# Integer Formulation and Data Analysis of a Real-World Course Timetabling Problem

Khang T.T.M. Nguyen
Faculty of Software Engineering
University of Information Technology
Ho Chi Minh city, Vietnam
khangnttm@uit.edu.vn

Nuong T.H. Tran
Faculty of Mathematics
University of Science
Ho Chi Minh city, Vietnam
tthnuong@math.hcmus.edu.vn

***Abstract*—Belonging to the class of hard combinatorial optimization problems, educational timetabling problems are considered to be challenging and attractive to operation research community in recent years. In this paper, we investigate a course timetabling problem in practice by introducing an integer formulation and data analysis of this problem. Fourteen data instances are taken from Faculty of Information Technology, University of Science in Vietnam. Thirteen measurements are used to analyze the hardness of these instances.**

***Keywords-integer formulation, course timetabling, curriculum-based course timetabling.***

## I. Introduction

The university timetabling problems in general are known to be NP-complete problems [1]. It often involves in assignment of a set of courses to a given number of periods and rooms subject to some hard and soft constraints. Hard constraints must be satisfied to produce a feasible solution and soft constraints should be satisfied as much as possible. Problem description and constraints are often varied a lot based on concrete requirements of each specific institution. The intention of this paper is investigating a real-world course timetabling problem. In order to make the problem statement being clear and concise, we introduce an integer formulation to represent problem’s constraints and objective function. In addition to that, an analysis of the hardness of fourteen practical data instances of the considered problem is done using thirteen measurements.

The paper is organized as follow: section II describes basic concepts used to state the problem, section III gives the integer formulation of the problem and section IV shows the data analysis.

## II. Problem Description

The general aim of the educational timetabling problem is to find an appropriate assignment of a set of courses (i.e., events) into a limited number of periods and rooms in such a way that satisfying a number of pre-defined constraints. The problem considered in this paper includes 8 hard constraints and 9 soft constraints. Each soft constraint is associated with a weight factor, which is listed right next to the constraint described below, to represent it level of importance. The major concepts of the problem are course and curriculum. A course is a group of lectures that have the same lecturer and attending classes. The main information of a course includes the number of periods in a week that this course must happen, the lecturer teaching this course, the classes attending this course (also the number of students of these classes). Note that some courses may be pre-assigned by the institution's staff. Note that a course may be spitted into distinct groups of consecutive lectures called block elements.The last concept is curriculum, which is a group of courses that should not be overlapped due to the special requirements of the university. In our considered problem, a course is allowed to belong to more than one curriculum.

This real-world problem includes 10 hard constraints, which must be satisfied and 10 soft constraints, which should be satisfied as much as possible. Statements and formulation of these constraints and the objective function of the problem are given in the next sections.

## III. Problem Formulation

### *A. Model's Parameters*

In this subsection, parameters of the model are described. The number of the parameters is not small due to the fact that the considered timetabling is a real-world and complex one and requires several definitions for being formulated.

– $P = \{1,2,..,PeriodMax\}$: set of studying periods in a week.

- $Q$: set of studying day in a week, $U_q$: set of periods belonging to the $q^{th}$ day
- $V$: set of learning session in a week (e.g., morning session, afternoon session)
- $S_v = \{s_1^v, s_2^v, .., s_{|S_v|}^v\}$: set of periods belonging to session $v \in V$
- $T$, $C$, $R$, $NS_c$, $RC_r$: set of lecturers, set of classes, set of rooms, the number of student of class $c$, respectively
- Matrix $AT_{|T|\times|P|}$, $AC_{|C|\times|P|}$ and $AR_{|R|\times|P|}$ respectively represents the available periods of lecturers, classes and rooms:

$AT_{tp} = \begin{cases} 1, if\ lecturertis\ available\ at\ period\ p \\ 0, otherwise \end{cases}$, with $t \in T, p \in P$.

$AC_{cp} = \begin{cases} 1, if\ classtis\ available\ at\ period\ p \\ 0, otherwise \end{cases}$, with $c \in C, p \in P$

$AR_{rp} = \begin{cases} 1, if\ room\ r\ is\ available\ at\ period\ p \\ 0, otherwise \end{cases}$, with $r \in R, p \in P$

- Matrix $PT_{|T|\times|P|}$ represents periods that each lecturer prefers:

$PT_{tp} = \begin{cases} 1, if\ lecturer\ t\ prefers\ to\ teach\ at\ period\ p \\ 0, otherwise \end{cases}$, with $\in T, p \in P$.

- $G$ : set of room groups
- $H_g$ : set of rooms belonging to room group $g$.
- $A$ : set of courses thatneed to bescheduled.
- $TA_a$ : the lecturer of course $a$.
- $CA_a$ : set of classes attending course $a$.
- $B$ : set of blocks, a block is a group of consecutive periods in the same session of an course.
- $PB = \{(b, p, r), b \in B, p \in P, r \in R\}$ : pre-assignment information of courses, each triple $(b, p, r) \in PB$ let us know that the first assigned period of block $b$ is pre-assigned to period $p$ and room $r$.
- $B_a$ : set of blocks belonging to courses $a$, with $a \in A$.
- $L_b$: set of periods of block $b$, with $b \in B$.
- $TB_t$: set of blocks that lecturer $t$ teaches, with $t \in T$.
- $CB_c$: set of blocks that class $c$ attends, with $c \in C$.
- $M$: set of curriculum – each curriculum is a set of courses that should not be overlapped.
- $N_m$: set of courses belonging to curriculum $m$.

B. *Model's Variables*

Here we use two kinds of variables: auxiliary variables and decision variables, both of them are binary.

- Decision variable: $x_{bpr}$ with $b \in B, p \in P, r \in R$:
  + $x_{bpr} = 1$if the first period of block $b$ is assigned to period $p$ and room $r$, $x_{bpr} = 0$ if otherwise.
- Auxiliary variables (auxiliary variables): $l_{bpr}$, $y_{tbpr}$, $z_{cbpr}$, with $b \in B, p \in P, r \in R, t \in T, c \in C$ which:
  + $l_{bpr} = 1$ if there exists a period of block $b$ which is assigned to period $p$ and room $r$, $l_{bpr} = 0$ if otherwise.
  + $y_{tbpr} = 1$if lecturer $t$ teaches block $b$ at period $p$ and room $r$, $y_{tbpr} = 0$ if otherwise.

$z_{cbpr} = 1$if class $c$ attends block $b$ at period $p$ and room $r$, $z_{cbpr} = 0$ if otherwise.

C. *Hard Constraint Formulation*

1. *All lectures that belong to the same block must assign to consecutive periods and the same room.*

$$\forall b \in B, \forall p \in P, \forall h \in \{1,2, .., L_b - 1\}, \forall r \in R, x_{bpr} - l_{b(p+h)r} \le 0$$

2. *Each lecturer must not teach more than one block at the same period.*

$$\forall t \in T, \forall p \in P, \forall r \in R, \sum_{b\in B}\sum_{r\in R} y_{tbpr} \le 1$$

3. *Each class must not attend more than one block at the same period*

$$\forall c \in C, \forall p \in P, \forall r \in R, \sum_{b\in B}\sum_{r\in R} z_{cbpr} \le 1$$

4. *Each room must not be assigned to more than one block at the same period*

$$\forall p \in P, \forall r \in R, \sum_{b \in B} l_{bpr} \leq 1$$

5. *All lecturers, classes and rooms must not be assigned to periods that they are not available*

– Lecturer:

$$t \in T, b \in B, p \in P, r \in R, y_{tbpr} \leq AT_{tp}$$

– Class:

$$c \in C, b \in B, p \in P, r \in R, z_{cbpr} \leq AC_{cp}$$

– Room:

$$b \in B, p \in P, r \in R, l_{bpr} \leq AR_{rp}$$

6. *All blocks of all courses must be assigned*

$$\forall b \in B, \sum_{p \in P, r \in R} x_{bpr} = 1$$

7. *All periods assigned to the same block must belong to the same session*

$$\forall b \in B, \forall r \in B, \forall v \in V, \forall i \in [1, |S_v|]: x_{bs_i^v r}(s_{|S_v|}^v - s_i^v - L_b + 1) \geq 0$$

8. *Different blocks of the same courses must be assigned to distinct days.*

$$\forall a \in A, \forall r \in R, \forall q \in Q: \sum_{b \in B_a} \sum_{p \in U_q} x_{bpr} \leq 1$$

9. *Pre-assignment of courses must not be violated*

$$\forall (b, p, r) \in PB: x_{bpr} = 1$$

10. *Blocks must assigned to rooms having enough capacity*

$$\forall b \in B, \forall r \in R, \forall p \in P: \sum_{c \in C}(z_{cbpr} \times NS_c) \leq RC_r.$$

*D. Soft Constraint Violation Formulation*

Soft constraints are not able to be formulated as inequalities like hard constraints, due to the fact that soft contraints are allowed to be violated. Therefore, in this paper, we modeled the total violation of each constraint throughout a function. Value of this function will be added into objective function of a feasible solution to evaluate its quality.

1. *Each lecturer should not be assigned to different room groups on the same day.*

$$d_1 = \sum_{q \in Q} \sum_{t \in T} \sum_{g, g' \in G, g \neq g'} \sum_{r, r' \in H_g} \sum_{p, p' \in U_q, p \neq p'} \sum_{b, b' \in B} (y_{tbpr} \times y_{tb'p'r'})$$

2. *Each lecturer should not be assigned to splitted periods in the same session*

Total violation (denoted as $d_2$) of this constraint in a solution is the number of $(v, t, b, b', r, r', p)$ in the solution, with$v \in V, t \in T, b \in B, b' \in B, b \neq b', r \in H_g, r' \in H_g, p \in S_v \wedge (p+1) \in S_v \wedge (p+2) \in S_v$ and:

$$y_{tbpr} = 1 \wedge y_{tb(p+1)r} = 0 \wedge y_{tb'(p+2)r'} = 1$$

3. *Each lecturer should not be assigned to splitted periods in the same day*

Total violation (denoted as $d_3$) of this constraint in a solution is the number of $(q, t, b, b', r, r', p)$ in the solution, with$q \in Q, t \in T, b \in B, b' \in B, b \neq b', r \in H_g, r' \in H_g, p \in U_q \wedge (p+1) \in U_q \wedge (p+2) \in U_q$ and:

$$y_{tbpr} = 1 \wedge y_{tb(p+1)r} = 0 \wedge y_{tb'(p+2)r'} = 1$$

4. *The number of periods that each lecturer is assigned to per day should not be greater than 9:*

$$d_4 = \sum_{q \in Q, t \in T} max(0, \sum_{b \in B} \sum_{r \in R} \sum_{p \in U_q} y_{tbpr} - 9)$$

5. *Lectures should be assigned to periods that they prefer*

$$d_5 = \sum_{t \in T} \sum_{r \in R} \sum_{b \in B} \sum_{p \in P} (y_{tbpr} \times (1 - PT_{tp}))$$

6. *Classes should not be assigned to different room groups on the same day*

$$d_6 = \sum_{q \in Q} \sum_{c \in C} \sum_{g, g' \in G, g \neq g'} \sum_{r, r' \in H_g} \sum_{p, p' \in U_q, p \neq p'} \sum_{b, b' \in B} (z_{cbpr} \times z_{cb'p'r'})$$

7. *Courses that belongs to the same curriculum should not be assigned to different room group on the same*

$$d_7 = \sum_{q\in Q}\sum_{c\in C}\sum_{g,g'\in G,g\neq g'}\sum_{r,r'\in H_g}\sum_{p,p'\in U_q,p\neq p'}\sum_{m\in M}\sum_{a\in N_m}\sum_{b,b'\in B_a}(z_{cbpr}\times z_{cb'p'r'})$$

8. *The number of sessions that each lecturer is assigned to should be minimized*

$$d_8 = \sum_{t\in T}(\sum_{v\in V} Y_{tv} - \frac{\sum_{p\in P} AT_{tp}}{\sum_{b\in TB_t} L_b})$$

Whereas $Y_{tv} = \begin{cases} 1\ if\ \sum_{p\in S_v}\sum_{b\in B}\sum_{r\in R} y_{tpbr} > 0 \\ 0\ otherwise \end{cases}$

and $\frac{\sum_{p\in P} AT_{tp}}{\sum_{b\in TB_t} L_b}$ is the approximated values of the minimum number of sessions that each lecturer must teach each week.

9. *The number of sessions that each classes is assigned to should be minimized*

$$d_9 = \sum_{c\in C}(\sum_{v\in V} Z_{cv} - \frac{\sum_{p\in P} AC_{cp}}{\sum_{b\in TC_c} L_b})$$

Whereas $Z_{cv} = \begin{cases} 1\ if\ \sum_{p\in S_v}\sum_{b\in B}\sum_{r\in R} z_{cpbr} > 0 \\ 0\ otherwise \end{cases}$

and$\frac{\sum_{p\in P} AC_{cp}}{\sum_{b\in TC_c} L_b}$ is the approximated values of the minimum number of sessions that each lecturer must attend each week.

10. *Each class should not be assigned to splitted periods in the same session*

Total violation (denoted as $d_{10}$) of this constraint in a solution is the number of $(v, c, b, b', r, r', p)$ in the solution, with $v \in V, c \in C, b \in B, b' \in B, b \neq b', r \in H_g, r' \in H_g, p \in S_v \wedge (p+1) \in S_v \wedge (p+2) \in S_v$ and

$$\boldsymbol{z_{cbpr} = 1 \wedge z_{cb(p+1)r} = 0 \wedge z_{cb'(p+2)r'} = 1}.$$

*E. Objective Function*

An optimal solution $X^*$ is an assignment of all blocks of all courses into appropriate periods and rooms, i.e., decide values of all decision variables $x_{bpr}$, in such a way that all hard constraints are satisfied and the objective function value $f(X)$, which is often used by timetabling community [2], reaches minima:

$$f(X) = \sum_{i=1}^{10} w_i d_i$$

Whereas $w_i$ is weight of the i$^{th}$ soft constraint.

## IV. Data Analysis

In this section, we describe the fourteen real-world data instances collected from the Faculty of Information Technology, Ho Chi Minh city University of Science in Vietnam. In order to give a general analysis of the complexity of these instances, we introduce thirteen measurements as follow:

*A:* the number of block elements, *T*: the number of lecturers, *C*: the number of classes, R: the number of rooms, P: the number of pre-assigned block elements, Cr: the number of curriculum.

*CoL:* the number of un-ordered pairs of block elements that are taught by the same lecturer.

CoC: the number of un-ordered pairs of block elements that have the same attending class.

CoCr: the number of un-ordered pairs of block elements that belong to the same curriculum.

Co: the number of un-ordered pairs of block elements that are taught by the same lecturer or have the same attending class or belong to the same curriculum.

*AvT*: percentage of lecturers' availability.

$$AvT = \frac{\sum_{i=1}^{T} \text{NumberOf AvailablePeriodsOf Teacher} T_i}{T \times PpD \times D}$$

Whereas PpD=12 is the total number of studying period per day and D=6 is the total number of studying day per week.

*AvC*: percentage of classes' availability.

$$AvC = \frac{\sum_{i=1}^{C} \text{NumberOf AvailablePeriodsOf Class} C_i}{C \times PpD \times D}$$

*AvR*: percentage of rooms' availability

$$AvR = \frac{\sum_{i=1}^{R} \text{NumberOf AvailablePeriodsOf Room}R_i}{R \times PpD \times D}$$

TABLE I. VALUES OF THE THIRTEEN PROPOSED MEASUREMENTS ON THE FOURTEEN CONSIDERED DATA INSTANCES

| | A | T | C | R | P | Cr | CoL | CoC | CoCr | Co | AvT(%) | AvC(%) | AvR(%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Data 1 | 88 | 44 | 13 | 46 | 34 | 11 | 22 | 130 | 130 | 282 | 67 | 100 | 100 |
| Data 2 | 77 | 46 | 11 | 47 | 13 | 2 | 32 | 229 | 229 | 490 | 68 | 100 | 100 |
| Data 3 | 80 | 44 | 12 | 52 | 16 | 6 | 45 | 325 | 325 | 695 | 68 | 100 | 99 |
| Data 4 | 86 | 51 | 20 | 52 | 13 | 4 | 45 | 344 | 344 | 733 | 53 | 100 | 100 |
| Data 5 | 83 | 50 | 19 | 52 | 27 | 5 | 25 | 192 | 192 | 409 | 59 | 100 | 99 |
| Data 6 | 99 | 54 | 25 | 52 | 11 | 8 | 59 | 417 | 417 | 893 | 27 | 100 | 99 |
| Data 7 | 88 | 44 | 10 | 46 | 18 | 10 | 39 | 211 | 211 | 461 | 59 | 100 | 100 |
| Data 8 | 95 | 44 | 14 | 45 | 46 | 11 | 23 | 131 | 131 | 285 | 70 | 100 | 100 |
| Data 9 | 84 | 45 | 11 | 47 | 22 | 10 | 29 | 232 | 232 | 493 | 67 | 100 | 100 |
| Data 10 | 89 | 48 | 13 | 51 | 38 | 11 | 40 | 279 | 279 | 598 | 70 | 100 | 100 |
| Data 11 | 51 | 31 | 19 | 45 | 0 | 2 | 25 | 45 | 45 | 115 | 33 | 100 | 100 |
| Data 12 | 82 | 49 | 13 | 46 | 18 | 8 | 31 | 187 | 187 | 405 | 70 | 100 | 100 |
| Data 13 | 80 | 46 | 10 | 46 | 10 | 4 | 43 | 287 | 287 | 617 | 58 | 100 | 100 |
| Data 14 | 95 | 42 | 14 | 46 | 35 | 4 | 36 | 149 | 149 | 334 | 69 | 100 | 99 |

From values of A, T, C and R measurements, we can see that the size of those data instances seems to be not small, this is one of reasons why we think that metaheuristics should be promising approach for solving this problem. In addition to that, values of CoL and CoC measurements state that the conflict between block elements based on hard constraints relevant to lecturers and classes is not trivial. Finally, since AvC and AvR measurements are almost equal to 100, the hardness of time availability of these instances mostly focuses on lecturers' availability.

## REFERENCES


[1] TB Cooper and JH Kingston, "The Complexity of Timetable Construction Problems" in the Practice and Theory of Automated Timetabling, ed. EK Burke and P Ross, pp. 283-295, Springer-Verlag (Lecture Notes in Computer Science), 1996. Basser Department of Computer Science, University of Sydney, Australia, 1996.

[2] TB Cooper and JH Kingston, "The Complexity of Timetable Construction Problems" in the Practice and Theory of Automated Timetabling, ed. EK Burke and P Ross, pp. 283-295, Springer-Verlag (Lecture Notes in Computer Science), 1996. Basser Department of Computer Science, University of Sydney, Australia, 1996.