Nikolay Anisimov: *How to Keep Your Agents Really Busy*

Agent capacity model will help you to route several multichannel interactions to one agent

Guess how many customer interactions a contact center agent can handle simultaneously? One or ten? Can the agent work with several customers concurrently? If yes, how we should distribute interactions to such agents?

In this article we will answer all of these questions and introduce a logical mechanism known in the contact center industry as Agent Capacity Model, enabling us to solve all of these problems.

When one interaction is not enough

Normally, an agent, as a human being, can do only one thing at a time. That is true. This is also good if the agent handles a customer interaction, say, communicating with a customer by telephone. However, some interactions do not fully occupy the agent's capacity and leave some room for other activities.

For example, if an agent participates in one chat session with a customer, agent is doing nothing for some time, waiting for customer replies. Clearly, it would be nice to engage the agent in another useful activity to occupy these idle time intervals. Say, working a second chat session with another customer. It is known that well-trained agent could work with two or even more chat sessions without sacrificing the quality of service.

The problem is how can we distribute a second chat interaction to

an agent? Indeed, traditional methods of distribution of new interactions will consider the agent 'busy' because they already have an interaction.

Let us consider another important situation when an agent is handling an email from a customer. Suppose also that at some moment a voice call arrives from another customer and there are no available agents to route this call to. It is clear that it would be reasonable to distribute the call to the agent who is working with email, otherwise the call could be abandoned and the customer will be lost.

> *we need more sophisticated mechanisms of routing interactions and contacts that will take into consideration media types and agents' abilities*

Indeed, the nature of working with email is such that email can wait for a couple of minutes while a voice call cannot. In other words, a voice call is a real-time interaction while email is a deferred one. Again, conventional methods of call distribution will not help, as they consider the agent 'busy'.

Moreover, with the expansion of the *rich contact experience* paradigm, working with each customer becomes

more complex as the task may comprise exchanging multiple interactions of different media types in any order or even in parallel.

Therefore, we need more sophisticated mechanisms of routing interactions and contacts that will take into consideration media types and agents' abilities. These mechanisms should be flexible and customizable, allowing the contact center manager to tune-up routing rules in accordance with business requirements.

In the next section, we introduce a simple mechanism that allows to distribute multichannel interactions to agents who already are working with customer interactions and contacts.

Agent capacity and media interruption

Let us consider two basic notions of agent capacity and media interruption. Agent capacity is defined as a maximum number of interactions of some media type an agent can handle at once. For example, if a maximum number of chat sessions for some agent is equal to three, then the agent can handle three chats concurrently without degrading quality of service. If such an agent works with one chat session, the agent is not completely occupied and we could add two more chat interactions. For voice calls, it is

reasonable to set the number to one, prohibiting the distribution of a new call to an agent who already has one. Of course, these numbers are customizable and depend on nature of the business and the personal traits of agents.

But what do we do when we have interactions of different media types? In this case we could use the notion of media interruption. Within this concept, interaction with one media can or cannot interrupt interaction of another media type. For example, it is reasonable to allow interruption of working with email by receiving a voice call but not vice versa. In this case, the agent postpones working with the email and starts talking to a customer. When the voice call is wrapped up, the agent resumes

> *it is reasonable to allow interruption of working with email by receiving a voice call but not vice versa*

working with the email.

More strictly, the agent capacity model is defined with aid of two components: an interruption matrix and a maximum capacity vector. The interruption matrix is a matrix of size $N \times N$ where N is a number of interaction media types. Each element of the matrix x_{ij} could be equal to "y", "n", or "a" saying that interaction of i-th type could be interrupted by interaction of j-th type ("y") or not ("n"). Value "a" means that a new interaction could be added to an existing one and processed in parallel.

Consider for example an interruption matrix **INT** and a maximum capacity vector **MAX**:

$$INT = \begin{bmatrix} n & n & n \\ y & a & y \\ n & n & a \end{bmatrix}, MAX = \begin{bmatrix} 1 \\ 5 \\ 3 \end{bmatrix}$$

The matrix **INT** describes a relationship between three media types – "voice", "email", and "chat". The first row corresponds to voice type and says that voice interactions cannot be interrupted by any other interaction (including voice). The

second row corresponds to email type and specifies that email can be interrupted by voice and chat. Moreover, new email could be added to another email. The third row says that a chat session cannot be interrupted by voice and email. However, chat could be added to another chat and processed in parallel.

The maximum capacity vector **MAX** specifies maximum number of interactions of each type the agent can process simultaneously. This vector says that maximum number of voice calls is one, emails is five and chats is three.
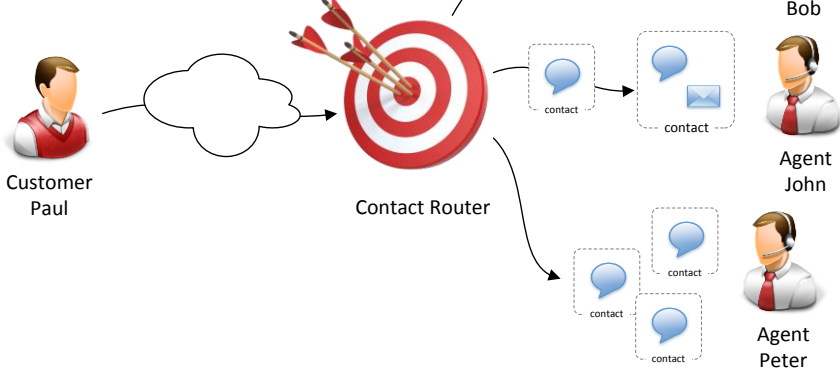
A current status of an agent could be described by a matrix **AS** of size $N \times M$ where M is a number of contacts the agent is working with. The next example says that an agent works with two contacts. The first contact contains only one voice call. The second contact contains one email and one chat session with the same customer.

$$AS = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

An agent capacity model contains logic in the form of procedure that decides whether the agent can accept a new contact or not. Suppose a new contact contains one email and one chat and defined as vector C:

$$C = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

To determine if the agent can accept the contact we make the following calculations:

Step 1: For each none-zero element of the vector C we check if this media type can interrupt or be added to agent's current interactions using the interception matrix and the maximum capacity vector. If for some interactions we find that it cannot interrupt existing agent interactions, then the agent is considered unavailable.

Step 2: If all new interactions with the contact can interrupt or be added to agent's interactions we then check if the new number interactions does not exceed maximum numbers from MAX vector. If that is true, the agent considered to be available for this contact. Otherwise, the agent is not available and cannot receive this contact.

For the above example we can easily determine that the agent is not available for a new contact C because his contacts a contain voice interaction that cannot be interrupted, neither by email nor by chat.

The next picture clarifies the routing of contacts in accordance with the agent capacity model. Suppose all three agents have the same interruption matrix and maximum capacity vector as introduced above. Customer Paul initiates a contact with an initial chat interaction. All agents are working with different customers' contacts.

Agent Bob works with a contact containing an active voice call and cannot be interrupted. Agent Peter works with three contacts, each in a chat session. Therefore, he cannot receive the fourth chat as it would exceed his maximum value. Agent John works with a contact containing email and chat. Clearly, he could receive additional contacts with chat. Therefore the contact router distributes the new contact to John.

∗ ∗ ∗

At Bright Pattern, we are advocating a new paradigm of *rich contact experience* and believe that contacts between customers and contact centers will become richer and more complex, dramatically improving the quality of customer service. To make this happen we introduce a new approach to arranging agents' work by distributing to them multiple cross-channel interactions, therefore improving their productivity and utilization.