  - rather, uncertain data lies within some uncertainty set
  - aside: if probability distribution of uncertain data is known, **stochastic optimization** may be the better choice

Robust Optimization Assumptions

- Probability distribution of uncertain data is not known beforehand
  - rather, uncertain data lies within some uncertainty set
  - aside: if probability distribution of uncertain data is known, **stochastic optimization** may be the better choice
- Γ-Robustness built around deviation from average (construction of uncertainty set) [Bertsimas and Sim].

Robust Optimization Assumptions

- Probability distribution of uncertain data is not known beforehand
  - rather, uncertain data lies within some uncertainty set
  - aside: if probability distribution of uncertain data is known, **stochastic optimization** may be the better choice
- Γ-Robustness built around deviation from average (construction of uncertainty set) [Bertsimas and Sim].
- All the uncertain parameters are independent, e.g. chance of all VMs running at max load simultaneously is 0.

Robust Model for the VM Consolidation Problem

- Objective: Minimize power consumption (normalized) and the number of migrations

Robust Model for the VM Consolidation Problem

- Objective: Minimize power consumption (normalized) and the number of migrations

  minimize $f = \alpha \cdot \frac{\sum_{j=1}^{n} P_j}{\sum_{j=1}^{n} P_{ini_j}} + (1 - \alpha) * \frac{\sum_{k,j} (z_{jk}^{-} > + z_{jk}^{<-})}{m}$

  where α controls weighted average between power consumption and number of migrations.

Robust Model for the VM Consolidation Problem

- **Modeling Uncertainty for PMs**
  - Power consumption of physical machine $uncP_j$ is symmetrically distributed $[-\Delta P_j, +\Delta P_j]$

Robust Model for the VM Consolidation Problem

- **Modeling Uncertainty for PMs**
  - Power consumption of physical machine $uncP_j$ is symmetrically distributed $[-\Delta P_j, +\Delta P_j]$
  - Calculated with
    $P_j = P_{\text{idle},j} \cdot y_j + (P_{\max,j} - P_{\text{idle},j}) \cdot u_{ij} + uncP_j \cdot y_j, \quad i = CPU$

Robust Model for the VM Consolidation Problem

- **Modeling Uncertainty for PMs**
  - Power consumption of physical machine $uncP_j$ is symmetrically distributed $[-\Delta P_j, +\Delta P_j]$
  - Calculated with
    $P_j = P_{\text{idle},j} \cdot y_j + (P_{\max,j} - P_{\text{idle},j}) \cdot u_{ij} + uncP_j \cdot y_j, \quad i = CPU$
  - Guaranteed to be satisfied with constraint
    $P_{idle,j} \cdot y_j \leq P_j \leq P_{\max,j} \cdot y_j \quad \forall j$

Robust Model for the VM Consolidation Problem

- **Modeling Uncertainty for PMs**
  - Power consumption of physical machine $uncP_j$ is symmetrically distributed $[-\Delta P_j, +\Delta P_j]$
  - Calculated with
    $P_j = P_{\text{idle},j} \cdot y_j + (P_{\max,j} - P_{\text{idle},j}) \cdot u_{ij} + uncP_j \cdot y_j, \quad i = CPU$
  - Guaranteed to be satisfied with constraint
    $P_{idle,j} \cdot y_j \leq P_j \leq P_{\max,j} \cdot y_j \quad \forall j$
  - Finally, Γ robust tuned with:
    $\sum_{j=1}^{n} \left| \frac{uncP_j}{\Delta P_j} \right| \leq \Gamma, \quad \left| \frac{uncP_j}{\Delta P_j} \right| \leq 1, \quad \forall j, \quad \Gamma \in \{0, \ldots, \text{n}\}$
- We continue and do the same for other variables.

Problem Variables

Input parameters	
m	Total number of VMs
n	Total number of servers
x_{jk}^{O}	Is 1 if VM k is allocated to server j before consolidation, and 0 otherwise
P_{ini_j}	Initial power consumption for server j
$P_{idle,j}$	Idle power consumption for server j
$P_{max,j}$	Maximum power consumption of server j
r_{ik}	Amount of resource i needed to allocate VM k
$rovh_{ik}$	Overhead for resource i to migrate VM k
s_{ij}	Amount of resource i available at server j
$tdown_k$	Downtime for the migration of VM k
SLA_k	SLAs for the applications running on VM k
η_{ij}	Overbooking of resource i at server j
Γ	Protection level over the uncertain variables
M	A large number

Decision variables	
y_j	Is 1 if server j is active after consolidation, 0 otherwise
x_{jk}^{N}	Is 1 if VM k is allocated to server j after consolidation, and 0 otherwise
$allocR_{ij}$	Resource i allocated to server j after consolidation
w_{ij}	Is 1 if resource i on server j is overbooked after consolidation
P_j	Power consumption at server j after consolidation
$z_{jk}^{->}$	Is 1 if VM k migrates from server j
$z_{jk}^{<-}$	Is 1 if VM k migrates to server j
$uncP_j$	Uncertain power at server j
$uncR_{ik}$	Uncertain demand of resource i for VM k
$uncROV_{ik}$	Uncertain overhead for resource i to migrate VM k

MILP Optimization Problem

$$min\, f = \alpha \cdot \frac{\sum_{j=1}^{n} P_j}{\sum_{j=1}^{n} P_{ini_j}} + (1-\alpha) \cdot \frac{\sum_{k,j} \frac{(z_{jk}^{->} + z_{jk}^{<-})}{2}}{m} \tag{5}$$

subject to

$$P_j = P_{idle,j} \cdot y_j + (P_{max,j} - P_{idle,j}) \cdot u_{ij} + uncP_j \cdot y_j, \qquad i = CPU \tag{6}$$

$$P_{idle,j} \cdot y_j \leq P_j \leq P_{max,j} \cdot y_j \qquad \forall j \tag{7}$$

$$\sum_{k=1}^{m} (r_{ik} + uncR_{ik}) \cdot x_{jk}^{N} - s_{ij} \leq M \cdot w_{ij}, \tag{8}$$

$$s_{ij} - \sum_{k=1}^{m} (r_{ik} + uncR_{ik}) \cdot x_{jk}^{N} \leq M \cdot (1 - w_{ij}),$$

$$allocR_{ij} \geq \sum_{k=1}^{m} (r_{ik} + uncR_{ik}) \cdot x_{jk}^{N} - (M \cdot w_{ij}), \tag{9}$$

$$allocR_{ij} \geq s_{ij} - (M \cdot (1 - w_{ij})), \quad \forall i,\ \forall j.$$

$$u_{ij} = \frac{allocR_{ij}}{s_{ij}}, \quad \forall i,\ \forall j$$

$$\sum_{j=1}^{n} \left| \frac{uncP_j}{\Delta P_j} \right| \leq \Gamma, \quad \left| \frac{uncP_j}{\Delta P_j} \right| \leq 1, \quad \forall j, \quad \Gamma \in \{0, \ldots, \mathrm{n}\}, \tag{10}$$

$$\sum_{k=1}^{m} \left| \frac{uncR_{ik}}{\Delta R_{ik}} \right| \leq \Gamma, \quad \left| \frac{uncR_{ik}}{\Delta R_{ik}} \right| \leq 1, \quad \forall i, \forall k, \quad \Gamma \in \{0, \ldots, \mathrm{m}\} \qquad ,$$

$$\sum_{k=1}^{m} \left| \frac{uncROV_{ik}}{\Delta ROV_{ik}} \right| \leq \Gamma, \quad \left| \frac{uncROV_{ik}}{\Delta ROV_{ik}} \right| \leq 1, \quad \forall i, \forall k, \quad \Gamma \in \{0, \ldots, \mathrm{m}\}$$

$$\sum_{k=1}^{m} (r_{ik} \cdot x_{jk}^{O} + (r_{ik} + uncR_{ik} + rovh_{ik} + uncROV_{ik}) \cdot z_{jk}^{<-} - (r_{ik} + uncR_{ik} + rovh_{ik} + uncROV_{ik}) \cdot z_{jk}^{->}) \leq \eta_{ij} \cdot (s_{ij} \cdot y_j) \quad \forall j, \forall i \tag{11}$$

$$x_{jk}^{O} + x_{jk}^{N} + z_{jk}^{->} + z_{jk}^{<-} \leq 2, \tag{12}$$

$$x_{jk}^{O} - (x_{jk}^{N} + z_{jk}^{->}) \leq 0, \quad x_{jk}^{O} + x_{jk}^{N} \geq b_{jk},$$

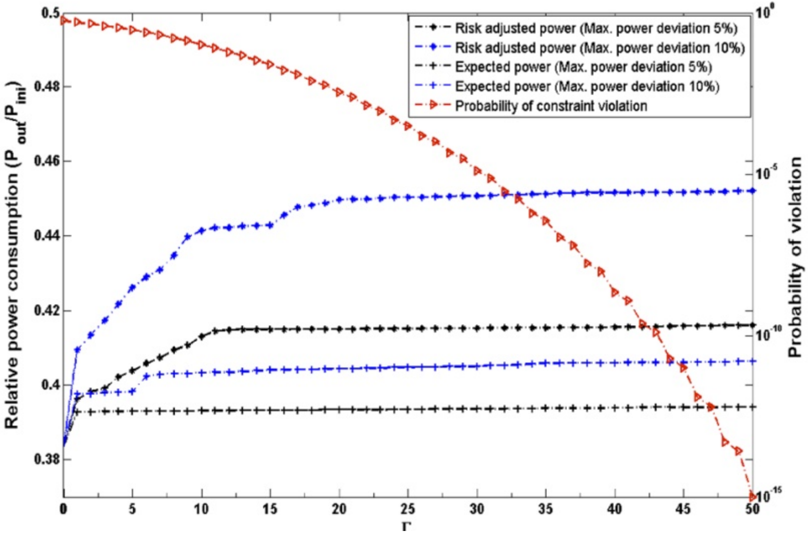
$$x_{jk}^{N} - (x_{jk}^{O} + z_{jk}^{<-}) \leq 0, \quad z_{jk}^{->} + z_{jk}^{<-} \leq b_{jk},$$

$$x_{jk}^{N} \leq y_i \leq \sum_{j=1}^{n} x_{jk}^{N}, \qquad \sum_{j=1}^{n} x_{jk}^{N} = 1, \quad \forall j, \forall k.$$

$$tdown_k \cdot z_{jk}^{->} \leq SLA_k, \tag{13}$$

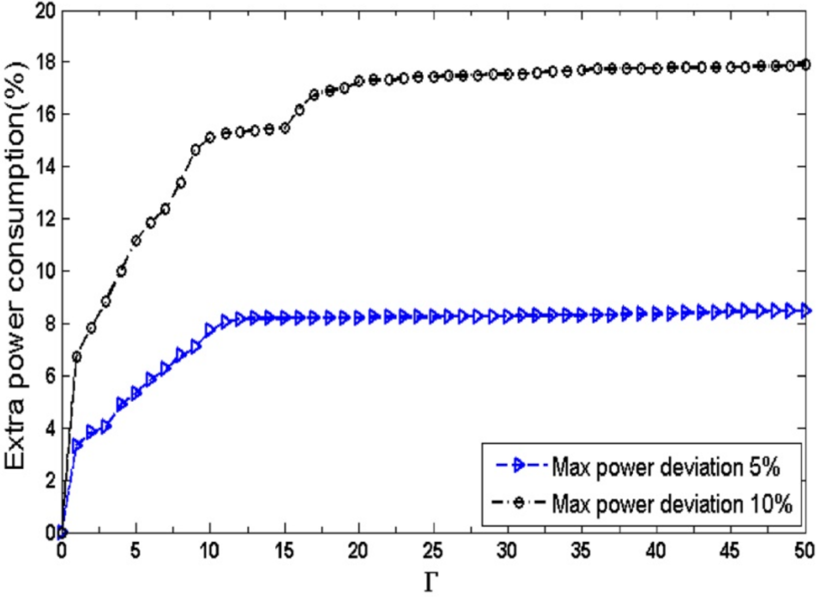
$$tdown_k \cdot z_{jk}^{<-} \leq SLA_k, \forall j, \forall k.$$

Results



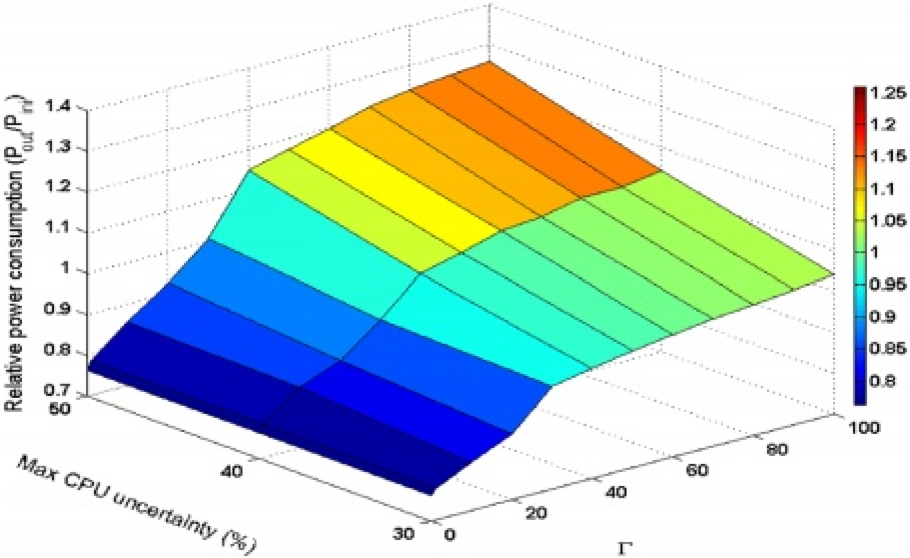
Relative power consumption (P_{out}/P_{ini})
Probability of violation
Γ
Risk adjusted power (Max. power deviation 5%)
Risk adjusted power (Max. power deviation 10%)
Expected power (Max. power deviation 5%)
Expected power (Max. power deviation 10%)
Probability of constraint violation
0.5
0.48
0.46
0.44
0.42
0.4
0.38
10^0
10^{-5}
10^{-10}
10^{-15}
0
5
10
15
20
25
30
35
40
45
50

Results



Extra power consumption(%)
20
18
16
14
12
10
8
6
4
2
0
0
5
10
15
20
25
30
35
40
45
50
Γ
Max power deviation 5%
Max power deviation 10%

Results



Relative power consumption (P_{out}/P_{in})
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
50
40
30
Max CPU uncertainty (%)
0
20
40
60
80
100
Γ
1.25
1.2
1.15
1.1
1.05
1
0.95
0.9
0.85
0.8

References I

T. Bauschert, C. Bsing, F. D'Andreagiovanni, A. M. C. A. Koster, M. Kutschka, and U. Steglich.
Network planning under demand uncertainty with robust optimization.
IEEE Communications Magazine, 52(2):178–185, February 2014.

A. Ben-Tal, L. El Ghaoui, and A.S. Nemirovski.
Robust Optimization.
Princeton Series in Applied Mathematics. Princeton University Press, October 2009.

Dimitris Bertsimas and Melvyn Sim.
The price of robustness.
Operations Research, 52(1):35–53, 2004.

Simone Gast and Alexander Keimer.
Robust optimization in network flows.
03 2009.

References II

Robayet Nasim, Enrica Zola, and Andreas J. Kassler.
Robust optimization for energy-efficient virtual machine consolidation in modern datacenters.
Cluster Computing, 21(3):1681–1709, Sep 2018.