Modification to Appendix A of Performance Baseline Text on MIMO Latency

Gojko Babic, Raj Jain, Arjan Durresi, Justin Dolske

Raj Jain is now at
Washington University in Saint Louis
Jain@cse.wustl.edu
http://www.cse.wustl.edu/~jain/
Overview

- Motivation: Why MIMO?
- Zero Delay Switch
- MIMO Latency Definition
- Examples
- MIMO Measurements
- User Perceived Delay
Single Bit Latency

- **Non-Zero Delay Switch:**
  
  \[ \text{bit in} \rightarrow \text{Latency} \rightarrow \text{bit out} \]
  
  \[ \text{time} \quad t_{in} \quad t_{out} \]

- **Zero Delay Switch:**
  
  \[ \text{bit in} \quad \text{bit out} \]
  
  \[ \text{time} \quad t_{in} \quad t_{out} \]
Usual Frame Latencies Metrics

First bit of the frame enters

Frame input time

Last bit of the frame enters

LILO

FIFO

FILO

First bit exits

Frame output time

Last bit of the frame exits
In cut-through switches LIFO latency can be negative.
FIFO latency may be small but the later cells may delayed considerably.
Zero-Delay Switch I

- Input Rate = Output Rate
- A fiber of length $k$ km has a latency of $5k$ $\mu$s.
- Fiber of zero length has zero latency.

First bit of cell enters to First bit of cell exits

Cell input time

Last bit of cell enters to Last bit of cell exits

Cell output time
Zero-Delay Switch II

- Input Link Rate > Output Link Rate
- Need to buffer input.

First bit enters $\rightarrow$ Last bit enters $\rightarrow$ Last bit exits

Cell input time

First bit enters $\rightarrow$ First bit exits

Cell output time
Zero-Delay Switch III

- Input Link Rate < Output Link Rate

- The zero-delay switch is intelligent to avoid underruns
MIMO Latency Definition

- MIMO Latency = \textit{FILO Latency} - \textit{NFOT}

- \textit{FILO latency} = Time between the first bit entry and the last bit exit

- \textit{NFOT} = Nominal Frame Output Time: the time a frame needs to pass through the zero-delay switch, calculated as:
  Initially \textit{NFOT} = 0 and time \( t \) is measured from the arrival of the first bit of the first cell. For each cell with its first bit arriving at time \( t \)
  \( \Rightarrow \textit{NFOT} = \max\{t, \textit{NFOT}\} + CT. \)

- \( CT = \text{Max}\{\text{cell input, cell output time}\} \)
Example 1

- Input rate > Output rate
- $CT = \text{Cell Output Time} = 4$
- 2nd cell at 5: $\text{NFOT} = \max\{5, 4\} + 4 = 9$

First bit of cell arrives → First bit of cell transmitted

First bit of cell arrives → First bit of cell transmitted
Example 2

- Input rate > Output rate
- CT = Max{1, 4} = 4
- 2nd Cell arrival at 2: NFOT = max {2, 4} + 4 = 8
Input rate < Output rate

First bit of cell arrives

First bit of cell transmitted

First bit of cell arrives

First bit of cell transmitted
Equivalent MIMO Latency Definition

- When Input Link Rate \( \leq \) Output Link Rate:
  - CIT \( \geq \) COT
  - NFOT = Frame Input Time
  - MIMO = \( FILO \) Latency - NFOT
    = \( FILO \) Latency - Frame Input Time
    = LILO

\[
MIMO = \begin{cases} 
  \text{FILO - NFOT, if Input rate} > \text{Output rate} \\
  \text{LILO, otherwise}
\end{cases}
\]
Practical MIMO Measurements

- Contemporary ATM Monitors provide measurements data at the cell level:
  - Cell Transfer Delay (CTD)
  - Cell Inter-Arrival Time

- From the next slide:
  \[ \text{FILO} = \text{First Cell Transfer Delay} + \]
  \[ + \text{First Cell to Last Cell inter-arrival time} \]

- Then, calculate NFOT and obtain MIMO as:
  \[ \text{MIMO} = \text{FILO} - \text{NFOT} \]


Practical MIMO Measurements (Cont.)

\[ \text{FILO} = a + b \]

- \(a\) = First Cell Transfer Delay
- \(b\) = First Cell to Last Cell inter-arrival time
Practical MIMO Measurements (Cont.)

- If Input Link Rate ≤ Output Link Rate:
  - MIMO = LILO = Last cell’s Transfer Delay - CIT

![Diagram showing MIMO and CIT relationships]
User Perceived Delay

- The user starts waiting as soon as the first bit starts entering the system until the last bit exits the network.
- So, user perceived performance is reflected by FILO Latency.
- MIMO latency measures only the switch’s contribution to the delay.
User Perceived Delay (Cont)

- a. Input Link Rate is equal to Output Link Rate
- b. Input Link Rate is greater than Output Link Rate
**Summary**

- Usual frame latencies are not appropriate for ATM systems.
- User perceive FILO latency as network delay.
- MIMO measures the switch component of FILO.
- MIMO Latency can easily measured using contemporary ATM monitors.
Motion

- Adopt the text under heading “Modifications to Appendix A of Performance Testing Baseline Text on MIMO latency” of 97-0834 to replace Appendix A of Performance Testing Baseline Text.