

Optimal design of line replaceable units

Joni Driessen¹ | Joost de Kruijff¹  | Joachim Arts²  | Geert-Jan van Houtum¹ 

¹School of Industrial Engineering,
Eindhoven University of Technology, Eindhoven,
The Netherlands

²Luxembourg Centre for Logistics and Supply
Chain Management, University of Luxembourg,
Luxembourg, Luxembourg

Correspondence

Joachim Arts, Luxembourg Centre for Logistics
and Supply Chain Management, University of
Luxembourg, Luxembourg, Luxembourg.
Email: joachim.arts@uni.lu

Funding information

Nederlandse Organisatie voor Wetenschappelijk
Onderzoek, Grant/Award Number: 407-12-001

Abstract

A line replaceable unit (LRU) is a collection of connected parts in a system that is replaced when any part of the LRU fails. Companies use LRUs as a mechanism to reduce downtime of systems following a failure. The design of LRUs determines how fast a replacement is performed, so a smart design reduces replacement and downtime cost. A firm must purchase/repair a LRU upon failure, and large LRUs are more expensive to purchase/repair. Hence, a firm seeks to design LRUs such that the average costs per time unit are minimized. We formalize this problem in a new model that captures how parts in a system are connected, and how they are disassembled from the system. Our model optimizes the design of LRUs such that the replacement (and downtime) costs and LRU purchase/repair costs are minimized. We present a set partitioning formulation for which we prove a rare result: the optimal solution is integer, despite a nonintegral feasible polyhedron. Second, we formulate our problem as a binary linear program (BLP). The article concludes by numerically comparing the computation times of both formulations and illustrates the effects of various parameters on the model's outcome.

KEYWORDS

column generation, graph theory, integer programming, line replaceable units

1 | INTRODUCTION

System failures are major frustrations for their users. The consequences of failures can vary from discomfort, to disutility, to direct cost penalties such as downtime cost. In particular, industries that rely on systems to render a service or to manufacture a product experience high downtime cost. For example, the downtime cost of a computer system of a brokerage company is roughly between \$100 000 and \$1 000 000 per hour (CNET News, 2001) and the downtime cost in the semiconductor industry is in the order of magnitude of \$100 000 per hour (Parent, 2000). Together, the cost of downtime and maintenance can constitute up to 70%–80% of a system's total life cycle costs (Öner et al., 2007). As a consequence, it is crucial for a company's profitability to minimize the cost of downtime and maintenance. Significant cost reductions can be realized during the design phase of a system

with relatively little effort. Asiedu and Gu (1998) show that 70%–85% of the total life cycle costs are determined during the design phase, even though costs accrued in this phase only accounts for 10%–20% of the life cycle costs (Norman, 1990; Öner et al., 2007; Saranga & Kumar, 2006). Hence, practice and research explore various system design concepts that enable a reduction of the after-sales costs including downtime and maintenance costs. Such concepts include common components (Briant & Naddef, 2004; Driessen et al., 2017; Thonemann & Brandeau, 2000), reliability and redundancy optimization (Öner et al., 2013; Xie et al., 2014), and a smart design of line replaceable units (LRUs) (Parada Puig & Basten, 2015).

A LRU is a collection of connected system parts that can be easily replaced when one of the parts in the LRU fails. An example of such a LRU is the wheel of a car that can be quickly replaced by a spare wheel in case of a tire

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Naval Research Logistics* published by Wiley Periodicals LLC.

puncture. The design of LRUs has a particularly large effect on downtime and maintenance, because they directly determine long a maintenance intervention lasts (Birolini, 2007; Dhillon, 1999; Kumar et al., 2012; Muckstadt, 2004). It is crucial for companies to design the LRUs in a smart way that minimizes the maintenance and downtime costs. Discussions with our project partners, ASML (a manufacturer in the semiconductor industry) and Dutch Railways (a maintenance, repair and overhaul [MRO] company in the railway industry), indicate that this problem is highly relevant for their industries. Similarly, two case studies (Parada Puig & Basten, 2015; van Geel, 2018) at Thales (a manufacturer in the defense industry) and a case study (van Deursen, 2020) at Canon (a manufacturer of industrial printing equipment) report the same relevance, and other examples where LRUs are deliberately designed include Durand (2001), Brasseur et al. (2012), Air France Industries KLM Engineering and Maintenance (2013), and Klauke et al. (2015). On a more general note, the problem of designing LRUs is relevant for firms that are responsible for operational aspects (downtime and the maintenance) of systems. Such a firm may be an original equipment manufacturer (OEM) that closes service contracts with its customers, but it may also be a firm that sells MRO as a service to the user. Examples of OEMs that consider the design of LRUs include PACCAR or Volkswagen Group (trucking industry), Airbus (aviation industry), and the aforementioned manufacturers ASML, Thales, and Canon; while examples of MROs that consider the design of LRUs include the aforementioned Dutch Railways and Air France Industries KLM Engineering & Maintenance (aviation industry).

LRUs are designed based on a given system design in practice. At an OEM, who designs the systems itself, the engineers design the system and subsequently they define the LRUs. The problem of defining the LRUs for a given design of the system is called LRU design. If development time allows, the OEM's design department may redesign their system design based on the outcome of a LRU design, and subsequently determine the LRUs again. A MRO does not design the system itself and thus starts with a given design of the system, when designing the LRUs. Similar to the OEM, the outcome of the LRU design may lead a MRO to modify the initial system design and derive the LRUs from this revised design. This interactive process between system design and LRU design is a powerful concept to reduce the total expected costs for an OEM or MRO.

A system consists of various critical parts that are all connected to each other. As soon as one of these parts fails, the entire system is down and the company incurs downtime cost. To reduce the downtime cost, it is essential to quickly restore the system to a functioning state by replacing the failed part. Firms typically group parts in LRUs such that each LRU can be replaced quickly if any of the parts in this LRU fails. As designers make larger LRUs, the purchase costs typically increase, because the LRU contains more parts and thus more value. The purchase cost can refer to the purchase cost

of a new LRU or to the purchase cost of a repair depending on the context. The failure of any part in a LRU triggers the failure of the entire LRU. The LRU's failure rate therefore equals the sum of the failure rates of the parts in the LRU.¹ Larger LRU incur higher purchase cost and fail more often than smaller LRUs. Now, the challenge is to optimally design the LRUs such that the average costs per time unit are minimized. We address this problem in this article when a LRU design is required to partition the parts into LRUs. The first reason for considering this partition constraint comes from interviews with our project partners. If parts are partitioned into LRUs, replacement ambiguity is avoided. Replacement ambiguity is the situation in which a part fails and it is contained in more than one LRU. The engineer can then decide what LRU to replace and this leads to each system being maintained differently, which is undesirable from an asset configuration management perspective. Second, partitioned LRUs simplify the OEM's/MRO's production and functional testing procedures of LRUs (van Geel, 2018, p. 19). A third reason for a partitioning constraint is that it is in line with common practice in engineering for reliability and maintenance Birolini (2007, p. 154). Nevertheless, the problem can also be solved without the partition constraint as shown in the appendix. In the remainder (Section 5) we numerically illustrate that the partition restriction leads to very limited increased costs, while it significantly increases the practical applicability of the LRU design.

Our LRU design problem relates to multi-component maintenance research with structural dependencies. Structural dependency between parts occurs when some parts have to be replaced or removed before the failed part can be replaced. Practitioners frequently face this type of dependency, but the academic field studying it "is wide open ... and there have only been a few articles published on this topic" (Nicolai & Dekker, 2008). The problem of designing LRUs naturally falls into this class, and only two papers address this problem. Thomas (1986) poses the question whether to replace the entire car, the engine or just the piston rings in case the piston rings need replacement. More recently, Parada Puig and Basten (2015) have revisited the question posed by Thomas (1986) and propose a model to design LRUs based on a narrow set of potential LRUs. Both works start from a bill of materials structure, that is, the system's structure is a tree. The major issue with tree structure is that it does not capture the structural dependencies between parts. Tree structures cannot model the connections that exist and that have to be broken in order to replace a failed part, that is, the connections between parts need not adhere to a tree structure. By contrast, we focus on the connections (cables, hoses, bolt-nut connections etc.) between various parts in a system, that is, we consider the structural dependencies between parts. We explicitly incorporate disassembly sequences that exist

¹This holds when the part's time between failures is exponentially distributed and represents a good approximation in other circumstances.

for maintenance based on connections. Literature typically considers part based disassembly sequences: Part B must be disassembled before part A can be disassembled (De Fazio & Whitney, 1987; Gupta & Krishnan, 1998; Lambert, 2007). This means that *all* connections that part B has with other parts have to be broken before the connections of part A can be broken that enable part A's replacement. We capture these disassembly dynamics by considering the connections between parts. The connection oriented disassembly sequence has a major benefit over the part based disassembly sequence: it allows a *subset* of a part's connections to be broken. For instance, if we want to replace the part A, connection based disassembly sequences allow a subset of part B's connections to be broken; whereas the part based disassembly sequence forces all connections of part B to be broken. Only breaking a subset of connections is common in practice—as observed at our project partners, for example, part B can be tilted after unmounting the bolts and leaving a hose still attached to this part. Subsequently, a part A can be reached and replaced without fully disconnecting part B.

Modeling the disassembly of a system based on its connections and the subsequent disassembly sequence is new and it enables us to accurately model the time needed to replace any LRU. Thus, we endogenize the replacement time and therewith the replacement cost of a LRU. This contrasts with Parada Puig and Basten (2015) where the replacement time and cost are exogenously given. Furthermore, our modeling considers the full set of potential LRUs, because a LRU—in our model—can be any combination of parts in the system, contrasting Thomas (1986) and Parada Puig and Basten (2015) who must pre-define all potential LRUs and their corresponding parameter values.

Another line of related research studies LRUs (which are called modules) from a systems engineering perspective, where “a module is a unit whose parts are powerfully connected among themselves, and relatively weakly connected to parts in other units” (Baldwin & Clark, 2000). This literature stream typically describes a system in terms of parts that are connected to each other, and these connections are commonly depicted in a design structure matrix (DSM) (Steward, 1981). However, this approach neglects the disassembly sequence dynamics that exist for the replacement of LRUs (or modules). Papalambros and Michelena (1997) manage to relate this line of research to the area of optimization. Most research in the DSM stream aims to define measures of modularity and optimize these. Such measures typically focus on the connections between parts, and the measures prefer a high number of intra-LRU connections and a low number of inter-LRU connections; see Newcomb et al. (1998), Sharman and Yassine (2004), Sosa et al. (2007), and Wilschut et al. (2017). Optimization of the defined measures is typically done by using genetic algorithms (Meier et al., 2006; Yu et al., 2007) or simulated annealing (Thebeau, 2001). The aforementioned research focuses on single product DSMs, whereas Alvaro

and Harrison (2011) and Kim et al. (2021) apply this approach to a product family.

Work on (dis)assembly sequencing also has similarities to our work, because this stream models the (dis)assembly sequence that exists between parts in much detail (De Fazio & Whitney, 1987; Gupta & Krishnan, 1998; Lambert, 2007). Research in this area optimizes the (dis)assembly sequence. We do not optimize this sequence, but we consider it to be given and focus on optimizing the design of LRUs.

Finally, our work relates to several operations research studies that consider the impact of modular design on operations. These studies are often combinatorial in nature and aim to design product configurations such that the demand for end products is met and the average costs per time unit are minimized; see for example Swaminathan and Tayur (1998), Thonemann and Brandeau (2000), and Briant and Naddef (2004). The structure in their problems superficially resembles ours, because we also study configurations of parts, which are LRUs in our case. The main difference is that we model the connections between parts and the disassembly sequences that exist for maintenance, while research in this stream does not.

In this article, we make the following contributions: we present (i) a novel way to represent a system with multiple parts that are connected to each other, and we incorporate the disassembly sequences that exist for maintenance based on connections rather than parts. Modeling the connections and disassembly sequences enables us to endogenize the downtime cost due to the replacement of a LRU containing the failed part. Next, we use our system description to define an optimization model—called LRU DESIGN—that minimizes the sum of the replacement and purchase cost by optimizing the LRU designs.

We provide (ii) a set partitioning formulation of LRU DESIGN that allows for branch-and-price algorithms. Next, we prove (iii) that an optimal solution to the set partitioning formulation is integer. This result is rather remarkable, because the feasible polyhedron is not integral. There exist two problems that also possess this property: a minimax transportation problem (Ahuja, 1986) and a multi-period machine assignment problem (Zhang & Bard, 2006). The majority of research typically shows the existence of an optimal integer solution by proving that the feasible polyhedron is integral (e.g., through total unimodularity), see for instance Hillier and Brandeau (1998), Ball et al. (2003), Churchill et al. (2012), Gamvros and Raghavan (2012). Our integrality result cannot be established in this way due to the nonintegral polyhedron. We define a so-called LRU cycle and prove that a solution that contains such a LRU cycle is suboptimal. Subsequently, we study the matrix encoding of an optimal solution to prove that an optimal solution is integer. We believe that our proof approach is applicable and promising to other problems that can be formulated as set partitioning problems, because one's main effort would be to prove suboptimality of partitions that contain cycles.

Fourth (iv), we focus on additive failure rates such that we obtain linear expressions that support the implementation of our models. We specify the set partitioning formulation under such additive failure rates and we formulate LRU DESIGN as a binary non-linear program (BNLP), which we then transform into a BLP.

Finally, (v) we illustrate that the set partitioning formulation is suitable for large instances, and we study the effects of various parameters on the model's outcome. Moreover, we numerically show that the relative cost impact of introducing a partitioning constraint is very limited, while we strongly improve practical applicability of our model.

The rest of this article is organized as follows. In Section 2, we discuss our system representation including the disassembly sequences, and the optimization model LRU DESIGN. We present a set partitioning formulation of LRU DESIGN in Section 3, and we prove that an optimal solution is integer for this formulation. In the succeeding part of the article, we focus on additive failure rates (for implementation convenience) and we present the set partitioning formulation under this assumption. Moreover, we discuss a binary non-linear programming (BNLP) formulation of LRU DESIGN in Section 4, which we subsequently linearize to obtain a BLP. Finally in Section 5, we numerically compare the computation times of the BLP formulation to the set partitioning formulation, we illustrate the effects of various parameter perturbations on the model's outcome, and illustrate the limited cost impact of the partitioning constraint in our problem. We offer concluding remarks in Section 6.

2 | MODEL

First this section explains how a system with parts and connections can be represented by means of an example. The generalization then follows from the example and we present the optimization model called LRU DESIGN.

2.1 | An illustrative example

Consider a laptop repair shop that repairs laptops by removing failed parts from the laptop and replacing the failed parts with new ones. The repair shop's objective is to design the LRUs such that it minimizes the cost of repair time and procurement of new LRUs. We consider the illustrative example of a laptop, because this system is technologically simple and many people have some familiarity with it. Bear in mind however that the model was designed for and has greater financial impact for large and technologically complex systems. Unfortunately, such a system lacks the familiarity of general readership.

The laptop example is based on data for a Dell Precision 7710 laptop (Dell Inc., 2016). Each part has a purchase cost and a failure rate (in failures per year). Table 1 lists the estimated purchase cost and fictitious failure rate for all parts (in

TABLE 1 Part identifier list.

Identifier	Part name	Part cost (\$)	Failure rate (failures/year)
A	Battery	180	0.3
B	Hard disk drive	170	0.2
C	Keyboard	45	0.001
D	WLAN card	50	0.15
E	Palm rest	45	0.001
F	Speakers	14	0.05
G	Heat sink	75	0.1
H	4 GB video card	250	0.1
I	Display housing	40	0.001
J	Display front cover	20	0.001
K	Display bezel	170	0.25
L	Motherboard	270	0.25
M	Computer base	50	0.001

failures per year). The purchase cost of a part is its price found online on websites such as <https://www.amazon.com>.

Each of the parts is connected to other parts, for example, the palm rest is screwed to the computer base, the palm rest is wired to the motherboard, and the palm rest is screwed to the keyboard. Thus, there exist connections $\{E, M\}$, $\{E, L\}$, and $\{C, E\}$. In the event the palm rest fails and one wishes to replace it individually, one has to break all the connections that the palm rest has with all other parts: $\{E, M\}$, $\{E, L\}$, and $\{C, E\}$. Breaking each connection takes a certain amount of time, which translates into costs by multiplying the time with a cost rate, for example, the salary rate of the repair man or downtime penalty. When the failed palm rest has been disconnected from the system, a new and identical palm rest from stock is installed into the system by reconnecting all the connections that have been broken previously (in order to remove the failed palm rest). This re-establishing of connections also costs time and can be translated into costs as well. Finally, a new palm rest is purchased to replenish the stock.

All information about parts, connections, failure rates, purchase costs, and the costs of breaking and re-establishing connections can be represented in a weighted un-directed graph. The parts correspond to vertices, and the part connections correspond to the edges. Furthermore, the failure rates and the purchase costs are attributes of the vertices, and the costs for breaking and re-establishing a connection correspond to the weight of an edge in the graph. For the laptop example, this graph can be found by analyzing the Owner's manual (Dell Inc., 2016), and is given in Figure 1. The cost of breaking and re-establishing a connection is an estimate and is depicted on the edges.

We call the graph in Figure 1 the *connection graph*. The connection graph may suggest that we only need to break connections $\{E, L\}$, $\{E, M\}$, and $\{C, E\}$ in order to remove the palm rest. However, the Owner's manual states that in order to disconnect the palm rest, one must first break the

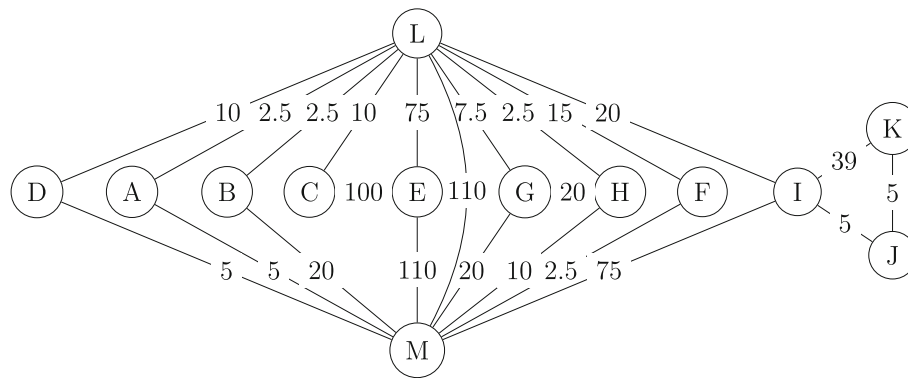


FIGURE 1 The laptop's connection graph.

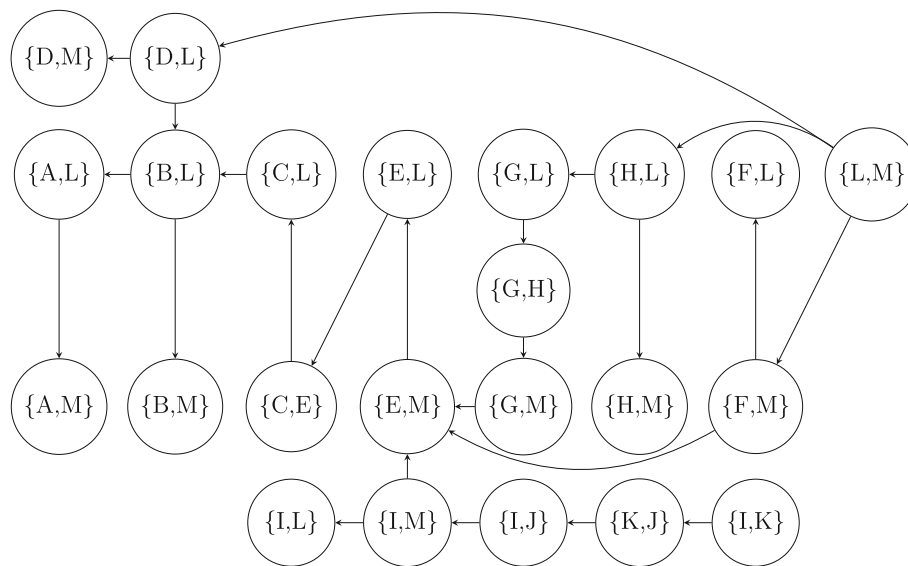


FIGURE 2 The laptop's precedence graph.

connections that enable one to remove the keyboard (C), the hard disk drive (B), and the battery (A); that is, there is a disassembly sequence. This implies that there exists a collection of connections that needs to be broken prior to breaking the connections $\{E, L\}$, $\{E, M\}$, or $\{C, E\}$ (Dell Inc., 2016). Therefore there is a predecessor–successor relationship for breaking (and re-establishing) the connections depicted in Figure 1. We model such predecessor–successor relationships in a separate directed graph, which we call the *precedence graph*. An arc in the precedence graph from an edge $\{E, M\}$ to $\{E, L\}$ implies that connection $\{E, L\}$ must be broken before connection $\{E, M\}$ can be broken. Figure 2 shows the precedence graph for the laptop (Dell Inc., 2016).

The combination of the precedence graph (Figure 2) with the connection graph (Figure 1), enables us to list *all* connections that need to be broken for the replacement of an arbitrary part. For example, replacement the palm rest requires one to break $\{E, M\}$, $\{E, L\}$, and $\{C, E\}$ (see Figure 1), but to break connection $\{C, E\}$ one must first break the set of connections $\{\{A, L\}, \{A, M\}, \{B, M\}, \{B, L\}, \{C, L\}\}$ (see Figure 2). Similarly, one can determine all connections

that need to be broken prior to $\{E, M\}$ and $\{E, L\}$. Finally, one must break all connections $\{\{A, L\}, \{A, M\}, \{B, M\}, \{B, L\}, \{C, E\}, \{C, L\}, \{E, L\}, \{E, M\}\}$ in order to remove the palm rest (E). Analogously, one must break connections $\{\{A, L\}, \{A, M\}, \{B, M\}, \{B, L\}, \{C, E\}, \{C, L\}\}$ to replace the keyboard.

If one decides to replace the palm rest (E) together with the keyboard (C), that is, define a LRU Q that contains C and E. However, this implies that the engineer has to break all connections $\{\{A, L\}, \{A, M\}, \{B, M\}, \{B, L\}, \{C, L\}, \{C, E\}, \{E, L\}, \{E, M\}\}$ upon the failure of either the palm rest (E) or the keyboard (C). As a consequence, one must break the expensive edges $\{E, L\}$ and $\{E, M\}$ more often than when the palm rest and the keyboard are separate LRUs. Furthermore, the LRU Q has a higher purchase cost as well as a higher failure rate compared to the palm rest and the keyboard individually. Therefore, it is better to keep the palm rest and the keyboard as separate LRUs instead of combining these two into one LRU Q . It is now of interest to find the optimal design of LRUs that minimizes the sum of the replacement and purchase costs, based on the connection graph in Figure 1 and the precedence graph in Figure 2.

2.2 | A generic model

The example above illustrates how a system is built up and what relationships parts and connections have. The approach used for the laptop also applies to more complicated systems such as a bogie in a train (Bombardier/Dutch Railways), a positioning module in a lithography system (ASML), a truck engine (PACCAR/Volkswagen Group), or a jet engine (Pratt & Whitney). Consider a system that consists of multiple parts, and assume that maintenance is done upon the failure of a part. Moreover, assume that one can accurately and instantaneously determine which part has failed, when the system fails as a whole. The system is defined by two graphs: a weighted undirected connection graph G and a directed precedence graph D . The graph $G = (V, E)$ is characterized by the set of vertices V and the set of edges E . The former set V corresponds to the parts in the system, and the latter set E corresponds to the connections between parts. Furthermore, each part in G has a purchase cost $\ell : V \rightarrow \mathbb{R}_+$, where $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x > 0\}$, and the failure rate for any subset $Q \subseteq V$ is given by $\lambda : 2^V \rightarrow \mathbb{R}_+$. We also assume that the failure rate function is subadditive, that is, $\lambda(R) + \lambda(T) \leq \lambda(R \cup T)$, $\forall R, T \subseteq V$. The cost to break a connection are given by the edge costs $w : E \rightarrow \mathbb{R}_+$. We use the terms part and vertex interchangeably, as well as the terms connection and edge. At the end of this section, we discuss what happens when LRUs are repaired rather than purchased, see Remark 1.

Besides the connection graph G , the precedence graph $D = (E, A)$ is an unweighted acyclic directed graph that captures the disassembly sequences of the connections $e \in E$. The set A corresponds to the set of arcs, and an arc $(i, j) \in A$ from edge i to edge j exists if and only if edge j has to be broken *before* edge i can be broken. We assume that arcs can only connect adjacent edges, that is, all arcs in A satisfy $(\{u, v\}, \{v, x\}) \in A : u, v, x \in V$ and $u \neq v \neq x$. The graph D determines a set $H(e)$ of successor edges for each edge $e \in E$. This set $H(e)$ consists of all edges including the edge $e \in E$ that must be disconnected in order to break e , and it can be determined by using the polynomial time Algorithms 1 and 2 in Appendix A. We remark that $H(e)$ is a directed tree rooted at $e \in E$.

Further, we assume that G is connected, without loss of generality. If G is not connected, there do not exist arcs $(\{u, v\}, \{x, y\}) \in A$ such that u, v are in one connected component and x, y are in the other connected component. This follows because all arcs in A satisfy $(\{u, v\}, \{v, x\}) \in A : u, v, x \in V$ and $u \neq v \neq x$. Hence, if G were disconnected, we apply our model to each connected component of G with the precedence graph induced by the connected component. We define a LRU design as partition S of the vertices V . Connections need to be broken in order to replace a LRU $Q \in S$ from the system. First, define the set $B(Q) = \{\{u, v\} \in E : u \in Q, v \in V \setminus Q\}$, as the set of all edges that connect the LRU to the other parts of the system not in the LRU. That is, the

set $B(Q)$ contains the edges that cross the LRU's boundary. Next, for the removal of LRU Q , one must break all the edges $e \in B(Q)$, as well as all the edges that need to be broken prior to breaking any edge $e \in B(Q)$. Hence, $\Gamma(Q) = \bigcup_{e \in B(Q)} H(e)$ is the set of edges that need to be broken in order to replace a LRU $Q \in S$. Note that $\Gamma(Q)$ may contain edges between vertices in Q . This is a model feature as it allows us to model LRUs such as a chain between two cogwheels. If each link in a chain is a vertex that is connected to the adjacent vertices, then it is necessary to break the edge between two links in order to remove the chain from the cogwheels. A detailed example of this is provided in Appendix B.

Each LRU Q has a purchase cost and failure rate. The purchase cost of a LRU is given by the sum of the purchase cost of all parts in the LRU, that is, the LRU's purchase cost is given by $\sum_{v \in Q} \ell(v)$. We relax this assumption in Remark 1. The total failure rate of a LRU $Q \in S$ is denoted by $\lambda(Q)$.

Next, we derive the cost expression for a LRU $Q \in S$. Upon the failure of LRU Q , one breaks all edges $e \in \Gamma(Q)$ resulting in the cost $\sum_{e \in \Gamma(Q)} w(e)$. Moreover, replacement LRU is purchased at cost $\sum_{v \in Q} \ell(v)$. The average cost per time unit of LRU Q then satisfies

$$\omega(Q) = \lambda(Q) \left(\sum_{e \in \Gamma(Q)} w(e) + \sum_{u \in Q} \ell(u) \right). \quad (1)$$

As a LRU design S is a partition of V and $Q \in S$, the total cost per time unit of a LRU design is given by

$$\pi(S) = \sum_{Q \in S} \omega(Q). \quad (2)$$

The LRU DESIGN problem can now be simply stated as: What is the LRU Design S that minimizes $\pi(S)$?

LRU DESIGN has the property that each LRU Q in optimal solution S^* to LRU DESIGN is a connected subgraph of G .

Lemma 1. *Each LRU $Q \in S^*$ is a connected subgraph of G , for any optimal solution S^* to LRU DESIGN.*

Proof. Let S^* be an optimal solution to LRU DESIGN, and let \mathcal{J} be the finite set of connected components in the subgraph induced by a LRU $Q \in S^*$. The set \mathcal{J} partitions Q , \mathcal{J} is finite because Q is finite, and $|\mathcal{J}| \geq 1$. The case of $|\mathcal{J}| = 1$ implies that Q is connected, which satisfies our claim. Thus, we consider the case $|\mathcal{J}| \geq 2$ in the remainder, and observe that in this case $\exists \{u, v\} \in E : u \in \mathcal{J}_1, v \in \mathcal{J}_2$ with $\mathcal{J}_1, \mathcal{J}_2 \in \mathcal{J}$. Note further that $\Gamma(\mathcal{J}) \subseteq \Gamma(Q)$ for a $\mathcal{J} \in \mathcal{J}$ as $\mathcal{J} \subset Q$. Thus,

$$\begin{aligned} \sum_{\mathcal{J} \in \mathcal{J}} \omega(\mathcal{J}) &= \sum_{\mathcal{J} \in \mathcal{J}} \lambda(\mathcal{J}) \sum_{e \in \Gamma(\mathcal{J})} w(e) + \sum_{\mathcal{J} \in \mathcal{J}} \lambda(\mathcal{J}) \sum_{u \in \mathcal{J}} \ell(u) \\ &\leq \sum_{\mathcal{J} \in \mathcal{J}} \lambda(\mathcal{J}) \sum_{e \in \Gamma(Q)} w(e) + \sum_{\mathcal{J} \in \mathcal{J}} \lambda(\mathcal{J}) \sum_{u \in \mathcal{J}} \ell(u) \end{aligned}$$

$$\begin{aligned}
&< \sum_{\mathcal{J} \in \mathcal{F}} \lambda(\mathcal{J}) \sum_{e \in \Gamma(Q)} w(e) + \sum_{\mathcal{J} \in \mathcal{F}} \lambda(\mathcal{J}) \sum_{u \in Q} \ell(u) \\
&\leq \lambda(Q) \sum_{e \in \Gamma(Q)} w(e) + \lambda(Q) \sum_{u \in Q} \ell(u) = \omega(Q),
\end{aligned}$$

where the first inequality follows from the fact that each $\Gamma(\mathcal{J}) \subseteq \Gamma(Q)$, $\forall \mathcal{J} \in \mathcal{F}$. The second inequality follows from the fact that $\mathcal{J} \subset Q$, $\forall \mathcal{J} \in \mathcal{F}$, and thus $\sum_{u \in \mathcal{J}} \ell(u) < \sum_{u \in Q} \ell(u)$, $\forall \mathcal{J} \in \mathcal{F}$. The last inequality holds because \mathcal{F} partitions Q and the failure rate function λ is sub-additive. Hence, we obtain $\sum_{\mathcal{J} \in \mathcal{F}} \omega(\mathcal{J}) < \omega(Q)$, where \mathcal{F} partitions Q . However, this contradicts the optimality of S^* . Therefore, each LRU $Q \in S^*$ is a connected subgraph of G , for any optimal solution S^* to LRU DESIGN. ■

We will use Lemma 1 throughout this article. We conclude this section with two remarks that enable further generalization of LRU DESIGN.

Remark 1. We assumed that we do not repair a LRU, and thus purchase a new one. If we relax this assumption and repair a failed part of a LRU offline, we incur a total repair cost per time unit of $\lambda(V) \sum_{v \in V} q(v)$, where $q(v)$ is the repair cost of part v . However, if we repair a part of a LRU, we have to test the entire LRU to see whether it functions again. This means that we have to test each part in the LRU, and thus $\ell(v)$ now represents the cost of testing part $v \in V$ offline. Larger LRUs, now, have more parts that need to be tested before the LRU is certified as repaired. The total repair cost per time unit is sunk as $\lambda(V) \sum_{v \in V} q(v)$ is independent of the LRU design, but we still have the testing cost per time unit of LRU Q given by $\lambda(Q) \sum_{u \in Q} \ell(u)$. Hence, our model LRU DESIGN still applies.

3 | SET PARTITIONING FORMULATION

We formulate LRU DESIGN as a set partitioning problem that allows for column generation (branch-and-price) algorithms. Then, we prove in Section 3.1 that an optimal solution to the relaxed master program is integer, even though the feasible polyhedron is *not* integral. Finally, we present the column generating procedure in Section 3.2 for solving the set partitioning formulation of LRU DESIGN.

A LRU design S consists of various non-intersecting LRUs $Q \in S$ that have been selected. Let $\mathcal{S} = 2^V$ be the power set of V from which LRUs can be selected; \mathcal{S} contains all possible LRUs. Then, $S \subset \mathcal{S}$, and our objective is to determine which solution S is optimal via column generation. A LRU $Q \in \mathcal{S}$ can equivalently be represented as a $(0, 1)$ column with $|V|$ elements, where a 1 indicates that a vertex is in the LRU Q

and a 0 denotes that the vertex does not belong to the LRU Q . Hence, we consider the matrix entries z_{vQ} that equal 1 if $v \in Q$ and 0 otherwise. Then, a column from the matrix $Z = (z_{vQ})$ corresponds to LRU Q , and we denote this column by Z_Q . Note that a column Z_Q and the LRU $Q \subseteq V$ are equivalent representations of a LRU.

Let x_Q be the indicator variable that denotes whether a LRU $Q \in \mathcal{S}$ is selected for the LRU design S . We denote \mathbf{x} as the vector consisting of all entries x_Q . Given \mathbf{x} , we can straightforwardly derive the solution S to LRU DESIGN by $S = \{Q \in \mathcal{S} : x_Q > 0\}$. We remark that S can equivalently be represented as the submatrix $\tilde{Z} = \{Z_Q : x_Q > 0\}$ of Z . Our objective is to determine the LRU design in terms of x_Q such that the average costs per time unit are minimized, and each part $v \in V$ is included in exactly one LRU. We capture this in the master problem (M):

$$(M) \quad \min_{\mathbf{x}} \sum_{Q \in \mathcal{S}} \omega(Q)x_Q, \quad (3a)$$

$$\text{s.t.} \quad \sum_{Q \in \mathcal{S}} z_{vQ}x_Q = 1, \quad \forall v \in V, \quad (3b)$$

$$x_Q \in \{0, 1\}, \quad \forall Q \in \mathcal{S}. \quad (3c)$$

Recalling that $\omega(Q)$ is the average cost per time unit of using LRU Q , the objective function (3a) minimizes the costs of using the selected LRUs, while constraints (3b) enforce that each part $v \in V$ is included in exactly one LRU $Q \in \mathcal{S}$. The set \mathcal{S} is exponentially large, so straightforward optimization is not tractable. Therefore, we propose to solve the LP relaxation of M by column generation. We relax the integrality of x_Q to obtain the LP relaxation of the master problem called LPM:

$$(LPM) \quad \min_{\mathbf{x}} \sum_{Q \in \mathcal{S}} \omega(Q)x_Q \quad (4a)$$

$$\text{s.t.} \quad \sum_{Q \in \mathcal{S}} z_{vQ}x_Q = 1, \quad \forall v \in V, \quad (4b)$$

$$0 \leq x_Q \leq 1, \quad \forall Q \in \mathcal{S}. \quad (4c)$$

Subsequently, we present our procedure for solving LPM in Section 3.2.

3.1 | Integrality and polyhedral structure of LPM

We prove that an optimal solution to LPM is integer by considering a so-called LRU cycle. We show that if a given fractional solution contains a LRU cycle, there exists a feasible solution to LPM without the LRU cycle and strictly lower costs. This implies that an optimal solution does not contain a LRU cycle.

Let $\tilde{\mathbf{x}}$ be a fractional solution to LPM with $\tilde{S} = \{Q \in \mathcal{S} : \tilde{x}_Q > 0\}$ (or equivalently $\tilde{Z} = \{Z_Q : \tilde{x}_Q > 0\}$) and such that each $Q \in \tilde{S}$ is a connected subgraph of G . Furthermore, let x^* be an optimal solution to LPM with $S^* = \{Q \in \mathcal{S} : x_Q^* > 0\}$

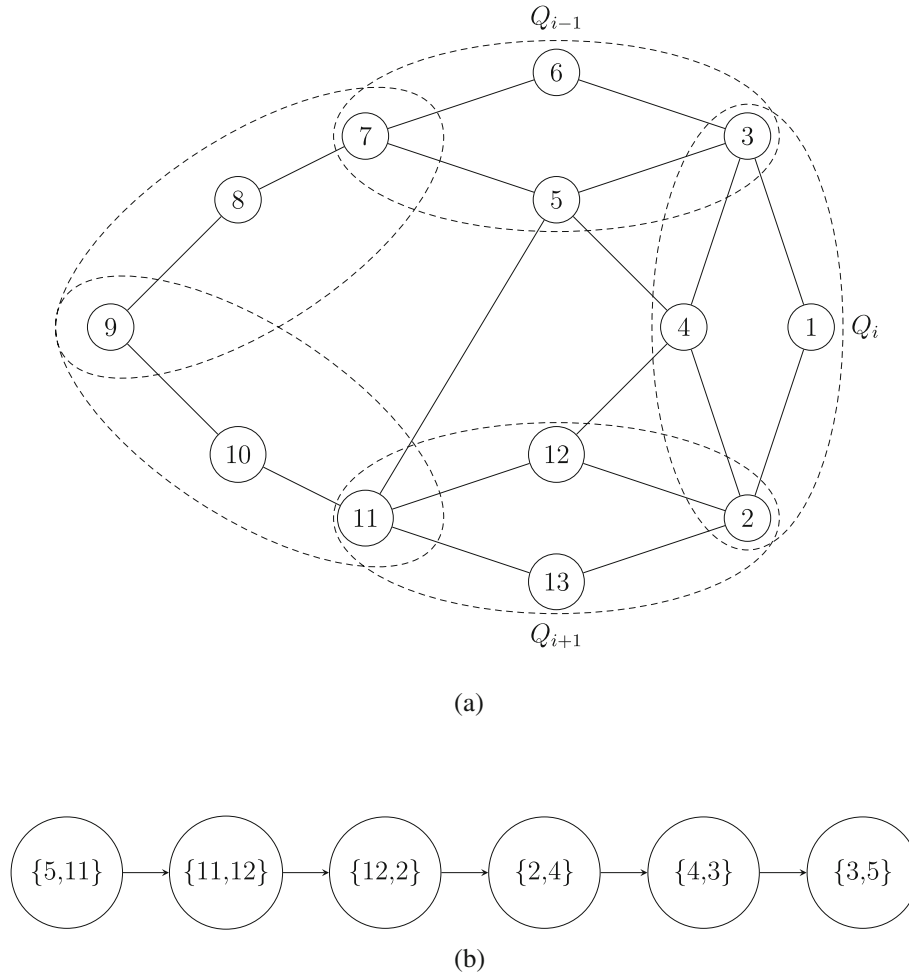


FIGURE 3 The input graphs and a LRU cycle C . (A) Connection graph. (B) Precedence graph.

(or equivalently $\mathcal{Z}^* = \{Z_Q : x_Q^* > 0\}$) and that also has connected LRUs $Q \in S^*$. Note that x^* exists by Lemma 1.

Definition 1. A LRU cycle is a collection of LRUs $C = \{Q_1, Q_2, \dots, Q_n\}$ such that each Q_i is connected, $n \geq 3$ and for all $1 \leq i \leq n$ we have $Q_i \cap Q_{i+1} \neq \emptyset$, $(Q_i \cap Q_{i+1}) \setminus (Q_{i+1} \cap Q_{i+2}) \neq \emptyset$, $(Q_{i+1} \cap Q_{i+2}) \setminus (Q_i \cap Q_{i+1}) \neq \emptyset$, with $n + 1 \equiv 1 \pmod n$ and $n + 2 \equiv 2 \pmod n$.

For an example of a LRU cycle, we refer the reader to Figure 3. We remark that there can exist a solution containing a LRU cycle such that the solution is an extreme point of the feasible polyhedron of LPM, and thus the feasible polyhedron of LPM is not integral. Next, we prove that an optimal solution to LPM does not contain a LRU cycle: Theorem 1.

Theorem 1. An optimal solution x^* to LPM does not contain a LRU cycle.

The proof of this theorem uses a technical lemma that can be found in the Appendix D as Lemma 2 and the following two concepts. Define the set of edges that is broken for a LRU X but not for a LRU Y by $\mathcal{F}(X, Y) = \Gamma(X) \setminus \Gamma(Y)$. Furthermore, modular arithmetic is used for the indices of LRUs that form

a cycle Q_i, Q_{i-1} , and Q_{i+1} with $1 \leq i \leq n$, $n + 1 \equiv 1 \pmod n$, and $Q_0 \equiv Q_n$.

Proof. We show that a solution to LPM that contains a LRU cycle is suboptimal. Let \tilde{x} be a solution to LPM such that each $Q \in \tilde{S}$ is connected and there exists a LRU cycle $C = \{Q_1, Q_2, \dots, Q_n\} \subseteq \tilde{S}$ with minimal n . Note that a solution that contains a LRU cycle must be fractional. We prove that there exists a feasible solution x' to LPM with $S' = \{Q \in \mathcal{S} : x'_Q > 0\}$ in which the LRU cycle C does not exist and $\pi(S') < \pi(\tilde{S})$.

Let $W_j = \min \left\{ \sum_{e \in \mathcal{F}(Q_i \cap Q_{j+1}, Q_j)} w(e), \sum_{e \in \mathcal{F}(Q_i \cap Q_{j-1}, Q_j)} w(e) \right\}$ for each LRU Q_j . Next, we consider a specific LRU $Q_i = \operatorname{argmin}_{Q_j \in C} \{W_j\}$ and we assume that $W_i = \sum_{e \in \mathcal{F}(Q_i \cap Q_{i+1}, Q_i)} w(e)$ (later we consider $W_i = \sum_{e \in \mathcal{F}(Q_i \cap Q_{i-1}, Q_i)} w(e)$). We create an alternative solution x' by partitioning Q_i in $Q_i \cap Q_{i+1}$ and $Q_i \setminus Q_{i+1}$. That is, let the alternative solution x' be identical to \tilde{x} except for the entries $x'_{Q_i} = 0$, $x'_{Q_i \setminus Q_{i+1}} = \tilde{x}_{Q_i}$, and

$$\begin{aligned}
& x'_{Q_i \cap Q_{i+1}} = \tilde{x}_{Q_i}. \text{ We have} \\
& \pi(S') - \pi(\tilde{S}) \\
& \leq \tilde{x}_{Q_i} \left(\lambda(Q_i \cap Q_{i+1}) \sum_{e \in \Gamma(Q_i \cap Q_{i+1})} w(e) + \lambda(Q_i \cap Q_{i+1}) \right. \\
& \quad + \sum_{u \in Q_i \cap Q_{i+1}} \ell(u) \lambda(Q_i \setminus Q_{i+1}) \sum_{e \in \Gamma(Q_i \setminus Q_{i+1})} w(e) \\
& \quad + \lambda(Q_i \setminus Q_{i+1}) \sum_{u \in Q_i \setminus Q_{i+1}} \ell(u) - \lambda(Q_i) \sum_{e \in \Gamma(Q_i)} w(e) \\
& \quad \left. - \lambda(Q_i) \sum_{u \in Q_i} \ell(u) \right) \\
& < \tilde{x}_{Q_i} \left(\lambda(Q_i \cap Q_{i+1}) \sum_{e \in \Gamma(Q_i \cap Q_{i+1})} w(e) + \lambda(Q_i \setminus Q_{i+1}) \right. \\
& \quad \left. \sum_{e \in \Gamma(Q_i \setminus Q_{i+1})} w(e) - \lambda(Q_i) \sum_{e \in \Gamma(Q_i)} w(e) \right) \\
& \leq \tilde{x}_{Q_i} \left(\left[\sum_{e \in \Gamma(Q_i \cap Q_{i+1})} w(e) - \sum_{e \in \Gamma(Q_i)} w(e) \right] \lambda(Q_i \cap Q_{i+1}) \right. \\
& \quad \left. + \left[\sum_{e \in \Gamma(Q_i \setminus Q_{i+1})} w(e) - \sum_{e \in \Gamma(Q_i)} w(e) \right] \lambda(Q_i \setminus Q_{i+1}) \right),
\end{aligned}$$

where the first inequality follows from the subadditivity of λ and the second inequality holds as $\lambda(Q_i \cap Q_{i+1}) \sum_{u \in Q_i \cap Q_{i+1}} \ell(u) + \lambda(Q_i \setminus Q_{i+1}) \sum_{u \in Q_i \setminus Q_{i+1}} \ell(u) < \lambda(Q_i) \sum_{u \in Q_i} \ell(u)$, because $Q_i \cap Q_{i+1}$ and $Q_i \setminus Q_{i+1}$ partition Q_i , $\lambda(Q) > 0$ for all $Q \subseteq V$, and $\ell(v) > 0$ for all $v \in V$. The last inequality follows after rearranging terms and using the subadditivity of the failure rate function λ . We continue by proving that the right hand side of the last equality is less than zero; that is, we show that $\sum_{e \in \Gamma(Q_i \cap Q_{i+1})} w(e) \leq \sum_{e \in \Gamma(Q_i)} w(e)$ and $\sum_{e \in \Gamma(Q_i \setminus Q_{i+1})} w(e) \leq \sum_{e \in \Gamma(Q_i)} w(e)$. We have

$$\begin{aligned}
& \pi(S') - \pi(\tilde{S}) \\
& < \tilde{x}_{Q_i} \left(\left[\sum_{e \in \Gamma(Q_i \cap Q_{i+1})} w(e) - \sum_{e \in \Gamma(Q_i)} w(e) \right] \lambda(Q_i \cap Q_{i+1}) \right. \\
& \quad \left. + \left[\sum_{e \in \Gamma(Q_i \setminus Q_{i+1})} w(e) - \sum_{e \in \Gamma(Q_i)} w(e) \right] \lambda(Q_i \setminus Q_{i+1}) \right) \\
& = \tilde{x}_{Q_i} \left(\left[\sum_{e \in \Gamma(Q_i \cap Q_{i+1})} w(e) - \sum_{e \in \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})} w(e) \right. \right. \\
& \quad \left. \left. - \sum_{e \in \Gamma(Q_i) \setminus \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})} w(e) \right] \lambda(Q_i \cap Q_{i+1}) \right. \\
& \quad \left. + \left[\sum_{e \in \Gamma(Q_i \setminus Q_{i+1})} w(e) - \sum_{e \in \mathcal{F}(Q_i \cap Q_{i+1}, Q_{i+1})} w(e) \right. \right. \\
& \quad \left. \left. - \sum_{e \in \Gamma(Q_i) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_{i+1})} w(e) \right] \lambda(Q_i \setminus Q_{i+1}) \right),
\end{aligned}$$

$$\begin{aligned}
& - \sum_{e \in \Gamma(Q_i) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_{i+1})} w(e) \left] \lambda(Q_i \setminus Q_{i+1}) \right) \\
& \leq \tilde{x}_{Q_i} \left(\left[\sum_{e \in \Gamma(Q_i \cap Q_{i+1})} w(e) - \sum_{e \in \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})} w(e) \right. \right. \\
& \quad \left. \left. - \sum_{e \in \Gamma(Q_i \cap Q_{i+1}) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_i)} w(e) \right] \lambda(Q_i \cap Q_{i+1}) \right. \\
& \quad \left. + \left[\sum_{e \in \Gamma(Q_i \setminus Q_{i+1})} w(e) - \sum_{e \in \mathcal{F}(Q_i \cap Q_{i+1}, Q_{i+1})} w(e) \right. \right. \\
& \quad \left. \left. - \sum_{e \in \Gamma(Q_i \setminus Q_{i+1}) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_i)} w(e) \right] \lambda(Q_i \setminus Q_{i+1}) \right) \\
& \leq \tilde{x}_{Q_i} \left(\left[\sum_{e \in \Gamma(Q_i \cap Q_{i+1})} w(e) - \sum_{e \in \mathcal{F}(Q_i \cap Q_{i+1}, Q_i)} w(e) \right. \right. \\
& \quad \left. \left. - \sum_{e \in \Gamma(Q_i \cap Q_{i+1}) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_i)} w(e) \right] \lambda(Q_i \cap Q_{i+1}) \right. \\
& \quad \left. + \left[\sum_{e \in \Gamma(Q_i \setminus Q_{i+1})} w(e) - \sum_{e \in \mathcal{F}(Q_i \cap Q_{i+1}, Q_i)} w(e) \right. \right. \\
& \quad \left. \left. - \sum_{e \in \Gamma(Q_i \setminus Q_{i+1}) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_i)} w(e) \right] \lambda(Q_i \setminus Q_{i+1}) \right) = 0.
\end{aligned}$$

The equality holds, because $\mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1}) \subset \Gamma(Q_i)$ and $\mathcal{F}(Q_i \cap Q_{i+1}, Q_{i+1}) \subset \Gamma(Q_i)$. The second inequality holds by Lemma 2. The last inequality follows because $Q_i = \operatorname{argmin}_{Q_j \in C} \{W_j\}$ and we assumed that $W_i = \sum_{e \in \mathcal{F}(Q_i \cap Q_{i+1}, Q_{i+1})} w(e)$. Hence, x' is a solution without the LRU cycle C and satisfies $\pi(S') < \pi(\tilde{S})$.

Next, consider the case $W_i = \sum_{e \in \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})} w(e)$. Then, we create the solution x' by partitioning Q_i in $Q_i \cap Q_{i-1}$ and $Q_i \setminus Q_{i-1}$, and we follow the same procedure as above using Lemma 2 and $W_i = \sum_{e \in \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})} w(e)$. Hence, the solution x' does not have the LRU cycle C and satisfies $\pi(S') < \pi(\tilde{S})$. ■

The result of Theorem 1 is illustrated by means of Example 1.

Example 1. Suppose we have the connection graph G with $w(e) = 1$, $\forall e \in E$ and the precedence graph D from Figure 3. Furthermore, we consider the solution \tilde{x} with a LRU cycle $C \subseteq \tilde{S}$ as drawn by the dashed ellipses in Figure 3. We illustrate our procedure for splitting LRU a Q_i into $Q_i \cap Q_{i+1}$ and $Q_i \setminus Q_{i+1}$.

In this example, we have

$$\begin{aligned}
\Gamma(Q_i) &= \{\{2, 12\}, \{2, 13\}, \{2, 4\}, \{3, 4\}, \\
&\quad \{3, 5\}, \{3, 6\}, \{4, 5\}, \{4, 12\}\},
\end{aligned}$$

$$\Gamma(Q_{i+1}) = \{\{1, 2\}, \{10, 11\}, \{4, 12\}, \{5, 11\}, \\ \{11, 12\}, \{12, 2\}, \{2, 4\}, \{4, 3\}, \{3, 5\}\},$$

$$\Gamma(Q_{i-1}) = \{\{1, 3\}, \{4, 5\}, \{7, 8\}, \{5, 11\}, \{11, 12\}, \\ \{2, 12\}, \{2, 4\}, \{3, 4\}, \{3, 5\}\},$$

$$\Gamma(Q_i \cap Q_{i-1}) = \{\{1, 3\}, \{4, 3\}, \{3, 5\}, \{3, 6\}\},$$

$$\Gamma(Q_i \setminus Q_{i-1}) = \{\{1, 3\}, \{2, 12\}, \{2, 13\}, \{2, 4\}, \\ \{3, 4\}, \{3, 5\}, \{4, 5\}, \{4, 12\}\},$$

$$\mathcal{F}(Q_i \cap Q_{i+1}, Q_{i+1}) = \{\{2, 12\}, \{2, 13\}\},$$

$$\mathcal{F}(Q_i \cap Q_{i-1}, Q_i) = \{\{1, 3\}\},$$

$$\mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1}) = \{\{3, 6\}\}.$$

One can use the above expressions to verify that the procedure in the proof of Theorem 1 yields a solution x' such that $\pi(S') < \pi(\mathcal{S})$.

Next, we introduce the concept of totally balanced matrices and Theorem 2 that—in combination with Theorem 1—helps us to prove that an optimal solution to LPM is integer. The definition of a totally balanced matrix can for instance be found in Anstee and Farber (1984) or Hoffman et al. (1985).

Definition 2. A binary matrix \mathcal{R} is totally balanced if it does not contain a square submatrix R that has no identical columns and the sum of each row and column equals to two.

Given this definition, we can now introduce Theorem 2, stating that the matrix encoding of an optimal solution to LPM is totally balanced. Note that we do not consider the constraint matrix of LPM—as is mostly done—but we study the matrix encoding of an optimal solution to LPM.

Theorem 2. *If an optimal solution x^* does not contain a LRU cycle, then \mathcal{Z}^* is totally balanced.*

Proof. By the definition of a totally balanced matrix the statement of this theorem is equivalent to: given an optimal solution x^* —with S^* or equivalently \mathcal{Z}^* —that does not contain a LRU cycle, there does not exist a binary $k \times k$ submatrix R of \mathcal{Z}^* with $k \geq 3$, no identical columns, and such that the sum of each row and column of R equals to two. We prove this by the contraposition, that is, we prove that if such a submatrix exists, the solution contains a LRU cycle.

To be more precise, we prove the following statement: Given a binary $k \times k$ matrix R with $k \geq 3$, no identical columns, and such that each row and column sum to two, there exists an $n \times n$ submatrix \hat{R} of R with minimal $n \geq 3$, no identical columns, and such that each row and column of \hat{R} sum to two. If we interpret the columns of \hat{R} as LRUs and the rows of \hat{R} as vertices, then the LRUs—corresponding to the columns of \hat{R} —are a LRU cycle.

Consider an $n \times n$ submatrix \hat{R} of R with minimal $n \geq 3$, no identical columns, and such that each row and column of \hat{R} sum to two. Such a submatrix \hat{R} exists, because R satisfies the same conditions. We will rename the rows and columns of \hat{R} such that we can easily show that the columns (LRUs) of \hat{R} are a LRU cycle. This renaming procedure is as follows.

The first row is v_1 and Q_1 and Q_2 are the columns such that $\hat{r}_{v_1, Q_1} = \hat{r}_{v_1, Q_2} = 1$. This follows without loss of generality, because \hat{R} is binary and the sum of each row equals two. Furthermore, note that all other values of v_1 are zero. Next, let v_2 be the second row such that $\hat{r}_{v_2, Q_2} = 1$. This is feasible because the sum of column Q_2 is two. Moreover, all other rows (except v_1 and v_2) have the value 0 in column Q_2 . We also remark that $\hat{r}_{v_2, Q_1} = 0$, since otherwise all other values in column Q_1 (except for \hat{r}_{v_1, Q_1} and \hat{r}_{v_2, Q_1}) are zero and this means that columns Q_1 and Q_2 are identical, which is a contradiction.

Next, we label the column Q_i such that $\hat{r}_{v_{i-1}, Q_i} = 1$ for each $i = 3, \dots, n$, and we call the row v_i that satisfies $\hat{r}_{v_i, Q_i} = 1$, for all $i = 3, \dots, n$. This can be done due to the following reasoning. The columns Q_j with $1 < j < i - 1$ are such that $\hat{r}_{v_{i-1}, Q_j} = 0$, because each column Q_j already sums to two. Unless $i = n$, we have $\hat{r}_{v_{i-1}, Q_1} = 0$ because otherwise we would have a $i \times i$ submatrix for which each row and column sum equal 2, $i \geq 3$, and where no identical columns exists. But this would contradict the fact that n is minimal. Hence, we can label Q_i such that $\hat{r}_{v_{i-1}, Q_i} = 1$. Moreover, all rows v_j with $1 \leq j < i - 1$ are such that $\hat{r}_{v_j, Q_i} = 0$, because each row v_j already sums to two (by considering columns Q_k with $k < i$). Therefore, we can call a row v_i such that $\hat{r}_{v_i, Q_i} = 1$.

Finally, we let $r_{v_n, Q_1} = 1$ such that the row and column sum of each row and column of \hat{R} equals 2.

Given the renaming of the columns and rows of \hat{R} , we have for all $1 \leq i \leq n$ that $\{v_i\} = Q_i \cap Q_{i+1}$ with $n+1 \equiv 1 \pmod{n}$. Furthermore, this implies that $(Q_i \cap Q_{i+1}) \setminus (Q_{i+1} \cap Q_{i+2}) = \{v_i\} \setminus \{v_{i+1}\} \neq \emptyset$ and $(Q_{i+1} \cap Q_{i+2}) \setminus (Q_i \cap Q_{i+1}) = \{v_{i+1}\} \setminus \{v_i\} \neq \emptyset$ with $n+1 \equiv 1 \pmod{n}$ and $n+2 \equiv 2 \pmod{n}$. This implies that the LRUs Q_1, Q_2, \dots, Q_n are a LRU cycle.

Hence, we proved that the submatrix \hat{R} corresponds to a LRU cycle. Hence, if an optimal solution x^* does not contain a LRU cycle, it will not contain such a submatrix and, thus, the matrix \mathcal{Z}^* is totally balanced. ■

Given Theorems 1 and 2, our final and rather unusual result follows relatively easily in Theorem 3.

Theorem 3. *All extreme optimal points of the feasible polyhedron of LPM are integer.*

Proof. Let x^* be an optimal solution to LPM. Each LRU $Q \in \mathcal{S}^*$ is a connected subgraph of G by Lemma 1 and x^* does not contain a LRU cycle by Theorem 1. Then, the matrix \mathcal{Z}^* is totally balanced by Theorem 2. Consequently, the polyhedron $\mathcal{P} = \{x : \mathcal{Z}^*x = 1, x \geq 0, x \in \mathbb{R}^{|\mathcal{S}^*|}\}$ is integral (Fulkerson et al., 1974). Hence, x^* is either integer or a convex combination of integer solutions to LPM, and thus we obtain our desired result. ■

In a column generation approach, a restricted version of LPM is solved repeatedly after new columns are added (this will be detailed in Section 3.2). If the restricted versions of LPM are solved by the simplex algorithm, we obtain an optimal solution x^* that is an extreme point of the polyhedron $\mathcal{P} = \{x : \mathcal{Z}x = 1, x \geq 0, x \in \mathbb{R}^{|\mathcal{S}^*|}\}$, but is also an extreme point of the polyhedron $\mathcal{P} = \{x : \mathcal{Z}^*x = 1, x \geq 0, x \in \mathbb{R}^{|\mathcal{S}^*|}\}$ spanned by the submatrix \mathcal{Z}^* . Theorem 3 now implies that x^* is integral, because x^* is an extreme point of \mathcal{P} . Hence, solving LPM with the simplex algorithm yields an optimal integer solution, and this solution is thus also optimal for M.

If LPM is solved by the simplex algorithm, we obtain an optimal solution x^* that is an extreme point of the polyhedron $\mathcal{P} = \{x : \mathcal{Z}x = 1, x \geq 0, x \in \mathbb{R}^{|\mathcal{S}^*|}\}$, but is also an extreme point of the polyhedron $\mathcal{P} = \{x : \mathcal{Z}^*x = 1, x \geq 0, x \in \mathbb{R}^{|\mathcal{S}^*|}\}$ spanned by the submatrix \mathcal{Z}^* . Theorem 3 now implies that x^* is integral, because x^* is an extreme point of \mathcal{P} . Hence, solving LPM with the simplex algorithm yields an optimal integer solution, and this solution is thus also optimal for M.

We would like to stress that our result in Theorem 3 is rather unusual, because the polyhedron \mathcal{P} of LPM is *not* integral. This contrasts with much other research that focuses on proving integrality of the polyhedron to conclude that an optimal integer solution can be found (if the objective function is convex), for example, via totally unimodular constraint matrices. We demonstrate that—for nonintegral polyhedra—analysis of the objective function can be used to establish the existence of an optimal integer solution to a relaxed problem when the constraint matrix will not guarantee an integral polyhedron. The crucial analysis of the objective function is in the proof of Theorem 1. We believe that our proof approach can be used more generally for partitioning problems where the objective function is convex. One is only required to show that cycles in solutions are suboptimal.

3.2 | Solving LPM and M

Given Theorem 3, we move our attention to solving LPM, for which we apply column generation. Hence, we consider

a feasible subset of LRUs (or columns) $\tilde{\mathcal{S}} \subseteq \mathcal{S}$ for LPM. This results in the *Restricted Master Program* (RLPM). For RLPM, we generate profitable LRUs (columns) by solving the pricing problem of RLPM:

$$c^* = \min_{Q \in \tilde{\mathcal{S}}} \left\{ \omega(Q) - \sum_{v \in Q} r_v \right\}, \quad (5)$$

where r_v are dual variables for the partitioning constraints of RLPM. We want to find a LRU $Q \in \tilde{\mathcal{S}}$ with minimal reduced cost. After we solve the pricing problem (7a), we add the obtained LRU to $\tilde{\mathcal{S}}$ and we again solve LPM with the new $\tilde{\mathcal{S}}$. Next, we solve the pricing problem again, and we repeat this procedure until the pricing problem does not return a profitable LRU (column), that is, we terminate when $c^* \geq 0$. This means that there does not exist a LRU (column) that is worthwhile to add to our LPM, and we have obtained the optimal solution.

The computation of the pricing problem is difficult in general, but can be done with a standard solver when the failure rate function λ is additive, that is, when $\lambda(S) = \sum_{i \in S} \lambda(\{i\})$. We provide this formulation below.

3.2.1 | Pricing problem for additive failure rates

Problem (5) can be expressed as follows when λ is additive ($\lambda(S) = \sum_{i \in S} \lambda(\{i\})$). The binary auxiliary variable k^e denotes whether edge $e \in E$ needs to be disconnected to remove LRU Q . The binary decision variable γ_v indicates whether node v is included in the LRU Q . Now we rewrite Problem (5) to obtain

$$c^* = \min_{r, k} \sum_{e \in E} k^e w(e) \sum_{v \in V} \gamma_v \lambda(\{v\}) + \sum_{u \in V} \gamma_u \ell(u) \sum_{v \in V} \gamma_v \lambda(\{v\}) - \sum_{v \in V} \gamma_v r_v, \quad (6a)$$

$$\text{s.t. } \gamma_u - \gamma_v \leq k^e, \quad \forall \{u, v\} \in E, \\ \forall e \in H(\{u, v\}), \quad (6b)$$

$$\gamma_v, k^e \in \{0, 1\}. \quad (6c)$$

Problem (5) can be linearized by applying the McCormick linearization method (McCormick, 1976). In particular, let $\eta_{ev} = k^e \gamma_v$ and $\delta_{uv} = \gamma_u \gamma_v$. Note that η_{ev} denotes whether the LRU contains part $v \in V$ and that edge $e \in E$ needs to be broken for the LRU to be removed. Similarly, δ_{uv} represents whether the LRU contains both parts $u, v \in V$. Then a binary linear formulation of (5) is given below in (7a).

$$c^* = \min_{k, \gamma, \eta, \delta} \sum_{e \in E} \sum_{v \in V} \eta_{ev} \lambda(v) w(e) + \sum_{u, v \in V} \delta_{uv} \ell(u) \lambda(v) - \sum_{v \in V} \gamma_v r_v, \quad (7a)$$

$$\text{s.t. } \gamma_u - \gamma_v \leq k^e, \quad \forall \{u, v\} \in E, \\ \forall e \in H(\{u, v\}), \quad (7b)$$

$$\eta_{ev} \leq k^e, \quad \forall e \in E, \quad \forall v \in V, \quad (7c)$$

$$\eta_{ev} \leq \gamma_v, \forall e \in E, \forall v \in V, \quad (7d)$$

$$\eta_{ev} \geq k^e + \gamma_v - 1, \quad \forall e \in E, \forall v \in V, \quad (7e)$$

$$\delta_{uv} \leq \gamma_u, \quad \forall u, v \in V, \quad (7f)$$

$$\delta_{uv} \leq \gamma_v, \quad \forall u, v \in V, \quad (7g)$$

$$\delta_{uv} \geq \gamma_u + \gamma_v - 1, \quad \forall u, v \in V, \quad (7h)$$

$$\eta_{ev}, \delta_{uv} \geq 0, \gamma_v, k^e \in \{0, 1\}, \gamma_v, k^e \in \{0, 1\}. \quad (7i)$$

After we solve the pricing problem (7a), we add the obtained LRU to $\tilde{\mathcal{S}}$ and we again solve LPM with the new $\tilde{\mathcal{S}}$. Next, we solve the pricing problem again, and we repeat this procedure until the pricing problem does not return a profitable LRU (column), that is, we terminate when $c^* \geq 0$. This means that there does not exist a LRU (column) that is worthwhile to add to our LPM, and we have obtained the optimal solution.

4 | BINARY PROGRAMMING FORMULATION AN ADDITIVE FAILURE RATE FUNCTION

In this section, we consider only additive failure rates λ , that is, $\lambda(Q) = \sum_{v \in Q} \lambda(\{v\})$. This special case is convenient when implementing the model and solving it with a standard solver like CPLEX or Gurobi. We formulate LRU DESIGN as a BNLP and linearize it so that it can be solved by a standard integer program solver. We use this linearized BLP as a benchmark to compare our set partitioning formulation with.

For the BNLP, we first relax the fact that $\emptyset \notin S$. In the foregoing, S was a partition of V . For the BNLP (and later the BLP) we consider a solution S' satisfying $|S'| = |V|$, where S' may contain empty LRUs, and we have that $S = \{Q \in S' : Q \neq \emptyset\}$. Next, we index each LRU in S' by $i \in \{1, \dots, |V|\}$, that is, we have LRUs $Q_i \in S'$ that are indexed by i . Furthermore, we create a binary variable y_{vi} that indicates whether a part $v \in V$ is assigned to LRU Q_i , $i \in \{1, \dots, |V|\}$:

$$y_{vi} = \begin{cases} 1 & \text{if } v \in Q_i \\ 0 & \text{otherwise} \end{cases}, \quad \forall v \in V, \forall i \in \{1, \dots, |V|\}.$$

We denote Y as the matrix consisting of all entries y_{vi} . Note that we can derive S' easily from Y . We also define the auxiliary binary variable k_i^e that denotes whether edge $e \in E$ needs to be broken in order to replace LRU Q_i , and organize them in a matrix K . We determine the value of k_i^e by considering all the edges $b \in B(Q_i)$. We determine $B(Q_i)$ by considering the edges $\{u, v\} \in E$ such that $y_{ui} - y_{vi} = 1$ or $y_{vi} - y_{ui} = 1$, that is, one of the end points of $\{u, v\}$ belongs to LRU Q_i while the other end point does not. This corresponds to the definition of $B(Q_i)$. Subsequently, we consider each edge $e \in E$ that needs to be broken before breaking $\{u, v\}$; that is, for each $\{u, v\} \in B(Q_i)$ we consider all $e \in H(\{u, v\})$. Hence, the variable k_i^e satisfies $y_{ui} - y_{vi} \leq k_i^e$, $\forall \{u, v\} \in E, \forall e \in H(\{u, v\}), \forall i \in \{1, \dots, |V|\}$. Note that k_i^e may take the value of one, even if

an edge $e \in E$ is fully contained within a LRU Q_i . We denote K as the matrix consisting of all entries k_i^e . We use the variable k_i^e in our objective function (8a), since it represents whether an edge has to be broken ($k_i^e = 1$) in order to remove LRU Q_i . Furthermore, the objective function implies $k_i^e = 0$ if edge e is not broken for the replacement of Q_i . Next, we use k_i^e , the edge weights, and the failure rate of Q_i (expressed using y_{vi}) to determine the cost for replacing Q_i . The total purchase cost of the LRU Q_i is derived by using y_{vi} , and we multiply this by the total failure rate of Q_i (which also depends on y_{vi}). This results in a BNLP formulation of LRU DESIGN:

$$\begin{aligned} \text{(BNLP)} \quad & \min_{Y, K} \sum_{i=1}^{|V|} \sum_{e \in E} k_i^e w(e) \sum_{v \in V} y_{vi} \lambda(v) \\ & + \sum_{i=1}^{|V|} \sum_{u \in V} y_{ui} \ell(u) \sum_{v \in V} y_{vi} \lambda(v), \end{aligned} \quad (8a)$$

$$\text{s.t.} \quad \sum_{i=1}^{|V|} y_{vi} = 1, \quad \forall v \in V, \quad (8b)$$

$$\begin{aligned} y_{ui} - y_{vi} &\leq k_i^e, \quad \forall \{u, v\} \in E, \\ \forall e \in H(\{u, v\}), \forall i &\in \{1, \dots, |V|\}, \end{aligned} \quad (8c)$$

$$y_{vi}, k_i^e \in \{0, 1\}. \quad (8d)$$

Constraints (8b) ensure that each part $v \in V$ is included in exactly one LRU, and constraints (8c) enforce the definition of the auxiliary variable k_i^e .

The BNLP is a problem with a quadratic objective function. Therefore, the BNLP can be linearized by the McCormick reformulation (McCormick, 1976); that is, we introduce new variables $\rho_{ev}^i = y_{vi} k_i^e$ and $\sigma_{uv}^i = y_{ui} y_{vi}$. The variable ρ_{ev}^i denotes whether LRU Q_i contains part $v \in V$ and whether edge $e \in E$ needs to be broken in order to replace the LRU Q_i . Analogously, σ_{uv}^i denotes whether two parts $u, v \in V$ are both contained in the same LRU Q_i . Substituting ρ_{ev}^i and σ_{uv}^i into the above implies that we need to add constraints that enforce the interpretation we gave. Furthermore, we optimize over Y , K , ρ , and σ , where ρ and σ correspond to the 3-D arrays with entries ρ_{ev}^i and σ_{uv}^i , respectively. Hence, we obtain the BLP formulation of LRU DESIGN:

$$\begin{aligned} \text{(BLP)} \quad & \min_{Y, K, \rho, \sigma} \sum_{i=1}^{|V|} \sum_{e \in E} \sum_{v \in V} \rho_{ev}^i \lambda(v) w(e) \\ & + \sum_{i=1}^{|V|} \sum_{u, v \in V} \sigma_{uv}^i \ell(u) \lambda(v), \end{aligned} \quad (9a)$$

$$\text{s.t.} \quad \sum_{i=1}^{|V|} y_{vi} = 1, \quad \forall v \in V, \quad (9b)$$

$$\begin{aligned} y_{ui} - y_{vi} &\leq k_i^e, \quad \forall \{u, v\} \in E, \quad \forall e \in H(\{u, v\}), \\ \forall i &\in \{1, \dots, |V|\}, \end{aligned} \quad (9c)$$

$$\rho_{ev}^i \leq y_{vi}, \quad \forall e \in E, \forall v \in V, \quad \forall i \in \{1, \dots, |V|\}, \quad (9d)$$

$$\rho_{ev}^i \leq k_i^e, \forall e \in E, \forall v \in V, \quad \forall i \in \{1, \dots, |V|\}, \quad (9e)$$

$$\rho_{ev}^i \geq y_{vi} + k_i^e - 1, \quad \forall e \in E, \forall v \in V, \\ \forall i \in \{1, \dots, |V|\}, \quad (9f)$$

$$\sigma_{uv}^i \leq y_{ui}, \quad \forall u, v \in V, \quad \forall i \in \{1, \dots, |V|\}, \quad (9g)$$

$$\sigma_{uv}^i \leq y_{vi}, \quad \forall u, v \in V, \forall i \in \{1, \dots, |V|\}, \quad (9h)$$

$$\sigma_{uv}^i \geq y_{ui} + y_{vi} - 1, \quad \forall u, v \in V, \\ \forall i \in \{1, \dots, |V|\}, \quad (9i)$$

$$\rho_{ev}^i, \sigma_{uv}^i \geq 0, \quad (9j)$$

$$y_{vi}, k_i^e \in \{0, 1\}. \quad (9k)$$

We acknowledge that tighter MIP formulations than BLP will exist due to extensive studies on binary quadratic programming and Boolean quadratic polytopes (Bonami et al., 2018; Boros et al., 1992; Boros & Hammer, 1993; Charfreitag et al., 2022; Deza et al., 1997; Jünger & Mallach, 2021; Padberg, 1989; Rehfeldt et al., 2023; Rendl et al., 2010). The formulation above is a straightforward formulation that will appeal to many practitioners of operations research due to its simplicity.

5 | NUMERICAL EXPERIMENTS

In this section, we use the BLP formulation and the set partitioning formulation (LPM) of LRU DESIGN under additive failure rate functions to gain some insight into the size of instances that can be solved to optimality. We shed a light on the objective value gap between both approaches, and we explore the effects of parameter perturbations on our model's outcomes. Furthermore, we numerically illustrate—based on randomly generated instances of practical size—that the relative cost increase from introducing a partitioning constraint is very limited, and its magnitude decreases as the number of parts grows. We have implemented all optimization model formulations in JuMP (Dunning et al., 2017; Lubin & Dunning, 2015), which is a mathematical optimization package of Julia (Balbaert et al., 2016), and we solved all problems using Gurobi 7.0.1 on an Intel i5-4300U @2.50GHz processor with 16GB RAM and running Ubuntu 16.04 LTS.

In Section 5.1, we explain the used instance generator for our experiments. In Section 5.2, we study the difference between the computation times (in seconds) of the binary linear programming formulation (BLP) and the set partitioning formulation (LPM). Also, we discuss the relative difference in objective value between both models. Furthermore, we shed some light on how the downtime cost per time unit affects the number of LRUs used in an optimal solution. Moreover, we study how the system's complexity affects the total annual costs by considering the number of connections between parts, and the number of predecessor–successor relationships

that exist. Finally, we show that the cost effect of introducing a partitioning constraint is very limited for practically sized instances.

5.1 | Instance generator

An instance is described by the graphs G and D . We vary the number of vertices $|V|$, the number of edges $|E|$, and the number of arcs $|A|$ in our numerical experiments. We relate the number of edges in G to the number of vertices by $|E| = \delta|V|$, where δ is the average vertex degree in the graph G . Similarly, we relate the number of arcs in D to the number of edges by $|A| = \delta_E|E|$, where δ_E is the average out degree of an edge $e \in E$. All other parameters such as $\lambda(v) > 0$ and $\ell(v) > 0$ for all $v \in V$, and $w(e) > 0$ for all $e \in E$ are randomly generated, as well as a graph's layout in terms of the edge set E and the arc set A .

The graphs G and D are generated in the following way. For G , we have a set of vertices V and a number of unique edges $|E| = \delta|V|$, and we create a spanning tree with $|V| - 1$ edges. We add an arbitrary vertex $v \in V$ to a set of considered vertices \tilde{V} , and we select a new vertex $u \in V \setminus \tilde{V}$ and connect it to an arbitrary vertex $z \in \tilde{V}$ by adding the edge $\{u, v\}$ to the edge set E . We keep doing this until $\tilde{V} = V$. Subsequently we add remaining edges randomly to our graph and we terminate once we have $|E|$ edges in G . Second, we generate the precedence graph D . We (randomly) assign an index to each edge $e \in E$ and denote this index by $I(e)$, and the minimum and maximum values assigned are 1 and $|E|$, respectively. We start with $A = \emptyset$ and add an arc in each iteration. An iteration starts by selecting two random edges $\{u, v\}, \{v, x\} \in E : u, v, x \in V$ and $u \neq v \neq x$. If $I(\{u, v\}) \leq I(\{v, x\})$ we create an arc $(\{u, v\}, \{v, x\})$ and add it to A , otherwise we create an arc $(\{v, x\}, \{u, v\})$ and add it to A . We repeat this procedure until $\delta_E|E| = |A|$, and upon termination we have obtained a set A that has a topological sorting and thus the precedence graph D is acyclic.

5.2 | Computational results

Next, we discuss the computational results for our model. The generation of a random graph follows the procedure from Section 5.1, and we let $|V| \in \{10, 20, 30, 40, 50, 60\}$, $\delta \in \{2, 3, 4\}$, and $\delta_E \in \{0.5, 1, 1.5\}$. For each combination $(|V|, \delta, \delta_E)$, we generate 10 random instances, resulting in a total of 540 instances.

We use a time limit of 600 s for the BLP formulation and also for the set partitioning formulation. This time limit is relatively low because we solve a large number of instances, thereby making it feasible to perform the entire numerical study in a reasonable amount of time. If an instance has not been solved to optimality within 600 s, we say that it is inefficient. If all instances of a certain parameter combination $(|V|, \delta, \delta_E)$ are inefficient, we write—as an entry for the

TABLE 2 Average computation times (s) of both formulations.

BLP		V					
δ	δ_E	10	20	30			
2	0.5	4.45 ₁₀	121.02 ₁₀	–			
2	1	5.16 ₁₀	144.59 ₁₀	–			
2	1.5	3.39 ₁₀	85.38 ₁₀	577.21 ₃			
3	0.5	7.18 ₁₀	316.42 ₉	–			
3	1	6.22 ₁₀	308.73 ₁₀	–			
3	1.5	4.98 ₁₀	151.64 ₁₀	–			
4	0.5	11.00 ₁₀	479.45 ₆	–			
4	1	8.56 ₁₀	502.25 ₄	–			
4	1.5	6.62 ₁₀	246.39 ₉	–			
LPM		V					
δ	δ_E	10	20	30	40	50	60
2	0.5	0.21 ₁₀	1.64 ₁₀	13.02 ₁₀	37.32 ₁₀	111.75 ₁₀	216.53 ₁₀
2	1	0.61 ₁₀	3.79 ₁₀	30.97 ₁₀	88.65 ₁₀	314.11 ₉	533.70 ₂
2	1.5	0.19 ₁₀	1.65 ₁₀	9.66 ₁₀	36.46 ₁₀	124.71 ₁₀	228.21 ₉
3	0.5	0.37 ₁₀	2.59 ₁₀	22.13 ₁₀	60.02 ₁₀	105.39 ₁₀	331.38 ₇
3	1	0.97 ₁₀	8.08 ₁₀	37.98 ₁₀	232.89 ₁₀	331.50 ₅	–
3	1.5	0.65 ₁₀	3.68 ₁₀	17.76 ₁₀	93.86 ₁₀	289.00 ₁₀	514.14 ₃
4	0.5	0.50 ₁₀	3.39 ₁₀	26.50 ₁₀	114.09 ₁₀	304.42 ₉	410.97 ₃
4	1	1.37 ₁₀	10.06 ₁₀	48.56 ₁₀	294.81 ₁₀	432.31 ₂	–
4	1.5	1.29 ₁₀	6.25 ₁₀	32.54 ₁₀	211.55 ₁₀	521.40 ₃	–

Note: The subscripts indicate that number of instances that were solved to optimality for a given setting.

combination. We determine the average computation time of both formulations based on the efficient instances. The results are presented in Table 2, where the computation times are given in seconds, and the subscripts indicate the number of efficient instances. Furthermore, we have not reported computation times for the BLP with $|V| \geq 40$, since we have found no efficient solutions within the time limit.

We observe that, given the time limit of 600 s, the BLP formulation can only solve small size instances up to 20 vertices (parts), while the set partitioning (SP) formulation can solve medium size to large instances up to 60 vertices (parts). Furthermore, we see that the set partitioning formulation solves instances faster than the BLP formulation. This effect is amplified when the instances become larger, that is, when $|V|$ and δ increase. Real-life instances are typically medium to large sized instances and can have 50 vertices (parts). Furthermore, such instances may possess many and complex connections and predecessor–successor relationships. This makes the BLP formulation unsuitable for practical purposes. Hence, it is worthwhile to invest extra time to implement the set partitioning formulation (LPM) with a pure pricing algorithm. Furthermore, the computation times illustrate that the set partitioning formulation of LRU DESIGN is particularly useful as a feedback mechanism for the company's design department. The engineers can quickly assess many design alternatives (in terms of the connection graph and precedence graph) and their effects on the optimal LRU design and the corresponding (after-sales) costs.

TABLE 3 Objective value gap between both formulations.

		V			
δ	δ_E	10	20	30	40
2	0.5	0.0%	0.0%	18.9%	257.6%
2	1	0.0%	0.0%	20.3%	349.1%
2	1.5	0.0%	0.0%	29.2%	369.0%
3	0.5	0.0%	0.0%	96.7%	523.1%
3	1	0.0%	0.1%	115.9%	520.6%
3	1.5	0.0%	0.0%	110.3%	474.2%
4	0.5	0.0%	0.0%	150.1%	580.2%
4	1	0.0%	7.8%	127.5%	218.9%
4	1.5	0.0%	4.6%	95.5%	191.1%

Not solving the BLP formulation to optimality may still result in solutions that are near optimal. Therefore, we also perform a gap analysis by studying the relative difference in the objective values of both formulations. We only study this difference for parameter combinations for which each instance solved by LPM is efficient, that is, instances with $|V| \leq 40$. We define the objective value gap

$$\beta = \frac{\pi(S_b) - \pi(S_l)}{\pi(S_l)} \times 100\%,$$

with $\pi(S_b)$ and $\pi(S_l)$ denoting the costs of the best feasible solution found after 600 s of the BLP and set partitioning formulation (LPM), respectively. The results are shown in Table 3.

We observe that the relative objective value difference rapidly grows as the number of vertices grows. This implies that the inefficient instances, produced by the BLP, do not result in a competitive solution compared to the set partitioning formulation (LPM).

Next, we numerically study the effect of the cost of one time unit of system downtime on the number of LRUs that is used in an optimal LRU design. In the remainder of this section, we use the same instance generator as discussed in Section 5.1, and we generate 1000 instances per parameter setting (δ , δ_E) and keep $|V| = 20$. For a given instance, we vary the edge weights by multiplying all edge weights of the instance by a constant factor $q \in \{0.1, 1, 10\}$. A higher value for q means that it is more expensive to break edges. If the time for breaking an edge remains constant, it means that the cost rate per time unit for breaking an edge increases, and thus we can capture a higher downtime cost per time unit by varying q . This way, we create three classes of instances (i) low downtime cost per time unit ($q = 0.1$); (ii) moderate downtime cost per time unit ($q = 1$); (iii) and high downtime cost per time unit ($q = 10$). We keep the parameter values for δ and δ_E constant at $\delta = 3$ and $\delta_E = 1$. We focus on the number of LRUs $|S^*|$ in an optimal solution S^* . The results are presented in Figure 4.

Based on Figure 4, the instances where the downtime cost per time unit is low, have many small LRUs (each part is a LRU in itself in the extreme case). These solutions prefer small LRUs because they have lower purchase costs. As the

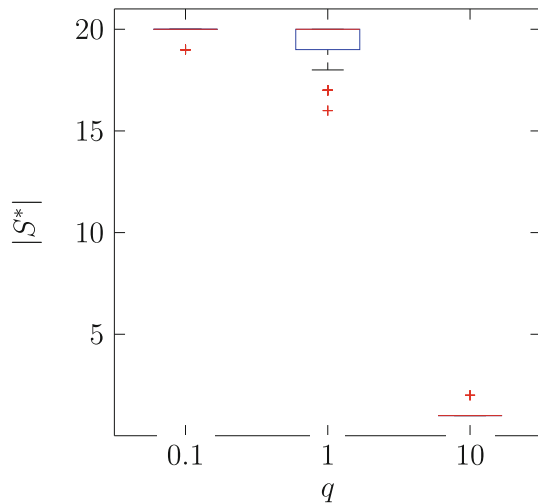


FIGURE 4 Effect of edge weights on the number of LRUs in S^* .

cost for a single time unit of downtime increases, we see that the optimal solution prefers fewer LRUs that each become larger, because such larger LRUs enable faster replacement and thus lower downtime costs. This explains, for example, why we observe that the consumer electronics industry with low values for q has rather small LRUs. On the other end of the spectrum, capital intensive industries such as the semiconductor industry or the aviation industry have high values for q , and they tend to opt for larger LRUs which enable faster replacement. Both these phenomena are confirmed by the numerical results of our model.

The second effect that we study considers the complexity of the system, and how this affects the costs of the optimal LRU design. We restrict our attention, for now, on the number of edges in the connection graph G that describes system complexity. We vary how strongly various parts are connected to each other by altering δ . A low (high) value of δ corresponds to lesser (more) connected parts. We are interested in the effect that the number of connections in G has on the costs, because this provides a justification of whether to avoid many connections between parts in order to reduce the total cost. For our analysis, we keep δ_E and q constant at $\delta_E = 1$ and $q = 1$. The results are presented in Figure 5.

We observe that more connections in the connection graph G result in cost increases, because we need to disconnect more edges in order to remove a LRU. This has an important managerial implication, as engineers should be urged to reduce the number of connections in systems to be developed. Thus, it may be wise for a company to invest extra in a system's design such that the number of connections in G is reduced. An example wherein few number of connections lead to low costs is a bicycle. A typical connection graph of bicycles has few connections, and consequently relatively low replacement cost because we only need to disconnect few connections in case a part fails.

Finally, we also study the effect that system complexity has on the costs of the optimal solution $\pi(S^*)$, when we consider the number of predecessor–successor relationships. A lower

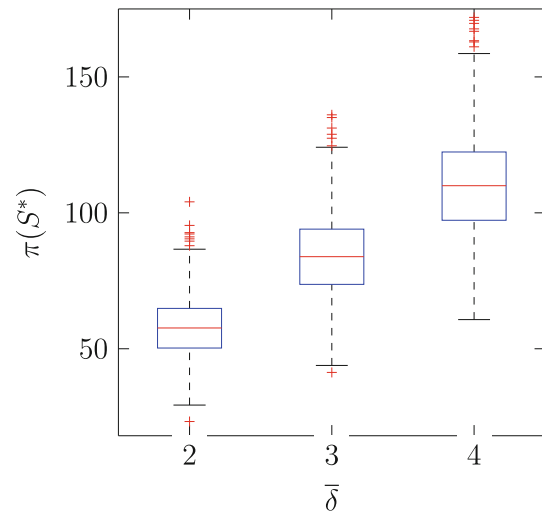


FIGURE 5 Effect of the number of connections in G on $\pi(S^*)$.

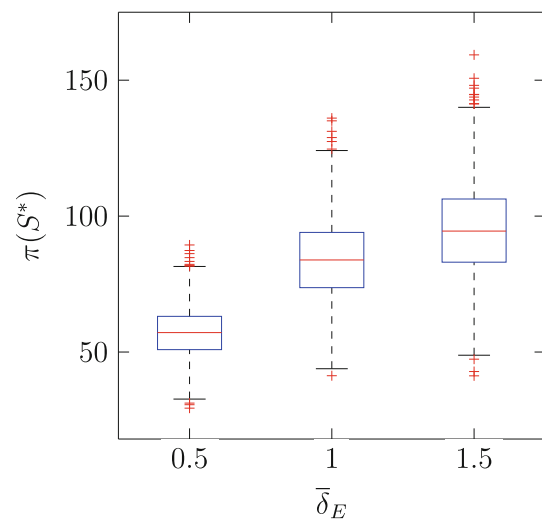


FIGURE 6 Effect of the number of predecessor–successor relationships in D on $\pi(S^*)$.

value of δ_E indicates that fewer predecessor–successor relationships exist in the precedence graph D . Similar to the foregoing, we keep the other parameters constant at $\delta = 3$ and $q = 1$. The results for different values of δ_E are depicted in Figure 6 and we observe a similar behavior to changes in δ .

The costs increase as the number of predecessor–successor relationships increases, because we need to disconnect more connections upon the failure of a LRU. Consequently, the costs of an optimal solution $\pi(S^*)$ increase when the number of connections in D increases (as δ_E increases). The managerial implications of our results also align with those for δ : managers should urge their designers to avoid predecessor–successor relationships in order to reduce costs. This objective may be easier to attain than avoiding connections in the connection graph G by careful design. Hence, the results confirm that careful design (in terms of G and D) is crucial to reduce the overall costs.

Finally, we numerically illustrate the cost effect of introducing a partitioning constraint. Therefore, we consider LRU

TABLE 4 Relative cost differences between LRU DESIGN and C-LRU DESIGN.

δ	δ_E	V = 20		V = 30		V = 40	
		avg (%)	max (%)	avg (%)	max (%)	avg (%)	max (%)
2	0.5	0.74	1.56	1.07	3.63	1.20	4.00
2	1	1.62	4.36	2.23	5.49	2.52	4.20
2	1.5	0.26	1.05	0.19	0.62	0.19	0.51
3	0.5	0.56	1.36	0.63	1.73	0.54	1.11
3	1	0.84	3.53	0.87	2.45	1.48	3.12
3	1.5	2.37	13.56	0.56	2.56	0.36	2.34
4	0.5	0.00	0.00	0.13	1.20	0.08	0.26
4	1	1.58	6.07	1.02	3.78	0.47	1.39
4	1.5	4.22	13.48	1.80	6.36	1.32	3.08
All		1.34	13.56	0.94	6.36	0.91	4.20

DESIGN and the variant without a partition constraint called C-LRU DESIGN. Further details on C-LRU DESIGN are included in Appendix C. Both, C-LRU DESIGN and the numerical results, are based on Driessen (2018). We study the average relative cost difference for a given parameter combination $\Delta\pi = \frac{\pi(S^*) - \pi_c(S_c^*)}{\pi_c(S_c^*)} \times 100\%$, where S^* and S_c^* are the optimal solutions of LRU DESIGN and C-LRU DESIGN, respectively. We limit our analysis to instances with $|V| \in \{20, 30, 40\}$, because the set partitioning formulation of LRU DESIGN is efficient for each parameter combination.

The results in Table 4 illustrate that we sacrifice little costs when we introduce a partitioning constraint. Combining this with the practical benefits that LRU DESIGN has over C-LRU DESIGN, we conclude that LRU DESIGN is highly usable for practical purposes.

6 | CONCLUSIONS

We considered an OEM or MRO that is concerned with the (re)design and maintenance of a system. If the system does not operate, the company loses money, customer goodwill or has to pay customers a downtime penalty. Therefore, the company is interested in lowering the cost for non-functioning systems by designing LRUs that can be removed quickly. Furthermore, the LRUs should not be too large, because this increases a LRU's total failure rate and the LRU's purchase cost (or repair cost). Thus, the company has to determine what the optimal LRU design is that balances the replacement cost and the purchase cost (or repair cost) of LRUs.

We presented a novel model for representing the connections between parts in a system, also capturing the existing disassembly sequences. We used the system representation to derive an optimization model LRU DESIGN that minimizes the replacement cost and the purchase cost (or repair cost) by optimizing the LRU design. Our optimization model was constrained such that a LRU design is a partition of the parts, and we saw that this constraint has strong practical benefits (avoiding replacement ambiguity, simplification of

the production and testing processes, and consistency with general maintenance practice), while sacrificing very little extra costs compared to the case where we relax the partition constraint. We formulated the problem as a BLP and as a set partitioning problem. We proved a result infrequently encountered in research: an optimal solution to the set partitioning formulation is integer, despite a nonintegral feasible polyhedron. This result follows from proving the suboptimality of LRU cycles and relating this to the matrix encoding of an optimal solution. Furthermore, the set partitioning formulation reduces the computation times and makes it useful as a feedback mechanism to assess various design alternatives and their effects on the optimal LRU design and the corresponding (after-sales) costs. Moreover, the set partitioning formulation is suitable to solve large instances, while the binary linear programming formulation is not. In addition to the computation times, we observed that optimal solutions to LRU DESIGN have larger LRUs when the cost per time unit of system downtime increases, because this enables faster replacement and thus avoids large downtime cost. Finally, we found that managers should urge their designers to reduce the number of connections and predecessor–successor relationships in a system's design.

ACKNOWLEDGMENTS

The authors would like to thank Willem van Jaarsveld for his suggestions and remarks that aided in establishing Theorem 3. Moreover, we would like to express our gratitude to Gerhard Woeginger for better positioning the integrality result in the literature of combinatorial optimization. Finally, we thank the reviewers for their constructive feedback and one reviewer for the generalization to subadditive failure rate functions. The authors also express their gratitude to ASML and Dutch Railways for their cooperation in this research. Furthermore, we gratefully acknowledge the support of The Netherlands Organisation for Scientific Research.

DATA AVAILABILITY STATEMENT

No new data was generated in the course of this research.

ORCID

Joost de Kruijff  <https://orcid.org/0000-0002-3348-4019>

Joachim Arts  <https://orcid.org/0000-0002-3749-6965>

Geert-Jan van Houtum  <https://orcid.org/0000-0003-4225-5434>

REFERENCES

- Ahuja, R. K. (1986). Algorithms for the minimax transportation problem. *Naval Research Logistics Quarterly*, 33(4), 725–739.
- Air France industries KLM Engineering & Maintenance. (2013). Engines accessories repair & LRU management. http://www.afiklmem.com/AFIKLMEM/pdf/AFIKLMEM_EnginesLRU.pdf
- Alvaro, J. R. A., & Harrison, M. K. (2011). Optimal component sharing in a product family by simultaneous consideration of minimum

- description length and impact metric. *Engineering Optimization*, 43(2), 175–192.
- Anstee, R. P., & Farber, M. (1984). Characterizations of totally balanced matrices. *Journal of Algorithms*, 5(2), 215–230.
- Asiedu, Y., & Gu, P. (1998). Product life cycle cost analysis: State of the art review. *International Journal of Production Research*, 36(4), 883–908.
- Balabaert, I., Sengupta, A., & Sherrington, M. (2016). *Julia: High performance programming*. Packt Publishing.
- Baldwin, C. Y., & Clark, K. B. (2000). *Design rules: The power of modularity* (Vol. 1). MIT Press.
- Ball, M. O., Hoffman, R., Odoni, A. R., & Rifkin, R. (2003). A stochastic integer program with dual network structure and its application to the ground-holding problem. *Operations Research*, 51(1), 167–171.
- Birolini, A. (2007). *Reliability engineering* (Vol. 5). Springer.
- Bonami, P., Günlük, O., & Linderoth, J. (2018). Globally solving non-convex quadratic programming problems with box constraints via integer programming methods. *Mathematical Programming Computation*, 10(3), 333–382.
- Boros, E., Crama, Y., & Hammer, P. L. (1992). Chvátal cuts and odd cycle inequalities in quadratic 0–1 optimization. *SIAM Journal on Discrete Mathematics*, 5(2), 163–177.
- Boros, E., & Hammer, P. L. (1993). Cut-polytopes, Boolean quadric polytopes and nonnegative quadratic pseudo-Boolean functions. *Mathematics of Operations Research*, 18(1), 245–253.
- Brasseur, A., Leppert, W., & Pradille, A. (2012). Inside the 747-8 new environmental control system. *Boeing Aero Magazine*, 1, 19–25.
- Briant, O., & Naddef, D. (2004). The optimal diversity management problem. *Operations Research*, 52(4), 515–526.
- Charfreitag, J., Jünger, M., Mallach, S., & Mutzel, P. (2022). *McSparse: Exact solutions of sparse maximum cut and sparse unconstrained binary quadratic optimization problems*. In 2022 Proceedings of the Symposium on Algorithm Engineering and Experiments (ALENEX) (pp. 54–66). SIAM.
- Churchill, A. M., Lovell, D. J., Mukherjee, A., & Ball, M. O. (2012). Determining the number of airport arrival slots. *Transportation Science*, 47(4), 526–541.
- CNET News. (2001). California power outages suspended—For now. <https://www.cnet.com/news/california-power-outages-suspended-for-now/>
- De Fazio, T., & Whitney, D. (1987). Simplified generation of all mechanical assembly sequences. *IEEE Journal on Robotics and Automation*, 3(6), 640–658.
- Dell Inc. (2016). Dell precision 17 7000 series (7710) owner's manual. http://topics-cdn.dell.com/pdf/precision-m7710-workstation_Owner's_manual_en-us.pdf
- Deza, M. M., Laurent, M., & Weismantel, R. (1997). *Geometry of cuts and metrics* (Vol. 2). Springer.
- Dhillon, B. S. (1999). *Engineering maintainability: How to design for reliability and easy maintenance*. Gulf Professional Publishing.
- Driessen, J. P. C. (2018). *Life cycle costs optimization for capital goods* (PhD thesis). Eindhoven University of Technology.
- Driessen, J. P. C., Arts, J. J., & van Houtum, G. J. (2017). *Optimal commonality and reliability: A life cycle costs perspective*, BETA publicatie [Working papers]. Technische Universiteit Eindhoven.
- Dunning, I., Huchette, J., & Lubin, M. (2017). Jump: A modeling language for mathematical optimization. *SIAM Review*, 59(2), 295–320.
- Durand, Y. (2001). The maintainability approach at Alstom transport. *World congress on Railway Research (WCRR 2001)*.
- Fulkerson, D. R., Hoffman, A. J., & Oppenheim, R. (1974). *On balanced matrices*. In M. L. Balinski (Ed.), *Pivoting and extension Mathematical Programming Study* (Vol. 1, pp. 120–132). Springer.
- Gamvros, I., & Raghavan, S. (2012). Multi-period traffic routing in satellite networks. *European Journal of Operational Research*, 219(3), 738–750.
- Gupta, S., & Krishnan, V. (1998). Product family-based assembly sequence design methodology. *IIE Transactions*, 30(10), 933–945.
- Hillier, M. S., & Brandeau, M. L. (1998). Optimal component assignment and board grouping in printed circuit board manufacturing. *Operations Research*, 46(5), 675–689.
- Hoffman, A. J., Kolen, A. W. J., & Sakarovitch, M. (1985). Totally-balanced and greedy matrices. *SIAM Journal on Algebraic Discrete Methods*, 6(4), 721–730.
- Jünger, M., & Mallach, S. (2021). Exact facial odd-cycle separation for maximum cut and binary quadratic optimization. *INFORMS Journal on Computing*, 33(4), 1419–1430.
- Kim, J., Saidani, M., & Kim, H. M. (2021). Designing an optimal modular-based product family under intellectual property and sustainability considerations. *Journal of Mechanical Design*, 143(11), 112002.
- Klauke, S. S., Laurent, K., & Leomant, A. (2015). Incremental development for maintenance. *Airbus Fast Magazine*, 3, 32–35.
- Kumar, U. D., Crocker, J., Knezevic, J., & El-Haram, M. (2012). *Reliability, maintenance and logistic support: A life cycle approach*. Springer Science & Business Media.
- Lambert, A. J. D. (2007). Optimizing disassembly processes subjected to sequence-dependent cost. *Computers & Operations Research*, 34(2), 536–551.
- Lubin, M., & Dunning, I. (2015). Computing in operations research using Julia. *INFORMS Journal on Computing*, 27(2), 238–248.
- McCormick, G. P. (1976). Computability of global solutions to factorable nonconvex programs: Part I—convex underestimating problems. *Mathematical Programming*, 10(1), 147–175.
- Meier, C., Yassine, A. A., & Browning, T. R. (2006). Design process sequencing with competent genetic algorithms. *Journal of Mechanical Design*, 129(6), 566–585.
- Muckstadt, J. A. (2004). *Analysis and algorithms for service parts supply chains*. Springer Science & Business Media.
- Newcomb, P. J., Bras, B., & Rosen, D. W. (1998). Implications of modularity on product design for the life cycle. *Journal of Mechanical Design*, 120(3), 483–491.
- Nicolai, R. P., & Dekker, R. (2008). *Optimal maintenance of multi-component systems: A review*. In *Complex system maintenance handbook* (pp. 263–286). Springer.
- Norman, G. (1990). Life cycle costing. *Property Management*, 8(4), 344–356.
- Öner, K., Franssen, R., Kiesmüller, G., & van Houtum, G.-J. (2007). *Life cycle costs measurement of complex systems manufactured by an engineer-to-order company*. In R. G. Qui, D. W. Russell, & W. G. Sullivan (Eds.), *The 17th International Conference on Flexible Automation and Intelligent Manufacturing* (pp. 569–589). Elsevier.
- Öner, K. B., Scheller-Wolf, A., & van Houtum, G. J. (2013). Redundancy optimization for critical components in high-availability technical systems. *Operations Research*, 61(1), 244–264.
- Padberg, M. (1989). The Boolean quadric polytope: Some characteristics, facets and relatives. *Mathematical Programming*, 45, 139–172.
- Papalambros, P. Y., & Michelen, N. F. (1997). *Model-based partitioning in optimal design of large engineering systems*. In *Multidisciplinary design optimization: State-of-the-art* (pp. 209–226). SIAM.
- Parada Puig, J. E., & Basten, R. J. I. (2015). Defining line replaceable units. *European Journal of Operational Research*, 247(1), 310–320.
- Parent, K. (2000). AvantiCom moves on pilot project. <http://www.edn.com/electronics-news/4362659/AvantiCom-Moves-on-Pilot-Project>

- Rehfeldt, D., Koch, T., & Shinano, Y. (2023). Faster exact solution of sparse MaxCut and QUBO problems. *Mathematical Programming Computation*, 15, 445–470.
- Rendl, F., Rinaldi, G., & Wiegele, A. (2010). Solving max-cut to optimality by intersecting semidefinite and polyhedral relaxations. *Mathematical Programming*, 121, 307–335.
- Saranga, H., & Kumar, U. D. (2006). Optimization of aircraft maintenance/support infrastructure using genetic algorithms-level of repair analysis. *Annals of Operations Research*, 143(1), 91–106.
- Sharman, D. M., & Yassine, A. A. (2004). Characterizing complex product architectures. *Systems Engineering*, 7(1), 35–60.
- Sosa, M. E., Eppinger, S. D., & Rowles, C. M. (2007). A network approach to define modularity of components in complex products. *Journal of Mechanical Design*, 129(11), 1118–1129.
- Steward, D. V. (1981). The design structure system: A method for managing the design of complex systems. *IEEE Transactions on Engineering Management*, 28(3), 71–74.
- Swaminathan, J. M., & Tayur, S. R. (1998). Managing broader product lines through delayed differentiation using vanilla boxes. *Management Science*, 44(12), S161–S172.
- Thebeau, R. E. (2001). *Knowledge management of system interfaces and interactions from product development processes* (Master's thesis). Massachusetts Institute of Technology.
- Thomas, L. (1986). A survey of maintenance and replacement models for maintainability and reliability of multi-item systems. *Reliability Engineering*, 16(4), 297–309.
- Thonemann, U. W., & Brandeau, M. L. (2000). Optimal commonality in component design. *Operations Research*, 48(1), 1–19.
- van Deursen, R. W. H. (2020). *A tool to define line replaceable units to improve maintenance* (Master's thesis). Eindhoven University of Technology.
- van Geel, G. C. J. (2018). *Improving the LRU definition process* (Master's thesis). Eindhoven University of Technology.
- Wilschut, T., Etman, L. F. P., Rooda, J. E., & Adan, I. J. B. F. (2017). Multilevel flow-based markov clustering for design structure matrices. *Journal of Mechanical Design*, 139(12), 121402.
- Xie, W., Liao, H., & Jin, T. (2014). Maximizing system availability through joint decision on component redundancy and spares inventory. *European Journal of Operational Research*, 237(1), 164–176.
- Yu, T., Yassine, A. A., & Goldberg, D. E. (2007). An information theoretic method for developing modular architectures using genetic algorithms. *Research in Engineering Design*, 18(2), 91–109.
- Zhang, X., & Bard, J. F. (2006). A multi-period machine assignment problem. *European Journal of Operational Research*, 170(2), 398–415.

How to cite this article: Driessen, J., de Kruijff, J., Arts, J., & van Houtum, G. -J. (2023). Optimal design of line replaceable units. *Naval Research Logistics (NRL)*, 1–21. <https://doi.org/10.1002/nav.22146>

APPENDIX A: DERIVING THE SUCCESSOR COLLECTION $H(e)$

We determine $H(e)$ for all edges $e \in E$ in polynomial time by the following polynomial algorithms, where Algorithm 2 is called by Algorithm 1.

Algorithm 1. Derive $H(e)$ for all edges $e \in E$

```

1: procedure REMOVAL_EDGES( $E, A$ )
2:    $\hat{E} \leftarrow E$ 
3:    $\tilde{E} \leftarrow \text{DEGREE}(D(\hat{E}, A))$ 
4:   while  $\tilde{E} \neq \emptyset$  do
5:     for all  $e \in \tilde{E}$  do
6:        $H(e) \leftarrow \{e\}$ 
7:     end for
8:     for all  $e \in \tilde{E}$  do
9:       for all  $(e, z) \in A$  do
10:         $H(e) \leftarrow H(e) \cup H(z)$ 
11:      end for
12:     end for
13:      $\hat{E} \leftarrow \hat{E} \setminus \tilde{E}$ 
14:      $\tilde{E} \leftarrow \text{DEGREE}(D(\hat{E}, A))$ 
15:   end while
16:   return  $H(e)$  for all  $e \in E$ 
17: end procedure

```

Algorithm 2. Determine all edges $e \in E$ that have no successors in D

```

1: procedure DEGREE( $D$ )
2:    $\tilde{E} \leftarrow \emptyset$ 
3:   for all  $e \in E$  do
4:     if  $\delta^{out}(e) = 0$  then
5:        $\tilde{E} \leftarrow \tilde{E} \cup \{e\}$ 
6:     end if
7:   end for
8:   return  $\tilde{E}$ 
9: end procedure

```

APPENDIX B: LRU WITH INTERNAL LINKS THAT MUST BE BROKEN

The precedence graph determines which edges must necessarily be broken before a certain edge can be broken. Therefore it is possible to define a LRU Q such that certain connections within LRU Q must be broken in order to be able to detach LRU Q from the rest of the system. While such situations do not happen often in optimal designs in practice, it is sometime inevitable. Why this cannot physically be avoided is illustrated by a simple example.

Consider two cog-wheels with a chain between them such that motion in one cog is transmitted to the other cog. Every bicycle has such a system to transmit motion from the pedals to the rear-wheel and a picture of such a system is shown below in Figure B1. The chain consists of many links. Now one may wish to define the chain as a LRU. To remove the chain from the two-cog-wheels on which it sits, one must necessarily break the connection between two links in the chain, despite the fact that such a connection is internal to the LRU/chain.

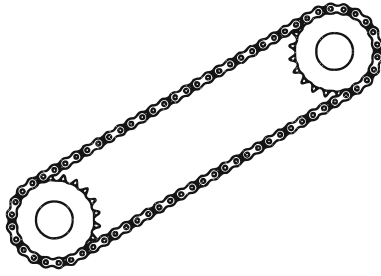


FIGURE B1 Two cog-wheels connected with a chain.

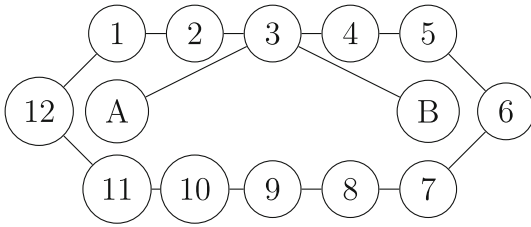


FIGURE B2 Connection graph for two cog-wheels with a chain.

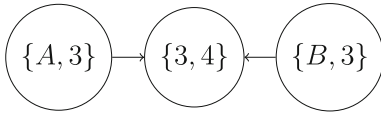


FIGURE B3 Precedence graph for two cog-wheels with a chain.

This example can be formalized in the connection and precedence graph shown in Figures B2 and B3 respectively. In these graphs A and B are the cog-wheels and nodes 1 to 12 correspond to the links in the chain. In this example the chain LRU is given by $Q = \{1, 2, \dots, 12\}$ so that $\Gamma(Q) = \{\{A, 3\}, \{B, 3\}, \{3, 4\}\}$. The fact that $\{3, 4\}$ is internal to chain reflects the fact that it is physically impossible to remove the chain from the cog-wheels without first breaking a connection between two links in the chain. Thus the precedence graph models the physical necessity of breaking certain connections before it is possible to break/access other connections. The fact that $\Gamma(Q)$ may contain connections that are internal to an LRU Q is a modeling feature that allows us to model systems appropriately. However, we do observe that most LRUs in optimal designs do not have internal connections that need to be broken.

APPENDIX C: LRU DESIGN WITHOUT PARTITIONING CONSTRAINT

In this appendix, we explain the model of designing LRUs without considering a partitioning constraint. We call this problem C-LRU DESIGN. Let a system be represented by a connection graph $G = (V, E)$ and a precedence graph $D = (E, A)$, with V , E and A corresponding to the vertex set, edge set and arc set, respectively. Furthermore, we let $\lambda(v) > 0$ and $\ell(v) > 0$ be the failure rate and purchase cost of part $v \in V$, and we define $w(e) > 0$ as the cost of breaking edge $e \in E$. An arc $(\{u, v\}, \{v, x\}) \in A$ denotes that we have to break $\{v, x\}$ prior to breaking $\{u, v\}$.

A part $v \in V$ belongs to at least one LRU. That is, part v is replaced upon its own failure, but it may also be replaced upon the failure of a part $u \in V : u \neq v$. Therefore, we represent a LRU differently from LRU DESIGN. We let a LRU be a tuple characterized by a replacement set and a failure set, that is, $Q = (R_Q, F_Q)$ where Q is the LRU, $R_Q \subseteq V$ is the replacement set, and $F_Q \subseteq R_Q$ is the failure set. The failure of a part in the failure set triggers replacement of the LRU. The replacement set is replaced if any of the vertices in the failure set fails. Furthermore, we assume that a part $v \in V$ belongs to exactly one failure set, that is, the failure sets partition V .

Next, we study what happens when a LRU Q fails, or technically what happens when a part $v \in F_Q$ fails. In this case, we have to break all edges $e \in \Gamma(R_Q)$, where $\Gamma(R_Q) = \bigcup_{e \in B(R_Q)} H(e)$. The failure rate of LRU Q is given by $\sum_{v \in F_Q} \lambda(v)$, because all parts $v \in F_Q$ induce the replacement of R_Q . Similarly, the total purchase cost of LRU Q is given by $\sum_{v \in R_Q} \ell(v)$. This yields the following average cost per time unit for LRU Q :

$$\omega_c(Q) = \sum_{e \in \Gamma(R_Q)} w(e) \sum_{v \in F_Q} \lambda(v) + \sum_{u \in R_Q} \ell(u) \sum_{v \in F_Q} \lambda(v).$$

We are interested in determining the optimal LRU design. Let S_c be a collection of LRUs such that $\emptyset \notin S_c$ and each part $v \in V$ is included in at least one replacement set and in exactly one failure set; that is, $\bigcup_{Q \in S_c} F_Q = V$, $F_Q \cap F_{Q'} = \emptyset$ for all $Q, Q' \in S_c : F_Q \neq F_{Q'}$, and $F_Q \subseteq R_{Q'}$ for each LRU $Q \in S_c$. The average costs per time unit of a LRU design S_c are given by

$$\begin{aligned} \pi_c(S_c) &= \sum_{Q \in S_c} \omega_c(Q) = \sum_{Q \in S_c} \sum_{b \in \Gamma(R_Q)} w(b) \sum_{v \in F_Q} \lambda(v) \\ &+ \sum_{Q \in S_c} \sum_{u \in R_Q} \ell(u) \sum_{v \in F_Q} \lambda(v). \end{aligned} \quad (C1)$$

Next, we define C-LRU DESIGN as: What is the LRU Design S_c that minimizes $\pi_c(S_c)$?

For implementation, we use the binary programming formulation of C-LRU DESIGN as explained in Driessen (2018), which is similar to the binary programming formulation of LRU DESIGN.

APPENDIX D: LEMMA USED IN THE PROOF OF THEOREM 1

The proof of Theorem 1, which states that an optimal solution to LPM does not contain a LRU cycle, uses Lemma 2. This appendix contains the statement, an interpretation and the proof of that lemma. Recall that we define the set of edges that is broken for a LRU X but not for a LRU Y by $\mathcal{F}(X, Y) = \Gamma(X) \setminus \Gamma(Y)$. Furthermore, recall that we use modular arithmetic for the indices of LRUs that form a cycle Q_i , Q_{i-1} , and Q_{i+1} with $1 \leq i \leq n$, $n+1 \equiv 1 \pmod{n}$, and $Q_0 \equiv Q_n$.

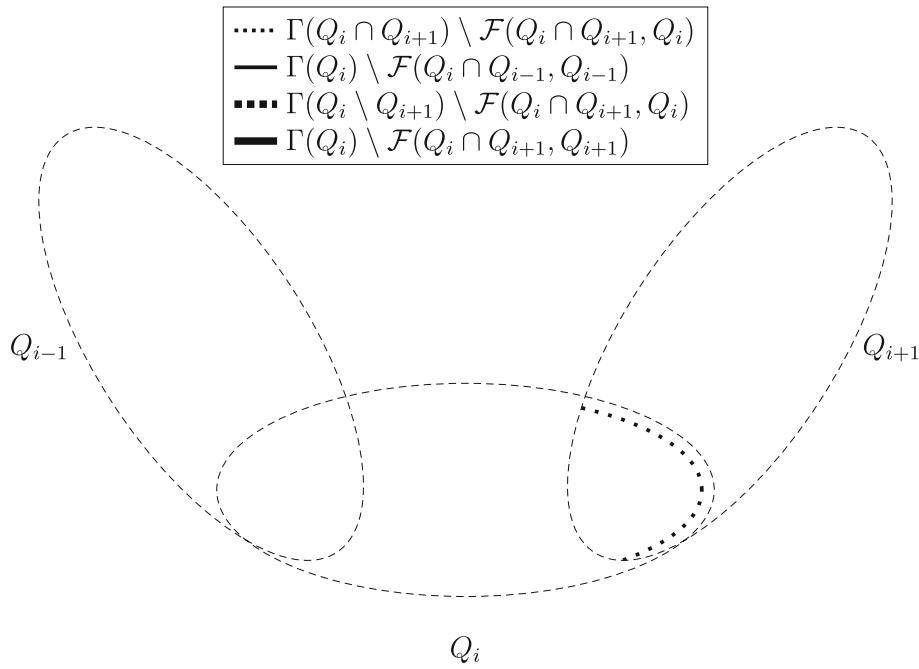


FIGURE D1 Interpretation of Lemma 2. The edges that cross the lines shown in the legend are contained in the corresponding sets of the legend.

Lemma 2. *Given a solution x to LPM that contains a LRU cycle $C = \{Q_1, Q_2, \dots, Q_n\} \subseteq S$ with minimal n , we have $\Gamma(Q_i \cap Q_{i+1}) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_i) \subseteq \Gamma(Q_i) \setminus \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})$ and $\Gamma(Q_i \setminus Q_{i+1}) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_i) \subseteq \Gamma(Q_i) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_{i+1})$. Furthermore, we have $\Gamma(Q_i \cap Q_{i-1}) \setminus \mathcal{F}(Q_i \cap Q_{i-1}, Q_i) \subseteq \Gamma(Q_i) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_{i+1})$ and $\Gamma(Q_i \setminus Q_{i-1}) \setminus \mathcal{F}(Q_i \cap Q_{i-1}, Q_i) \subseteq \Gamma(Q_i) \setminus \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})$.*

Lemma 2 can be interpreted by considering Figure D1, which is a simplified part of a solution in which we depict only the LRUs Q_{i-1} , Q_i , and Q_{i+1} , and we assume that $A = \emptyset$. The first claim in Lemma 2 states that the set of edges that are broken for LRU Q_i as well as for the intersection $Q_i \cap Q_{i+1}$ (the edges that cross the thin dotted line in Figure D1) is a subset of the edges that are broken for Q_i except for the edges that are broken for $Q_i \cap Q_{i-1}$ but not for Q_{i-1} (the edges that cross the thin solid line in Figure D1). Second, the set of edges that is broken for both Q_i and $Q_i \setminus Q_{i+1}$ (the edges that cross the thick dotted line in Figure D1) is a subset of the edges that are broken for Q_i except for the edges that are broken for $Q_i \cap Q_{i+1}$ but not for Q_{i+1} (the edges that cross the thick solid line in Figure D1).

The third and fourth claim from Lemma 2 can be interpreted similarly to the first two claims, where Q_{i+1} is replaced by Q_{i-1} and vice versa.

The same interpretation—as illustrated in Figure D1—holds when $A \neq \emptyset$, but this makes exposition cumbersome. The proof of Lemma 2 is given below, for arbitrary A .

Proof of Lemma 2. Let x be a solution to LPM that contains a LRU cycle $C = \{Q_1, Q_2, \dots, Q_n\} \subseteq S$ with minimal n . First, we

prove that if $e = \{u, v\} \in \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})$ then $u, v \in Q_{i-1}$. Subsequently, we use this result to show that $\Gamma(Q_i \cap Q_{i+1}) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_i) \subseteq \Gamma(Q_i) \setminus \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})$. Similarly, we show that if $e = \{u, v\} \in \mathcal{F}(Q_i \cap Q_{i+1}, Q_{i+1})$ then $u, v \in Q_{i+1}$, and we use this to prove that $\Gamma(Q_i \setminus Q_{i+1}) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_i) \subseteq \Gamma(Q_i) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_{i+1})$.

Let $e = \{u, v\} \in \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})$. If $u \in Q_{i-1}$ and $v \notin Q_{i-1}$, it follows directly that $e \in \Gamma(Q_{i-1})$ as $e \in B(Q_{i-1})$. This contradicts that $e \in \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})$. If $u, v \notin Q_{i-1}$, then there exists a (sub)path (e_1, e_2, \dots, e) in the precedence graph with $e_1 = \{u_1, v_1\} : u_1 \in Q_i \cap Q_{i-1}$. Consequently, there is an $e_j = \{u_j, v_j\} : u_j \in Q_{i-1}, v_j \notin Q_{i-1}$ in this path, due to the assumption that $(\{\bar{u}, \bar{v}\}, \{\bar{v}, \bar{x}\}) \in A$ for $\bar{u}, \bar{v}, \bar{x} \in V$. Therefore, $e_j \in \Gamma(Q_{i-1})$ and thus $e \in \Gamma(Q_{i-1})$. Subsequently, $e \notin \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})$, which is a contradiction. Therefore, if $e = \{u, v\} \in \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})$, then $u, v \in Q_{i-1}$.

Next, we show that $\Gamma(Q_i \cap Q_{i+1}) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_i) \subseteq \Gamma(Q_i) \setminus \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})$. Let $e = \{u, v\} \in \Gamma(Q_i \cap Q_{i+1})$ and suppose that $e \in \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})$. There exists a (sub)path (e_1, e_2, \dots, e) in the precedence graph with $e_1 = \{u_1, v_1\} : u_1 \in Q_i \cap Q_{i+1}$. Also, $u_1 \notin Q_{i-1}$, because C is a LRU cycle with minimal n (implying that $Q_{i-1} \cap Q_i \cap Q_{i+1} = \emptyset$, otherwise one can omit Q_i and obtain a LRU cycle with fewer LRUs, which is a contradiction). Furthermore, $u, v \in Q_{i-1}$ as $e = \{u, v\} \in \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})$ (see above). Hence, there exists $e_j = \{u_j, v_j\} : u_j \in$

$Q_{i-1}, v_j \notin Q_{i-1}$ in this path, as $(\{\bar{u}, \bar{v}\}, \{\bar{v}, \bar{x}\}) \in A$ for $\bar{u}, \bar{v}, \bar{x} \in V$. Therefore, $e_j \in \Gamma(Q_{i-1})$ and thus $e \in \Gamma(Q_{i-1})$. Consequently, $e \notin \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})$ which is a contradiction. Hence, if $e \in \Gamma(Q_i \cap Q_{i+1})$, then $e \notin \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})$. Next, we consider $\Gamma(Q_i \cap Q_{i+1}) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_i)$, and have that $\Gamma(Q_i \cap Q_{i+1}) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_i) \subseteq \Gamma(Q_i)$ as $\mathcal{F}(Q_i \cap Q_{i+1}, Q_i) = \Gamma(Q_i \cap Q_{i+1}) \setminus \Gamma(Q_i)$. Furthermore, each $e \in \Gamma(Q_i \cap Q_{i+1}) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_i)$ satisfies $e \in \Gamma(Q_i \cap Q_{i+1})$ and thus $e \notin \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})$, as we proved in the foregoing. Hence, we have $\Gamma(Q_i \cap Q_{i+1}) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_i) \subseteq \Gamma(Q_i) \setminus \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})$.

Let us now prove that if $e = \{u, v\} \in \mathcal{F}(Q_i \cap Q_{i+1}, Q_{i+1})$, then $u, v \in Q_{i+1}$. The proof is identical to proving that if $e \in \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})$ then $u, v \in Q_{i-1}$, but Q_{i-1} is replaced by Q_{i+1} . We continue by proving that $\Gamma(Q_i \setminus Q_{i+1}) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_i) \subseteq \Gamma(Q_i) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_{i+1})$. Let $e = \{u, v\} \in \Gamma(Q_i \setminus Q_{i+1})$ and suppose that $e \in \mathcal{F}(Q_i \cap Q_{i+1}, Q_{i+1})$. There exists a (sub)path (e_1, e_2, \dots, e) in the precedence graph with $e_1 = \{u_1, v_1\} : u_1 \in Q_i \setminus Q_{i+1}$. Moreover, $u, v \in Q_{i+1}$ as $e = \{u, v\} \in \mathcal{F}(Q_i \cap Q_{i+1}, Q_{i+1})$ (from foregoing). Hence, there exists

$e_j = \{u_j, v_j\} : u_j \in Q_{i+1}, v_j \notin Q_{i+1}$ in this path, as $(\{\bar{u}, \bar{v}\}, \{\bar{v}, \bar{x}\}) \in A$ for $\bar{u}, \bar{v}, \bar{x} \in V$. Therefore, $e_j \in \Gamma(Q_{i+1})$ and thus $e \in \Gamma(Q_{i+1})$. As a consequence, $e \notin \mathcal{F}(Q_i \cap Q_{i+1}, Q_{i+1})$, which is a contradiction. Hence, if $e \in \Gamma(Q_i \setminus Q_{i+1})$, then $e \notin \mathcal{F}(Q_i \cap Q_{i+1}, Q_{i+1})$. Next, we consider $\Gamma(Q_i \setminus Q_{i+1}) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_i)$. We observe that $\mathcal{F}(Q_i \cap Q_{i+1}, Q_i) = \mathcal{F}(Q_i \setminus Q_{i+1}, Q_i)$, because $Q_i \cap Q_{i+1}$ and $Q_i \setminus Q_{i+1}$ partition Q_i , and thus the edges existing between $Q_i \cap Q_{i+1}$ and $Q_i \setminus Q_{i+1}$ are the same. Hence, $\Gamma(Q_i \setminus Q_{i+1}) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_i) = \Gamma(Q_i \setminus Q_{i+1}) \setminus \mathcal{F}(Q_i \setminus Q_{i+1}, Q_i) \subseteq \Gamma(Q_i)$, where the last inequality holds as $\mathcal{F}(Q_i \setminus Q_{i+1}, Q_i) = \Gamma(Q_i \setminus Q_{i+1}) \setminus \Gamma(Q_i)$. Finally, we have $\Gamma(Q_i \setminus Q_{i+1}) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_i) \subseteq \Gamma(Q_i) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_{i+1})$, since each $e \in \Gamma(Q_i \setminus Q_{i+1})$ is such that $e \notin \mathcal{F}(Q_i \cap Q_{i+1}, Q_{i+1})$.

Finally, we can do the same analysis as above, but replace Q_{i-1} by Q_{i+1} and vice versa. Then, we obtain $\Gamma(Q_i \cap Q_{i-1}) \setminus \mathcal{F}(Q_i \cap Q_{i-1}, Q_i) \subseteq \Gamma(Q_i) \setminus \mathcal{F}(Q_i \cap Q_{i+1}, Q_{i+1})$ and $\Gamma(Q_i \setminus Q_{i-1}) \setminus \mathcal{F}(Q_i \cap Q_{i-1}, Q_i) \subseteq \Gamma(Q_i) \setminus \mathcal{F}(Q_i \cap Q_{i-1}, Q_{i-1})$. ■