Limited Agents Mix-Training in Operations – A Call Center Case

Ditila Ekmekçiu

Department of Finance, Teleperformance Albania, AMS shpk, Tirana, Albania

Abstract: ***The operations of an inbound call center are difficult to handle; there is important uncertainty in the arrival rates estimates, and the operation is usually depending on severe service level restrictions. This paper is inspired by the work with an outsourcing call center where projects involve an inbound help desk depending on a service level agreement (SLA). Services of support are very specialized and an important training investment is needed and this kind of investment cannot be transferred to other projects. In this work, it will be analyzed the possibility of cross training a group of agents in order to serve calls from two different projects. This process is called a partial merging. This paper looks for quantifying the advantages of partial merging and characterizing the circumstances under which merging is more profitable. It will be seen that low levels of cross training produce important profit.***

Keywords: mix-training, call center, uncertain arrival rates, sla

1. Introduction

A call center is composed by a set of resources (generally staff, computers and telecommunication supplies) that guaranties the delivery of services through the phone. The working environment of a typical large call center could be imagined as very big room with numerous open-space workstations, in which people with earphones sit in front of computer terminals, providing services/products to “unseen” customers (Ekmekçiu, 2015). During the last years, the growth of the call center’s industry has been very high. In particular, in Albania, the number of outsourcing call centers is more than 300 with 20.000 employees (Mapo online, Call-center: Biznesi që vijon lulëzimin). Large range call centers are technologically and organizationally sophisticated operations and have been the object of important academic research. Call centers can be categorized into many different dimensions: functionality (help desk, emergency, telemarketing, tele selling, technical support, market research, information providers, etc.), size (from a few to several thousands of agent workstations), geography (single- or several locations, that can be domestic, nearshore or offshore), agents qualifications (low-skilled or highly-trained, single or multiskilled etc.), industry (telecommunication, finance, TV and internet providers, travel, marketing, sport etc.), line of business (inbound, outbound, back office etc.), and so on (Ekmekçiu et al 2015).

The paper is precisely related to a research project with an outsourced Albanian call center. This operation includes providing help desk support to big corporate entities. While the range of services differs from account to account, many accounts demand 24/7 support and virtually all accounts depend on some shape of Service Level Agreement. There are different types of SLAs, but the most typical determines a minimum level of the Telephone Service Factor. A TSF SLA determines the fraction of calls that should be answered within a given time. For example, a 90/110 SLA specifies that 90% of calls must be answered within 110 seconds. A very significant point is that the service level is applied to an extended period, generally a week or month. Consequently, the workstation is generally staffed so that at some times the service level is under obtained, sometimes over obtained, and reaches the target for the entire extended period (week or month). The main challenge implicated with staffing this call center is meeting a certain, fixed SLA with a variable and uncertain arrival rate pattern (Ekmekçiu et al 2016).

The number of calls given in each ½ hours period is very variable with various sources of uncertainty. Beyond the day of week seasonality these projects go through a very important time of day seasonality. The volumes increase strongly in the morning with a considerable growth of calls between 7 and 11 AM. The volumes have the tendency to lower during lunch break, and then a second peak happens in the afternoon; this peak is usually lower comparing with the morning peak. In this context agents need considerable training on the systems for which they give support. An agent might need up to 3 months of training before taking calls, but the general period of training is 3-4 weeks. The most of this training is specific for each project. Then it is needed for the agent’s different months to obtain full productivity with an additional training on the job. As the costs of training are high, the normal procedure is to train agents for one single project. The main staffing challenge is to reach a minimum staffing cost that obtains the SLA goal with high probability. The plan should clearly set before the arrival rate uncertainty is shown.

As the management has few alternatives to arrange the manpower during the flow of the day (utilizing the sooner dismissal, the overtime etc.) these kinds of actions are usually very restricted. In this paper, it will be analyzed a specific protecting strategy; a different approach that is called partial merging. In partial merging a small group of agents make a cross training to handle calls from two projects. Then the call center can be seen as a model of skill-based routing. These agents have both skills, while the other agents continue having one skill. It is obvious that to cross train all agents will grow the service level of the call center for a given level of staffing. The assumption is that cross training a small number of agents can bring an important portion of the profit and the goal is to achieve the level of cross training that diminishes staffing costs, at the same time by satisfying the

Licensed Under Creative Commons Attribution CC BY

service level restriction with high probability.

2. Literature Review

We can find a summary of the literature oriented to call centers in (Gans et al. 2003). This accurate tutorial and analysis gives a detailed overview of the operations, vocabulary and research in call centers. The work sums up the most important academy research on different areas linked to the research in call center, counting the management of the capacity and scheduling. In (Mandelbaum and Zeltyn 2004; Garnett et al. 2002; Whitt 2006) are given accurate analysis of models including abandonment and Erlang A models. We can find an empirical analysis of the data in call center in some papers of (Brown et al. 2005; Mandelbaum et al. 2001). These works include an accurate statistical analysis of data from a small call center.

The authors examine different ordinary hypothesis utilized in queuing models, and discover between other things that talk time in call centers pursue a lognormal distribution instead of the usually taken for granted exponential one. A part of call centers according to (Avramidis et al., 2004) show arrival behavior with higher variability compared to a Poisson process and large correlations among periods within a day. In the paper (Wallace and Whitt 2005) there are six different call types and each agent is qualified to handle a given number of these types. Here it is utilized a model based on simulation to find the perfect level of cross training. The principal awareness of the paper is that cross training in low levels gives the most of the profit. Particularly, they find that qualifying each agent in two skills gives the mass of the profit, meanwhile supplementary training has a nearly low reward. Even though the discovery of this work is similar, the models are very distinct. Their optimal solution has each agent cross trained in two skills, meanwhile my model considers that only a little group of agents are cross trained. In my paper, cross training is very costly and cross train each agent is not efficient. The (Wallace and Whitt 2005) model does not consider abandonment that is important for my scenario.

Other studies for the scheduling and staffing of multiskill call centers involve stochastic fluid models, research methods with the approximations of loss-delay and algorithm with cutting- plane mixed with simulation (Harrison and Zeevi 2005; Avramidis et al. 2006; Cezik and L'Ecuyer 2006). These ways are designed for big call centers with different skill classifications and group of customers.

3. Merging Model

In this section I present the model of partial merging. I consider that in the standard case the call center in restricted by project and every project behaves with a different Erlang A queuing system. Every project i sustain calls that come with the λi rate. Correlated with every call there is an average talk time equal to $1/\mu i$. another assumption is that the caller's patience is exponentially distributed with mean $1/\Theta i$. The parameter of patience symbolizes the time a caller has the will to wait on hold. Every caller will hang up (abandon the queue) if they are not served within their parameter of patience. For the goal of the paper the Telephone Service Factor is the key performance metric for this queue and determined as:

$$TSF = P\{W \leq T\}$$

The staffing decision in the steady state then includes forecasting the arrival rate λi and establishing the staff level to obtain the indicated SLA with an adequate level of probability. My analysis is based on modifying the existing practice; particularly cross training a group of agents to handle 2 projects. In this model, I consider that the routing system that is skill based is set as follows:

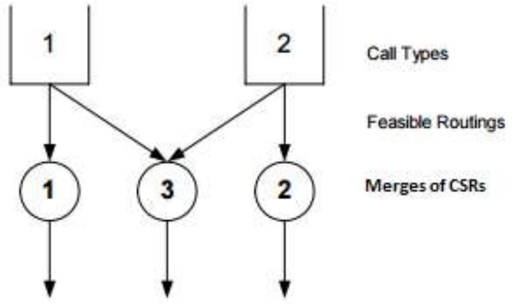

Figure 1: Basic routing structure

In this case, there are two types of calls, one for every project, and 3 agent pools. The first pool has the first skill and can handle the first type of calls. The same for the second pool. The third pool is cross trained and therefore can handle the calls of each kind. I implement an elementary routing model. A call that is coming is routed to a one-skill agent if one of them is available. Only in cases where all one-skill agents are busy, the call is routed to a multi-skill agent. In the case where one-skill agents become available, they handle the call of the caller that is waiting the longest in their respective queue. If there is no call waiting they switch in the waiting status. In the case that a multi-skill agent becomes available, they handle the call of the largest queue.

4. Partial Merging in Steady State

4.1. The TSF Response Function

In the individual queue, individual resource merge case, there is an analytical expression for the service level seen as a function of staffing and arrival rates and it is simple to generate a scheme of the TSF as a function of staffing. From the TSF function it is achieved an s-shaped curve, where we can see the service level increasing fast at the beginning, then settling and in the end demonstrating descending returns when the staff increases. In the merging case the situation is significantly more difficult. There are not available known expressions to calculate the service level. Depending on the intuition it is expected that the service level is getting higher in the number of one-skill agents and of multi-skill agents. To prove this intuition, I utilize simulation to build the following chart of the TSF being a function of the number of agents. In this type of simulation, I take for granted that every queue gets calls at a rate of hundred per hour, in every case talk time is in average twelve minutes, the patience of the caller is in average is 350

Licensed Under Creative Commons Attribution CC BY

seconds, and that the service level depends on a hold time of 120 seconds. I modify the number of agents assigned to every base merge and the number of agents assigned to the merge of the multi0skill agents separately. For every combination of the staffing, I simulate the operations for 2 days and complete 25 replications:

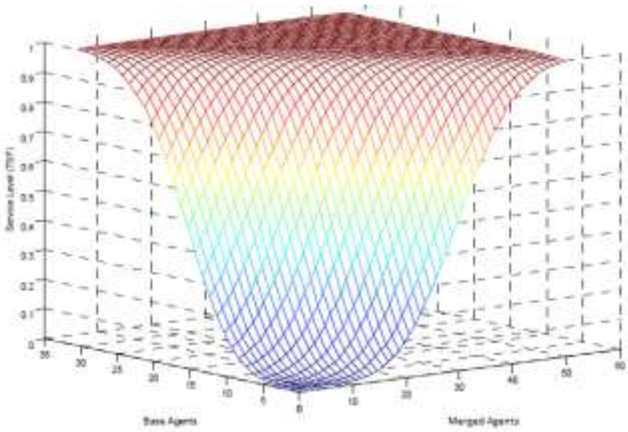

Figure 2: The TSF response function for a partial merging model

The chart demonstrates a large level of TSF at 100% when it is large the total number of agents. Likewise, a small level at TSF of 0% exists when it is small the total number of agents. Between the two levels the surface presents a profile s-shaped. In figure 3 we can see on outline graph of this data in two dimensions. The outline graph demonstrates a series of iso-service level lines, mixture of agents that give the same service level.

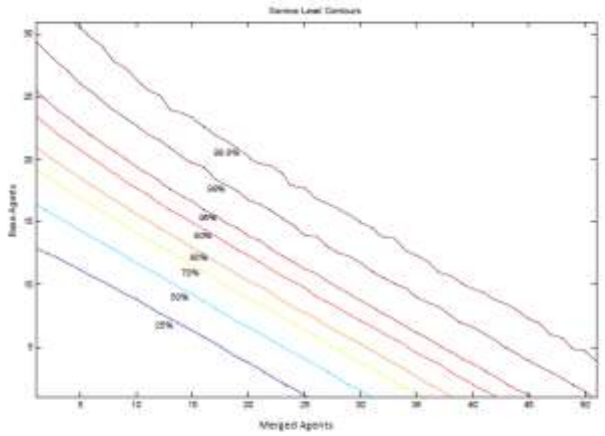

Figure 3: The merged TSF outline graph

To get a 95% service level, for example, we require approximately 25 agents in every merge of 50 agents in total. Nevertheless, in a genuine merged mode, the same service level might be obtained with a total of 45 merges agents only. Even though it is complicated to see, close examination demonstrates that the iso-service lines have a convex rounded shape, so are not straight. This is additionally shown in the following figure where I demonstrate the TSF of 80% outline with a line that connects the end points. The convexity of the outline indicates that the mixture of the cost minimizing of merged and one-skill agents can be in the inside.

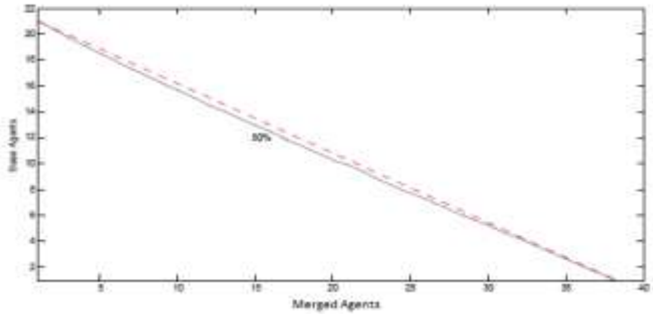

Figure 4: Outline of 80% TSF

4.2. Symmetric Projects

During this examination, I analyze the impact of merging on steady state performance with symmetric projects. Take into consideration 2 statistically equal projects and each of them staffed with 36 agents and getting calls at a rate that is constant λ. The distribution of talk time is exponential with mean twelve minutes and the abandonment mean time is 350 seconds. The measurement of the service level is done against a hold time of 2 minutes. I examine the situation where we have a constant total number of agents, but every project provides between 0 and 36 agents to the merge. The first chart demonstrates the service level for every level of merging when the arrival rate (λ) is 200 calls/hour. The abandonment rate is schemed in the second chart. In every situation, I scheme abandonment rate and TSF for one of the projects, knowing that the projects are symmetric so they have an identical curve. The generation of the data is done by simulating 5 days of operations above 50 replications. It is shown in every curve the sample average in addition to a 90% confidence interval.

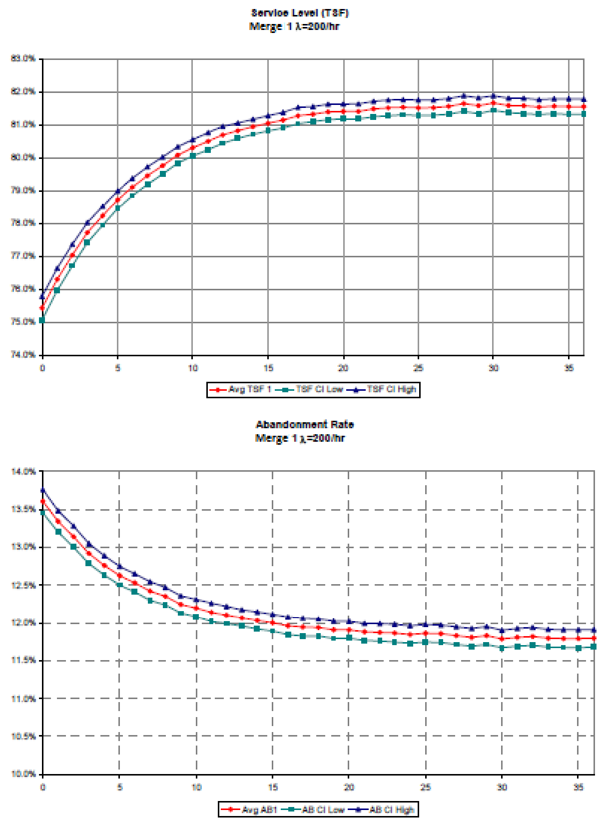

Figure 5: Impact of merging on TSF and abandonment with fixed staffing levels

Licensed Under Creative Commons Attribution CC BY

These charts show that a small level of merging produces improvement, but also that the return on cross training descends fast. In every situation, cross training ten agents gives the largeness of the benefit and more than 15 agents cross-trained give a very low benefit. I have repeated this examination with different arrival rates (180 and 220 calls/hour) and achieved very much alike results. In every situation, cross training can improve TSF with 5/6%, and meanwhile the highest improvement is in the case of medium volume (200 calls/hour). The abandonment rate decreases as follows: 1% in the situation of high volumes, 1.8% in medium situation and 2.3% in low situation.

4.3. Steady State Differential Rates

The former analysis shows that fair benefits are gained when agents receive a cross training, and the quantity of improvement counts on the unused capacity in the system. Still in that analysis the projects had both identical arrival rate. The case when the arrival rates are different can be more interesting as can be the case if rates are contingent to forecast error. I permit variation on arrival rates separately from target by ± 10% in the next analysis. The total number of staffing is fixed at 72, which means that in the no merging case every project has 36 agents. This staff level appears in a nearly 76% service level with no merging. The coming tables sum up the resulting TSF amounts under different arrival rate combinations. The number of agents merged is indicated by the numbers along the top (0-35). In Table 1 I analyze the impact on the mixed TSF, and the TSF of every single project. We notice that the global TSF is continually improved by merging, and the intensity of improvement is based on the quantity of unused capacity in the system. When the two projects have lower plan volume, the global TSF's improvement is by 2.8% with only 5 agents. If the two projects have higher plan volume, the TSF's improvement is also by 2.8%. The biggest benefit is achieved when the projects have distinctive rates; when one project is high and the other low we achieve a 5.2% benefit in global TSF. The improvement decreases quickly with the number of cross trained agents; the most gain is achieved from the first few agents. Results from cross training more than 15 agents are not significant, and in a lot of cases are not statistically different from zero.

Table 1: Impact of merging on global TSF. * shows statistical significance at the 0.9 level

TSF Total

λ_1	λ_2	0	5	10	15	20	25	30	35
180	180	87.9%	90.8%	91.9%	92.5%	92.7%	92.8%	92.9%	92.8%
180	200	81.5%	85.3%	86.9%	87.6%	88.0%	88.1%	88.1%	88.0%
180	220	73.6%	78.8%	80.9%	81.9%	82.4%	82.5%	82.5%	82.5%
200	200	76.1%	79.3%	80.8%	81.6%	81.9%	82.0%	82.1%	81.9%
200	220	68.8%	72.3%	74.2%	75.0%	75.3%	75.6%	75.6%	75.5%
220	220	62.3%	65.1%	66.6%	67.5%	67.8%	68.0%	68.0%	68.1%

Δ TSF Total

λ_1	λ_2	5	10	15	20	25	30	35
180	180	2.8% *	1.1% *	0.6% *	0.2% *	0.1% *	0.0%	0.0%
180	200	3.8% *	1.6% *	0.7% *	0.3% *	0.1% *	0.0%	0.0%
180	220	5.2% *	2.2% *	1.0% *	0.5% *	0.1% *	0.0%	-0.1% *
200	200	3.2% *	1.5% *	0.8% *	0.3% *	0.1% *	0.1% *	-0.2% *
200	220	3.6% *	1.9% *	0.8% *	0.3% *	0.2% *	0.0%	-0.1% *
220	220	2.8% *	1.5% *	0.9% *	0.3% *	0.3% *	0.0%	0.1%

When I analyze the data at the single project level (Tables 2-3), the results are even more interesting When the arrival rate of every project are similar the gains are distributed uniformly. The maximum benefit happens when the arrival rates are different; and that benefit accumulates excessively to the understaffed project. In the case when volumes are at opposite extremes, the benefit of the understaffed project is of an 11% increase in TSF from only five cross trained agents. The TSF is increased by another 10 pts by cross training 10 agents bringing TSF to almost 80%. In the case of important discrepancy, the overstaffed project can go through reduction in performance, but this recession is considerably smaller than the increase to the other project and global TSF increases always. The most important case is when volumes have a maximum discrepancy and the TSF of the overstaffed project lessens by 2.2% with five agents cross trained. It must be noticed that this project's baseline TSF was 86%, much higher than the standard objective of 80%. However, this result raises an alertness for merging projects with very high (like 90%) TSF goals. In the case of a smaller discrepancy the recession was very modest, around 0.9% with the cross training of 10 agents, where the busy project might notice an increase on the order of 4 pts from only the cross training of 5 agents. We gained comparable results for the abandonment rate for every project; merging decreases the maximum wait time callers face, and thus diminishes the proportion of callers kept on hold over their patience level. The increase is the most significant when a capacity discrepancy happens.

Table 2: Impact of merging on low volume project TSF

TSF Merge 1

λ_1	λ_2	0	5	10	15	20	25	30	35
180	180	86.8%	90.4%	91.7%	92.4%	92.7%	92.8%	92.8%	92.8%
180	200	86.8%	87.5%	87.7%	87.8%	87.7%	87.6%	87.4%	87.3%
180	220	86.8%	84.7%	82.9%	82.1%	81.4%	80.9%	80.6%	80.2%
200	200	75.5%	78.9%	80.5%	81.4%	81.8%	82.0%	82.0%	82.0%
200	220	75.5%	75.3%	75.1%	74.9%	74.7%	74.4%	74.3%	74.0%
220	220	61.9%	65.0%	66.4%	67.6%	68.0%	68.1%	68.2%	68.2%

Δ TSF Merge 1

λ_1	λ_2	5	10	15	20	25	30	35
180	180	3.5% *	1.3% *	0.7% *	0.2% *	0.1% *	0.0%	0.0%
180	200	0.7% *	0.2% *	0.1%	0.0%	-0.1% *	-0.2% *	-0.1% *
180	220	-2.2% *	-1.7% *	-0.8% *	-0.7% *	-0.5% *	-0.3% *	-0.4% *
200	200	3.4% *	1.6% *	1.0% *	0.4% *	0.1% *	0.1%	-0.1% *
200	220	-0.2%	-0.2%	-0.3% *	-0.2% *	-0.2% *	-0.1% *	-0.2% *
220	220	3.1% *	1.5% *	1.1% *	0.4% *	0.2% *	0.1%	0.0%

Table 3: Impact of merging on high volume project TSF

TSF Merge 2

λ_1	λ_2	0	5	10	15	20	25	30	35
180	180	89.0%	91.1%	92.1%	92.5%	92.8%	92.9%	92.9%	92.8%
180	200	76.7%	83.3%	86.1%	87.5%	88.2%	88.5%	88.7%	88.7%
180	220	62.7%	73.9%	79.3%	81.8%	83.2%	83.8%	84.1%	84.3%
200	200	76.7%	79.7%	81.0%	81.7%	82.0%	82.0%	82.2%	81.9%
200	220	62.7%	69.6%	73.4%	75.1%	75.9%	76.6%	76.8%	76.8%
220	220	62.7%	65.2%	66.7%	67.4%	67.6%	67.9%	67.8%	68.0%

Δ TSF Merge 2

λ_1	λ_2	5	10	15	20	25	30	35
180	180	2.1% *	0.9% *	0.4% *	0.3% *	0.1% *	0.0%	-0.1% *
180	200	6.6% *	2.8% *	1.4% *	0.7% *	0.4% *	0.1% *	0.0%
180	220	11.3% *	5.3% *	2.5% *	1.5% *	0.6% *	0.3% *	0.2% *
200	200	3.0% *	1.3% *	0.6% *	0.3% *	0.0%	0.1% *	-0.2% *
200	220	7.0% *	3.8% *	1.7% *	0.9% *	0.7% *	0.2% *	0.0%
220	220	2.6% *	1.5% *	0.6% *	0.2% *	0.3% *	-0.1%	0.2% *

In general, this analysis demonstrates that partial merging produces important gains in steady state. The increase is the highest when a capacity discrepancy happens and the under-capacity project gets the higher gain. In the next section, we analyze how arrival rate uncertainty affects the merging analysis.

Licensed Under Creative Commons Attribution CC BY

4.4. Steady State but Uncertain Arrival Rate

In this examination, I continue to analyze the impact of merging when projects have a constant rate, but now I permit uncertainty in the arrival rate. Particularly I take for granted that the calls in every merge will come with a constant rate, but the fulfilled rate is a random variable. Assume that the arrival rates are independent and distributed identically normal random variables with mean 200 and standard deviation 20. I analyze how partial merging impacts the TSF expected and abandonment rate. The coming graph show the results of a simulation experiment.

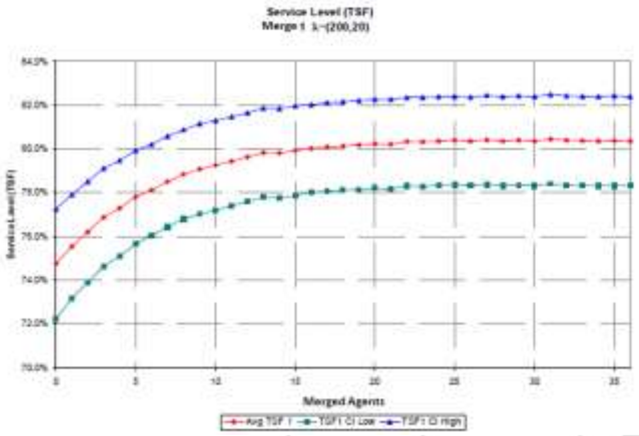

Figure 6: Impact of merging on TSF with fixed staffing levels and uncertain arrivals - TSF confidence levels

In this case, the generated service curve is almost equal to the schemes for steady state arrivals at 200 calls/hour but the level of TSF is lower in the case of uncertainty: 74.2% vs. 75.4% with no merging and 80.5% vs. 81.5% in the case of full merging. Thus, not a considerable shift, this demonstrates one of the effects of uncertain arrival rate. As a result of the nature of the TSF curve the response of volume changes is unproportioned; higher volume brings a larger shift in the service level that result then lower volume. Thus, even if volume changes around the mean symmetrically, the TSF that will result is going to be lower in the case of uncertainty than in the case of the corresponding mean value. Much alike results were discovered for abandonment: the curves in the uncertainty case have an alike shape as in the certainty case, with a fairly higher abandonment rate at all levels. A curious event is given in Figures 7 and 8. In Figure 7 we see that the standard deviation of the global service level is basically not affected by merging, staying at an approximately constant level s little over 8%.

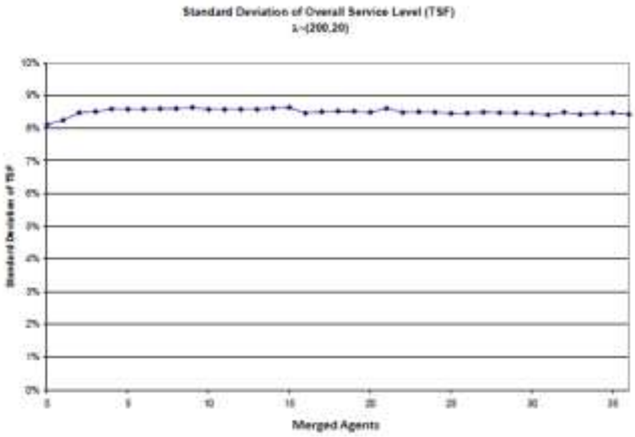

Figure 7: Standard deviation of the global TSF

Nevertheless, Figure 8 shows the standard deviation of the service level for merge 1 gets lower as the merging level gets higher, at least for the first few agents.

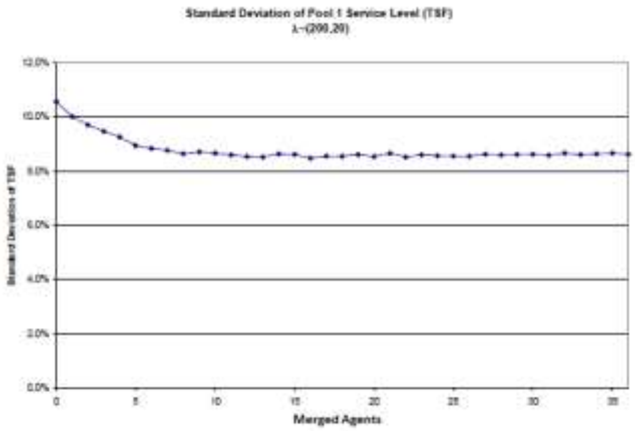

Figure 8: Standard deviation of the individual project TSF

In the no merging case the service level in every merge is independent from the service level in the other merge. As merging gets higher the service levels in every merge turn into dependent random variables.

5. Summary and Conclusions

This paper analyzes the topic of partial cross training in call centers that should reach an overall service level. I manage the case where every project has a unique skill-set; this might not be ideal but is the case in different projects I examined. The present examination concentrates on merging in steady state operations and defines the improvement from partial merging. My current research is concentrated on finding the optimal cross training level; which is the level at which the profit of cross training is balanced by the extra cost of cross training. My current research is concentrated on developing algorithms to find the optimal cross training level with both steady state and non-stationary arrival rates

References

[1] Avramadis, A., A. Deslauriers, and P. L'Ecuyer 2004. Modeling daily arrivals to a telephone call cente Management Science 50(7):896–908.

Licensed Under Creative Commons Attribution CC BY

[2] Avramidis, A., W. Chan, and P. L'Ecuyer 2007. Staffing multi-skill call centers via search methods and a performance approximation.

[3] Brown, L., N. Gans, A. Mandelbaum, A. Sakov, S. Haipeng, S. Zeltyn and L. Zhao 2005. Statistical Analysis of a Telephone Call Center: A Queueing-Science Perspective. Journal of the American Statistical Association 100(469):36-50..

[4] Cezik, M, and P. L'Ecuyer 2007. Staffing multiskill call centers via linear programming and simulation. Working Paper [online] 34p

[5] Ekmekçiu, 2015. Optimizing a call center performance using queueing models – an Albanian Case. 5th International Conference - "Compliance of the Standards in South-Eastern European Countries with the Harmonized Standards of European Union", 15-16 June, 2015 Peja, Republic Of Kosovo

[6] Gans, N., G. Koole and A. Mandelbaum 2003. Telephone call centers: Tutorial, review, and research prospects. Manufacturing & Service Operations Management 5(2):79-141.

[7] Garnett, O., A. Mandelbaum and M. I. Reiman 2002. Designing a Call Center with impatient customers. Manufacturing & Service Operations Management 4(3):208-227.

[8] Harrison, J. and A. Zeevi 2005. A method for staffing large call centers based on stochastic fluid models. Manufacturing & Service Operations Management 7(1):20- 36..

[9] Mandelbaum, A. and S. Zeltyn. 2004. Service Engineering in Action: The Palm/Erlang-A Queue, with Applications to Call Centers.

[10] Mapo online, Call-center: Biznesi që vijon lulëzimin

[11] Wallace, R. B. and W. Whitt 2005. A Staffing Algorithm for Call Centers with Skill-Based Routing. Manufacturing & Service Operations Management 7(4):276- 294..

[12] Whitt, W. 2006. Sensitivity of Performance in the Erlang A Model to Changes in the Model Parameters. Operations Research 54(2):247-260.

Licensed Under Creative Commons Attribution CC BY