

“Why Can’t I Talk to the Person I Spoke with Last Time?”

We bet you’ve heard this question from a caller in your center. At Vanguard, we see this requirement more frequently as call centers adopt more sophisticated routing features. Almost every advanced call routing RFP that we’ve seen in the past twelve months had a requirement to connect callers with the last agent contacted. Last agent routing is possible through a combination of 1) caller input (e.g., entering a policy number on the touchtone pad) or network services (i.e., ANI) and 2) agent information in your CTI database. Automatically routing customers to the last agent they spoke with is a clear service differentiator.

This is the first of several articles on last agent routing. It addresses the key question of how often repeat callers actually reach the last agent and the associated service level for varying size centers and routing strategies. We’ll be analyzing additional features and combinations of size and routing strategies in the future.

In this article we analyze the impact of implementing last agent routing from an operational perspective using a modeling and simulation technique. We focus on how last agent routing performs for both callers who want to talk to the agent they spoke with last and the first time caller requiring the next available agent, and the effect on service levels in your center. One of the main findings of our analysis shows that different size centers perform quite differently when routing calls to the last agent. For this first article, we analyzed the performance of groups of 5, 20, and 50 simultaneously logged in agents. These groups represent a single agent queue that can take a mixture of general and repeat callers via an advanced ACD feature and/or in combination with a CTI routing engine.

Call Routing Strategies and Modeling Assumptions

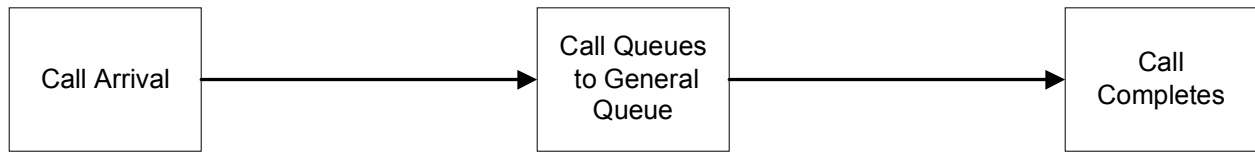
We looked at three different call routing strategies for three agent group sizes:

1. Route all calls to a general queue where they are answered in the order in which they arrived. This routing strategy was used to benchmark the overall service level associated with each agent group size.
2. Route new calls to the general queue and repeat calls to the last agent contacted without any overflow from specific agent routing to the general queue. Repeat callers have a higher priority than general queue callers.
3. Route new calls to the general queue and repeat calls to the last agent contacted with overflow to the general queue. In this routing strategy, calls that queue for a specific agent are overflowed to the general queue after a pre-set period of time (i.e., wait, 1, 10, 20 and 30 seconds for a specific agent). Repeat callers waiting for the last agent have a higher priority than general callers.

The following diagrams depict the three routing strategies:

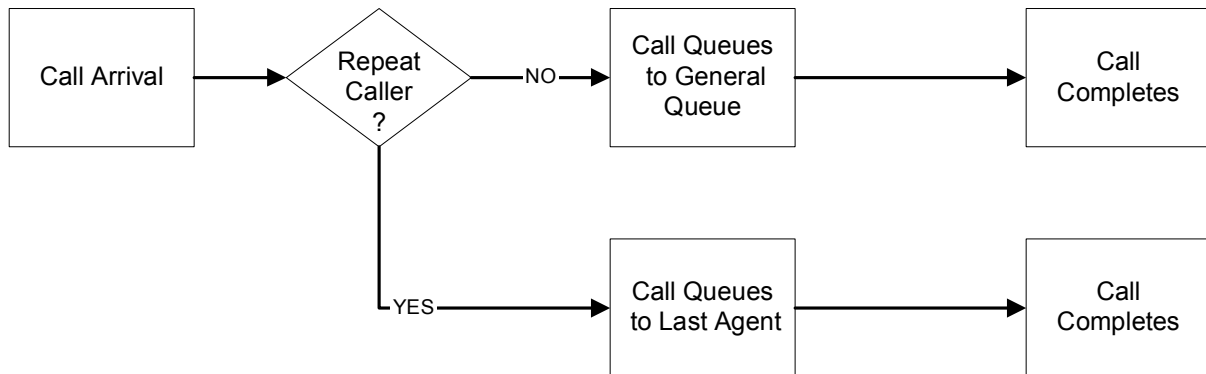
Routing Strategy 1: Route all calls to a general queue where they are answered in the order in which they arrived.

General Queue Only Call Routing Model



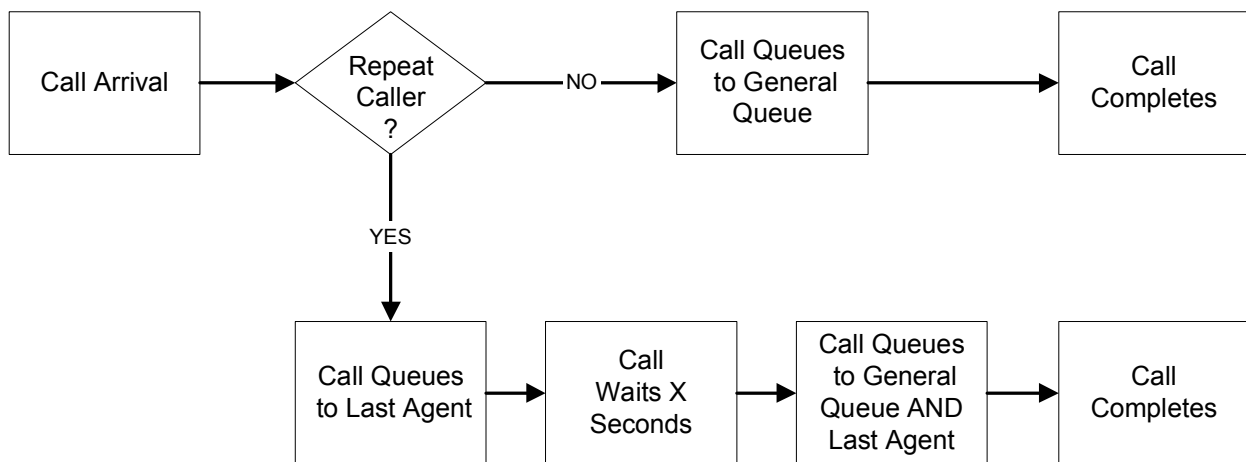
Routing Strategy 2: Route new calls to the general queue and repeat calls to the last agent contacted without any overflow.

General Queue and Last Agent Routing With No Overflow



Routing Strategy 3: Route new calls to the general queue and repeat calls to the last agent contacted with overflow to the general queue.

General Queue and Last Agent Routing With Timed Overflow



The following table outlines the assumptions and scenarios we used.

Assumptions:					
<ul style="list-style-type: none"> • We ran the models and associated scenarios with 5, 20 and 50 agent call centers. • In the baseline scenario (i.e., general queue routing only) we used call volumes that provided an ASA of under 20 seconds regardless of the call center size. • Average call handle times of three and six minutes were used in each scenario. • A repeat caller rate of 15% was applied to the total call volume (i.e., out of 100 calls, 15 are repeat callers wanting to speak to the last agent contacted). • All agents were members of the general queue. Agents in the general queue take calls based on availability. • Calls available to overflow to the general queue were simultaneously queued to individual agents. 					
Scenarios					
Model 1 (Baseline)	Model 2 (Last Agent Only)	Model 3A (Timed Overflow)	Model 3B (Timed Overflow)	Model 3C (Timed Overflow)	Model 3D (Timed Overflow)
One general queue	One general queue and last agent routing with no overflow. Repeat callers queued to the last agent have a higher priority than general callers.	One general queue and last agent routing. There is a one second queue time before overflow to the general queue.	One general queue and last agent routing. There is a ten second queue time before overflow to the general queue.	One general queue and last agent routing. There is a twenty second queue time before overflow to the general queue.	One general queue and last agent routing. There is a thirty second queue time before overflow to the general queue.

Analysis Results

First, the smaller the agent population, the greater the likelihood that a repeat caller will reach the last agent contacted due to the lower occupancy rate. As shown in the chart below (Figure 1), the likelihood of a repeat caller getting to speak with the last agent is significantly higher for a 5 agent group than for 20 agent and 50 agent groups; the likelihood of an agent group consisting of 5 agents is 30% higher than a larger group consisting of 50 agents regardless of the wait time before overflow. To meet the same service level, smaller call centers need to be staffed at a lower occupancy rate due to the random nature of call arrivals (i.e., agents have more time between calls than in the larger centers). This lower occupancy rate increases the likelihood that the 'right' agent will be available to take the repeat call compared to a larger call center where agents are much 'busier' (i.e., have shorter time periods between calls).

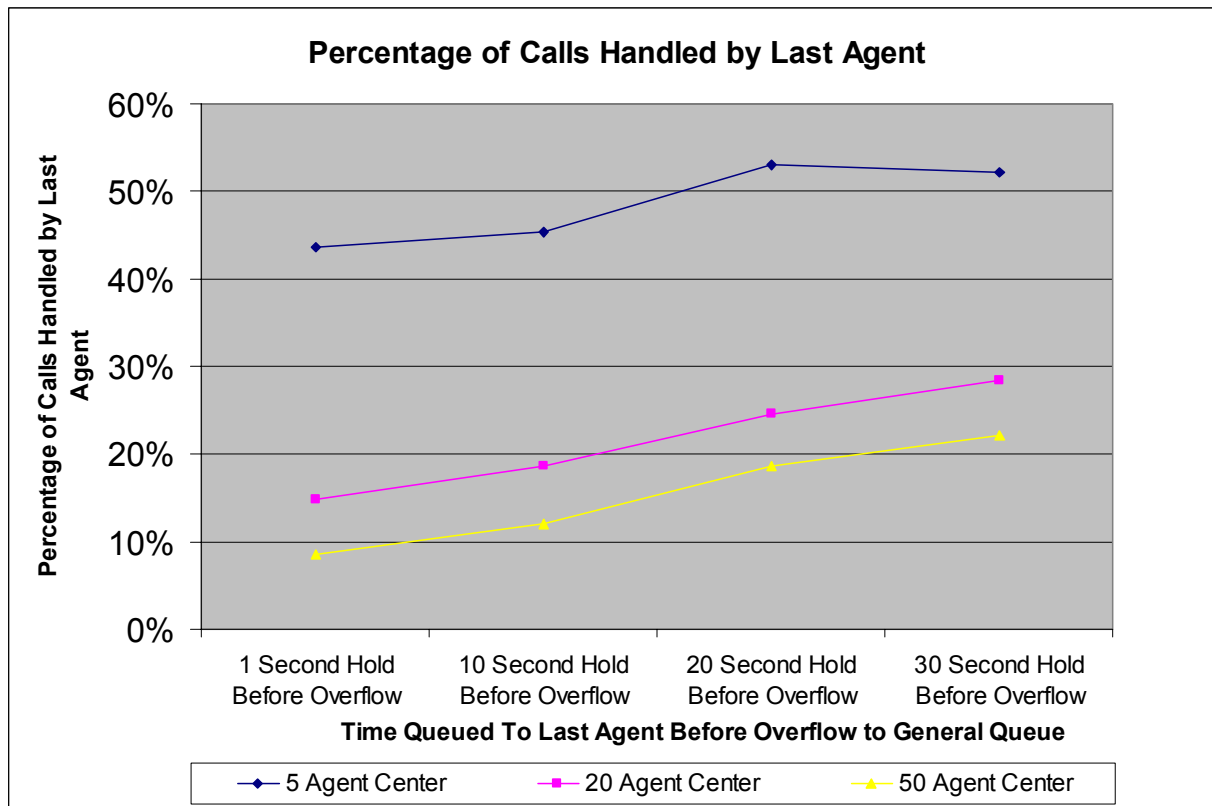


FIGURE 1

Next, we found that there is a law of diminishing returns when holding a repeat caller in queue for the 'last agent' contacted. Having an overflow timer set between one and ten seconds appears to provide the best balance between meeting service levels and increasing the likelihood of routing a call to a specific agent. The following table presents these results for each routing strategy and various size agent groups.

- Having callers wait until they are connected to the last agent contacted can increase the callers wait time from under 20 seconds to over 3 minutes, if no overflow is allowed.
- The best strategy is to overflow callers after a short duration to keep ASA to a reasonable level while significantly increasing the likelihood of connecting the caller with the last agent.
 - ◆ For the five agent group size, the 1 to 10 second overflow strategies connect repeat callers with the last agent over 44% of the time while keeping their ASA to less than 15 seconds.
 - ◆ The percentage of calls connected to the last agent increases by increasing the wait time before overflowing to the general queue. However this increased likelihood diminishes once we move beyond a 20 second overflow.

		Agent Group Size					
		5 Agent Group		20 Agent Group		50 agent Group	
Routing Strategy		Repeat Caller ASA (sec)	Percent to Last Agent	Repeat Caller ASA (sec)	Percent to Last Agent	Repeat Caller ASA (sec)	Percent to Last Agent
	No Overflow	125.9	100%	161.3	100%	198.3	100%
	1 sec Overflow	11.64	44%	24.03	15%	22.03	8%
	10 sec Overflow	14.52	45%	28.23	19%	28	12%
	20 sec Overflow	23.52	53%	29.07	25%	33.55	19%
	30 sec Overflow	31.44	52%	37.05	28%	48.98	22%

Finally, our analysis did not find any *significant* variation in performance when the average call-handling times were increased from 3 to 6 minutes. However, for the non-overflow scenario, the average repeat caller ASA increased significantly when the call handle times were increased from 3 to 6 minutes, regardless of the size of the agent group.

What's Next?

Now that we have the foundation models built, we are thinking about other scenarios that would help contact center professionals better understand the dynamics of caller routing in their centers. For instance, will routing to the last agent reduce call handle times? If so, what is the impact on service levels? Would reducing call handling times make last agent routing more attractive in the larger call centers? What if repeat callers are a much higher (or lower) percentage of the total call volume? What is the effect on the conclusions that we've drawn?

Finally, what are the challenges that you are facing in complex call routing and how do you measure increased customer satisfaction and or loyalty that results from last agent routing? Any suggestions on scenarios that you would like to see us to run? If so, drop us an email at dpowis@vanguard.net or areg@vanguard.net and let us know. If we see any common requests we can apply them to the models and publish the results.

Areg Gharakhanian and David Powis are consultants at Vanguard Communications Corporation a consulting firm that specializes in contact center processes, operations and technology. Areg and David lead projects to help clients improve operational and technological efficiency in contact centers. For more, information see <http://www.vanguard.net> or call 973-605-8000.