\section{Experiments}\label{experiments}

In this section, we validate the correctness and tightness of our performance model by comparing with simulation results and other analytical methods. There is extensive research on the buffer sizing problem of the priority-aware NoC, representative methods include shaping delay analysis \cite{Manolache:2006:BSO:1131481.1131683}, FLBA and LLBA \cite{189}. We will only perform the comparison with LLBA in subsection 5.1 since it gives lower buffer space bound than FLBA and shaping delay analysis. There are also several analytical methods exist for the delay analysis of priority-aware NoC, examples include lumped link model \cite{707545}, dependency graph model \cite{708526}, contention tree model \cite{LuJS05}, Rate Monotonic (RM) method \cite{365629}, FLA \cite{Shi:2008:RCA:1397757.1397996}, LLA \cite{73} and DNC \cite{Qian489900}, etc. Among all these methods, LLA and DNC outperform the others when the tightness of delay bound is considered. Thus, we will only compare our delay analysis algorithms with LLA and DNC, as presented in subsection 5.2. We also present the simulation results to validate the correctness and tightness of our method in subsection \ref{sim}.

Throughout this section, we compare our analytical results with those obtained by other analytical methods and simulation under a $4\times 4$ mesh network. The applications we choose include the synthesis traffic pattern shown in Fig. \ref{topology} and two real-world applications, i.e. autonomous vehicle \cite{Shi2009} and errison radio \cite{Jafari1922089,LuJa08} applications. The original autonomous vehicle application consists of 38 flows mapped onto a $4\times 4$ mesh network \cite{Shi2009}, with each flow a distinct priority. We employ the same task mapping and priority assignment strategy as \cite{Shi2009}. The flow ID ($f\_i$), priority ($P\_i$), source address (Src), destination address (Dst), packet length ($F\_i$, in flits) and injection period ($I\_i$, in $10^6$ cycles) of each flow in this application are presented in Table \ref{vehicle}. The errison application is comprised of 16 IP cores and 26 communication flows. These flows are classified into nine groups with each group has different bandwidth requirement, as shown in Fig. \ref{trafficpattern}.

\begin{table}[htbp]

\centering

\caption{\label{vehicle}Flow specification of autonomous vehicle benchmark}

\begin{tabular}{|c|c|c|c|c|c||c|c|c|c|c|c|c|c|}

\hline

$f\_i$ & $P\_i$ & Src &Dst & $F\_i$ & $I\_i$ & $f\_i$ & $P\_i$ & Src &Dst & $F\_i$ & $I\_i$\\

\hline

1 & 24 & 6 & 16 & 512 & 10 & 2 & 1 & 9 & 2 & 38400 & 4\\

\hline

3 & 2 & 16 & 15 & 38400 & 4 & 4 & 3 & 2 & 6 & 512 & 4\\

\hline

5 & 4 & 15 & 6 & 512 & 4 & 6 & 5 & 1 & 2 & 38400 & 4\\

\hline

7 & 6 & 5 & 6 & 38400 & 4 & 8 & 7 & 9 & 10 & 38400 & 4\\

\hline

9 & 8 & 13 & 14 & 38400 & 4 & 10 & 9 & 4 & 7 & 38400 & 4\\

\hline

11 & 10 & 8 & 7 & 38400 & 4 & 12 & 11 & 12 & 11 & 38400 & 4\\

\hline

13 & 12 & 16 & 15 & 38400 & 4 & 14 & 13 & 2 & 7 & 2048 & 4\\

\hline

15 & 14 & 6 & 7 & 2048 & 4 & 16 & 15 & 10 & 7 & 2048 & 4\\

\hline

17 & 16 & 14 & 7 & 2048 & 4 & 18 & 17 & 3 & 10 & 2048 & 4\\

\hline

19 & 18 & 7 & 10 & 2048 & 4 & 20 & 19 & 11 & 12 & 2048 & 4\\

\hline

21 & 20 & 15 & 10 & 2048 & 4 & 22 & 21 & 7 & 11 & 8192 & 4\\

\hline

23 & 22 & 10 & 11 & 8192 & 4 & 24 & 23 & 11 & 5 & 4096 & 4\\

\hline

\end{tabular}

\end{table}

\begin{figure}

\centering

% Requires \usepackage{graphicx}

\includegraphics[scale=0.7]{fig9.pdf}\\

\caption{Traffic pattern of ericsson radio system application}\label{av}

\end{figure}

\begin{figure}

\centering

% Requires \usepackage{graphicx}

\includegraphics[scale=0.7]{fig10.pdf}\\

\caption{Traffic pattern of ericsson radio system application}\label{trafficpattern}

\end{figure}

\subsection{Buffer Space Comparison}\label{llacmp}

There are four flows (i.e. $f\_1$, $f\_2$, $f\_3$ and $f\_4$) in the network shown in Fig. \ref{topology}, with different priorities $P\_4>P\_1>P\_2>P\_3$ ($f\_4$ has the highest priority). We perform the comparison on a set of periodical traffic due to the restriction of LLBA method \cite{189}. The packet length (in flits) and injection period (in cycles) of flow $f\_i$ ($i=1,2,3,4$) are denoted as $F\_i$ and $I\_i$, respectively. To ease the analysis of buffer requirement, the LLBA method assumes the number of bits in a flit is the same as the physical channel width, and the latency of a router is one cycle \cite{189}. Thus, our performance model for the standard wormhole-switched router should be specialized, which is achieved by letting the service curve of BW, RC, VA, SA and LT stage be a pair of burst delay function $<\delta\_0(t),\delta\_0(t)>$. Under this condition, the service curve of the entire router is equal to the service curve provided by the ST stage, which is $<\beta\_{ST,R\_i^p}^l,\beta\_{ST,R\_i^p}^u>$. The traffic jitter for all the flows are assumed to be zero for brevity and clarity. In addition, we set the credit feedback delay $\sigma=0$ cycle in our model since the LLBA method does not consider the influence of credit feedback delay on the buffer requirement.

Let us suppose that all the flows have the same packet injection period $I\_i=50$ cycles and packet length $F\_i=8$ flits ($i=1,2,3,4$). The delay bounds computed with the LLA method \cite{73} for these four flows are 21, 22, 21 and 13 cycles, respectively. We can also obtain the total buffer size reserved for the four flows (i.e. $f\_1$, $f\_2$, $f\_3$ and $f\_4$) with LLBA method \cite{189}, which are 11, 12, 18 and 4 flits, respectively. The total buffer size required by these four flows can be obtained by summing up the buffer size reserved for each flow along its path, which is 45 flits. This value is the minimum buffer space required for all these four flows to avoid the back-pressure caused by flow control. By allowing the flow control to be triggered, our buffer sizing algorithm can be utilized to reduce this buffer size further as long as the deadline constraint is not violated. Figure \ref{LLBAvsRTC} demonstrates the normalized buffer requirement calculated with LLBA and our method under different deadline constraint. It can be found that our method can give much tighter buffer estimation than LLBA, and the improvement of our method over LLBA becomes more and more significant as the deadline constraint is relaxed.

\begin{figure}

\centering

% Requires \usepackage{graphicx}

\includegraphics[scale=0.55]{figures/bufopt.pdf}\\

\caption{Buffer requirement computed with our method for different deadline}\label{LLBAvsRTC}

\end{figure}

\subsection{Delay Bound Comparison}\label{dnccmp}

In this subsection, we present the numerical results to demonstrate the improvement of our method over the DNC method proposed in \cite{Qian489900}. The traffic pattern discussed in this subsection is shown in Fig. \ref{topology}. The priorities of these four flows (i.e. $f\_1$, $f\_2$, $f\_3$ and $f\_4$) in the network satisfy $P\_4>P\_1>P\_2>P\_3$. We perform the comparison on a set of periodical traffic. The packet length (in flits) and injection period (in cycles) of flow $f\_i$ ($i=1,2,3,4$) are denoted as $F\_i$ and $I\_i$, respectively. The router architecture we considered in this subsection adopts the lookahead pipeline \cite{jerger2009chip}, which removes the RC stage from the critical path of the router pipeline. Thus, our service model is customized by letting the service curve of RC stage to be $<\delta\_0(t),\delta\_0(t)>$. The RTC arrival curve $<\alpha\_{f\_i}^l,\alpha\_{f\_i}^u>$ of flow $f\_i$ can be obtained according to the method introduced in subsection \ref{traffic}. The DNC arrival curve for $f\_i$ is $\alpha\_{f\_i}=V\_i t+F\_i$, where $V\_i=F\_i/I\_i$ represents the average arrival rate.

We assume the VC buffer size of each router is $32$ flits, and the credit feedback delay $\sigma=0$ cycle. We change the injection rate $V\_i$ ($i=1,2,3,4$) from $1/3$ to $1/6$ (flits/cycle) and the packet length $F\_i$ from $1$ to $8$ flits. The end-to-end delay of flow $f\_3$ calculated with the DNC method and our method are presented in Fig. \ref{comparison}. By comparison, we find that our method can give a much tighter delay bound than the DNC-based method proposed in \cite{Qian489900}. The root cause for this improvement lies in the fact that our method utilizes the upper service curve to limit the output upper arrival curve, which further leads to a tighter leftover service curve for the low-priority flows.

\begin{figure}

\centering

% Requires \usepackage{graphicx}

\includegraphics[scale=0.8]{figures/rtcvsdnc.pdf}\\

\caption{Delay bound comparison with network calculus under different injection rate and packet length}\label{comparison}

\end{figure}

\subsection{Comparison with Simulation}\label{sim}

Our performance model is also verified by simulation. We modified the Booksim 2.0 \cite{6557149}, a cycle-accurate NoC simulator, to support the specified traffic pattern and injection process. The traffic patterns investigated in this subsection include the example shown in Fig. \ref{topology} and an real application provided by Ericsson Radio Systems \cite{LuJa08}\cite{Jafari1922089}. We adopt the optimized lookahead router \cite{jerger2009chip} to construct the mesh network. To fit this optimization, our service model is customized by letting the service curve of RC stage be $<\delta\_0(t),\delta\_0(t)>$. Other architecture and simulation parameters used in the simulation are listed in the Table \ref{arcpara}.

\begin{table}[htbp]

\centering

\caption{\label{arcpara}Architecture parameters used in the simulation}

\begin{tabular}{|c|c||c|c|}

\hline

network topology & $4\times 4$ mesh & routing algorithm & X-Y routing\\

\hline

credit delay & 0 cycle & channel width & 128 bits\\

\hline

buffer size & 32 flits & switch allocator & priority-based\\

\hline

link latency & 1 cycle & sampling period & $1\times 10^5$ cycles\\

\hline

clock cycle & 1 ns & warmup period & $3\times 10^5$ cycles\\

\hline

\end{tabular}

\end{table}

We first investigate the traffic pattern shown in Fig. \ref{topology}. Suppose that the injection process of each flow is the strictly periodical traffic without any jitter, and the priorities of these four flows satisfy $P\_4>P\_1>P\_2>P\_3$. However, even under this assumption, determining the worst-case delay by simulation for general scenarios is still a non-trivial task, because the worst-case delay of each flow depends on the traffic pattern and offset (i.e. the injection time of the first packet in a flow $f\_i$, denoted as $O\_i$) of all the flows in the network.

We set the injection rate $V\_i$ ($i=1,2,3,4$) to $1/6$ (flits/cycle), and change the packet length $F\_i$ from 2 flits to 9 flits. Under these conditions, we find that $f\_1$ will experience the maximum delay in the network if its offset $O\_1$ is equal to that of $f\_4$. Similarly, the worst-case interference of $f\_2$ occurs when its offset $O\_2=O\_1+(4+1)$, where $O\_1+(4+1)$ is the time instance that the first packet of $f\_1$ leaves router $R\_{16}$, since the basic latency of a lookahead router and physical channel are 4 cycles and 1 cycles, respectively. Flow $f\_3$ experiences the worst-case delay when $O\_3=O\_2+F\_1+2\times(4+1)$. Thus, we set $O\_1=O\_4=0$ cycle, $O\_2=5$ cycles, $O\_3=15+F\_1$ cycles, and run the simulation. The collected maximum end-to-end delay of these four flows under the given offset combination are compared with our RTC method, as shown in Fig. \ref{rtcvssim}. As indicated in this figure, for the given configurations, the delay bounds calculated with our method are indeed the upper bounds of simulation results, which verifies the correctness of our method.

\begin{figure}

\centering

% Requires \usepackage{graphicx}

\includegraphics[scale=0.5]{figures/rtcvssim.pdf}\\

\caption{Delay comparison with simulation under different packet length}\label{rtcvssim}

\end{figure}

For the errison application discussed in \cite{LuJa08}\cite{Jafari1922089}, we set the packet size to 128 bits, and collect the maximum end-to-end delay of each flow obtained by simulation. The comparison results between one simulation run and the delay bound calculated with our method are shown in Fig. \ref{ericsson}. The offset of each flow in this simulation run is zero cycle. We can see that the calculated delay bounds constrain the simulation results well, which verifies the correctness of our method. This comparison also demonstrates the ability of our method to analyze the real system with large number of flows.

\begin{figure\*}

\centering

% Requires \usepackage{graphicx}

\includegraphics[scale=0.45]{figures/ericsson.pdf}\\

\caption{Delay comparison with simulation}\label{ericsson}

\end{figure\*}