Role-Usage Role Mining Heuristics for Permission-Role-Usage Cardinality Constraints

Additional comparisons

Carlo Blundo
DISA-MIS
Università di Salerno, Italy

Stelvio Cimato
Dipartimento di Informatica
Università di Milano, Italy

Luisa Siniscalchi
Department of Computer Science
Aarhus University, Denmark

May 28, 2020

Our first heuristic PRUCC_1 starts with a pre-processing phase where each user is assigned with a number of roles not larger than the threshold mpr then, for each user u having more than mru roles (i.e., user u violates the constraint on the number of roles that can be assigned to each user) heuristic PRUCC_1 reassigns roles to u in such a way that u will receive no more than mru roles each of size at most mpr (to assign new roles to user u, we apply to u the strawman heuristic sketched in the paper). Therefore, in our heuristic we have first enforced mpr and then mru. We could devise another heuristic where such choise is inverted, that is we could first enforce mru and then mpr. We refer to such a *new* heuristic as PRUCC_3. We have experimentally shown that, on average, such a *new* heuristic and PRUCC_1 behave quite similarly (although variants `OR` and `OF` perform better for heuristic PRUCC_1). Indeed, the *new* heuristic generates role-sets not much different from the ones generated by PRUCC_1(although, in general, `OF` and `OR` PRUCC_1's variants perform better than the corrisponding PRUCC_3's variants). As test bed we used the same as the one described in the paper. Namely, we use the datasets described in the following table.

Dataset	$\lvert\mathcal{U}\rvert$	$\lvert\mathcal{P}\rvert$	$\lvert\mathcal{UPA}\rvert$	max#P	max#U
Americas large	3485	10127	185294	733	2812
Americas small	3477	1587	105205	310	2866
Apj	2044	1164	6841	58	291
Customer	10021	277	45427	25	4184
Domino	79	231	730	209	52
Emea	35	3046	7220	554	32
Firewall 1	365	709	31951	617	251
Firewall 2	325	590	36428	590	298
Healthcare	46	46	1486	46	45

With respect to the complexity measures *Role-set size* and *Weighted Structural Complexity* the *new* heuristic generates role-sets not much different from the ones generated by PRUCC_1. The results of our experiments are summarized in the next two tables where to each combination of heuristic-variant is associated a real number denoting the *rank* of such combination (we refer to the paper so see how rank is defined). On the negative side, it results that the new heuristic is often slower than PRUCC_1.

Dataset	PRUCC_1				PRUCC_2				PRUCC_3			
	OF	OR	UF	UR	OF	OR	UF	UR	OF	OR	UF	UR
Americas large	2.55	4.85	3.8	5.85	4.35	7.05	6.65	8.65	8.4	9.0	7.75	9.1
Americas small	3.14	5.0	7.08	8.28	5.24	7.38	8.02	10.16	4.76	5.7	6.16	7.08
Apj	6.22	7.78	7.06	8.14	8.22	8.94	8.96	9.94	3.52	3.9	2.46	2.86
Customer	7.48	8.52	6.28	7.54	9.16	10.04	6.98	8.32	4.26	4.08	2.66	2.68
Domino	4.675	3.925	5.625	5.3	5.55	5.925	5.975	6.25	8.525	8.375	8.8	9.075
Emea	4.0	4.7	4.0	4.7	4.0	5.5	4.0	5.1	10.3	10.7	10.3	10.7
Firewall 1	4.65	4.975	8.95	9.65	5.6	6.325	9.7	10.625	3.625	3.8	4.55	5.55
Firewall 2	5.3	5.3	5.3	5.3	5.3	5.3	5.3	5.3	8.9	8.9	8.9	8.9
Healthcare	5.975	6.05	5.975	6.525	6.1	8.25	7.7	8.725	5.65	5.75	5.575	5.725
Average Rank	**4.888**	**5.678**	**6.008**	**6.809**	**5.947**	**7.19**	**7.032**	**8.119**	**6.438**	**6.689**	**6.351**	**6.852**

Table 1: Heuristics ranking on $|\mathcal{R}|$

Dataset	$PRUCC_1$				$PRUCC_2$				$PRUCC_3$			
	OF	OR	UF	UR	OF	OR	UF	UR	OF	OR	UF	UR
Americas large	3.9	6.3	3.0	4.85	5.0	7.9	4.5	6.95	9.0	10.1	7.5	9.0
Americas small	4.88	6.04	7.98	9.1	4.86	7.0	7.28	8.92	4.36	5.34	5.74	6.5
Apj	6.86	8.12	5.48	6.24	8.24	8.58	6.44	7.98	5.62	5.92	4.26	4.26
Customer	3.58	4.02	7.36	8.42	4.56	5.32	8.58	9.58	4.78	5.58	8.28	7.94
Domino	2.85	2.825	6.25	5.95	3.725	4.95	6.55	6.8	8.625	8.825	10.275	10.375
Emea	4.0	4.7	4.0	4.7	4.0	5.5	4.0	5.1	10.3	10.7	10.3	10.7
Firewall 1	3.425	3.925	5.3	5.925	5.575	6.125	6.025	6.9	8.125	8.425	8.875	9.375
Firewall 2	2.5	2.5	2.5	2.5	7.3	7.3	7.3	7.3	9.7	9.7	9.7	9.7
Healthcare	6.475	6.75	6.475	7.1	6.7	8.775	8.025	9.1	4.55	4.7	4.55	4.8
Average Rank	**4.274**	**5.02**	**5.372**	**6.087**	**5.551**	**6.828**	**6.522**	**7.626**	**7.229**	**7.699**	**7.72**	**8.072**

Table 2: Heuristics ranking on WSC

We would like to point out that, in some particular situations, the new heuristic is outperformed by PRUCC_1. In the first set of experiment, we set $mru = 4$ and we let mpr take values in

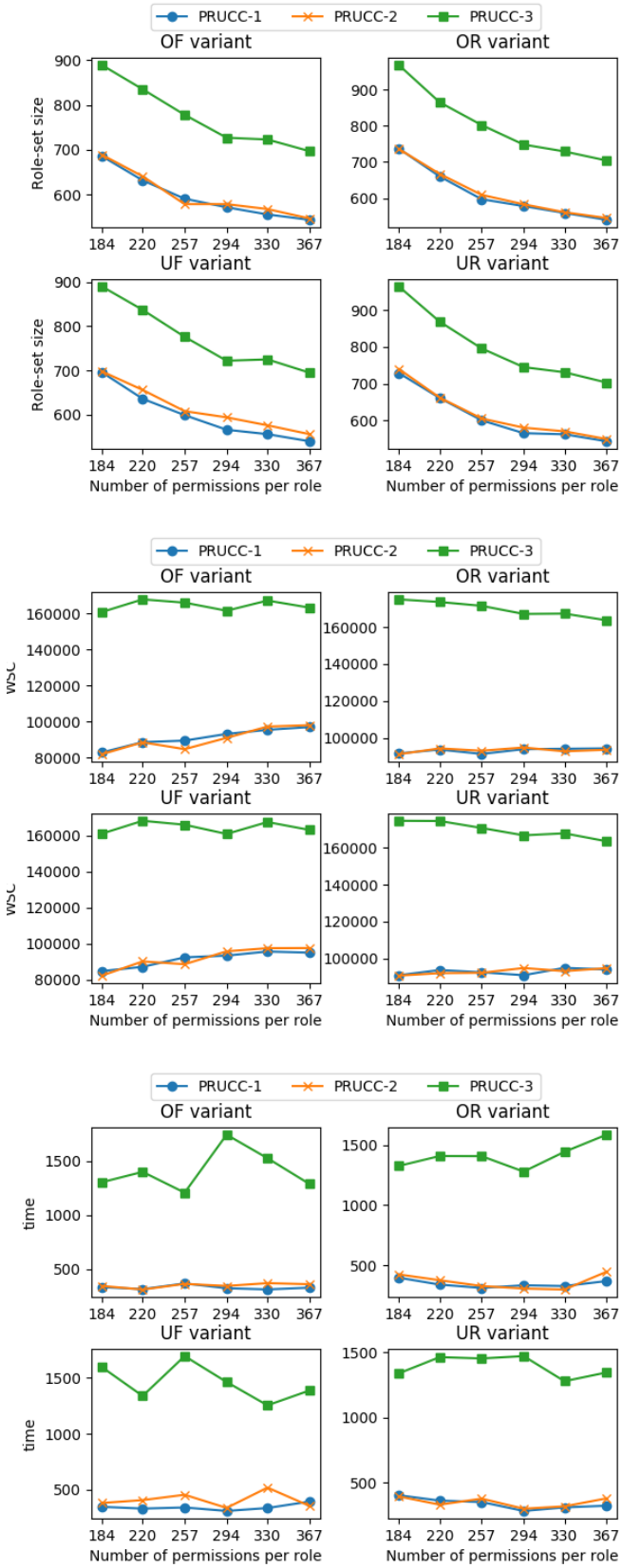
$$\{\lceil max\#P/mru \rceil \times i \mid i \in \{1, 1.2, 1.4, 1.6, 1.8, 2\}\}.$$

In the second set of experiment, we fix mpr to $\{\lceil \text{max\#P}/0.5 \rceil$ (i.e., any role can be assigned at most the 20% of the maximum number of permissions held by any user) and we let mru take values in

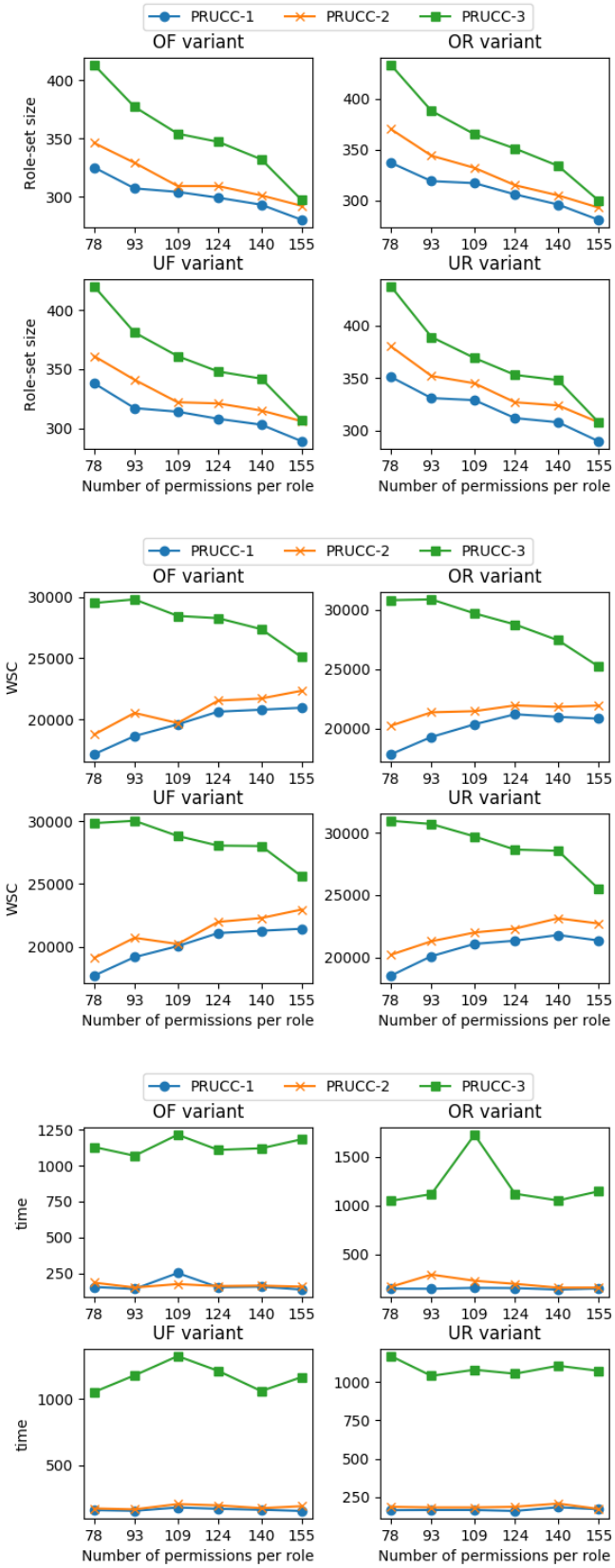
$$\{\lceil max\#P/mpr \rceil + i \mid i \in \{2, 4, 6, 8, 10\}\} = \{5, 7, 9, 11, 13, 15\}.$$

Results on the previuosly described set of experiments are reported in the following graphics.

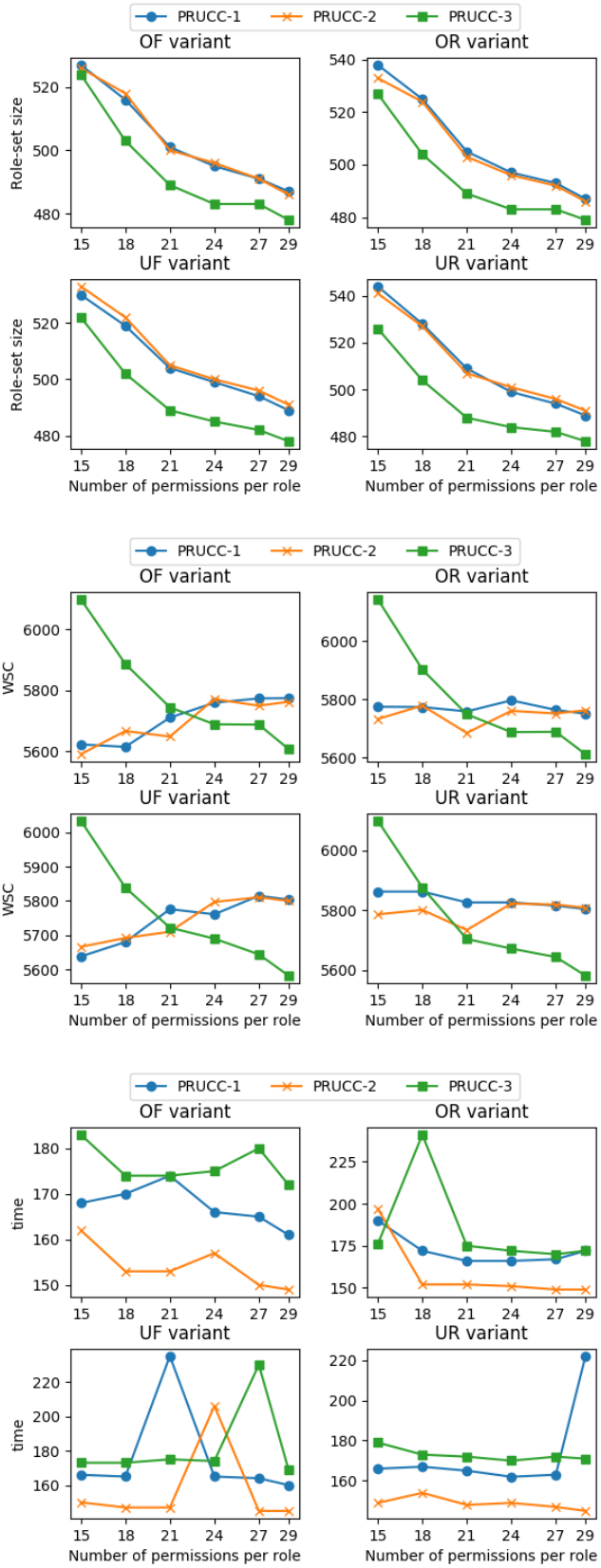
Dataset Americas Large $mru = 4$, $mpr \in [184, 220, 257, 294, 330, 367]$



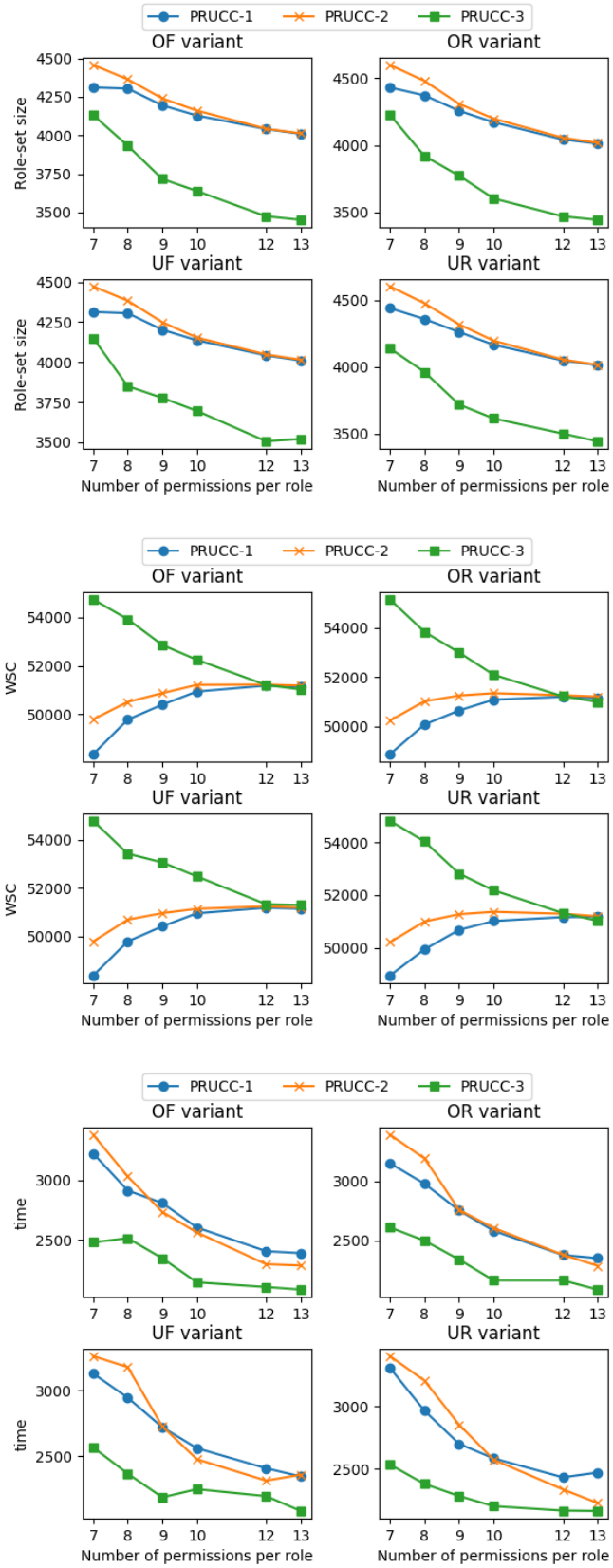
Dataset Americas Small $mru = 4$, $mpr \in [78, 93, 109, 124, 140, 155]$



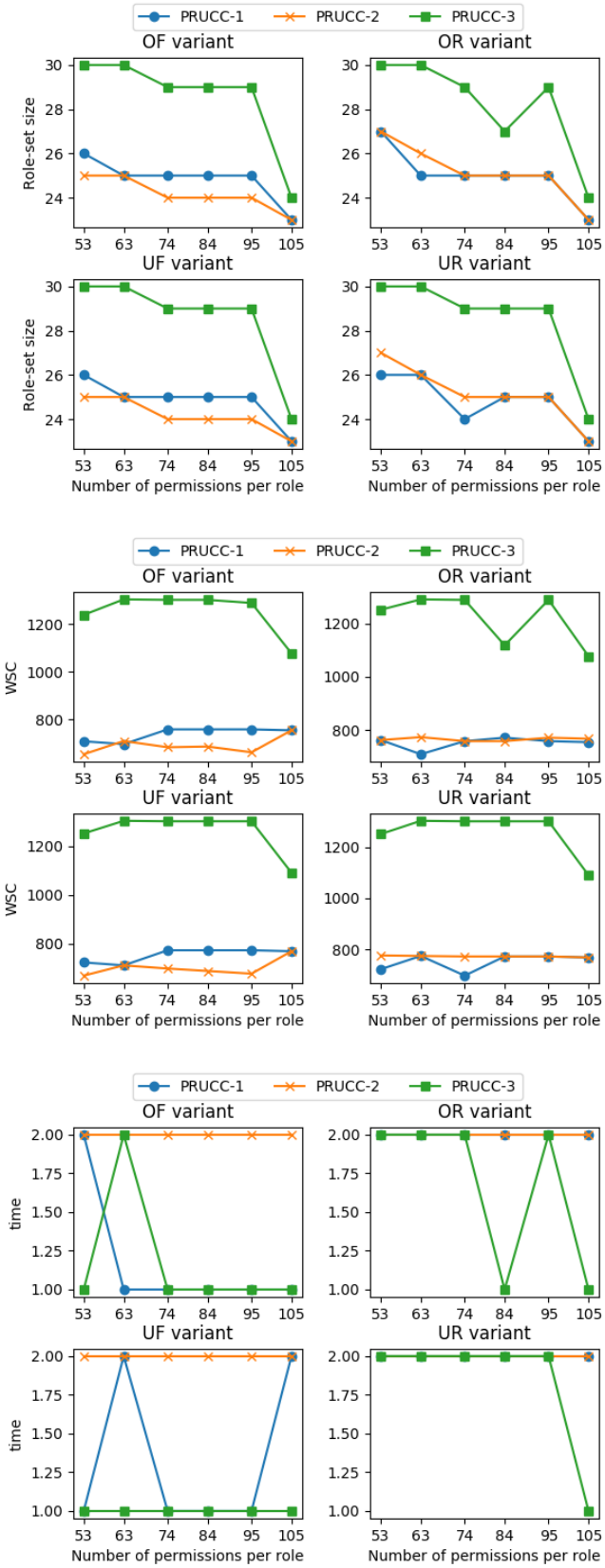
Dataset Apj $mru = 4$**,** $mpr \in [15, 18, 21, 24, 27, 29]$



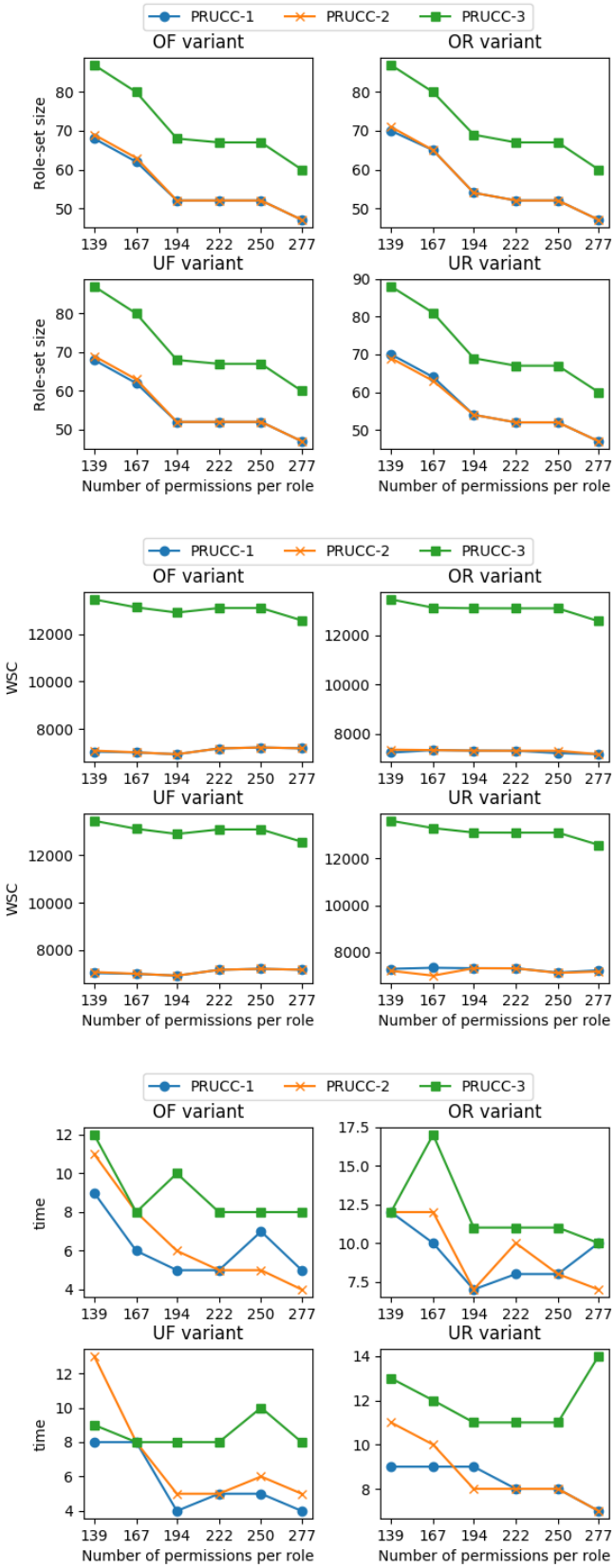
Dataset Customer $mru = 4$**,** $mpr \in [7, 8, 9, 10, 12, 13]$



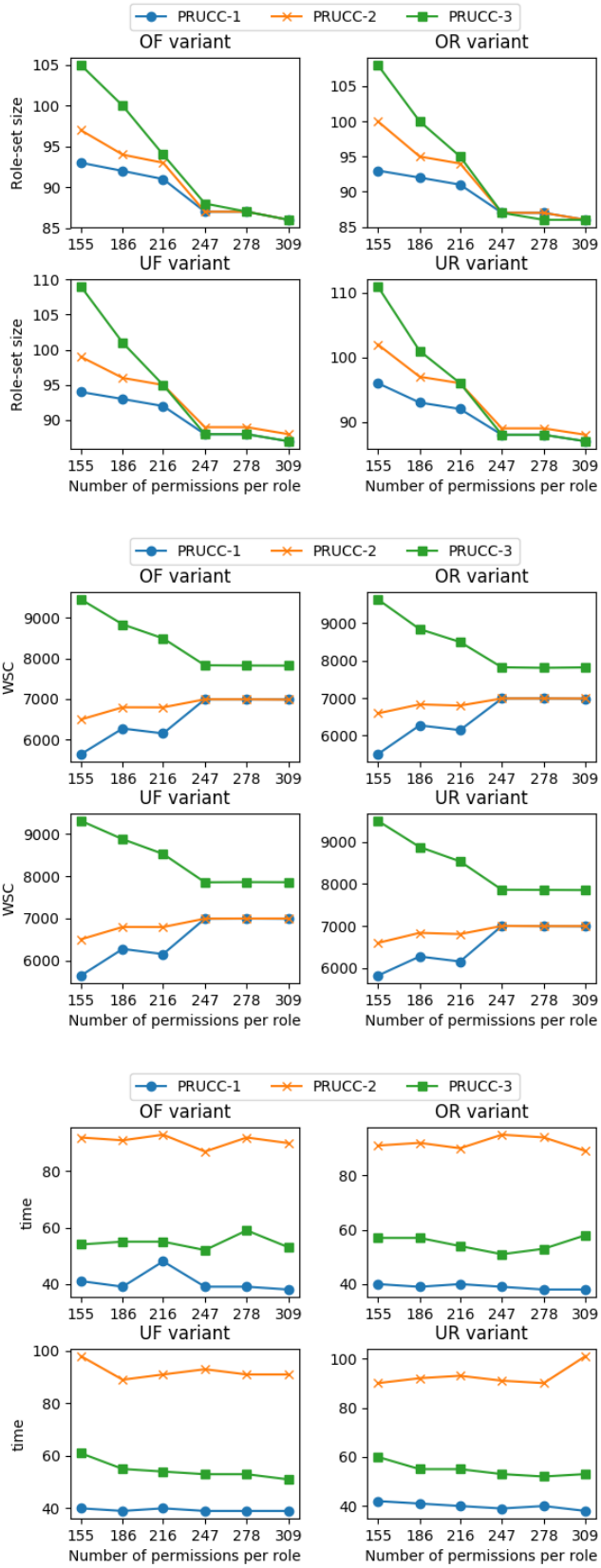
Dataset Domino $mru = 4$**,** $mpr \in [53, 63, 74, 84, 95, 105]$



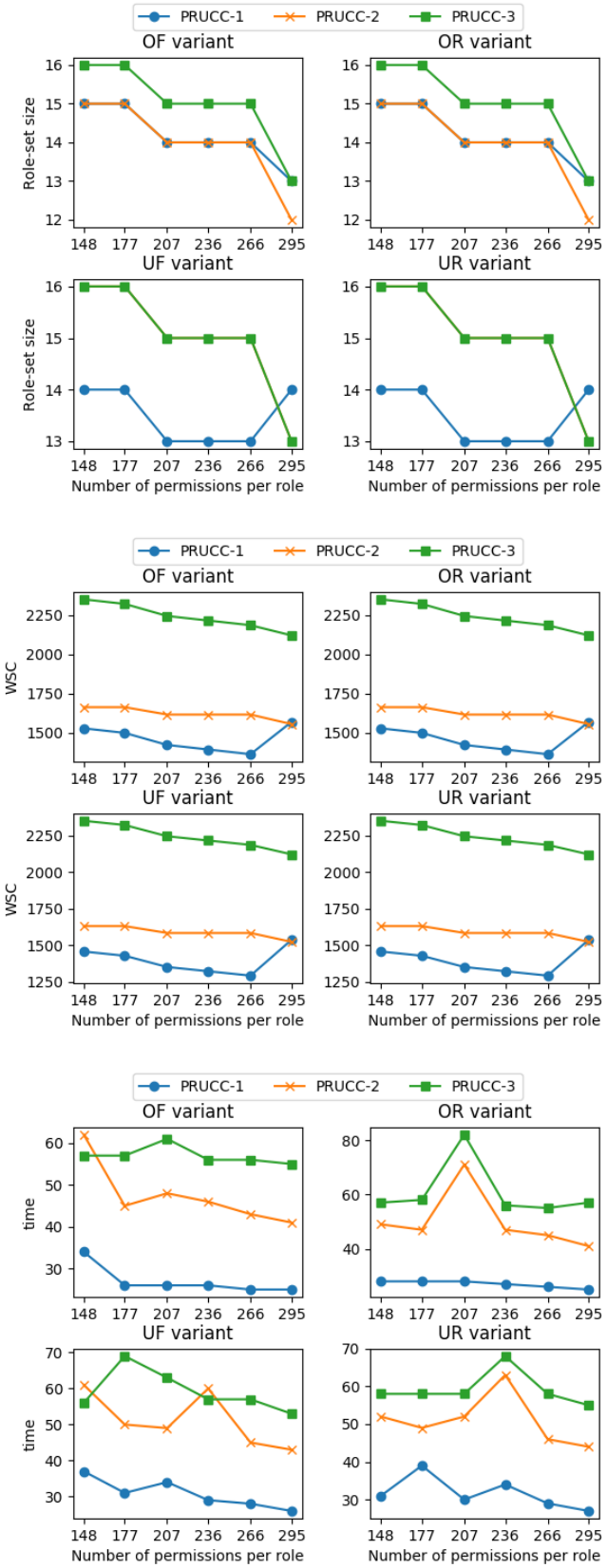
Dataset Emea $mru = 4$**,** $mpr \in [139, 167, 194, 222, 250, 277]$



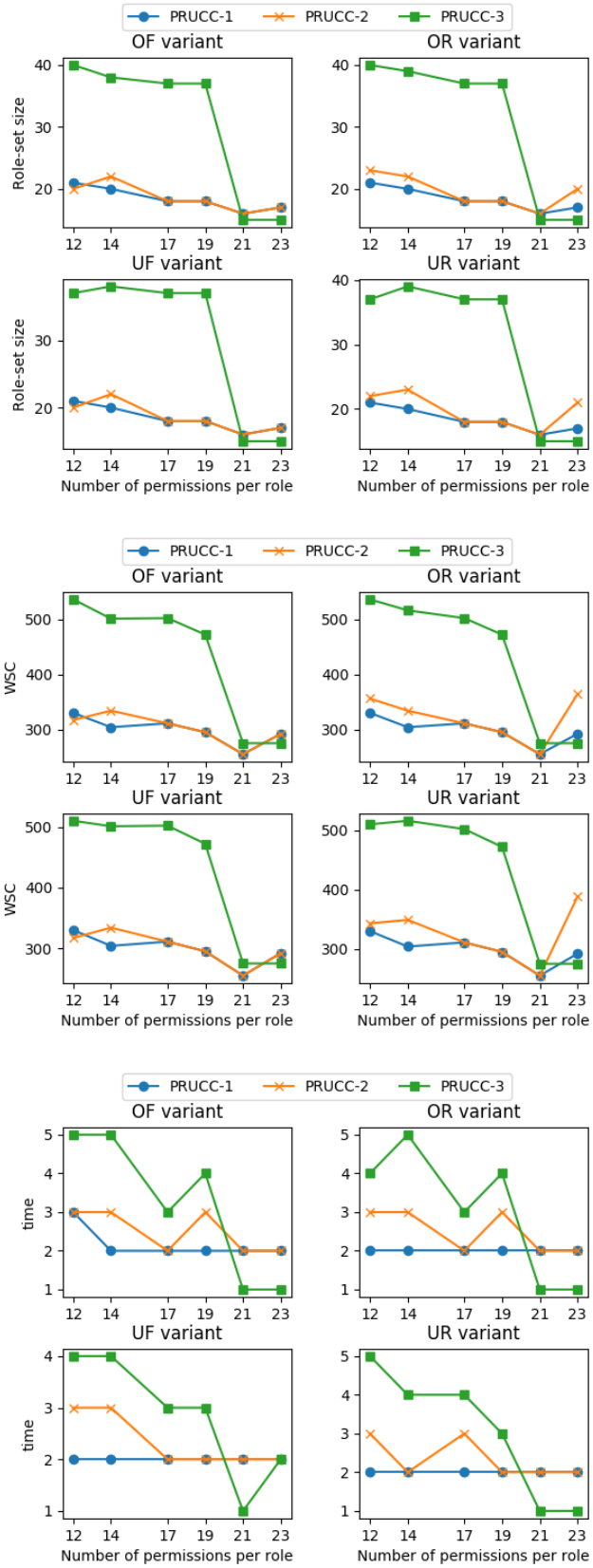
Dataset Firewall 1 $mru = 4$**,** $mpr \in [155, 186, 216, 247, 278, 309]$



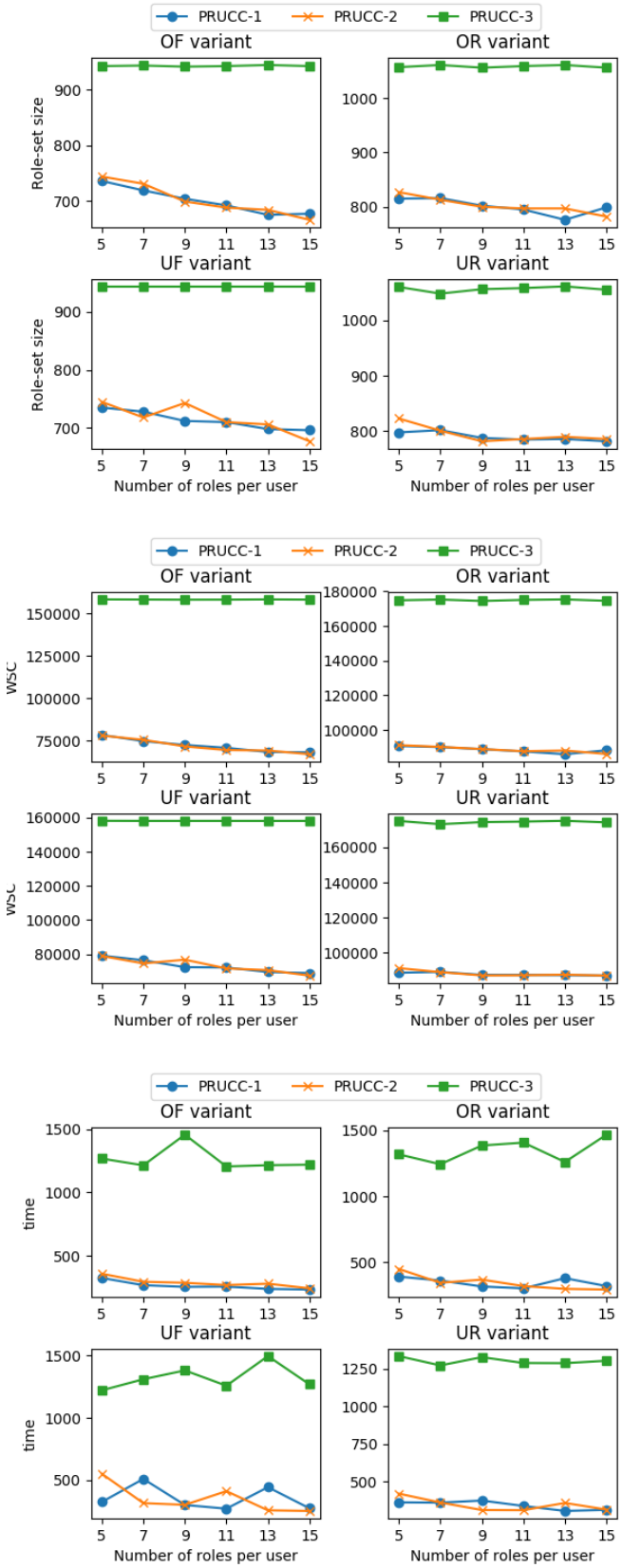
Dataset Firewall 2 $mru = 4$**,** $mpr \in [148, 177, 207, 236, 266, 295]$



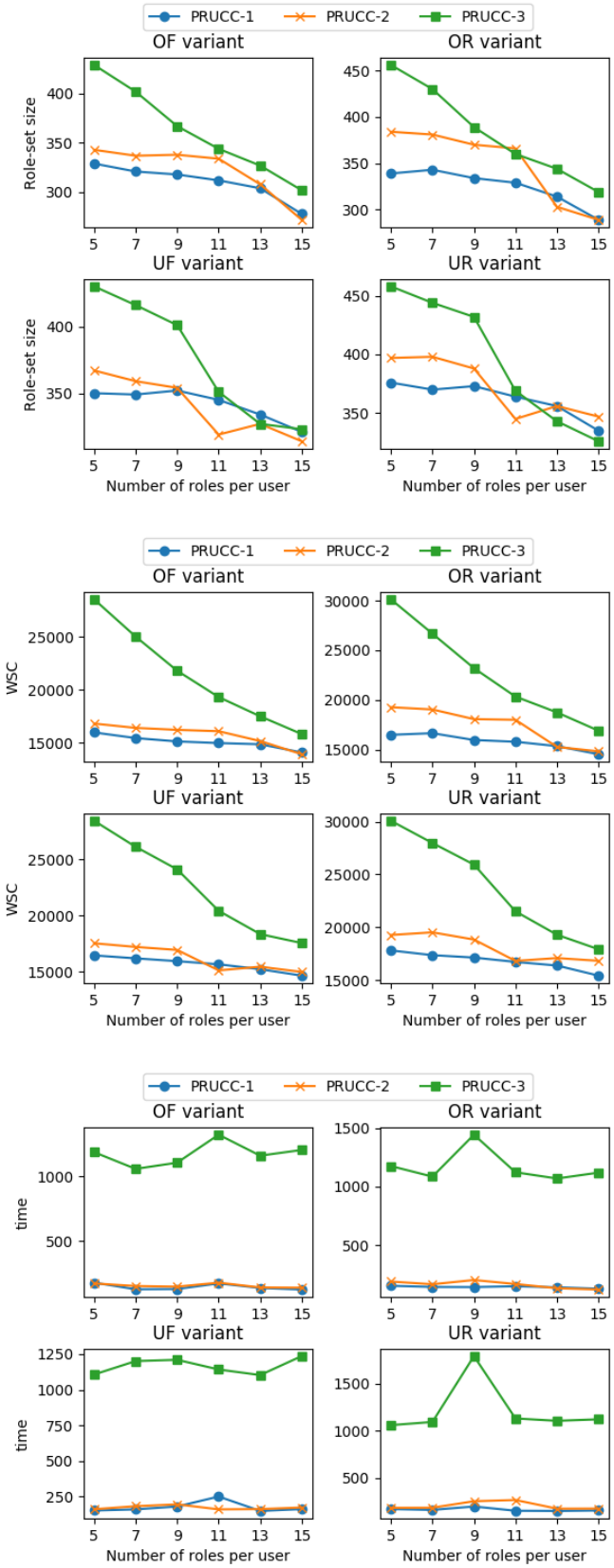
Dataset Healthcare $mru = 4$**,** $mpr \in [12, 14, 17, 19, 21, 23]$



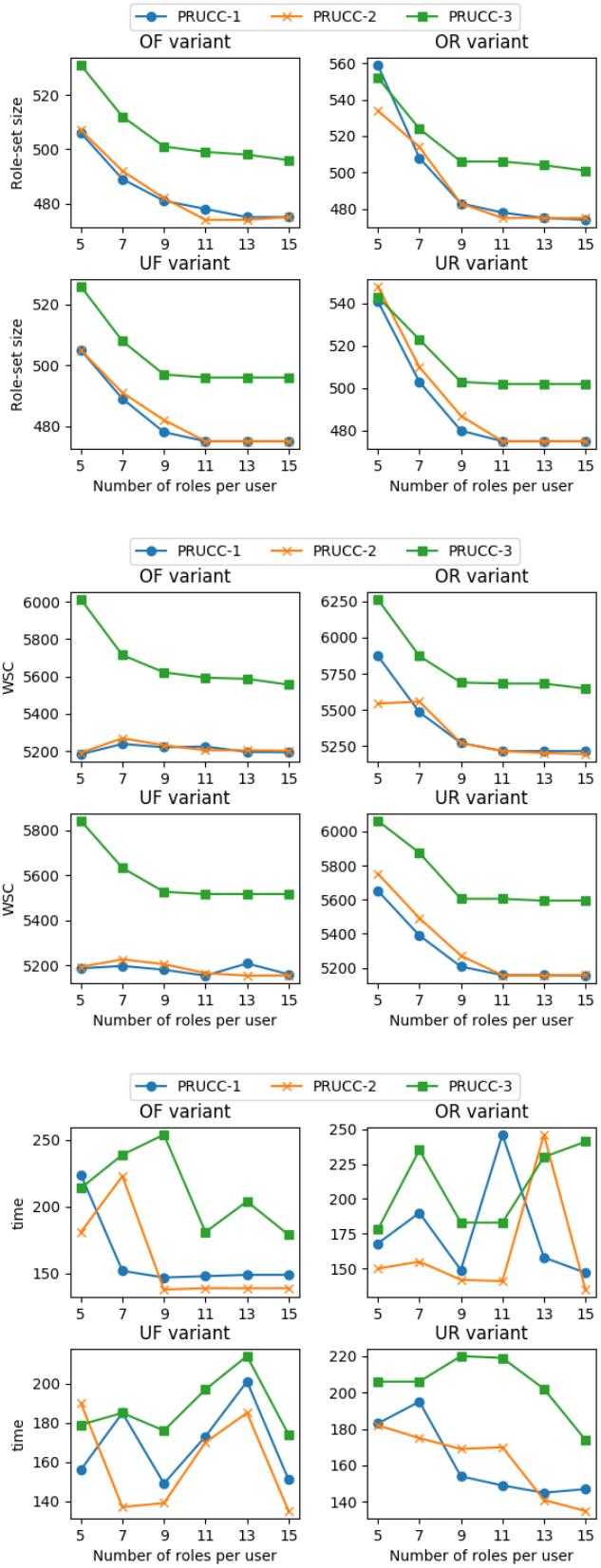
Dataset Americas Large $mpr = 147$**,** $mru \in [5, 7, 9, 11, 13, 15]$



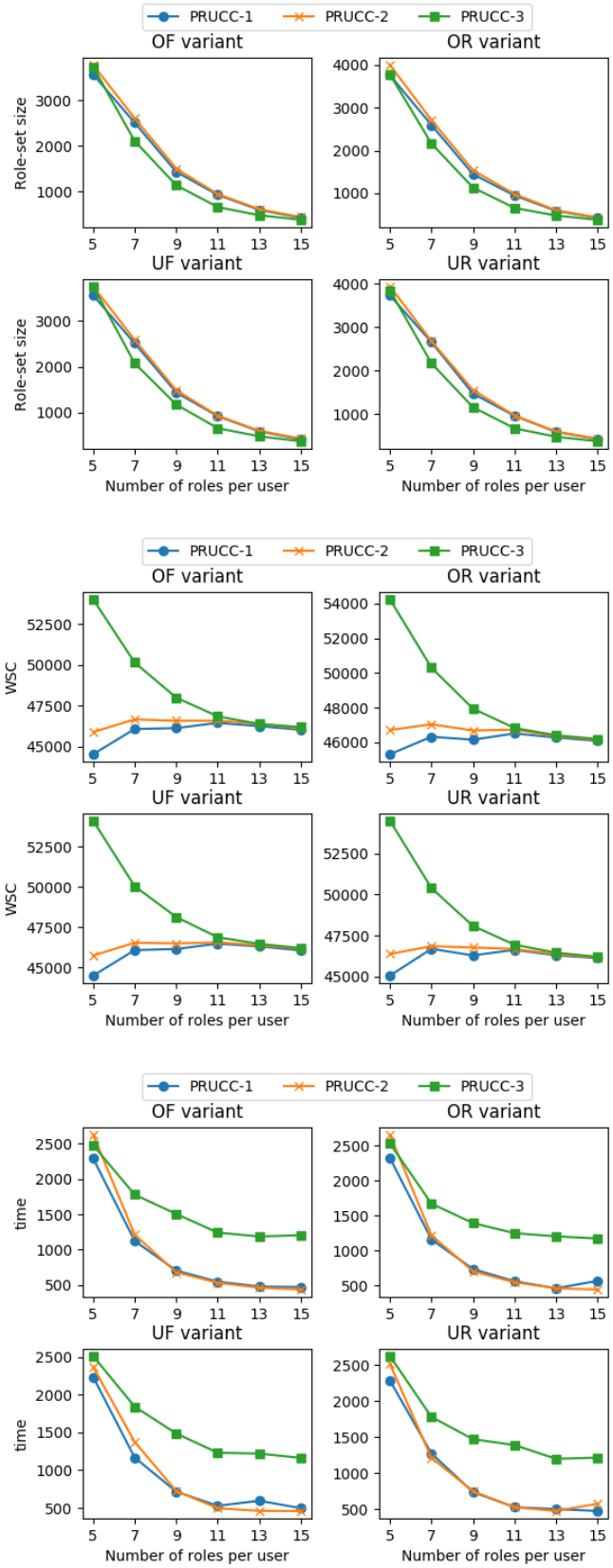
Dataset Americas Small $mpr = 62$**,** $mru \in [5, 7, 9, 11, 13, 15]$



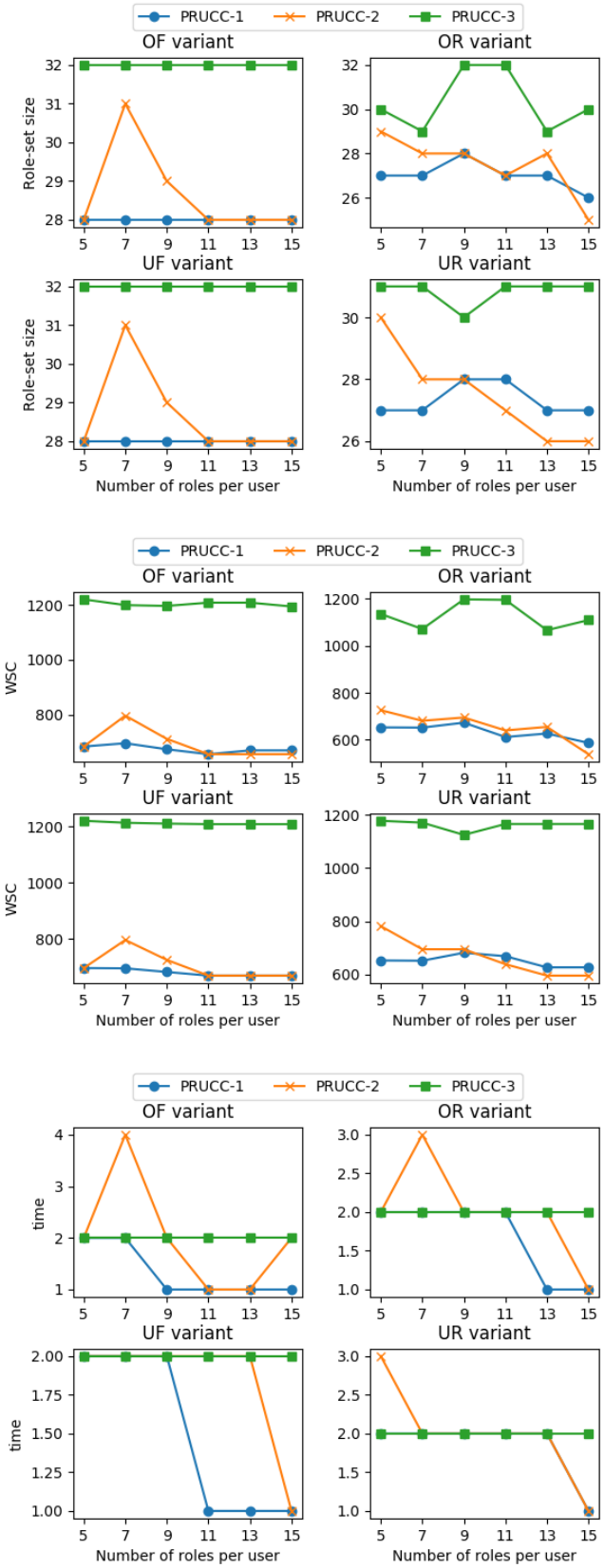
Dataset Apj $mpr = 12$**,** $mru \in [5, 7, 9, 11, 13, 15]$



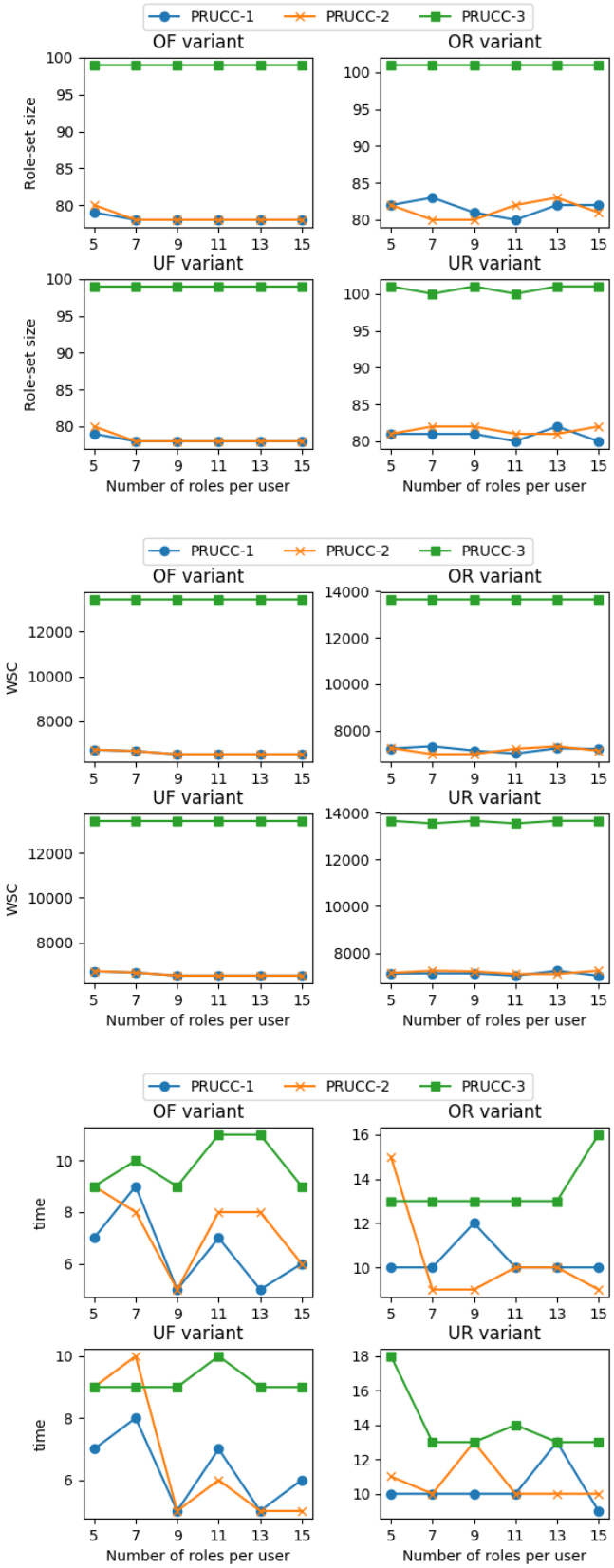
Dataset Customer $mpr = 5$**,** $mru \in [5, 7, 9, 11, 13, 15]$



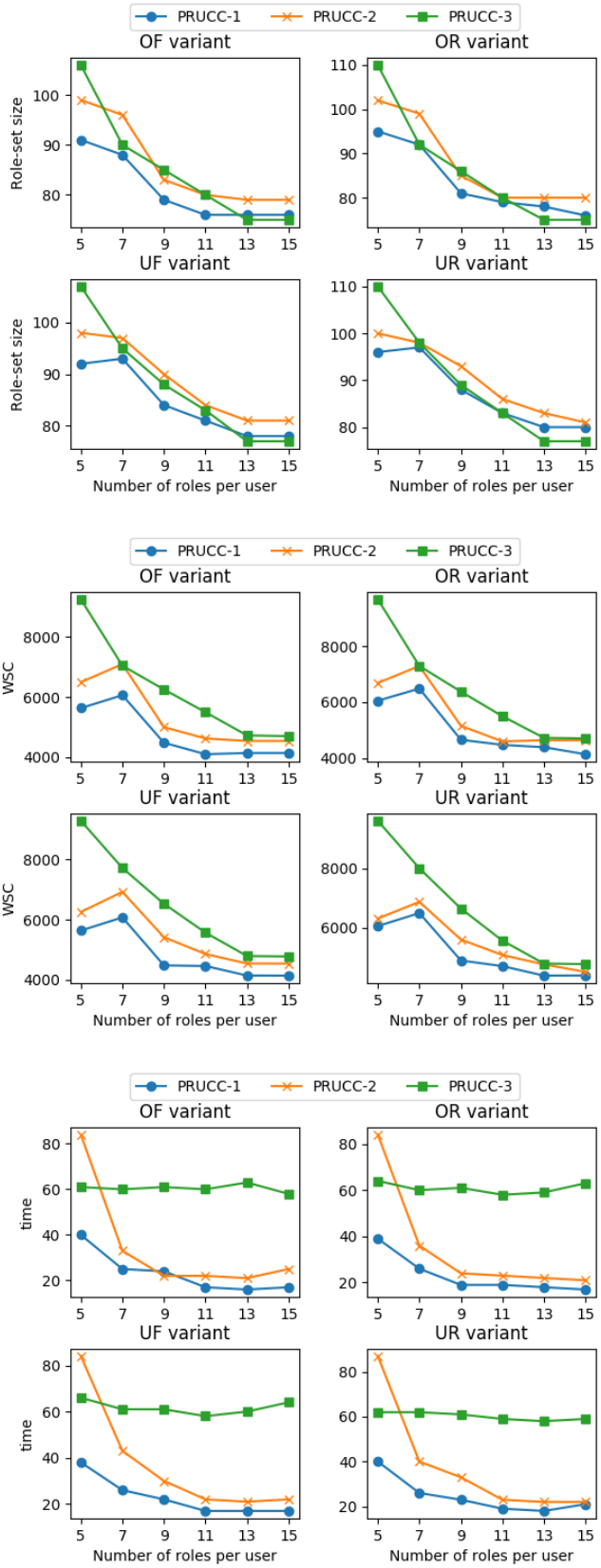
Dataset Domino $mpr = 42$**,** $mru \in [5, 7, 9, 11, 13, 15]$



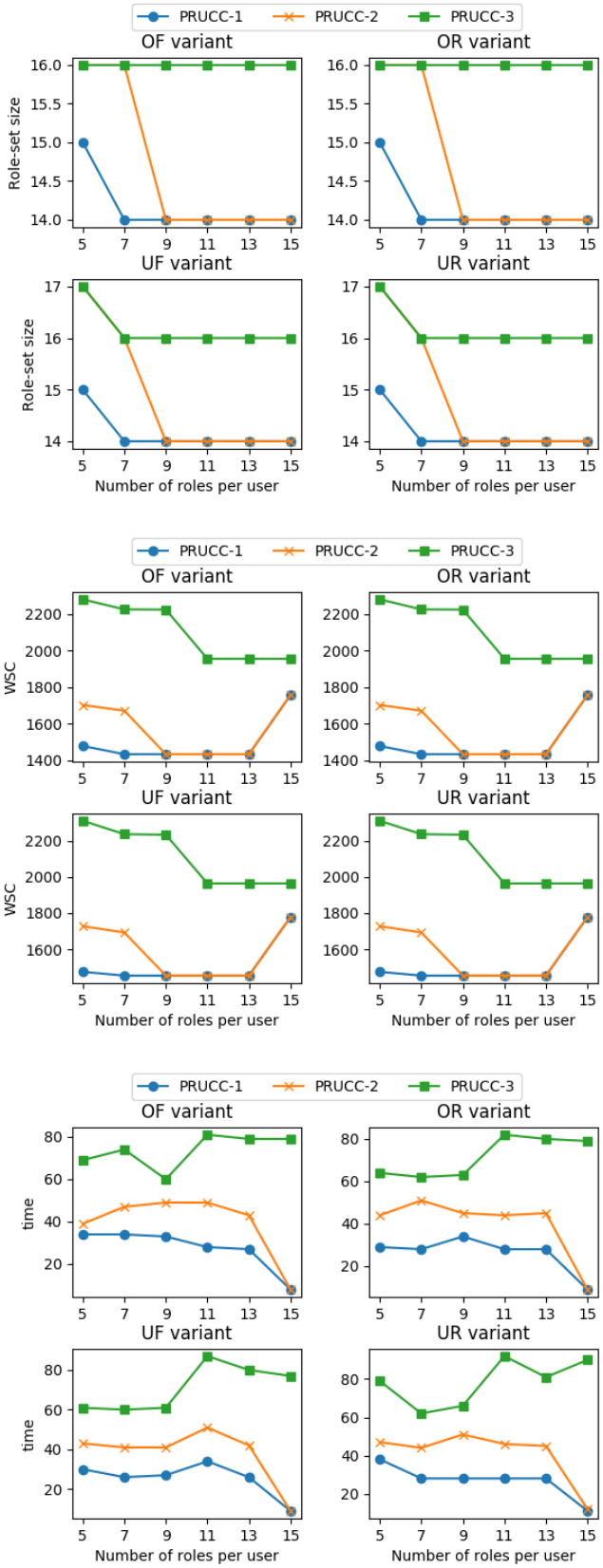
Dataset Emea $mpr = 111$**,** $mru \in [5, 7, 9, 11, 13, 15]$



Dataset Firewall 1 $mpr = 124$**,** $mru \in [5, 7, 9, 11, 13, 15]$



Dataset Firewall 2 $mpr = 118$**,** $mru \in [5, 7, 9, 11, 13, 15]$



Dataset Healthcare $mpr = 11$**,** $mru \in [5, 7, 9, 11, 13, 15]$