MVS SMF: SE2 and Onwards

Authors: John M. McCann, Morino Associates Inc.
David R. Boast, Morino Associates Inc.

Abstract

One of the more popular uses of SAS in recent years has been the processing of SMF based information for performance management and capacity planning. However, the format and content of SMF data has been continually changing. This paper attempts to identify those changes that have occurred from MVS/SE2 through MVS/SP1.3 with respect to processing considerations and information interpretation.

Introduction

The content of MVS SMF data has been continually changing with various releases of MVS, associated Program Products, and APARs and PTFs. This paper attempts to identify those changes for the people who are writing programs to process the SMF records. Even though a program may be written for the particular level of SMF that is currently being produced in an installation, a common requirement of long term planning is the processing of extensive amounts of historical information, usually data that was produced under previous levels of SMF. It is important to be aware of the historical changes that have occurred.

This paper is divided into four sections. The first section deals with the quite extensive changes introduced with MVS/SE2. Next, the considerations introduced by various Program Products will be discussed. The third section covers the MVS/SP1 modifications. The final topic is the authors' concerns over future changes that should be introduced.

While RMF is a separate Program Product, the various releases of RMF will be discussed with the release of MVS on which they run.

MVS/SE2

The SE2 version of MVS represents the second major architectural change to SMF since its introduction in OS release 18. (The first was the collecting of dataset information: record types 14, 15, 17, 18). SE2 introduced four new SMF record types: 23, 30, 32, and 90; it provided for more data capture; new reporting options were made available; and a new recording architecture was introduced. The new recording technique will not be dealt with in this paper. Many of the enhancements introduced in SE2 SMF were the direct result of SHARE and GUIDE resolutions.

MVS/SE2: New Record Types

The type 23 record, SMF Status, was introduced to record information concerning the new SMF architecture, i.e. buffer and record
activity, buffer high water mark, and recording suspension. The analysis of this record can identify problems with the number of buffers allocated for SMF recording.

The type 90 record, System Status, records operator command activity that affects the data being recorded by SMF. SET commands, information from IPL, and SMF switch cause various subtypes to be recorded. Besides providing an audit trail of the changes in an MVS system, the type 90 record has two subtypes that can be extremely important in the analysis of an environment: SET DATE/TIME and IPL PROMPT. When a SET DATE command is issued, the hardware time-of-day clock is NOT changed. Only the offset to local time (as recorded in the CVT field, CVTTZ) is changed. Problems then arise in analysing the data recorded after the SET command was issued. For data that is time-stamped via a STCK, (the hardware clock), such as CTF, CICS/CMF, LOGREC, etc., what is the relationship between the time in these records and the time in other records that recorded via the TIME SVC (uses CVTTZ offset)? For the RRF interval that was being recorded during the SET command, what is the duration of the interval? What time period does it actually represent?

The IPL PROMPT record can provide valuable information about system IPL's. The IPL reason and operator's name are recorded as well as the estimated down time as entered by the operator at IPL. The reason and name allow room for more detailed recording than the limited information recorded in LOGREC. However, the PROMPT subtype is only recorded if PROMPT(IPLR) is specified in the SMFPRMxx member in SYS1.PARMLIB. PROMPT should not be turned on without first establishing operator procedures for an installation, e.g., the first four characters of the IPL reason will be a coded number for long term tracking of IPL reasons. The use of the PROMPT facility can also affect the time it takes to have a system operational. The system IPL is halted while waiting for the operator reply to the PROMPT message. If the operator is unfamiliar with the proper format for the reply, or must spend considerable time calculating down time and looking up the reason codes, the availability of online systems could be adversely affected.

The type 30 record, Common Address Space Work, was the most significant change in SMF introduced by SE2. The new record provides for consolidated data collection by replacing the following record types:

- Type 4 - Step termination
- Type 5 - Job termination
- Type 20 - Job initiation
- Type 34 - TSO "step" termination
- Type 35 - TSO "job" termination
- Type 40 - Dynamic Allocation
There are five subtypes (see also SP13 below) of the type 30 record.
- Subtype 1 - Address space initiation - replaces type 20
- Subtype 2 - Interval step record
- Subtype 3 - Final interval record
- Subtype 4 - Step termination - replaces types 4, 34
- Subtype 5 - Job termination - replaces types 5, 35

The type 30 record follows the new section structure introduced with SE2. After the standard date, time, system and subsystem id at the beginning of the record, are a number of "section" fields. Each section group describes the offset to a section, the number of occurrences of the section, and the length of each section. This structure improves the maintainability of code that is written to process the records. If IBM and the user are consistent in following the structure, both will be more immune to changes in the record.

Perhaps the most significant change in the type 30 record was the capturing of the activity for dynamically allocated datasets. This was previously only available from the type 40 records. Many installations have stopped recording the type 40s because of the volume that are produced, particularly in a heavy TSO environment. Without the type 40 dynamic allocation information, only a small fraction of the EXCP activity for TSO was being captured for billing, capacity planning, and analytical modeling. Besides all the information in the records that it replaces, the type 30 records additional information (see below).

The final record introduced by SE2 is the type 32, TSO User Work Accounting. It will record by TSO command, the number of times invoked and the total resources consumed by all the invocations. To record this information, the TSO Command Package (5740-XT6) or it's successor, TSO Extensions (5665-285) is required. There are a number of problems introduced by the use of the type 32 record. They are described under SE2 Problems below.

SE2: Additional Data Capture

As described above, the types 23 and 90 records are completely new information for MVS. The type 32 record basically replaces the PCF command recording feature, but doesn't provide as much information. The type 30 record replaces other records, but also contains additional information that was not available in those records:

- The device segment of the type 4 and 40 records has been expanded in the type 30 record to include the DDNAME and largest blksize for the segment. While this enables the calculation of bytes transferred for fixed length records, what meaning does this calculation have for variable or undefined records, or instances where the blocksize is zero like many of the utility work datasets?
MVS SMF: SE2 and Onwards

- The type 30 record also captures specific and non-specific mount information for tape, dasd, and mss. This can eliminate much of the processing that was done against the type 21 EOV records or the type 14 and 15 dataset activity records.

- The type 30 record also records the TCB and SRB time spent under the initiator.

Besides the new record types, and the additional data recorded in them, the SE2 SMF provided for additional data capture in already existing records:

- Additional I/O activity is recorded for program fetch, JES2 spool activity, catalog management, and OPEN/CLOSE I/O activity. Installations converting from a base MVS 3.8 system to MVS SE2 or any of the MVS SP systems will record an increase in I/O activity. If the installation is billing for this activity to it's users, the users invoice will reflect an increased cost. From the users standpoint, this new system just costs them more.

- The ability to record information on started tasks and jobs with a TIME=1440 specification also enhances the data available with SE2.

- The type 14 and 15 data set activity records now record open time.

- RMF has a new subtype for the type 79 from RMF monitor II sessions; transaction activity is recorded in the subtype 8.

- There are additional fields that can be traced in the type 76, RMF Trace Record.

- The RMF type 72 record, Workload Activity, was expanded by 50 bytes to accomodate new fields for report performance groups.

SE2: New Recording Options

Two new recording options were introduced with SE2 and the RMF version that accompanied it: interval recording and report performance groups.

Interval recording is provided in the type 30 and 32 records. In the type 30 record, this is implemented via the subtype 2, interval record, and the subtype 3, final interval record. Each record identifies the address space and contains the resource consumption information that occurred during the interval, i.e. "delta" measurements. This can provide significantly improved accuracy of information for long running tasks. Data that would previously be lost at a system crash is now recorded up to the last interval. For TSO systems where the user logs on once in the morning and stays...
logged on the complete day, a system crash represented significant loss of revenue. Interval accounting also improves the accuracy of apportioning activity over the day. By previous methods, it was necessary to average the resource consumption of an eight hour TSO session over the hours it represented. TSO users, however, are not consistent users of the resources of the system. They have periods of different intensity. Interval recording can provide a better picture of the periods of high and low activity.

Report performance groups are implemented in the RMF type 72 record, Workload Activity. Address spaces, such as TSO users, may be in a common control performance group for SRM control purposes, but may be in separate report performance groups for identifying resource consumption activity by user group. Another common use of report performance groups is identification of major subsystems such as JES2, V TAM, and TCAM that might be in the same control performance group because the SRM actually exerts little or no control over them. The use of report performance groups can reduce the overhead of the SRM in maintaining multiple control performance groups. However, with new controls introduced such as storage isolation, the need for multiple control performance groups to separate workloads may still be required. The disadvantage of report performance groups for TSO work is the lack of period information. Many installations track their TSO service levels by TSO period one average time of ended transactions. Report performance groups have only one period for all work, and cannot be used to report separately on the "trivial" TSO work.

SE2: Problems

While the SE2 SMF provided the ability to record additional information on the status and activity of an MVS system, it also introduced a number of serious problems including the following:

- Real and pageable memory increase
- Significant increase in data volume
- Duplicate accounting
- Data interpretation
- Operation problems

There was both a real and a pageable memory increase for the new SMF. Not counting the requirements for the use of the new features in the type 30 and 32 records, SE2 required between 50K and 80K just for the new SMF structure. Use of the type 30 record increases the size of LSQA by 4K (that's per address space). Use of the type 32 record increases the size of LSQA from 5K if NODETAIL is specified to 14K for DETAIL recording. If the expanded table is used (see type 32 problems below), the increase can be from 8K to 24K per logged on TSO user!
Probably the most serious problem is the significant increase in the volume of data recorded to the SMF datasets. Use of the type 30, subtypes 1, 4, and 5 to replace the 4, 5, 20, 34, 35, and 40 records can be a 50% increase in the number of bytes recorded to SMF. Adding the new features of interval recording, TSO command recording, and RMF report performance groups can result in a more than doubling of SMF data. This assumes that the type 30 records replace the older record types and that type 40 records were being recorded. Many installations, even today, years after MVS SE2 was first available, still have programs that require the older series of records and must keep both the old and the new record types for a period of time. Use of the new features can have a drastic effect on all the programs that process the SMF data on a daily basis. That two volume tape is suddenly six or more volumes.

A number of installations have addressed the increased volume problem by judicious use of the IEFU83 and IEFU84 SMF exits. These two exits receive control when SMF is about to write a record and can prevent the writing of that record. Some techniques that have successfully been applied include the following:

- Do not allow the 30 subtype 4 record to be written for address spaces that have the interval records being written.
- Only record the type 30 interval records for long running address spaces were interval recording is of interest, e.g. TSO sessions; selected started tasks such as *MASTER*, JES2, NET, etc.; long running batch jobs such as IMS, CICS, etc. Do not record interval records for the short running address spaces. In batch this may be able to be identified by job class.
- Do not record any records for the very short duration started tasks such as the MOUNT command and the "S V" that operators issue after a VARY OFFLINE command. The cost of recording these and post processing can be greater than the resources consumed by the task itself.
- If Data set activity records, types 14, 15, 17, 18 are being used only as an audit trail to identify who "changed" a dataset, then don't record the type 14, "Open for input". Also only record the others for datasets not "owned" by the address space, e.g. if the first level qualifier of the dataset is not equal to the TSO userid.

SE2 provided more data, i.e. more chances to make mistakes. There is a potential problem with misinterpretation of SMF data because of the now duplicate recording that can occur with RMF report performance groups and interval information. Additional intelligence must be added to the SMF analysis programs to recognize report performance groups and not to use the type 30 subtype 4 if the subtypes 2 and 3 are used for the job.
Not actually a problem per se, but a serious shortcoming is the hardcoded blocksize of 4096 in IFASMFDP, the program that dumps SMF data from the MANx datasets to user specified files.

Interval recording provides additional accuracy, but it introduced its own problems:
- Interval determination
- Interval synchronization
- Missing or incorrectly apportioned information

There is no method within a record to determine the period of time an individual record represents. The start of an interval can only be determined from the end of the previous interval record.

Another "problem" might be identified as the non-synchronization of intervals for type 30 and 32 records. Interval recording is implemented by producing an interval record on a periodic basis for each address space that is eligible for interval recording. The duration of that interval and the specification of those address spaces eligible for interval recording (by type) are specified in the SMFPRMxx member in SYS1.PARMLIB. The "problem" is that the interval is calculated from the beginning of the step for the address space. Each address space writes its interval records at different times.

It is not possible to correlate the SMF intervals exactly with RMF intervals, or address spaces with each other. For example, it would be nice if the IMS control region and all its message regions had exactly the same ending time of an interval record. However, if all intervals were synchronized, think of the processing that must be done at the end of the interval. Every half hour, MVS would "go to sleep" for a few minutes, while every TSO address space was swapped in so that the interval record could be written.

Use of the type 30 interval records can result in missing or incorrectly apportioned data. For example:

- Initiator cpu time is only recorded in the 30 subtype 4, step termination record, not in the subtypes 2 and 3 interval records.
- OSAM I/O activity is added into the interval in which EOV or CLOSE processing is performed. For the IMS control region, this can reflect all of the OSAM I/O activity in the interval when IMS is terminated.

The maximum size of an SMF record is defined as 32767. With the exception of the RMF device record (type 74), few but very large shops ever reached this maximum. Since the type 30 record maintains a segment for every DD card ever used by a TSO session, the possibility is more prevalent that installations will write large records more often. This may not appear to be a problem except that the
largest blocksize definable in JCL is 32760. However, IBM has finally considered this an APARable situation and reduced the size of the maximum SMF record to below 32760. Another problem introduced by the large type 30 records is the continuation record. The handling of these records can be a non-trivial task.

The type 32 record has a number of serious problems:
- Significant data volume increased
- Lack of information recorded
- High overhead
- Data reduction problems

This additional record type, especially with interval recording activated, introduces another significant increase in the amount of data that must be recorded and processed.

There is a lack of information recorded, and a overhead to be paid for the information that is presented. When a command is recognized, the table IEEMB846 is searched SEQUENTIALLY. If the command is not found in the table, it is recorded as ***OTHER. As distributed with SE2, the table does not contain many of the commands that are standard in an installation. For example, some of the subcommands of EDIT and TEST are missing, IPCS commands are missing, some FIB commands (STATUS, CANCEL), and all of the AMS TSO commands are missing. Commands for the TSO Session Manager, HSM, and RACF program products are not included.

Information is lacking. It is nice to know that the CALL command was used, but what is important is what program was called. Similarly, for the EXEC command, what CLIST was invoked? In many installations that have attempted to use the type 32 records, a large majority of the commands were recorded as ***OTHER, i.e. commands not in the table. What are these commands? One method to determine commands not in the table is to run GTF for extended periods of time. Not only is the recording of this information a high overhead, the reduction of the GTF data can be an expensive and time consuming process.

Since an SMF record type 32 mask, with counting buckets for each command in the table, is created in each TSO user's LSQA, a tradeoff must be made between the number of commands not to be identified and the overhead of sequentially searching the table and the additional memory and swap load requirements. Aliases must be specified as separate entries in the table. This requires additional backend processing to combine the separate aliases to a single command name.

Probably the single biggest deficiency with the type 32 TSO command recording is the lack of information provided for SPF. As initially distributed, only the SPF command and subcommands of commands invoked under option 6 were recorded. With the more and more prevalent practice of all TSO work being done under SPF, the use of type
32 records for command recording seems to be nearly an overhead item. SPF has just recently been enhanced to record some additional information about the commands under SPF, but this is still far from complete. Nothing has been done to identify the use of user written panels and programs under Dialogue Management Services. Anyone using or planning to use the TSO command recording feature should review the G1575 and G20339 entries in INFO/SYS.

IBM Program Products

NJE introduced a number of changes in SMF records. These changes and the problems introduced by them are important to all JES2 installations since NJE became integrated in the SP1.3.0 release of MVS.

- New segment in type 26, JES2 Job Purge
- Route code structure changed in types 6 and 26
- New record type 57, JES2 Network SYSOUT Transmission

The type 26 record had a new network section added to record the NJE related information for a job. Installations that had been following the segmented structure of the type 26 record should have no problems with existing programs when introducing NJE (or JES2 SP1.3).

NJE changed the route code in the type 6 and 26 record to support the node number in the first half of the field. If an installation was using the special local routing facility of JES2, there is a possibility for confusion, because the first half of the field was also used to indicate special local routing.

The new type 57 record is used to record the transmission of a SYSOUT dataset to another node. All other job oriented records written by SMF have the job name and the timestamp that the job appeared on the MVS reader as a common identification field to tie the records together. The type 57 record does not have these. It only has the original and current JES job numbers. This makes it difficult to relate the type 57 record to the other records to get the complete accounting information for the job.

When a job goes through an intermediate NJE node, type 26 and 57 records are created for the job; at a node where printing occurs, type 6 and 26 records are created. These may be the only records for the job. This will cause special processing considerations for the SMF billing programs.

RACF added the group, user, and terminal id to the type 20, 30, and 32 records. There was an early problem in the type 20 record such that the terminal id was only recorded as four bytes. Record and/or segment lengths should always be checked before assuming that information in a record is present. An interesting situation is presented with the RACF recording on SMF. With the power of SAS, it
MVS SMF: SE2 and Onwards

is very easy to read all the type 20 records and do a PROC FRE0 on the group, user, and terminal ids and determine all the valid combinations. To some installations this might represent a problem because this information may be sensitive. RACF also produces two record types (80 and 81) for recording information particular to RACF. They are not addressed in this paper.

Various other IBM Program Products can produce information that is useful to the analysis of their environment. While details are not provided, a short list is as follows:

- **HSM** can produce two SMF record types. The first is written on a periodic basis and details information on complete system activity and activity by HSM managed volume. The second record is produced for each dataset HSM moves. It could be used as an audit trail and an analysis of HSM activity.

- **VSPC** will produce a number of records to reflect startup, shutdown, user logon and logoff, foreground processor activity, and security exceptions. Besides the accounting capability provided, these records provide a better insight into the activity of VSPC. The only problem is that VSPC uses some of the same record numbers as JES2 and JES3. The subsystem identification field should always be checked for all the 4x series of records (except, of course, type forty). Two is JES2, five is JES3, and six is VSPC.

- **DFEF** made changes to the 6x series of VSAM records. Processing of these may be considered as an alternate recovery method for VSAM only catalogs.

- **DFDS** added an indicator in the type 19 Direct Access Volume record for indexed VTOC's.

- **The CICS Measurement Facility (CMF)** introduced in CICS 1.5 can record system and transaction information to SMF via type 110 records. For installations with a high transaction volume, this can be another impacting increase to SMF volume. The use of a journal may be a better alternative.

**SP1.1 and RMF 2.3**

The only changes introduced in SP1.1 were the changes in the RMF records from RMF 2.3 that was introduced with SP1.1. Note that RMF 2.3 could actually run on SE2.

- The type 70 record, CPU Activity, didn't change format but had a processing consideration change. Before the 308x series of processors, the first cpu data section was usually for cpu zero,
the second for cpu one. With SP1.1 support of the 3081 and it's cpu ids of 0 and 2, it is necessary to interrogate the cpu id field. This is even more true of 3083 support in SP1 with it's single cpu having a cpu id of 2.

- The type 73 record, Channel Activity, had queue length distributions added for physical channels.

- The type 74, Device Activity, record had control unit and reserve delay counts added.

The big change for data analysis in RMF 2.3 was the ability to synchronize RMF intervals to the hour. This made comparison of RMF data between systems (for shared dasd analysis) and with other measurement techniques easier, but by no means foolproof. The clock synchronization problem between loosely coupled cpu's still exists.

SP1.3 and RMF 2.4

Again, most of the changes in SMF with SP1.3 were in RMF. They were made to support the new architectural features of SP1.3. However, there were some changes introduced in the other record types; these will be addressed first.

A new subtype was added to the type 30 record, subtype 6. This is for address spaces that do not go through "full function" start, e.g. PCAUTH. The record is incomplete, i.e., not all of the fields are valid. If interval recording is requested, the record is a cumulative interval record and will require special processing. Without interval recording, only the fact that these address spaces started is recorded. A problem existed with SP1.3.0 such that these records were not produced.

The type 90, System Status, record had a new subtype 14 added for the SET MPF command.

JES2 SP1.3 increased the support for the number of remotes and NJE nodes from the previous limit of 255. This caused a change in the route code fields in the types 6 and 26 records. Previously the fields were two byte long, the first byte indicating special routing or with NJE, node; the second byte was the remote number. For SP1.3, a new segment has been added to both records that has a four byte route code, two bytes for node and two bytes for remote number. If there are less than 255 remotes and/or nodes defined, both the new and old fields will be valid. However, if there are more than 255 entries of either type in the system, the previous field will be zeroes. Programs should check for the existence of the new route code segment and use the fields in it if present.
In SP1.3.1, JES3 started recording type 6 and 26 records for started tasks. Before this, even if these records were requested in the SHFPRMxx definition, they were not produced. This feature is being added to JES2 via an APAR. This can result in an additional increase in SMF data volume and changes to programs that were not prepared to deal with these records.

RMF had a number of changes to its records:

- The type 71 record, Paging Activity, had new fields added to record the page movement above and below the 16 megabyte line.
- The type 73 record, Channel Activity, had the number of outstanding requests added to the physical channel section.
- The type 74 record, Device Activity, had support added for the 3880-11, -13 devices in the form of number of exposures.
- The type 75 record, Page/Swap Data Set Activity, had a VIO eligible flag added.
- The type 76 record has even more fields that can be traced.
- The type 77 record, Enqueue Activity, had GRS support added. It now records the system ids for enqueues.
- The type 79 record, Monitor II Activity, had four new subtypes added: subtype 9 - Device; subtype 10 - Domain; subtype 11 - Page/Swap Data Set; and subtype 12 - Channel.

MVS/XA

While all the details for MVS/XA are not yet available, there is some information in INFO/SYS and the XA Conversion Notebook. The most important is that there is a new record structure for all records produced by RMF. This means that all programs that currently process RMF data will have to be rewritten.

Futures

While SMF has come a long way since its introduction in the 60's, it still has a number of major deficiencies. The following problems were identified in May 1980 by Brian Currah and Mario Morino. They are still very much applicable today.

With the recording of CICS/CMF transaction data to SMF, the possible future recording of IMS transaction data, larger processors, and other additional information that cannot be identified at this time, the volume of data recorded to the SMF datasets is going to be so enormous as to make it's use highly restricted, if not impossible.
There is the basic problem of date/time protection. The following protection facilities are needed:
- Verification of date/time entered at JPL to be within user-defined range of when system was last up.
- Verification of date/time entered by SET command.
- Synchronization of loosely coupled processor clocks.

A more sophisticated method is needed to route collected SMF data to a designated collection processor and therefore a centralized data depository within a CPU network. The problems are bad enough when the machines are in the same room, but even more complex with processors thousand of kilometers apart. The NJE/NJI environment aggravates this situation by not being able to tie all the information for a job together cleanly as it moves through the network. It has been suggested that an NJE/NJI job have its SMF job accounting information sent with it whenever it moves through the network.

Another problem exists with widely separated processors when the separation crosses geographical time zones. It is possible to submit a job in London, process it in New York, and print it in Tokyo. The onus of identifying the various time zone differences is on the post-processing program. The problem is exacerbated by the daylight savings or summer time programs that many countries establish at different times of the year. A code is needed in each SMF record that defines the geographical time zone of the system that recorded the information. A possible candidate is the GMT offset. This is also a requirement for any measurement data that uses date/time information from the TIME SVC since it may be compared with data from store clock recording of times.

The resolution of time recording in SMF is one hundredth of a second. With the advent of faster processors, this resolution will prove far too large for the recording of measurement timestamps. Already installations have found that they cannot always use the job/reader timestamp to tie records for a job together. With many jobs of the same name submitted by an automated process through the internal reader, the reader time stamps are identical. A similar but more severe problem exists in INS where the resolution of the clock is a tenth of a second. Some installations have found that a significant portion of their transactions run in zero cpu seconds. Will this happen with SMF and jobs? What revenue will be lost?

Summary

The processing of SMF data can provide a wealth of information for more effective management of installations. However, the processing of that data can be an expensive process. Not only is the installation faced with ever increasing data volumes, the format and content of the data being processed is continually changing. Programs written to process SMF data must be aware of not only the structure of the records being produced today, but because of the requirement...
for processing historical data, those programs must be aware of all the changes that have occurred. The manpower effort to maintain in-house programs through the extensive changes that are occurring in SMF can be considerable. This paper is an effort to make analysts aware of the changes that have occurred, and thereby reduce the costs to an installation.

Bibliography

OS/VS2 MVS SYSTEM PROGRAMMING LIBRARY: SYSTEM MANAGEMENT FACILITIES (SMF); GC28-1030; International Business Machines Corporation

MVS/SE2 SMF PERSPECTIVE, SESSION OS-34, GUIDE 50 PROCEEDINGS; MAY 23, 1980; BRIAN CURRAH, BDC COMPUTER SERVICES; MARIO MORINO, MORINO ASSOCIATES, INC.