Rule WLM660: Synchronous service time was high for the indicated structure

Finding: The synchronous service time for the indicated structure exceeded the guidance provided to CPExpert.

Impact: This finding can have a LOW IMPACT, MEDIUM IMPACT, or HIGH IMPACT on the signalling performance of the sysplex. The level of impact depends on the amount of delay to synchronous requests and how important the requests are.

Logic flow: This a basic finding. There are no predecessor rules.

Discussion: Signalling requests to a coupling facility can occur only if a subchannel to the coupling facility is available. If no subchannel is available, the cross-system extended services (XES) will either enter a CPU "spin loop" waiting for a subchannel to become available or queue the request until a subchannel is available. The type of action taken by XES depends on whether the request was specified as synchronous or asynchronous.

- Synchronous requests require that a response be received from the coupling facility before the requesting application continues execution. Synchronous requests would be used, for example, to request a lock. In this example, the application cannot proceed until the lock is granted.

  For synchronous requests, XES will either (1) satisfy the request if a subchannel is available, (2) enter CPU "spin-looping" until a subchannel is available and the request is satisfied, or (3) convert the synchronous request to an asynchronous request if the type of request permits the conversion.

- Asynchronous requests allow the requesting application to continue processing and be notified when the request is completed. For asynchronous requests, XES either starts or queues the request and returns control to the application issuing the request.

The type (synchronous or asynchronous) of request that is issued generally depends on the type of structure.

- Some requests can be satisfied only by synchronous requests (for example, signals generated by XES itself will always be synchronous and will not be converted to asynchronous requests).

- Some requests can be issued as either synchronous or asynchronous requests, depending on the application's use of the structure (for example, cache structure requests can be issued as either synchronous or asynchronous).

- Some requests are issued as asynchronous requests (for example, JES2 requests to the JES2 checkpoint will be issued as asynchronous requests).

- Some requests can be issued as synchronous but will be converted to asynchronous if the subchannels are busy[1] unless the application has indicated that the synchronous cannot be converted.

The time spent waiting for subchannels to become free for synchronous requests not only delays the request (and consequently delays the application waiting on the request), but wastes processor resources since the processor is in a CPU "spin-loop" waiting for the synchronous request to be satisfied.

The service time represents the time from when MVS issues a command for the coupling facility until the return from the command is recognized by MVS. The time includes time spent on the coupling facility links, the coupling facility processing time, and the time for MVS to recognize that the command was completed. The service time varies based on whether subchannels are available, the activity level of the coupling facility itself, and on the amount of data being processed.

IBM suggests that the service time for synchronous requests should be less than 250-350 microseconds, depending upon the length of the request. The service time for lock structures should be less than 250 microseconds, since lock structure requests are small.

CPExpert compares the synchronous service time (R744SSTM) against the **SYNCSRV** variable in USOURCE(WLMGUIDE). CPExpert produces Rule WLM660 when the synchronous service time is greater than the SYNCSRV guidance variable.

The default value for the SYNCSRV variable is 350, indicating that CPExpert should produce Rule WLM660 when synchronous service time is more than 350 microseconds. CPExpert subtracts 100 microseconds from the SYNCSRV guidance variable if evaluating a lock structure. Thus, CPExpert will produce Rule WLM660 when the service time for lock structures is greater than 250 microseconds.

[1]The application can specify which requests must be satisfied as synchronous and which can be converted to asynchronous. XES will automatically convert requests from synchronous to asynchronous if all signalling paths are busy, unless the application specifies that the conversion is not to be done.

The following example illustrates the output from Rule WLM660:

```
RULE WLM660: SERVICE TIME WAS HIGH FOR SYNCHRONOUS REQUESTS

   ISGLOCK:  The service time for this structure has exceeded the
   guidelines for synchronous requests.  Service time is accumulated from
   the time MVS issues a command for the coupling facility until the return
   from the command is recognized by MVS.  Service time is recorded for
   each structure used by each system.  You can alter the times used by
   CPExpert in making this finding by altering the SYNCSRV guidance
   variables in USOURCE(MVSGUIDE) if you are unable to make changes to
   reduce service time for the structure.

                                  TOTAL SYNC      AVERAGE SERVICE
   MEASUREMENT INTERVAL            REQUESTS     TIME (MICROSECONDS)
   10:45-11:00,06MAR1997             2,052              9,104
```

Suggestion: CPExpert suggests that you consider the following alternatives if Rule WLM660 is produced:

- Synchronous command processing is performed by the CP. You should make certain that sufficient CPU resources have been allocated to the coupling facility LPAR.

- Examine whether the structure activity is balanced between coupling facilities. You may wish to consider redistributing the structures among the coupling facilities if a significant imbalance exists.

- You should consider whether additional coupling facility links should be added between the MVS processor the coupling facility. Each coupling facility link will contribute two subchannels.

- If possible, you should consider influencing the exploiters of the coupling facilities to lower the activity rate to the coupling facilities. Taking other tuning actions (especially if indicated by other rules produced by CPExpert) may reduce the number of XCF signals. For example, signal activity can be lowered by (1) reducing lock contention, (2) reducing false lock contention, or (3) tuning the XCF to eliminate signals related to the expansion of a transport class size.

- If none of the above alternatives are appealing, you may wish to change the guidance to CPExpert by altering the **SYNCSRV** guidance variable in USOURCE(WLMGUIDE).

  IBM provides the following example service times based on measurements of CF lock requests for various combinations of sender CPCs and CFs. These measurements were reported in *S/390 MVS*

Parallel Sysplex Configuration, Volume 2: Cookbook, document Number SG24-2076-00.

Central Processor	Coupling facility	Service Time (microseconds)
9672R1 based	9674C01	250
9672R2/R3 based	9674C02/3	180
9672G3	9674C04	140
9672G4	9674C04	100
9672G4	9674C05	70
9021 711 based	9674C01	160
9021 711 based	9674C02/3	130
9021 711 based	9674C04	100
9021 711 based	9021 711 based	80

Reference: Washington System Center Flash 9609 ("CF Reporting Enhancements to RMF 5.1")

S/390 MVS Parallel Sysplex Configuration, Volume 2: Cookbook, document Number SG24-2076-00.

"Parallel Sysplex Performance: tuning tips and techniques," Kelley, Joan (IBM, Poughkeepsie, NY), SHARE 86, February 1996.

