OSF DCE SIG **R. Salz (OSF)**
Request For Comments: 24.0 **November 1992**

DCE 1.1 SERVICEABILITY PROPOSAL

1. INTRODUCTION

This document details OSF requirements for the serviceability support in DCE 1.1. It describes a message-reporting facility to be provided by OSF. All technology providers will be required to modify their code to use this facility at their existing message code-points. Additional messaging points will have to be added, so this document also presents criteria providers can use to determine where these code points are.

This document is OSF's response to the SIG requirements as detailed in [RFC11]. During the course of several meetings and discussions within OSF we prioritized the requirements and reviewed the design presented here so that reasonable goals could be set for DCE 1.1.

2. GOALS

The goal of serviceability is to make DCE easy to install, configure, and administer. The last area requires that DCE programs provide enough information so that an administrator can isolate faults and correct errors.

The primary method of accomplishing this is by having DCE programs detect error conditions and generate a message. All messages must be generated by calling a single routine. All components must be modified so that, *e.g.*, `printf()` calls are replaced by slightly different `dce_svc_printf()` calls. By requiring that all messages use this API, licensees can easily replace DCE serviceability facilities with their own, native, facilities.

All DCE messages must be consistent, providing a standard minimum amount of information. In order to guarantee this, all message calls will use a provided macro as their first argument. This macro will expand to several arguments. By using a macro, licensees can add additional information with minimal code changes to either DCE or customer source code.

All messages must meet internationalization requirements, as specified in [RFC23] (see also [Ogura] and [XPG4]). As a summary, however, all messages must be stored in message catalogs, they must not use fragmentation (*i.e.*, the `"%d unit%s"` trick[1]), and they will be displayed using the XPG4 language-independent `printf()` format string conventions. Messages will be uniquely identified by message number assigned from the status code space. In general, messages should not be re-used so that it will not be necessary to generate (and store) a UUID for each message.

Each message must have external explanatory text. This will be done using the message file described in [RFC23]. Each component will have a separate chapter in the *Problem Determination Guide* that is constructed primarily from the explanatory texts.

For DCE 1.1, we believe it is sufficient to provide only local log files written in human-readable or native binary format. We expect that we will also move from this binary format to IDL network representation (that is, ''pickling'') after it is available.

Only one process will write to each file, although the interface will be encapsulated so that this can be extended to allow for multiple writers.

''Remote Notification'' as described in [RFC11] is a non-goal for DCE 1.1. It *is* a goal, however, that nothing in this proposal preclude the future migration to DME's EVS or other network notification systems.

---

1. This refers to the common coding hack of indicating plurals by some such construct as `printf("%d unit%s", i, i==1 ? "" : "s")`.

Because messages are written under the control of an arbitrary format string, the data within the message is, essentially, opaque. This makes it very difficult to do any pre-generation filtering (as opposed to *post-mortem* data reduction tools such as `grep`). Because all message calls have a standard initial set of arguments, however, some limited filtering can be done. A publicized hook is available so that licensees can add whatever run-time filtering they wish.

It must be possible to specify the disposition of messages at a fairly fine level of detail. In addition, it must be possible to do this consistently across all components, providing boot-time defaults with an override, as well as dynamic remote reconfiguration.

While we want to detect and report faults as soon as possible, we recognize that this can have performance implications. For example, ''before-the-fact'' data capture is not a goal for this release. Because error recovery in the simple case is difficult, and automated distributed recovery is a black art, cascading error processing and extended error processing as outlined in [RFC11] is not a goal. It is sufficient to generate an accurate, detailed message, and terminate.

We will meet these goals by providing a new DCE serviceability component, and by requiring all existing code to use it. The functional outline and requirements it imposes are detailed below.

3. CODE POINT CRITERIA

In order to meet the DCE 1.1 serviceability goals, all providers will have to add messaging calls to their code. The following criteria should be used to determine these points:

(a) All servers must report when they are started, and when they have completed their initialization and are ready to perform work. They should also indicate when they are going off-line. These are informational messages and will use standard text to be supplied by OSF.

(b) All existing fatal exits must be reported as fatal errors.

(c) Errors which make it impossible to proceed must be detected as close as possible to the point of failure. This includes such ''impossible'' conditions such as failure to allocate (virtual) memory or failure to open a configuration file for reading, or a log file for writing.

(d) Conditions which might indicate system-level malfunction or poor performance must be reported. An example would be if the RPC runtime detects that it is making an excessive number of retransmits.

(e) Routine administrative actions should be reported as informational messages. This includes: creation, modification and deletion of tickets, threads, files, sockets, RPC endpoints, or other objects; message transfer, including name lookup, binding, and forwarding; directory maintenance, including caching, advertisting, skulking, and replication; and database maintenance, including replication or synchronization.

(f) It must be possible to determine that the system is operating reliably even if it is idle. Periodic ''ping'' messages are too expensive. We suggest a value-added program that collects and displays the statistics collected from the remote management interface (see below).

(g) It is *not* necessary for run-time programming errors to be detected and caught, as this could add unacceptable size and performance penalties. For example, internal subroutines need not validate their input parameters. Many routines in DCE are only called by other subroutines, and none of their input parameters is an arbitrary text string that would be entered by a user. These routines are not required to validate their input. DCE engineers might wish to use the `assert` macro to test for common obvious errors (such as an illegal null pointer); using `assert` means that these checks can easily be removed from production code. Routines that are not part of the exposed API also fall into this category.

(h) On the other hand, many other DCE routines process a string that could have been entered from a keyboard — directory name lookups are an obvious example. These routines must fail gracefully, and not core dump, for example, if given a syntactically incorrect path. Note that many components already address this to some extent; existing code should be converted to use the new service component and not removed.

It is not possible to use these criteria and assign an absolute ''serviceability value'' to some software. Nevertheless, we expect the providers to use them while making a best effort. At a minimum, if all existing message points have

been converted, all C library routine failures have been captured, and all program exits and administrative requests have been captured, then the requirements have been meet.

4. FUNCTIONAL DEFINITION

We partition DCE into pieces known as *components*. This matches the standard use of the term within the DCE development group, but at a finer grain. The DCE library is divided into the following components (name given in parentheses):

(a) Auditing (`aud`).

(b) DCE common library (`lib`).

(c) Directory/Naming (`dir`).

(d) Filesystem (`dfs`).

(e) RPC runtime (`rpc`).

(f) Security runtime (`sec`).

(g) Threads (`thr`).

Component names are limited to three characters chosen from the DCE status code ''rad40'' alphabet.[2] The names of all components are set by OSF, and are not subject to I18N requirements. Note that not all parts of the DCE API have serviceability requirements. For example, there seems to be little need to include the `utc_*()` functions. It is also possible that providers may identify additional components not listed above; this would be handled as described in the next paragraph. (For example, dividing directory into XDS, GDS, and CDS components.) The auditing component is primarily a handle to meet the auditing requirements of DCE 1.1; it has special semantics defined in [RFC25].

Some executables may share large portions of code that is in a private library and not part of the general DCE library. Providers can create a new component to capture this common code, after consultation with OSF to determine the name and scope of the component. For example, `cdsd` and `cdsclerk` might share a `cut` (''CDS utilities'') component.

In addition to the library, each server executable has at least one component.

Component-level logging does not provide enough serviceability control. Components are divided into *facilities*. Each facility corresponds to an architecturally-distinct part of the component. The number of facilities will vary greatly by component, and will be determined by the provider. For example, threads might have a ''mutex'' facility and a ''general'' facility, and little else. The RPC runtime, on the other hand, currently uses a 28-element debug table, where each entry would become a facility. Most components should fall between these two endpoints.

The same facility name may appear in more than one component — for example, every component should have a catch-all facility named ''general''. The facility should be a single-word mnemonic chosen by the provider. It is not subject to I18N requirements. Each facility also has a (short, one-line) description that *is* stored in a message catalog. OSF will also review all the facilities that are defined, and determine a standard name if any overlap is found.

Each message is uniquely part of a facility within a component.

Each message has a *severity level* attribute. The level is determined when the code-point is created and must be chosen from the following list (the manifest constant to be used is shown in parentheses):

(a) Fatal error exit (`SVC_CODE_FATAL_ERROR`).

---

2. Source licensees can refer to the implementation of `dce_error_inq_text()` for the definition of ''rad40''.

(b) Normal exit (`SVC_CODE_EXIT`).

(c) Correctable error (`SVC_CODE_WARNING`).

(d) Alarming condition (`SVC_CODE_ALARM`).

(e) Informational notice (`SVC_CODE_NOTICE`).

(f) Debug level 1 (`SVC_CODE_DEBUG1`), ..., Debug level 9 (`SVC_CODE_DEBUG9`).

Messages in the nine Debug levels would not normally appear in production code. They are used to capture the current `#ifdef DEBUG` messages found in the DCE sources. There is also a `SVC_CODE_NONE` which means that no logging should be done for a facility.

Within a given executable, messages can be routed according to their severity level. The reference implementation will only route to local files. This reasonably addresses end-user needs: ''send all fatal error messages to the console''.

Debug-level messages in an executable can be directed at the component level. For example, ''send all RPC runtime debugging messages to `/opt/dcelocal/adm/secd/foo` and all CMA debugging messages to `/opt/dcelocal/adm/secd/bar`''. This is an important facility for DCE developers and licensees.

It is also possible to vary the amount of debug information that is generated. This is done by attaching an ''active level'' to each facility. If a message is generated with a higher Debug Level than the current active level of the facility, then no output is generated.

5. DATA STRUCTURES AND DEVELOPER'S API

The following `typedef` defines a ''facility'':

```
typedef struct _dce_svc_facility_t {
    char *fac_name;             /* Facility name */
    char *fac_descr;            /* Facility description */
    unsigned32 fac_descr_msgid; /* Message ID of fac_descr */
    unsigned32 fac_level;       /* Active level for the
                                   facility */
} dce_svc_facility_t;
```

The `fac_descr_msgid` field is used to translate the `fac_descr` field of the facility. It may be zero to indicate that the message should not be translated.

These facilities are collected into the component table. By convention this is named `comp__svc_table`. The table is expected to be globally known throughout the component. For example (where `NNN` is the message ID):

```
dce_svc_facility_t thr__svc_table[] = {
    { "general", "General threads tracing", NNN,
      SVC_CODE_DEBUG_NONE },
    { "mutex", "Trace mutex calls", NNN,
      SVC_CODE_DEBUG_NONE },
    { NULL }
};

#define thr_svc_general 0
#define thr_svc_mutex   1
```

All components will be required to register their table with the service component:

```
dce_svc_handle_t
dce_svc_register (
    dce_svc_facility_t *table,
    const idl_char     *component_name,
    error_status_t     *st );
```

For example:

```
dce_svc_handle_t thr__svc_handle;

thr__svc_handle = dce_svc_register(thr__svc_table, "thr", &st);
```

A compile-time macro is also available to create a global handle variable from a table:

```
SVC_DEFINE_HANDLE(handle, table, component_name);
```

Once the table has been registered, the following routine is used for all service-related messages:

```
void
dce_svc_printf(
    DCE_SVC(handle, print_args),
    unsigned32 table_index,
    unsigned32 severity_level,
    unsigned32 message_index,
    ... );
```

Note that no status code is returned. The `table_index` specifies the offset into the service table associated with the registered `handle` parameter.

`DCE_SVC` is an opaque macro that provides initial arguments. Its first argument is the registered handle; the second describes the print arguments. The expansion of this macro is variable; for example, it might expand to the text string ‘‘`handle, __FILE__, __LINE__`’’. The ‘‘`...`’’ arguments indicate that this is a `printf()`-style routine. A sample call would be:

```
dce_svc_printf(DCE_SVC(thr__svc_handle, "x"), thr_svc_mutex,
    SVC_CODE_NOTICE, thr_s_created_mutex, &mutex);
```

This could generate a message like the following (terminal ‘‘\’’ denotes line-continuation/screen-wraparound that would normally appear as a single line on output):

```
1992-2-10-17:01:00+05:00 NOTE thr mutex Created mutex \
0x8088be62
```

Additional action can also be performed by OR’ing in any of the following with the ‘‘severity_level’’ parameter:

(a) Do not send message to log (`SVC_ACTION_NOLOG`).

(b) Send message to standard error (`SVC_ACTION_STDERR`).

(c) Send message to console (`SVC_ACTION_CONSOLE`).

(d) Abort program (`SVC_ACTION_ABORT`).

(e) Exit with failure status (`SVC_ACTION_EXIT_BAD`).

(f) Exit with succesful status (`SVC_ACTION_EXIT_OK`).

For example, the following will log an error message, display it to the user, and then exit:

```
dce_svc_printf(DCE_SVC(thr__svc_handle, "ld"), thr_svc_create,
    SVC_CODE_FATAL_ERROR | SVC_ACTION_STDERR
        | SVC_ACTION_EXIT_BAD,
    /* thr_s_no_malloc: "can't allocate %ld bytes for stack" */
    thr_s_no_malloc, size);
```

For debugging messages, the following macro can be used:

```
DCE_SVC_DEBUG((handle, table_index, debug_level, format, ...));
```

This macro can turn into null code in a production environment; in a debugging environment it decides if the message should be generated. Because the C pre-processor does not allow macros with varying numbers of arguments, the ''double parentheses trick'' must be used here. Note that `dce_svc_printf()` can also be used for debugging statements.

It is very important that debug messages not be used for errors that can occur during ordinary operation. For example, if the RPC runtime receives a corrupted packet this should be reported as a correctable error.

It is possible to add per-component filtering. This is done using the following data types and API:

```
/* This may change; as shown it is based on example given
   earlier. */
typedef struct _dce_svc_prolog_t {
    dce_svc_handle_t handle;
    int              line;
    char             *file;
    char             *format;
} *dce_svc_prolog_t;

/* Return true or false if message should be output or not. */
typedef boolean (*svc_filter_proc_t) (
    const dce_svc_prolog_t prolog,
    const char             *format,
    va_arg                 args );

/* External filter routine; parses byte string according to its
   own semantics, used to modify the filter_proc behavior. */
typedef void (*svc_filterctl_proc_t) (
    unsigned char *arg,
    int           length );

/* Attach a filter and a filter-control to a component. */
void
dce_svc_define_filter (
    svc_handle_t         *handle,
    svc_filter_proc_t    filter_function,
    svc_filterctl_proc_t filter_ctl_function,
    error_status_t       *status );
```

The `filter_function` points to a function that takes a pointer to a structure followed by the `printf()` format string and arguments. The exact elements of the prolog structure are not yet defined, but they correspond exactly to the prolog arguments created by the `DCE_SVC()` macro. A structure is used so that licensee-made changes in the macro need not affect filtering routine sources.

An additional hook is used to provide remote control of filtering by registering a control function. This function takes an array of bytes and parses it according to the licensee-supplied semantics to control the filtering.

6. MESSAGE DISPOSITION AND MANAGEMENT API

Each non-debug level can be routed separately. When a component first calls `dce_svc_register()`, a file is consulted to determine where each level is routed. Then an environment variable is consulted (see below).

In addition to these initialization mechanisms, the following routine can be called to specify where messages get routed:

```
void
dce_svc_routing (
    const char     *where,
    error_status_t *st );
```

The `where` parameter is divided into three parts, separated by colons: the level, a routing identifier and a routing parameter. (A single parameter is used, rather than three, because it will typically come from a single command-line flag or an environment variable.) OSF will implement at least the routing identifier `FILE`, where the parameter specifies the filename, as in `FILE:/tmp/log`. Vendors can provided value-added features such as `IPC:syslog` to indicate that the local `syslog` daemon should be used. A full specification would be `FATAL:FILE:/dev/console`.

All components except Audit will have a specified environment variable that can be used as the default `where` parameter. The variable will be named `SVC_`*`COMP`*, where *`COMP`* is the name of the component in all uppercase. All server executables should have a standard flag — we propose `-w` — that can be specified multiple times to set multiple routine requirements. For example:

```
cdsd -wFATAL:FILE:/dev/console -wNOTICE:FILE:/tmp/log
```

Debug-level messages are routed using the following routines:

```
void
dce_svc_debug_routing (
    const char     *where,
    error_status_t *st );

void
dce_svc_debug_set_levels (
    const char     *flags,
    error_status_t *st );
```

The first routine is used to set the debugging levels and routing. The second routine sets only the levels, typically after the routing has been set.

Debug flags are specified as a comma-separated list of facility name and numeric level separated by a period. An asterisk indicates all facilities. The settings are parsed in order, so ‘‘`*.1,cnauth.9`’’ could be used to obtain minimal debugging facilities in the RPC component, with extensive detail for the AUTH/CN facility. Again, after the config file an environment variable `SVC_`*`COMP`*`_DBG`, where *`COMP`* is the name of the component, specifies the level and disposition of debugging messages.

The remote interface (described below) will call these routines to dynamically change message disposition.

7. REMOTE INTERFACES

All servers will be required to export the OSF-written management routines that allow for queries and redirection of messages. If GDS is not converted to use threads then GDS will not be required to support the remote interface. A reasonable alternative would be to have an external program that exports the interface and forwards all requests to GDS using appropriate IPC.

All components will be under a single ACL that must be implemented by the manager. If a server already has the concept of a management ACL, then that can be used to control this interface; otherwise OSF will provide a default ACL manager that servers can use.

The interface will include the following:

```
[uuid(000cf72e-0688-1acb-97ad-08002b12b8f8), version(1)]
interface serviceability
{
typedef [ref, string] char *dce_svc_compname_t;

/*
** Calls dce_svc_routing() to actually set the routing.
*/
void
dce_svc_set_route (
    [in]        handle_t        binding_handle,
    [in]        long            where_size,
    [in, size_is(where_size)]
                byte            where[],
    [in]        long            severity,
    [out]       error_status_t  *status );

/*
** Returns an array of the names of all components that have
** been registered with the dce_svc_register() routine.
*/
void
dce_svc_inquire_components (
    [in, out]   long                    *comp_size,
    [in, out, size_is(*comp_size)]
                dce_svc_compname_t      *components,
    [out]       error_status_t          *status );

/*
** Returns the table registered with the indicated component
** name.
*/
void
dce_svc_inquire_table (
    [in]        handle_t                binding_handle,
    [in]        dce_svc_compname_t      component,
    [in, out]   long                    *tab_size,
    [in, out, size_is(*tab_size)]
                dce_svc_facility_t      table[],
    [out]       error_status_t          *status );

/*
** Calls the filter control routine for the component.
*/
void
dce_svc_filter_control (
    [in]        handle_t                binding_handle,
    [in]        dce_svc_compname_t      component,
    [in]        long                    arg_size,
    [in, size_is(arg_size)]
                byte                    *argument,
    [out]       error_status_t          *status );
```

```
/*
**  Calls dce_svc_debug_routing() to actually set the routing.
*/
void
dce_svc_set_dbg_route (
    [in]        handle_t                binding_handle,
    [in]        dce_svc_compname_t      component,
    [out]       error_status_t          *status );

/*
**  Return some operating statistics; data returned not yet
**  specified.  Probably an array of integer arrays, one per
**  component.
*/
void
dce_svc_inquire_stats (
    [in]        handle_t       binding_handle,
/*  [out]       ???            ???,             -- TBD */
    [out]       error_status_t *status );

}
```

Control programs should probably be enhanced to access this interface, although this is not a requirement.

Using this API, a new command, `dce_log`, will be provided. It can be used to remotely get and set the disposition of all messages for any DCE server. Typical uses might be as follows:

(a) Set the "mutex" facility to Debug Level 9:
`dce_log -s <entity> thr mutex 9`

(b) List the facilities within the threads component:
`dce_log -l <entity> thr`

(c) Send all fatal errors to the console:
`dce_log -r <entity> fatal FILE:/dev/console`

Entities are names within the namespace. All servers will be required to export a UUID and enter it into an OSF-specified place within the namespace.

APPENDIX A. PROVIDER CHECK-LIST

The following list is intended as a guide for providers so that they can estimate the amount of work that will be involved in converting their offering to meet the requirements described here:

(a) Determine if any new components need to be created; work with OSF to identify and name them and determine their scope.

(b) Divide the source into facilities, and create the service component table.

(c) Review all debugging, diagnostic, and logging messages. Assign them to a facility and give them the appropriate level.

(d) Create the message catalog and convert all text strings to use it. Add explanatory text to be used in the Problem Determination Guide.

(e) Add new messaging code points using the criteria described above.

(f) Add command-line support for routing messages, using the flags suggested above.

(g) Implement the ACL for the remote interface, using source provided by OSF.

APPENDIX B. SAMPLE CODE

The following library routine is intended to show how the various parts of the serviceability component are used. The `mem_s_*` values are message IDs defined elsewhere.

```
#include <dce.h>

dce_svc_facility_t mem_svc_table[] = {
    { "general", "General memory use", 0, SVC_CODE_NONE },
    { "alloc", "Memory allocations", 0, SVC_CODE_NONE },
    { "free", "Memory de-allocations", 0, SVC_CODE_NONE },
    { "copy", "String copying", 0, SVC_CODE_NONE },
    { NULL }
};

#define mem_svc_general 0
#define mem_svc_alloc   1
#define mem_svc_free    2

dce_svc_handle_t mem__svc_handle;

/*
**  Initialization routine.
*/
init(void)
{
    error_status_t      st;

    mem__svc_handle = dce_svc_register(mem__svc_table, "mem",
                          &st);
}

/*
**  Allocate memory.
*/
void *
allocate(int bytes)
{
    char        *new;

    DCE_SVC_DEBUG((mem_svc_table, mem_svc_alloc,
        SVC_CODE_DEBUG2, "Allocating %d bytes", bytes));
    if ((new = malloc(bytes)) == NULL) {
        dce_svc_printf(DCE_SVC(mem__svc_handle, "d"),
            mem_svc_alloc, SVC_CODE_FATAL_ERROR, mem_s_noalloc,
            strlen(p) + 1);
        exit(1);
    }
    DCE_SVC_DEBUG((mem__svc_handle, mem_svc_alloc,
        SVC_CODE_DEBUG1, "Allocated %d bytes at 0x%x", bytes,
        new));
    return new;
}
```

```
/*
**  Copy a string to allocated space.
*/
char *
strcopy(char *p)
{
    char        *new;

    DCE_SVC_DEBUG((mem__svc_handle, mem_svc_string,
        SVC_CODE_DEBUG1, "Copying string |%s|", p));
    if ((new = allocate(strlen(p) + 1)) == NULL) {
        dce_svc_printf(DCE_SVC(mem__svc_handle, "d"),
            mem_svc_alloc, SVC_CODE_FATAL, mem_s_noalloc,
            strlen(p) + 1);
        exit(1);
    }
    return strcpy(new, p);
}

/*
**  Free memory.
*/
void
unallocate(void *p)
{
    if (p == NULL) {
        dce_svc_printf(DCE_SVC(mem__svc_handle, NULL),
            mem_svc_free, SVC_CODE_WARNING, mem_s_free_null);
        return;
    }
    DCE_SVC_DEBUG((mem__svc_handle, mem_svc_alloc,
        SVC_CODE_DEBUG1, "Freeing memory at 0x%x", p));
    free(p);
}
```

REFERENCES

[Ogura] T. Ogura, ‘‘DCE 1.1 I18N Workbook’’, DCE Document, September 1992.

[RFC11] M. Hubbard, DCE-RFC 11.0, ‘‘DCE SIG Serviceability Requirements’’, August 1992.

[RFC23] D. Mackey, DCE-RFC 23.0, ‘‘DCE 1.1 Internationalization Guide’’ (to be published).

[RFC25] E. McDermott, DCE-RFC 25.0, ‘‘DCE Auditing Design and Strategy’’ (to be published).

[XPG4] *X/Open XPG4 Portability Guide*, message catalog routines.

AUTHOR’S ADDRESS

Rich Salz
Open Software Foundation
11 Cambridge Center
Cambridge, MA 02142
USA

Internet email: rsalz@osf.org
Telephone: +1-617-621-7253