Using Flexibility Envelopes for the Demand-Side Hierarchical Optimization of District Heating Networks

Audrey Blizard, Colin N. Jones, and Stephanie Stockar

Abstract—The demand-side control of district heating networks is notoriously challenging due to the large number of connected users and the high number of states to be considered. To overcome these challenges, this paper presents a hierarchical optimization scheme using the flexibility in heating demand provided by the users to improve the performance of the network. This hierarchical scheme relies on a low level controller to calculate the costs for a subsystem over a given set of potential pressure drops for that subsystem. The high level controller then uses these calculated costs to determine the optimal set of pressure drops for every subgraph of the partitioned network. The proposed hierarchical optimization scheme is demonstrated on a representative 20 user district heating network, resulting in a 67% reduction in bypass mass flow while ensuring all network users stay within 2 °C of their desired nominal temperatures.

I. INTRODUCTION

In 2022, nearly half of the energy used in buildings was for space and water heating [1]. District heating networks (DHNs) are a promising alternative to traditional methods of supplying this heat. By utilizing economy of scale and enabling the easier integration of renewable energy sources, DHNs are a valuable tool to reduce emissions and lower costs, especially in cities. However, existing control methods are not sufficient to capitalize on the full energy saving potential of these networks and advanced controllers are needed to further improve their performance.

Because DHNs are networked systems with many connected users, they provide more flexibility in how heat can be delivered as compared to single-user heating systems, while still ensuring the comfort of the connected users. This flexibility has been exploited by advanced controllers in a variety of ways [2]. For example, it has been used to better meet the electricity demands of the transmission system in a DHN fed by a combined heat and power system [3]. Additionally, the flexibility has been used to reduce total operating costs by considering variable pricing in the electricity supply [4] and incorporate additional renewable energy sources [5].

However, the focus of research on the control of DHNs has been on the interconnection between the network and the heat source to either reduce energy consumption, shift heat demand, or decrease operating costs, by aggregating the network into a single flexible consumer. Significantly less research has been conducted on demand-side management, which aims to modulate the individual user’s behaviors to reduce the total energy consumption of the network. Initial studies have shown that demand-side management is a promising method to reduce energy consumption, with one study showing a 35% reduction in energy consumption is possible [6]. This work proposes combining the flexibility provided by the buildings connected to the network with a demand-side control approach.

Centralized control of the individual users in DHNs is challenging, especially as the system scales due the large number of users and geographic dispersion [2]. Therefore, instead of relying on a centralized controller, a novel hierarchical control structure is proposed. In this structure, the network is decomposed into manageable subsystems to be optimized with assumptions introduced on the total pressure losses in the subsystems.

The low-level controller utilizes the flexibility of the local buildings to minimize the control objective for the local subsystem. Here, the available flexibility of a building will be quantified through the use of flexibility envelopes [7], which provide a measure of a building’s ability to temporarily accept a higher or lower heat supply while still remaining in an acceptable temperature range. The concept of flexibility envelopes has been proven successful for the integration of renewable energy sources [8] and in the control of buildings demand for peak shaving operations [9]. The high-level controller then uses the subsystems’ costs to determine the optimal set of pressure losses and mass flow rates for every subsystem. The resulting hierarchical structure is demonstrated on a 20 user DHN, with the objective to minimize bypass flow rate as compared to the nominal case.

II. SYSTEM MODEL

The three main elements of DHNs that must be modeled to accurately describe the thermal behavior of the network are 1) the heating plant, which supplies water with a certain mass flow rate and temperature, 2) the feeding/return network, which delivers the heated water to the users and returns the cooled water to the plant, and 3) the users, which extract heat from the network to meet their heat demands. It is convenient to represent a DHN as a directed graph \( G = (V, E) \), where \( E \) is the set of edges, which can be decomposed into four disjoint subsets, \( E = \{ E_F, E_R, E_{BY}, E_U \} \) for the feeding lines, return lines, bypass segments, and user segments, respectively, and \( V \) is the set of nodes connecting the
pipes in the network. Associated with the graph \( G \) are an adjacency matrix, \( \Gamma \), and incidence matrix, \( \Lambda \), which are both methods of representing the connections between the nodes and edges in a graph. In the graph, the plant is modeled by a pair of nodes \( v_{\text{root}}, v_{\text{term}} \) where indegree\( (v_{\text{root}}) = 0 \) and outdegree\( (v_{\text{term}}) = 0 \). A sample graph of a DHN is shown in Fig. 1, with the sets of edges highlighted. The behavior of the network components is modeled via a first principle approach as follows.

A. Thermal Model

A reduced version of the graph-based model, originally developed by the authors is used to model the thermodynamic response of the network pipes [10]. The temperature of an individual pipe segment is found using a well-mixed approach according to

\[
\frac{d}{dt} T = c_1 \cdot \dot{m} \cdot T_{\text{in}} + c_2 T_{\text{amb}} - (c_1 \cdot \dot{m} + c_2)T
\]

where \( \dot{m} \) is the mass flow rate through the pipe, \( T_{\text{in}} \) is the inlet temperature in the pipe segment, \( T_{\text{amb}} \) is the ambient temperature, \( T \) is the temperature in the pipe segment, and

\[
c_1 = \frac{1}{\rho V}
\]

\[
c_2 = \frac{h A_s}{\rho c_p V}
\]

where \( \rho \) and \( c_p \) are the density and constant pressure heat capacity of the operating fluid respectively, and \( V \) and \( h A_s \) are the volume and the heat transfer coefficient of the pipe.

Additionally, it is assumed the temperature of the user edge is a known constant, \( T_{\text{set}R} \). This assumption is valid as this temperature can be directly modulated via the speed the operating fluid is re-circulated through the heat exchanger, and the dynamics of the user edges are much faster than those of the rest of the network pipes, allowing these edges to be neglected in the dynamic temperature model.

From this assumption and the previously developed model, the state space representation of the entire network takes the form

\[
\frac{d}{dt} \mathbf{T}_N = A(\dot{n}_E) \mathbf{T}_E + B \begin{bmatrix} T_0 \\ T_{\text{set}R} \end{bmatrix} \]

where \( \mathbf{E}_N = \mathcal{E} \setminus \mathcal{E}_U \) is the set of non-user edges, \( T_E \) is the vector of temperatures in these pipes, \( T_0 \) is the supply temperature, and the \( A \) and \( B \) matrices can be found by placing the mass flow rates and \( c_1, c_2, c_3 \) coefficients based on the interconnections of the pipes in the network.

B. Fluids Model

The mass flow rate in the network can be fully resolved with three sets of equations: the pressure loss equations for each pipe segment, the network conservation of mass equation, and the network pressure balance equation [11]. The pressure loss, \( \Delta P \), in an individual pipe segment can be modeled by

\[
\Delta P = \zeta \dot{m}^2
\]

where \( \zeta \) is the total pressure loss coefficient, determined based on the pipe’s characteristics, such as length, diameter, and material.

The conservation of mass is enforced by

\[
\lambda \dot{n}_E = \dot{n}_V
\]

where \( \dot{n}_E \) is the vector of mass flow rates in the edges and \( \dot{n}_V \) is the vector of the mass flow rates in the nodes, given by

\[
\dot{n}_V(i) = \begin{cases} 
\dot{n}_0 & \text{if } v(i) = v_{\text{root}} \\
-\dot{n}_0 & \text{if } v(i) = v_{\text{term}} \\
0 & \text{otherwise}
\end{cases}
\]

where \( \dot{n}_0 \) is the supply mass flow rate.

Finally, the pressure balance in the network can be found using

\[
\Delta P_E = \Lambda^T \mathbf{P}_V
\]

where \( \Delta P_E \) is a vector of pressure losses in the edges and \( \mathbf{P}_V \) is a vector of the pressure at every node where \( \mathbf{P}_V(v_{\text{term}}) = 0 \) serves as the reference pressure.

C. Flexible Building Model

The users in the network are modeled as flexible consumers of heat, using a similar method to the one developed in Reyniers et al. [12]. The range of flexibility in the heat demand, called the flexibility envelope is given by

\[
C \Delta T_L \leq \int_{t_0}^{t_f} \Delta \dot{Q}_p(t) \, dt + \Delta \dot{Q}_p(t_0) \leq C \Delta T_U
\]

where \( C \) is the heat capacity of the building, \( \Delta T_L \) and \( \Delta T_U \) are the acceptable upper and lower temperature deviation from the nominal desired temperature \( (T_{B,\text{nom}}) \), \( \Delta \dot{Q}_p \) is the deviation from the nominal heat demand \( \dot{Q}_{\text{out}} \), needed to keep the building at its desired temperature, given by

\[
\Delta \dot{Q}_p = \dot{Q}_p - \dot{Q}_{\text{out}}
\]

and \( \Delta \dot{Q}_p(t_0) \) is the initial used flexibility value given by

\[
\Delta \dot{Q}_p(t_0) = C(T_{B,\text{nom}} - T_B(t_0))
\]
In this paper, it is assumed that $\dot{Q}_{\text{out}}$ is a known profile due to the limited range of $\Delta T_L$ and $\Delta T_U$. Additionally, any errors in this profile can be compensated for by updating $\Delta \dot{Q}_p(t_0)$ at each control step.

### III. HIERARCHICAL OPTIMIZATION OF DEMAND

To reduce communication overhead and ensure the problem is scalable, a novel hierarchical structure is proposed where the full network is decomposed into subsystems, each with a local controller. These local controllers serve to exploit the local building flexibility to optimize the subsystems behavior over a set of potential total pressure losses, $\Delta P_{\text{tot}}$ in each subsystem. By considering the flexibility at the local level, this hierarchical structure allows for improved performance within a subsystem while ensuring all other subsystems receive the flow needed to meet user demands via the high level controller. After the local controllers solve for the minimum cost behavior over a given set of $\Delta P_{\text{tot}}$ for each subnetwork, this information is passed to the high-level controller, which determines the best $\Delta P_{\text{tot}}$ value for each subsystem to minimize the total cost while maintaining a total pressure and mass flow balance in the network. A diagram of the proposed framework is shown in Fig. 2.

#### A. Partitioning of Network

The first step in designing the low-level controllers is partitioning the network into $n_g$ smaller sub-networks. In this paper, the normalized cut metric is used to recursively bi-partition the overall network graph into subsystems. This method is commonly used in the partitioning of large-scale systems for distributed control [13] and is easily integrated into the graph-based modeling approach used for the DHN. The normalized cut metric minimizes the cut cost as a fraction of the total edge weight [14], given by

$$N_{\text{cut}}(G_i, G_j) = \left( \sum_{i \in V_i, j \in V_j} w_{ij} \right) \times \frac{1}{\left( \sum_{i \in V_i, k \in V_j} w_{ik} \right) + \left( \sum_{j \in V_j, k \in V_i} w_{jk} \right)}$$  \hspace{2cm} (11)$$

where $G_i = (V_i, E_i), G_j = (V_j, E_j)$ are the two resulting subgraphs, and $w_{ij}$ is the edge weight from node $i$ to $j$.

The optimal normalized cut can be approximated by solving the Rayleigh quotient minimization problem, which is solved using the spectral decomposition of the graph Laplacian via the Fiedler vector [15]. The Fiedler vector is defined as the eigenvector associated with the second smallest eigenvalue of the graph Laplacian [16]. For a weighted, undirected graph, the graph Laplacian, $L$, can be found according to

$$L = I - \left( D^{-1} \right)^{1/2} \Gamma_{ud}(D^{-1})^{1/2}$$  \hspace{2cm} (12)$$

where $I$ is an appropriately sized identity matrix, $\Gamma_{ud}$ is the weighted, undirected adjacency matrix, and $(D^{-1})^{1/2}$ is the square root of the inverse of $D$, where $D$ is the weighted degree matrix given by

$$D_{ij} = \begin{cases} \sum_k w_{ik}, & \forall k \text{ s.t. } e_{ik} \in E \\ 0, & \text{otherwise} \end{cases}$$  \hspace{2cm} (13)$$

For the decomposition of DHNs presented here, the adjacency matrix for the undirected graph is created from the original directed graph according to $\Gamma_{ud} = \Gamma^T \Gamma$, and the edge weights are chosen as $1/\rho V$, the constant term in $c_1$, the coupling coefficient for the temperature in the pipes, indicating the impact the in-flow temperature has on the pipe temperature.

The normalized cut metric relies on removing edges to create two disjoint subgraphs, with no overlapping nodes. However, because the edges represent pipes, removing these edges impacts the pressure losses and temperature dynamics of the resulting subsystems. Therefore, these neglected edges are added back into the appropriate subsystem, based on the root and terminal nodes of each subsystem, resulting in subgraphs with overlapping sets of nodes.

#### B. Low-Level Optimization Problem

The partitioned network can be represented by a set of graphs $\{G^{(1)} \ldots G^{(n_g)}\}$, each with their own incidence matrix. For each such subsystem $j$, the following optimization problem is solved for a predetermined set of $\Delta P_{\text{tot}}^{(j)} = \{\Delta P_{\text{tot}}^{(j)} \ldots \Delta P_{\text{tot}}^{(j)}\}$:

$$\hat{c}_i^{(j)} = \min_{\hat{m}_i^{(j)}, \hat{c}_i^{(j)}>0} f(\hat{m}_i^{(j)}, T_{N_j}, \Delta \dot{Q}_p) \text{ s.t.}$$  \hspace{2cm} (14a)$$

$$\Delta P_{\dot{\xi}_i} = \zeta_\xi \hat{m}_i^{(j)}, \quad \zeta_{U_j} \subset \zeta_\xi$$  \hspace{2cm} (14b)$$

$$A \hat{m}_i^{(j)} = \hat{m}_j^{(j)}$$  \hspace{2cm} (14c)$$

$$\Delta P_{\dot{E}_j} = \Lambda^T P_{\gamma_j}(v_{\text{root},j}) = \Delta P_{\text{tot}}^{(j)}$$  \hspace{2cm} (14d)$$

$$\frac{d}{dt}T_{E_{N_j}} = A(\hat{m}_i^{(j)})T_{E_{N_j}} + B \begin{bmatrix} T_{O_j} \\ T_{\text{setR}} \end{bmatrix} \text{ s.t.} T_{\text{amb}}$$  \hspace{2cm} (14e)$$

$$\hat{Q}_p = \hat{m}_i^{(j)} c_p(T_{in} \varepsilon_{j} - T_{\text{setR}})$$  \hspace{2cm} (14f)$$

where $\hat{m}_i^{(j)}$ is the cost to be minimized. Equations (14b) to (14d) calculate the mass flow rates throughout the subgraph and Eqs. (14e) and (14f) calculates the heat delivered to the buildings in the partition, where $T_{O_j}$ is the local supply temperature and $T_{in} \varepsilon_{j} \subset T_{E_j}$. Finally, Eq. (14g) is a vector
inequality that ensures the heat delivered to all buildings in
the subsystem remain within their flexibility envelopes. The
problem is optimized over the friction coefficients for the
user edges, $\zeta_{ij}$. Each user has a control valve that, when
the position is modulated, changes the friction coefficient,
thereby changing the mass flow rate to the building. This
optimization is solved for the set $\Delta P_{\text{tot}j}$ giving a mapping
from $\Delta P_{\text{tot}j}$ to the optimal costs ($c^{(j)}$) and initial mass
flow rate ($\dot{m}_{ij}$), where $\dot{m}_{ij} \in \dot{m}_{\bar{e}}$, for each subsystem.
This information is transmitted to the high-level controller
to resolve the total network cost.

C. High-Level Optimization Problem

For the high-level controller, a reduced graph, $G_r = (V_r, E_r)$, is created, with a reduced incidence matrix $\Lambda_r$. The
root and terminal node of each subgraph are represented by
the nodes of the reduced graph, and the entire subsystem is reduced to a single edge with an associated set of pressure drops, costs, and initial mass flow rates triplets. From these triplets, the following integer-programming optimization problem is solved by the high-level controller:

$$\min_{I=\{i_1 \ldots i_{n_g}\}} \sum_{j=1}^{n_g} (\dot{m}_{Ij}^t)_{\bar{e},\bar{r}} \quad s.t. \quad \lambda^T \hat{P}_{\text{tot},r} < \epsilon$$

where $I$ is a vector of the selected pressure loss indexes and the vectors in Eq. (15a) and Eq. (15c) are vectors of the chosen values for each subsystem, appended with edges $e_{1 \ldots n_g}$, which are subsystems that do not contain any users, and epsilon is a small constant, indicating an imperfect pressure balance, due to the discretization of potential pressure drops. The result of this optimization problem is transmitted to the low-level controllers, and the optimal valve positions are implemented based on the desired friction coefficients. Additionally, the total initial mass flow rate, found from Eq. (15b) is transmitted to the heating plant. This optimization problem can then be solved in a receding horizon fashion to account for variations in heat demand, ambient temperature, and supply temperature from the forecasted conditions.

IV. RESULTS

A. Description of Case Study

The buildings’ nominal heating demands (Fig. 3) for February 2nd and corresponding OpenStock models were obtained from NREL’s ResStock [17] and ComStock [18] for a variety of buildings in Cook County, Illinois, presented in Table II. OpenBuild [19] was used to develop state-space models of each building, from which the heat capacity was calculated using its step responses. The ambient temperature in typical meteorological year 3 (TMY3) was provided by NREL in the 2021 data set [17].

The DHN graph was created using a variety of typical network characteristics, tabulated in Table I and an illustrative DHN layout, shown in Fig. 1 to contain a variety of network configurations seen in real-world DHNs. The network supply temperature, $T_0$ was assumed to be constant. The simulation was performed using the parallel computing toolbox in the Matlab environment to solve the low-level control problem simultaneously using CasADi [20] and IPOPT on a system equipped with an Intel Core i7-8565U CPU.

The network, partitioned into five sub-networks with the resulting reduced graph to be considered by the high-level

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Value</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density</td>
<td>$\rho$</td>
<td>971</td>
<td>kg/m$^3$</td>
</tr>
<tr>
<td>Specific heat capacity</td>
<td>$c_p$</td>
<td>4179</td>
<td>J/kgK</td>
</tr>
<tr>
<td>Supply temperature</td>
<td>$T_0$</td>
<td>80</td>
<td>C</td>
</tr>
<tr>
<td>Ambient temperature</td>
<td>$T_{\text{amb}}$</td>
<td>-19.5 – -13.9</td>
<td>C</td>
</tr>
<tr>
<td>Pipe diameter</td>
<td>$D$</td>
<td>0.15 – 0.40</td>
<td>m</td>
</tr>
<tr>
<td>Pipe length</td>
<td>$L$</td>
<td>10 – 100</td>
<td>m</td>
</tr>
<tr>
<td>Bypass length</td>
<td>$L_b$</td>
<td>3</td>
<td>m</td>
</tr>
<tr>
<td>Heat transfer coefficient</td>
<td>$h$</td>
<td>1.5</td>
<td>W/m$^2$K</td>
</tr>
<tr>
<td>Friction coefficient</td>
<td>$\lambda$</td>
<td>0.01</td>
<td>-</td>
</tr>
</tbody>
</table>

TABLE II

<table>
<thead>
<tr>
<th>$E$</th>
<th>Building ID</th>
<th>Type</th>
<th>Area ($m^2$)</th>
<th>$C$ (MJ/K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_1$</td>
<td>R-3561</td>
<td>House</td>
<td>160</td>
<td>78</td>
</tr>
<tr>
<td>$e_2$</td>
<td>R-80372</td>
<td>House</td>
<td>760</td>
<td>400</td>
</tr>
<tr>
<td>$e_3$</td>
<td>C-177428</td>
<td>Medical</td>
<td>7000</td>
<td>12562</td>
</tr>
<tr>
<td>$e_4$</td>
<td>C-343832</td>
<td>Retail</td>
<td>3500</td>
<td>8575</td>
</tr>
<tr>
<td>$e_5$</td>
<td>C-18740</td>
<td>Warehouse</td>
<td>1600</td>
<td>2393</td>
</tr>
<tr>
<td>$e_6$</td>
<td>C-123604</td>
<td>Office</td>
<td>700</td>
<td>1511</td>
</tr>
<tr>
<td>$e_7$</td>
<td>C-95364</td>
<td>Retail</td>
<td>7000</td>
<td>5088</td>
</tr>
<tr>
<td>$e_8$</td>
<td>R-20041</td>
<td>Apartment</td>
<td>80</td>
<td>513</td>
</tr>
<tr>
<td>$e_9$</td>
<td>R-28770</td>
<td>Apartment</td>
<td>110</td>
<td>265</td>
</tr>
<tr>
<td>$e_{10}$</td>
<td>R-22719</td>
<td>Apartment</td>
<td>110</td>
<td>657</td>
</tr>
</tbody>
</table>
controller is shown in Fig. 4. These partitions correspond with the major branches of the network and the cuts are made through the longest pipe segments, as longer edges decrease the coupling between neighboring nodes. Due to the chosen network configuration, the low-level controller assumes the local supply temperature is equal to be the global supply temperature, but this assumption can be modified based on network data.

The cost function for the subsystems in Eq. (14a) was chosen to be
\[
f_j(\dot{m}_{E_j}, T_{E_j}, \Delta Q_p) = \int_{t=0}^{t_f} \sum_{E_j \in E_{bj}} \dot{m}_{E_j} \, dt
\]
(16)
where \(E_{bj} \subset E_j\) are the bypass edges, which corresponds to minimizing the flow that is directly recirculated to the return network, but this cost can be modified to consider pumping costs, heat losses or any other desired network metric. The total pressure losses considered for an individual subsystem ranged between 0.5 Pa and 300 Pa. The 24 hour simulation was implemented in a receding horizon fashion where the control horizon one hour with a sampling rate of 10 minutes. Over this one hour optimization horizon, the pressure loss was held constant for the first control interval of 10 minutes, and was unconstrained during the remaining time. The removal of the pressure loss constraint after the first control interval allowed for the optimization to consider upcoming demands, without over-constraining the feasible solution set, as the solution beyond the current control step does not have to be consistent throughout the network. The resulting initial mass flow rate and \(\zeta_{U}\) values for the control interval were then used in a simulation for the total network response every ten minutes. The simulation accounts for the true pressure balance in the network, eliminating the issues that could be caused by the imperfect pressure balance seen in Eq. (15c) and ensuring the flexibility envelopes will not be exceeding during real operation due to this imperfect balance.

B. Results

In supplying the users with their nominal heat demand, where there was no flexibility in the heat supplied, the integral of the sum of the bypass mass flow rates was \(4.01 \times 10^5\) kg. In the optimized case, this cost was \(1.33 \times 10^5\) kg, a 67% reduction in total cost. The costs as a function of time can be seen in Fig. 5. It is also noted that, while the bypass mass flow rate decreased, the total mass supplied to the network increased by 2.2% from \(8.48 \times 10^5\) kg in the nominal case to \(8.66 \times 10^5\) kg in the optimized case. The resulting mass flow rates for all the buildings in the DHN are shown in Fig. 6. The used flexibility is shown in Fig. 7, where the horizontal lines represent the upper limits of the flexibility envelopes. The cost for the low level subsystems for each pressure loss value at optimization step 72 can be seen in Fig. 8. Here, invalid results occurred if the problem constraints can not be met while maintaining the desired pressure drop. The average simulation time for a single partition for the one hour optimization horizon was 7.4 minutes, allowing the presented optimization scheme to be successfully implemented in real time.
Fig. 8. Low level costs for considered pressure drops in all subsystems at optimization step 72.

V. CONCLUSIONS

This paper provides a framework for implementing a hierarchical optimization scheme for the demand-side control of large-scale district heating networks. It relies on the flexibility provided by individual network consumers to minimize the subsystems’ costs while ensuring the building temperatures remain in an acceptable range. The low-level controllers optimize this cost given a set pressure losses and find the mass flow rate provided to the subsystem. The high-level controller then selects the optimal mass flow, pressure pairs to optimize the entire network’s performance. The hierarchical framework was demonstrated in the open-loop optimization of a medium-scale 18 user DHN, with the goal of minimizing bypass flow. The results were compared to the case where the nominal heat demand was provided to each user and resulted in an 67% reduction in unused mass flow. Future work will be to expand the hierarchical framework to consider more complex network configurations with additional heat sources and pumps. Additionally, with the selected control objective, the total mass flow needed by the network increased due to the tendency for the optimization problem to heat the buildings to the upper limit of their flexibility envelope. Future work will look to add a penalty term to the cost function to prevent this undesired behavior and the associated initial mass flow spike. It will also look to consider additional cost elements such as pumping costs, heat losses, and the total heat supplied to the network.

REFERENCES