Appendix A: Performance of DSPs with Strategic Idleness Modification in Randomly Generated Open-shop System

As illustrated in Figure 1, XYZ's service system is similar to a three-stage tandem network with an open-shop network as the middle stage. The effect of the TBP in a three-stage tandem queue network has been established by Baron et al. (2014). Thus, here, we focus on open-shop service network. The main purpose of this section is to demonstrate that the SI modification can be combined with DSPs to provide nice tradeoff between macro- and micro-level SLMs in general open-shop service networks.

We considered a 10-station open shop: Each customer needs to be served in each station once. The stations can be visited in any order; there are no precedence constraints. Four stations are designated as potential bottlenecks: their utilizations are drawn from $Uniform\,(0.5, 0.9)$ distribution; the other six stations are non-bottlenecks with $Uniform\,(0.1, 0.5)$ distributed utilizations. Each station $j \in \{1, \ldots, 10\}$ has a random number of servers m_j, following a discrete $Uniform(1, 5)$ distribution. Let ρ_j be the utilization rate of station j. Customer service times in station j follow exponential distribution with parameter $\frac{10}{\rho_j m_j}$, where 10 is the average number of customers each station needs to serve per hour.

For the transient regime, we generated 100 workdays; the system starts empty at the beginning of the workday and works until all customers are processed. For every workday the number of customers, N, follows a discrete $Uniform(75, 85)$ distribution. One customer is scheduled to arrive every three minutes, during the first four hours of the workday. Their actual arrival times may uniformly deviate from their scheduled arrival times by ± 10 minutes. Then, all tests are performed using the same set of random numbers, stations, and customers to ensure the comparability of different policies.

The simulation for the steady-state regime was performed similarly as the transient-state regime, except that we generated 10,000 customers (i.e., 100,000 service requests) with a mean customer inter-arrival time of six minutes (inter-arrival times were drawn from $Uniform(0, 12)$ distribution).

We used the same three SLMs as in the XYZ case: the average system time, the proportion of customers with system time exceeding threshold T^s, and the number of red face incidents, where the latter was defined as a waiting time exceeding threshold T^{RF}. We also consider the measure of $E\,[W|RF]$ to examine if the TBP modification sacrifices the satisfaction of customers who already wait more than T^{RF}. To select reasonable values for these thresholds, we first simulate the performance of the non-idle LS policy, and use it as a benchmark. We set T^s as median (50-th percentile) of the system time; and $T^{RF}_{2.5\%}$ at the 97.5-th percentile of the waiting times, so that red faces are generated about 2.5% of the time under the non-idle LS policy. We define $T^{RF}_{5\%}$ and $T^{RF}_{10\%}$ in a similar fashion; clearly we have $T^{RF}_{2.5\%} > T^{RF}_{5\%} > T^{RF}_{10\%}$. Different thresholds for RFs lead to different rareness of RFs in the original system, from rare ($T^{RF}_{2.5\%}$) to somewhat common ($T^{RF}_{10\%}$).

Based on these T^s and T^{RF}s, we evaluate the other four non-idle DSPs and their TBP modifications, including Overtaking and Overtake-free TBPs. We remind that the threshold used in a TBP for deciding when to idle a customer is the one that minimizes the number of RFs, i.e., the incidents of waiting more than T^{RF}. The results for transient- and steady-states are displayed in Tables 7 and 8, respectively. These two tables have the same structure: Rows 4-8 contain the results for non-idle DSPs, rows 9-13 contain the results for the corresponding Overtake-free TBP modifications of each DSP, and rows 14-18 contain the results for the Overtaking TBP modifications of each DSP. Three column clusters (columns 2-6, 7-11, 12-16) contain results when the thresholds for RFs are $T^{RF}_{2.5\%}$, $T^{RF}_{5\%}$, and $T^{RF}_{10\%}$, respectively. The result in Tables 7 and 8 is a good representative of all networks we simulate (other results are available from the corresponding author upon request).

We first look at the result of transient state in Table 7. By comparing rows 4-8 with rows 9-13, we derive similar intuitions to those in Section 5.3. In the $T^{RF}_{2.5\%}$ case, comparing the performance of overtake-free TBP-modified policies with their non-idle counterparts, we see that the overtake-free TBP modification is very successful in reducing the number of RFs: all policies experience large reductions, with the top performers (LMOP, LCW) experiencing a reduction of over 96%. For LAW, LS, and SERP policies, the cost of these reductions in RFs is relatively modest: the average total system time increases by less than 10% and the probability of spending more than T^s time in the system increases by less than 17%. However, for LMOP and LCW policies, the cost is much higher: the average total system time increases by more than 22% and the probability of spending more than T^s time in the system increases by more than 30%. Further, when the threshold for RFs decreases, the advantage of LMOP and LCW policies in the number of RFs diminishes. For example, when the threshold for RFs decreases to $T^{RF}_{10\%}$, the non-idle versions of LMOP and LCW perform significantly worse than other TBPs in almost all measures. Even when the Overtake-free TBP modification reduces the number of RFs by over 50%, the resulting LMOP+TBP and LCW+TBP are still dominated by LS+TBP.

Comparing rows 9-13 with rows 14-18 in Table 7 shows that the performances of Overtake-free and Overtaking TBPs are similar. For example, in the $T^{RF}_{2.5\%}$ case, although when combined with LCW policy the Overtake-free TBP dominates the Overtaking TBP in all SLMs, when combined with other DSPs there is no clear preference between Overtaking and Overtake-free TBPs: a smaller number of RFs is always associated with a longer average total system time, and vice versa.

To illustrate the improvement that can be achieved by combining SI modification under different THs with DSPs compared with the non-idle version of DSPs, we plot the mean system time versus the number of RFs for $T^{RF}_{2.5\%}$, $T^{RF}_{5\%}$, and $T^{RF}_{10\%}$ under LS policy with Overtake-free TBP modification for $TH \in \{1, \dots, 20, \infty\}$ on Figure 6. Each curve represents a different value of T^{RF}. On each curve, from the right to the left, the TH increases. Note that, when TH increases, the TBP is triggered less often, so the performance of the LS policy gets closer to the non-idle policy. In the extreme case, when $TH = \infty$, the LS policy acts the same as the non-idle policy.

We see from Figure 6 that when the incidents of RFs are rare ($T^{RF}_{2.5\%}$), the LS policy with $TH^* = 10$ provides attractive tradeoff to reduce the number of RFs by close to 83% at the cost of increasing the mean system time by less than 10%. We can choose different THs in the LS policy for different tradeoffs, depending on the company's strategy. When the incident of RFs are more common, the SI modification is less successful and the maximum reduction of RFs for $T^{RF}_{5\%}$ and $T^{RF}_{10\%}$ are 74.6% and 42.1% respectively. Tradeoff curves for other DSPs look similar. Thus, in an open-shop network, the SI modification provides a similar tradeoff as it does in a tandem queue network in Baron et al. (2014).

We see from Table 8 that while the results for the steady-state regime are similar to the ones for the transient case above, some clear differences emerge. First, the Overtaking TBP dominates the Overtake-free TBP in the number of RFs. Of course, the cost of reducing the number of RFs is also high. In some extreme case, e.g., when the threshold for RFs is $T^{RF}_{2.5\%}$, the Overtaking TBP when combined with LAW policy can reduce the number of RFs by 97% at a cost of increasing the total system time by 68.4%. Whether it is a good tradeoff depends on the company's strategy. Second, when the threshold for RFs is small, i.e., the incidents of RFs are prevalent, the SI modification losses its advantage in reducing the number of RFs. For example, in the $T^{RF}_{10\%}$ case, the Overtake-free TBP increases the number of RFs when combined with LS, LMOP, LCW, and SERP policies. While the Overtaking TBP still manages to reduce the number of RFs for all DSPs, the percentage reduction is smaller than in the $T^{RF}_{2.5\%}$ and $T^{RF}_{5\%}$ cases. This observation is in line with the intuition in Baron et al. (2014): the TBP is able to reduce the number of RFs only when the incidents of RFs are sufficiently rare in the system.

Of course, one can repeat the above simulation for sufficient times to obtain an estimation of the system's expected performance under different TBPs with or without SI modification. However, due to the large number of randomly generated parameters, the time it takes to obtain such an estimation will be too long, and the insights would not change much. Moreover, one can use other distributions to generate the daily number of customers, arrival times of customers, utilizations, numbers of server, and service times for different stations. For example, we also run the same simulation using *Beta* distribution to generate the inter-arrival and service times. Nevertheless, the effect of SI modification remains the same. We remind that our DSP+TBP modification works well in the simulation of XYZ's service system, which is already a general service network with no presumptions on the inter-arrival or service time distribution.

To conclude, for a randomly generated open-shop system in both transient- and steady-states, introduction of SI into our DSPs allows us to reduce the number of incidents of RFs, at the cost of a modest increase in the amount of system time and proportion of customers with system time exceeding T^s. Thus, these conclusions seem robust for general stochastic open-shop networks.

	$T^{RF}_{2.5\%}$					$T^{RF}_{5\%}$					$T^{RF}_{10\%}$				
	System Time			RFs		System Time			RFs		System Time			RFs	
Policies \ Measures	Mean	Stdev	$\geq T^s$	$\geq T^{RF}_{2.5\%}$	$E[W\|RF]$	Mean	Stdev	$\geq T^s$	$\geq T^{RF}_{5\%}$	$E[W\|RF]$	Mean	Stdev	$\geq T^s$	$\geq T^{RF}_{10\%}$	$E[W\|RF]$
LAW+Non-idle	2:34:22	1:02:08	49.4%	1615	1:00:33	2:34:22	1:02:08	49.4%	4017	44:38	2:34:22	1:02:08	49.4%	9794	29:09
LS+Non-idle	2:30:40	0:52:52	50.0%	2000	1:03:35	2:30:40	0:52:52	50.0%	4003	48:39	2:30:40	0:52:52	50.0%	8007	33:34
LMOP+Non-idle	2:39:52	0:57:29	57.2%	1325	57:36	2:39:52	0:57:29	57.2%	4745	40:06	2:39:52	0:57:29	57.2%	12714	26:42
LCW+Non-idle	2:40:18	1:01:47	55.6%	1358	58:00	2:40:18	1:01:47	55.6%	4824	40:15	2:40:18	1:01:47	55.6%	12410	27:06
SERP+Non-idle	2:29:52	0:57:58	47.3%	1569	1:00:44	2:29:52	0:57:58	47.3%	3609	45:17	2:29:52	0:57:58	47.3%	8622	29:33
LAW+OTFreeTBP	2:44:13	1:04:20	56.9%	323	56:02	2:48:14	1:04:19	59.8%	1320	38:51	2:40:17	1:02:24	53.9%	5913	24:03
LS+OTFreeTBP	2:45:11	1:01:33	58.5%	341	55:37	2:43:59	0:59:38	58.7%	1017	39:08	2:36:38	0:55:22	54.8%	4638	25:30
LMOP+OTFreeTBP	3:20:18	1:09:44	76.1%	50	49:02	2:53:21	1:02:51	64.4%	594	36:36	2:40:07	0:58:13	57.0%	6317	22:01
LCW+OTFreeTBP	3:16:05	1:12:18	72.6%	44	48:48	2:55:36	1:06:07	63.8%	444	35:55	2:41:26	1:01:14	56.7%	5456	21:51
SERP+OTFreeTBP	2:37:54	1:02:23	52.6%	160	55:33	2:42:47	1:01:50	56.7%	727	36:56	2:34:39	0:58:58	51.2%	4557	22:34
LAW+OTTBP	3:11:28	1:13:21	69.8%	83	54:50	3:06:25	1:09:38	68.9%	366	35:05	2:50:55	1:04:17	62.1%	3366	22:32
LS+OTTBP	3:07:44	1:09:27	69.4%	13	48:12	2:55:32	1:03:31	65.8%	124	34:23	2:41:38	0:57:28	58.8%	2494	23:08
LMOP+OTTBP	3:17:25	1:14:13	72.4%	85	48:50	3:08:40	1:10:26	70.1%	681	32:13	2:49:55	1:00:55	63.8%	5782	20:35
LCW+OTTBP	3:19:53	1:15:43	72.6%	71	46:20	3:11:44	1:13:33	69.7%	615	31:03	2:48:32	1:03:55	61.3%	5373	19:43
SERP+OTTBP	2:49:44	1:05:48	59.9%	4	45:12	2:59:30	1:07:38	65.4%	60	30:11	2:40:30	1:00:51	56.0%	2273	19:36

Table 7 **The Average Performance of Different Scheduling Policies with and without SI in a Random Open-shop Network in Transient-State.**

	$T^{RF}_{2.5\%}$					$T^{RF}_{5\%}$					$T^{RF}_{10\%}$				
	System Time			RFs		System Time			RFs		System Time			RFs	
Policies \ Measures	Mean	Stdev	$\geq T^s$	$\geq T^{RF}_{2.5\%}$	$E[W\|RF]$	Mean	Stdev	$\geq T^s$	$\geq T^{RF}_{5\%}$	$E[W\|RF]$	Mean	Stdev	$\geq T^s$	$\geq T^{RF}_{10\%}$	$E[W\|RF]$
LAW+Non-idle	2:04:23	0:52:55	48.3%	2318	1:22:27	2:04:23	0:52:55	48.3%	5269	0:50:14	2:04:23	0:52:55	48.3%	11405	0:29:01
LS+Non-idle	2:03:12	0:46:35	50.0%	2500	1:20:02	2:03:12	0:46:35	50.0%	5000	0:52:43	2:03:12	0:46:35	50.0%	10000	0:31:45
LMOP+Non-idle	2:04:07	0:49:37	49.8%	2245	1:20:11	2:04:07	0:49:37	49.8%	5462	0:47:35	2:04:07	0:49:37	49.8%	11846	0:27:51
LCW+Non-idle	2:05:47	0:53:28	49.7%	2393	1:22:37	2:05:47	0:53:28	49.7%	5485	0:49:55	2:05:47	0:53:28	49.7%	12055	0:28:38
SERP+Non-idle	2:03:26	0:52:03	48.4%	2287	1:21:28	2:03:26	0:52:03	48.4%	5122	0:50:18	2:03:26	0:52:03	48.4%	10964	0:29:15
LAW+OTFreeTBP	2:20:55	0:56:36	62.0%	1212	0:49:46	2:25:53	0:54:39	67.8%	4319	0:32:37	2:17:36	0:53:17	61.3%	11400	0:22:39
LS+OTFreeTBP	2:25:21	0:54:13	67.0%	1526	0:53:28	2:26:47	0:52:08	69.4%	4426	0:36:30	2:12:49	0:49:00	58.9%	10239	0:25:14
LMOP+OTFreeTBP	2:33:14	0:59:13	70.2%	900	0:45:30	2:22:01	0:53:38	65.0%	4642	0:32:08	2:11:03	0:50:08	56.6%	11986	0:22:09
LCW+OTFreeTBP	2:21:35	0:58:10	61.3%	971	0:46:25	2:28:17	0:56:13	68.5%	4688	0:30:53	2:16:12	0:52:27	60.6%	12101	0:21:51
SERP+OTFreeTBP	2:18:57	0:55:29	61.2%	1031	0:49:28	2:21:49	0:54:42	64.0%	4277	0:33:17	2:08:20	0:52:11	52.8%	11129	0:22:57
LAW+OTTBP	3:29:30	1:05:16	93.5%	70	0:46:14	2:33:56	0:59:01	70.3%	1638	0:25:44	2:25:41	0:55:43	67.7%	9642	0:20:11
LS+OTTBP	3:31:12	1:02:44	93.6%	85	0:45:56	2:40:04	0:54:55	77.2%	1707	0:27:04	2:24:48	0:51:35	68.3%	8736	0:21:51
LMOP+OTTBP	2:19:25	1:07:03	56.3%	467	0:44:25	2:31:24	1:02:29	66.9%	3146	0:26:43	2:21:27	0:54:16	63.8%	10306	0:19:40
LCW+OTTBP	2:36:19	1:18:05	64.6%	450	0:43:11	2:19:30	1:01:55	58.2%	2933	0:25:37	2:12:55	0:54:09	56.6%	10481	0:19:01
SERP+OTTBP	2:54:14	1:07:13	80.5%	64	0:41:40	2:54:00	0:55:29	85.9%	1444	0:24:14	2:22:34	0:53:21	65.9%	9294	0:18:56

Table 8 **The Average Performance of Different Scheduling Policies with and without SI in a Random Open-shop Network in Steady-State.**

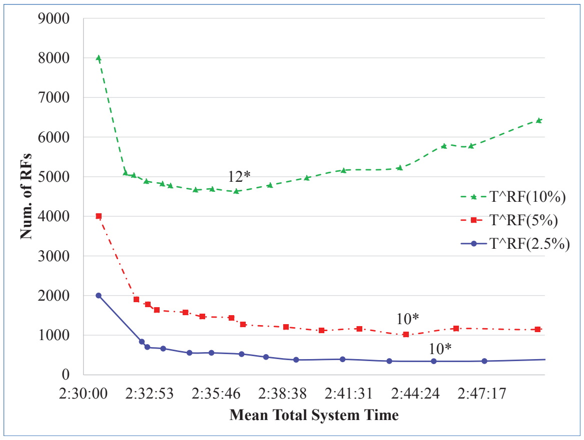

Figure 6 **Tradeoff curves of SI modification corresponding to $T^{RF}_{2.5}\%$, $T^{RF}_{5\%}$, and $T^{RF}_{10\%}$. Non-idle LS policy corresponds to the left-most point on each curve.**