On the Use of Regular Languages to Model Personnel Scheduling Problems (Appendices)

Appendix A: Compact Assignment Formulation

$$\min \quad \sum_{t\in[T]}\sum_{s\in\Sigma\setminus\{o\}}\left(z^-_{ts}u_{ts}+z^+_{ts}v_{ts}+\left(\sum_{e\in\mathcal{E}}(p^e_{ts}(1-x^e_{ts})+q^e_{ts}x^e_{ts})\right)\right) \tag{1}$$

s.t

$$\sum_{s\in\Sigma\setminus\{o\}} x^e_{ts}\le 1 \qquad \forall e\in\mathcal{E}, t\in[T] \tag{2}$$

$$x^e_{ts}+x^e_{(t+1)u}\le 1 \qquad \forall g\in\mathcal{G}, e\in\mathcal{E}_g, t\in[T-1], s\in\Sigma\setminus\{o\}, u\in\mathcal{F}_g(s) \tag{3}$$

$$\sum_{j=t}^{t+\gamma^{\max}_g}\sum_{s\in\Sigma\setminus\{o\}} x^e_{js}\le\gamma^{\max}_g \qquad \forall g\in\mathcal{G}, e\in\mathcal{E}_g, t\in[T-\gamma^{max}_g] \tag{4}$$

$$\sum_{s\in\Sigma\setminus\{o\}}\left(x^e_{ts}+x^e_{(t+k)s}-\sum_{j=t+1}^{t+k-1}x^e_{ts}\right)\ge 2-k \qquad \forall g\in\mathcal{G}, e\in\mathcal{E}_g, k\in\{2,\dots,\gamma^{min}_g\}, t\in[T-k] \tag{5}$$

$$\sum_{s\in\Sigma\setminus\{o\}}\left(x^e_{ts}+x^e_{(t+k)s}-\sum_{j=t+1}^{t+k-1}x^e_{jt}\right)\le 1 \qquad \forall g\in\mathcal{G}, e\in\mathcal{E}_g, k\in\{2,\dots,\delta^{\min}_g\}, t\in[T-k] \tag{6}$$

$$\sum_{s\in\Sigma\setminus\{o\}} x^e_{ts}\le y^e_w \qquad \forall g\in\mathcal{G}, e\in\mathcal{E}_g, w\in[W], \forall t\in\{7w-1,7w\} \tag{7}$$

$$y^e_w+y^e_{w+1}\le\nu^{\max}_g \qquad \forall g\in\mathcal{G}, e\in\mathcal{E}_g, w\in[W-1] \tag{8}$$

$$\sum_{j=7w-6}^{7w}\sum_{s\in\Sigma\setminus\{o\}}\tau_s x^e_{js}\le\tau^{\max}_g \qquad \forall g\in\mathcal{G}, e\in\mathcal{E}_g, w\in[W] \tag{9}$$

$$\sum_{j=t}^{t+6} x^e_{js}\le\rho^{\max}_{gs} \qquad \forall g\in\mathcal{G}, e\in\mathcal{E}_g, t\in[T-6], s\in\Sigma\setminus\{o\} \tag{10}$$

$$x^e_{ts}=0 \qquad \forall e\in\mathcal{E}, t\in\mathcal{O}^e, s\in\Sigma\setminus\{o\} \tag{11}$$

$$\sum_{e\in\mathcal{E}} x^e_{ts}+z^-_{ts}-z^+_{ts}=d_{ts} \qquad \forall t\in[T], s\in\Sigma\setminus\{o\} \tag{12}$$

$$x^e_{ts}\in\{0,1\} \qquad \forall e\in\mathcal{E}, t\in[T], s\in\Sigma\setminus\{o\} \tag{13}$$

$$y^e_t\in\{0,1\} \qquad \forall e\in\mathcal{E}, t\in[W] \tag{14}$$

$$z^-_{ts}, z^+_{ts}\ge 0 \qquad \forall t\in[T], s\in\Sigma\setminus\{o\}, \tag{15}$$

The binary variables x^e_{ts} indicate if an employee e is assigned to working shift s on day t. Variables z^-_{ts} and z^+_{ts} are respectively the under-staffing and over-staffing for working shift s on day t. The objective function (1) minimizes the costs of under- and over-staffing as well as the penalties associated with the non-satisfaction of preferred or undesired working assignments. Constraints (2) ensure that an employee cannot be assigned more than one working shift per day. If no working shift is assigned to an employee on a given day, then a day-off is assigned on this day. Constraints (3) and (4) guarantee that assignments of an employee with contract g respect the rules $\mathtt{r}^1_g$ and $\mathtt{r}^2_g$, respectively. Constraints (5) ensure that a schedule assigned to an employee does not have working stretches below the minimal number specified by the contract, except at the beginning or at the end of a schedule. Rule $\mathtt{r}^4_g$ for an employee with contract g is enforced by constraints (6) in a similar way. Constraints (7) and (8) ensure the weekend rules by using binary variables y^e_w that

indicate if employee e work at least one day of the wth weekend. Constraints (9) bound the weekly working time, and constraints (10) guarantee that a maximum number of working shifts of each type are assigned to each employee over seven rolling days. Constraints (11) set the supplementary days-off, while constraints (12) represent the demand satisfaction constraints for each day and each shift. Constraints (14) and (15) set the domain of variables.

Appendix B: Instances

Table 1 presents, for each instance, the number of weeks in the planning horizon (W), the number of nurses ($|\mathcal{E}|$), the number of shifts ($|\Sigma|$), and the number of employee contracts ($|\mathcal{G}|$).

Table 1 Sizes of the different instances.

	Instance1	*Instance2*	*Instance3*	*Instance4*	*Instance5*	*Instance6*	*Instance7*	*Instance8*	*Instance9*	*Instance10*	*Instance11*	*Instance12*	*Instance13*	*Instance14*	*Instance15*	*Instance16*	*Instance17*	*Instance18*	*Instance19*	*Instance20*	*Instance21*
W	2	2	2	4	4	4	4	4	4	4	4	4	4	6	6	8	8	12	12	26	26
$\|\mathcal{E}\|$	8	14	20	10	16	18	20	30	36	40	50	60	120	32	45	20	32	22	40	50	100
$\|\Sigma\|$	2	3	4	3	3	4	4	5	5	6	7	11	19	5	7	4	5	4	6	7	9
$\|\mathcal{G}\|$	1	5	11	3	4	10	9	18	23	25	33	58	120	13	34	6	9	8	25	40	89

Appendix C: Characteristics of the Different Models for the Modified Instances

Table 2 summarizes the main characteristics of the three formulations: `CA-MIP`, `P-RB-MIP'`, and `P-RB-MIP`, across the 21 instances of the personalized problem. For each instance and each formulation, the table reports the number of variables (Vars), the model construction time (T_{build} in seconds, including the time required to compute the various T-DFA in the RB-MIP formulations), and the value of the linear relaxation (LP).

Table 3 presents the same information for the anonymous version of the problem, comparing the formulations `CA-MIP`, `P-RB-MIP`, and `A-RB-MIP`. When the number of variables exceeds 500,000, we stop the construction of the model; we then report "$>$500,000" in the corresponding columns and use "-" for the construction time and linear relaxation value. An asterisk ("*") following the linear relaxation value indicates that it corresponds to the optimal value of the problem.

When modeling a personalized version of the problem, the three formulations (`CA-MIP`, `P-RB-MIP'` and `P-RB-MIP`) use the same x^e_{ts}, z^-_{ts} and z^+_{ts} variables. In addition, `CA-MIP` and `P-RB-MIP'` use the variables y^e_w for modeling $\mathtt{r}^5_g$ while both regular based formulations also use flow variables (one per transition for each unfolded DFA associated with an employee). The number of flow variables necessary may be large, hence we generally observe that RB-MIP formulations require more variables than a compact assignment formulation (approximately 10 times more in our case, e.g. increasing from about 150 to 900 for Instance1 and from $40,000$ to $400,000$ for Instance20). The comparison between `P-RB-MIP'` and `P-RB-MIP` also highlights this general trend, i.e. removing a subset of rules often reduces the size of the unfolded DFA and the number of variables used in the resulting formulations. For example, `P-RB-MIP` has more than twice as many variables as `P-RB-MIP'` for some instances (e.g., Instance10, Instance14 or Instance15). However for some specific instances, including the rules $\mathtt{r}^5_g$-$\mathtt{r}^7_g$ does not make the automata of `P-RB-MIP` much bigger (even sometimes slightly smaller, e.g., Instance 6 and 20) than the ones of `P-RB-MIP'`,

Table 2 Characteristics of different formulations when considering personal requests.

	CA−MIP			P−RB−MIP			A−RB−MIP		
Instance	Vars	T_{build}	LP	Vars	T_{build}	LP	Vars	T_{build}	LP
Instance1	**156**	**0.03**	415	844	0.06	468	951	0.06	**1,110***
Instance2	**420**	**0.05**	733	2,522	0.13	743	2,791	0.23	**1,336***
Instance3	**768**	**0.09**	1,003	5,343	0.25	1,003	6,903	0.6	**1,877**
Instance4	**656**	**0.07**	1,674	4,679	0.32	1,884	5,151	0.33	**3,077**
Instance5	**904**	**0.29**	695	6,932	0.34	1,404	8,235	0.83	**3,453***
Instance6	**1,360**	**0.15**	1,878	11,737	0.66	2,159	10,907	0.77	**5,147***
Instance7	**1,592**	**0.17**	976	13,351	0.58	1,223	16,237	1.61	**4,033***
Instance8	**2,668**	**0.29**	1,663	23,955	1	2,224	34,833	4.23	**7,484***
Instance9	**3,504**	**0.34**	1,532	32,238	1.3	1,546	31,950	3.57	**6,320***
Instance10	**4,192**	**0.44**	4,686	36,910	1.48	4,711	105,870	12.57	**9,642**
Instance11	**6,248**	**0.76**	3,533	52,639	2.48	3,536	56,358	3.31	**7,002**
Instance12	**12,056**	**1.73**	4,161	105,623	4.62	4,648	130,552	8.25	**15,283**
Instance13	**44,496**	**8.91**	**3,606**	>500,000	-	-	>500,000	-	-
Instance14	**4,476**	**0.66**	1,806	36,154	1.84	2,216	80,863	8.69	**5,462**
Instance15	**8,208**	**1.06**	11,184	78,993	3.91	11,394	219,119	30.44	**14,290**
Instance16	**3,016**	**0.46**	4,109	25,920	1.58	4,590	28,177	1.88	**9,965***
Instance17	**6,528**	**0.99**	6,495	65,558	3.47	6,893	111,164	7.38	**14,364***
Instance18	**5,304**	**0.77**	6,062	48,059	2.87	6,920	68,912	5.43	**15,451**
Instance19	**12,072**	**1.67**	4,910	112,191	5.61	5,488	261,933	20.43	**20,303**
Instance20	**42,614**	**5.09**	5,761	409,047	19.34	6,203	400,572	25.78	**31,177**
Instance21	**106,158**	**15.29**	**10,472**	>500,000	-	-	>500,000	-	-

Table 3 Characteristics of different formulations when considering personal requests.

	CA−MIP			P−RB−MIP'			P−RB−MIP		
Instance	Vars	T_{build}	LP	Vars	T_{build}	LP	Vars	T_{build}	LP
Instance1	**156**	**0.03**	400	1,228	0.08	**1,000***	178	0.09	**1,000***
Instance2	**420**	**0.05**	700	3,330	0.26	**1,200***	925	0.06	**1,200***
Instance3	**768**	**0.09**	1,000	8,763	0.69	**1,860**	4,224	0.25	**1,860**
Instance4	**656**	**0.07**	1,500	6,340	0.47	**3,050**	1,482	0.08	**3,050**
Instance5	**904**	**0.11**	300	10,707	0.93	**3,200***	2,317	0.15	**3,200***
Instance6	**1,360**	**0.15**	1,292	13,106	0.92	**4,800***	6,891	0.46	**4,800***
Instance7	**1,592**	**0.17**	625	20,637	1.57	**3,775**	8,164	0.6	**3,775**
Instance8	**2,668**	**0.28**	1,350	44,805	4.32	**7,150**	24,794	2.44	**7,150**
Instance9	**3,504**	**0.34**	1,405	39,354	3.03	**5,800***	20,246	1.57	**5,800***
Instance10	**4,192**	**0.42**	4,600	143,484	11.36	**9,550**	76,050	6.38	**9,550**
Instance11	**6,248**	**1.01**	3,400	69,906	4.12	**6,575**	44,195	2.8	**6,575**
Instance12	**12,056**	**1.73**	3,513	168,638	10.94	**14,750**	159,967	10.66	**14,750**
Instance13	**44,496**	**9.69**	**3,242**	>500,000	-	**-**	>500,000	-	-
Instance14	**4,476**	**0.45**	1,320	108,480	12.7	**4,503***	35,246	4.32	**4,503***
Instance15	**8,208**	**1.62**	10,505	303,449	48.08	**13,200**	187,812	31.23	**13,200**
Instance16	**3,016**	0.53	3,300	35,846	2.54	**8,702***	8,633	**0.47**	**8,702***
Instance17	**6,528**	**1.17**	6,000	141,736	10.09	**13,550**	33,172	2.25	**13,550**
Instance18	**5,304**	**0.6**	4,200	86,267	7.24	**13,403***	25,403	1.96	**13,403***
Instance19	**12,072**	**2.13**	2,565	345,112	28.77	**16,400**	209,093	19.45	**16,400**
Instance20	**42,614**	**4.8**	2,900	473,459	31.74	**18,350**	353,882	23.63	**18,350**
Instance21	**106,158**	**19.35**	**3,300**	>500,000	-	**-**	>500,000	-	-

resulting in P−RB−MIP having both less variables and a better linear relaxation than P−RB−MIP′. Note that the number of constraints is of the same order of magnitude across the three formulations.

The same observation can be made for the anonymous version of the problem formulated with an implicit RB-MIP model such as A−RB−MIP. In that case, there is one unfolded DFA for each type of contract. Therefore, the more employees sharing an identical contract, the less variables are necessary to express A−RB−MIP in comparison to the other formulations. For example, Instance 17 considers 32 employees sharing 9 different contracts which divides the number of variables by more than four in comparison with A−RB−MIP and R−RB−MIP. On the contrary, instances with (almost) one contract per employee (e.g., 34 contracts for 45 employees in Instance 15) naturally lead to formulations with a similar number of variables in both cases. Note also that considering personal requests leads to smaller unfolded automata. This is simply because mandatory days-off allow restricting some layers of the T-DFA to transitions representing days-off.

The model construction time is influenced not only by the number of variables but also by the time required to build the unfolded automata. As a result, T_{build} is typically higher for the RB-MIP formulation and tends to increase as more rules are included in the automata. Nevertheless, this construction time remains negligible compared to the solution times observed in Section 6.4.

In the personalized problem, the linear relaxations obtained with P−RB−MIP are significantly stronger than those of the compact assignment formulation CA−MIP. On the other hand, removing some rules strongly deteriorates the relaxation. While P−RB−MIP′ often still provides better relaxations than CA−MIP, the improvement is sometimes marginal (e.g., Instance11) or even nonexistent (e.g., Instance3). Finally, in the anonymous problem, the same order of magnitude can be observed between the linear relaxation values of CA−MIP and P−RB−MIP. Using an implicit formulation reduces the number of variables (see above) but does not improve LP, as can be observed in Table 3 and could also be established theoretically.

Appendix D: Characteristics of the Different Models for the Original Instances

As in Table 2 for the modified instances, Table 4 reports the same characteristics (Vars, T_{build}, and LP) for three different formulations applied to the original instances of [**?**]. CA−MIP corresponds to the model presented in [**?**], P−RB−MIP denotes the RB-MIP formulation where all rules are embedded in the unfolded automata, and P−RB−MIP′ refers to the RB-MIP formulation where constraints related to the total working time and the maximum number of shifts are not included in the automata but instead expressed as linear inequalities, as in CA−MIP.

Table 4 shows that P−RB−MIP often provides very strong linear relaxations, sometimes even optimal (Instance2 to Instance5). However, such a formulation becomes impractical for larger instances as the rules related to maximum working time and maximum number of certain shifts lead to extremely large unfolded automata. Thus, the limit of 500,000 variables is reached for every instance larger than the Instance9. In contrast, P−RB−MIP′ remains tractable on most of these larger cases, while still achieving strong relaxations, often very close to the one of P−RB−MIP Some exceptions exist (e.g., Instance9), but overall the hybrid formulation preserves most of the benefits of P−RB−MIP without incurring the prohibitive model sizes observed in larger instances. Note, however, that the model size remains logically much larger than for CA−MIP, and that many assignments (e.g., Instance13) or very long planning horizons (e.g., Instance20 and Instance21) have a substantial impact on the size.

Table 4 shows that P−RB−MIP often provides very strong linear relaxations, sometimes even optimal (see Instance2 to Instance5), and significantly better than those obtained with CA−MIP. However, this formulation becomes impractical for larger instances, as the rules related to maximum working time and maximum numbers of certain shifts generate

Table 4 Characteristics of different formulations for the original instances

	CA-MIP			P-RB-MIP'			P-RB-MIP		
Instance	Vars	T_{build}	LP	Vars	T_{build}	LP	Vars	T_{build}	LP
Instance1	**156**	**0.03**	408	914	0.05	512	1,325	0.09	**558**
Instance2	**420**	**0.02**	720	2,780	0.07	732	3,805	0.26	**828***
Instance3	**768**	**0.04**	1,000	5,852	0.16	1,000	13,554	0.94	**1,001***
Instance4	**656**	**0.03**	1,262	6,938	0.19	1,716	15,084	0.93	**1,716***
Instance5	**904**	**0.05**	711	9,909	0.27	1,140	41,388	3.64	**1,141***
Instance6	**1,360**	**0.06**	1,916	16,575	0.42	1,948	32,266	1.97	**1,949**
Instance7	**1,592**	**0.12**	927	19,802	0.47	1,053	129,084	9.29	**1,055**
Instance8	**2,668**	**0.11**	1,232	35,526	0.85	1,276	281,525	49.42	**1,297**
Instance9	**3,504**	**0.21**	39	47,833	1.07	39	391,201	23.69	**406**
Instance10	**4,192**	**0.25**	4,616	57,708	1.23	**4,629**	>500,000	-	**-**
Instance11	**6,248**	**0.36**	3,418	80,089	1.75	**3,427**	>500,000	-	**-**
Instance12	**12,056**	**0.7**	3,628	161,062	3.25	**4,031**	>500,000	-	**-**
Instance13	**44,496**	**3.61**	**519**	>500,000	-	-	>500,000	-	-
Instance14	**4,476**	**0.18**	1,255	70,026	1.68	**1,269**	>500,000	-	**-**
Instance15	**8,208**	**0.49**	**3,741**	150,644	3.15	3,724	>500,000	-	**-**
Instance16	**3,016**	**0.13**	3,153	59,305	1.37	**3,191**	>500,000	-	**-**
Instance17	**6,528**	**0.35**	5,707	156,633	3.4	**5,737**	>500,000	-	**-**
Instance18	**5,304**	**0.22**	4,176	152,084	3.54	**4,382**	>500,000	-	**-**
Instance19	**12,072**	**0.59**	2,690	346,097	7.52	**3,141**	>500,000	-	**-**
Instance20	**42,614**	**2.02**	4,092	>500,000	-	-	>500,000	-	**-**
Instance21	**106,158**	**6.49**	**11,898**	>500,000	-	-	>500,000	-	-

extremely large unfolded automata. As a result, the limit of 500,000 variables is reached for every instance beyond Instance9. In contrast, P-RB-MIP' remains tractable on most of these larger cases, while still achieving strong linear relaxations, often very close to those of P-RB-MIP. Some exceptions exist (e.g., Instance 9), but overall the hybrid formulation preserves most of the benefits of P-RB-MIP without incurring the prohibitive model sizes observed in larger instances. It should be noted, however, that the model size logically remains significantly larger than for CA-MIP, and that certain features, such as many possible assignments (e.g., Instance 13) or very long horizons (e.g., Instances 20 and 21), greatly increase the size. As for the modified instances, the construction time T_{build} is influenced both by the number of variables and by the time required to compute the unfolded automata. For P-RB-MIP', it nevertheless remains negligible compared to the overall solving time.