OSF DCE SIG **M. Mori (OSF)**
Request For Comments: 41.0 **R. Mackey (OSF)**
May 1993

RPC RUNTIME SUPPORT FOR I18N CHARACTERS — FUNCTIONAL SPECIFICATION

1. INTRODUCTION

RPC runtime library enhancements are required to support character conversion features being added to the IDL compiler. Together, the IDL improvements and runtime features provide applications the ability to communicate character data between clients and servers using different representations for their character data. This feature allows data to be passed to RPC stubs and automatically converted into an appropriate representation before it is received by the peer application. The runtime enhancements include functions that provide specific client/server code set management support such as conversion, evaluation of character and code set compatibility, and interfaces to local operating system code set management functions.

DCE 1.0.x only provides support for transparent conversion of character representation (**code sets**) within the intersection of characters that are supported in the ASCII and U.S. EBCDIC code sets (the Portable Character Set (**PCS**)). Applications using this feature must explicitly identify character data as `char` or `idl_char` in the interface definitions. The RPC runtime is compiled with the **code set** employed at a given host, just as the information about endianess and floating point format is. The application supplies character data through this type and the stub at the receiving machine inspects the encoding information and converts to the local format, if necessary.

The limit of support to ASCII and U.S. EBCDIC is adequate for some applications, but falls short of applications that need to be able to transmit and process other **character sets** (e.g., European, Chinese, Korean, Japanese, etc.). An added complication is that **character sets** may be represented by a variety of **code sets** (e.g., ISO 8859-1, pc850, ROMAN8, SJIS, and AJEC, etc.). For example, `<a-acute>` is represented `0xe1` in ISO 8859-1, `0xa0` in IBM's pc850, and `0xc4` in HP's ROMAN8. ISO 8859-1, IBM's pc850, and HP's ROMAN8 all encode the Latin-1 **character set**. Another example is the Japanese character for `one` (in kanji), which is represented as `0xB0EC` in AJEC, and can be represented as `0x88EA` in SJIS. So even though the **character set** is same, there is a need to convert from one **code set** to another. In fact, it is very common that one **code set** is represented differently from host to host in a **heterogeneous environments**, where several vendors' machines are connected together.

Prior to DCE 1.1, applications operating with data other than ASCII and U.S. EBCDIC were either restricted to **homogeneous code set environments** where all the character data representations were the same or were forced to implement their own **code set** recognition and conversion mechanism.

To aid programmers in developing applications that will behave correctly in both **homogeneous** and **heterogeneous environment**, DCE release 1.1 will include RPC runtime library enhancements and IDL compiler improvements to allow automatic conversion between various representations of characters.

The IDL compiler will be improved to have a new ACF feature. This feature, allows application developers to specify automatic conversion between different **code sets**.

RPC runtime library will be enhanced to include:

(a) **Code set** advertisement and retrieval routines.

(b) **Character and code set** compatibility evaluation routines.

(c) **Code set** conversion routines.

(d) Local operating system and *OSF Character and Code set Registry* access routines. (See FUNCTIONAL DEFINITION section about the OSF Character and Code set Registry.)

(e) Buffer sizing routines for marshalling and unmarshalling.

Using these improvements and enhancements, distributed applications can be written with *code set transparency*. *Code set transparency* means that application programs do not need to be aware of how their character data are encoded. Once an application sets a locale, the **code set** information will be retrieved internally. **Code set** compatibility evaluation will determine the binding between client and server. Then the **code set** is automatically converted to make a connection possible (between client and server), when necessary.

OSF's goal was to design these new features to be as flexible as possbile. We didn't want to hard code features like a *character and code set evaluation logic* or a *connection model* between client and server. Rather, emphasis was put upon a framework which allows these features to be more easily tailored to the needs of each application. For example, customized evaluation logic, or the selection of a particular connection model can be easily implemented.

RPC runtime library enhancement will include default **code set** compatibility evaluation routines. When these default routines are not sufficient for a particular application, they can be replaced by application developers.

See FUNCTIONAL DEFINITION sectin for the details of these enhancements.

2. TERMINOLOGY

(a) **CMIR**

Client Makes It Right. The client always converts data into the server's current locale's code set (when conversion is necessary).

(b) **Character Set**

A group of characters without any associated encoding. Examples of character sets are the English alphabet, Japanese Kanji, and the characters needed to write European languages. [RFC 27.0]

(c) **Coded Character Set, or Code set**

A mapping of the members of a character set to specific numeric code values. Examples include ASCII, ISO 8859-1 (Latin-1), JIS X0208 (Japanese Kanji). [RFC 27.0]

(d) **Dynamic**

Determine which code set will be used for exchanging data outside of **PCS** dynamically. It can be the server's code set, the client's code set, the universal code set, or the code set which both server and client has a converter to its primary code set.

(e) **Encoding Method**

In order to use computers in some countries, it may be necessary to combine multiple code sets in a data stream. An encoding method provides the rules for combining sets and recognizing the set to which a given character belongs. An implementation of an encoding method is a specific instantiation of an encoding method's rules. Examples of encoding method implementations are AJEC (Advanced Japanese EUC Code) or a specific version of Japanese SJIS.

Although code sets and encoding methods are different, they are treated nearly the same with respect to interoperability issues. Unless otherwise stated, the term **code sets** refers to both concepts. The distinction between **character** and **code sets**, however, is important. [RFC 27.0]

(f) **Heterogeneous Environment**

In the Heterogeneous Environment, all hosts support the DCE **Portable Character Set**. But, there is no single homogeneous **character set** which is shared across all hosts in a network. Each host may support its own **character set**(s) so long as it is a superset of the DCE **Portable Character Set**, and, may use its own **code set** which may be different from the **code set** that other hosts are using. Like the **Homogeneous character set Environment**, if the **code set** used by a server is different than the client **code set**, then the **code set** is converted to make the communication possible. The code set conversion can take place either on the server side, on the client side, or both. Since different **code sets** support different **character sets**, there will be a massive data loss when code set compatibility is not evaluated correctly.

(g) **Homogeneous character set Environment**

In the Homogeneous character set Environment, all DCE hosts within a network use the same **Character Set**. Each host may use a different **code set** to encode the **character set** locally. If the **code set**, used by a given server, is different than the **code set** of a client that wishes to bind with the server, then the **code set** is converted to make communication possible. This conversion can take place either on the server side, on the client side, or both. Even though only one **character set** is supported in the network, different **code sets** include different range of characters, so a minor data loss can be resulted by the code set conversion. For example, Japanese SJIS does not include JIS X0212 characters, while AJEC can encode them.

(h) **Homogeneous code set Environment**

This environment consists of a single **code set** which is supported by all hosts in a network. It assumes that a single **code set** is used to transfer text strings between hosts. No code set conversion is necessary, and there will be no data loss in a communication. This environment does not specify the **code set** to be used. The only restriction is that the **code set** used must be a superset of (contain at least the characters in) the DCE **Portable Character Set**.

(i) **I14Y**

Interoperability.

(j) **I18N**

Internationalization.

(k) **Portable Character Set (PCS)**

A minimum group of characters guaranteed to be supported on all compliant systems. The most commonly used PCS includes all the graphic characters in ASCII — that is, letters A-Z and a-z, digits 0-9, and common punctuation. The PCS may be encoded in multiple ways. [RFC 27.0]

(l) **RMIR**

Receiver Makes It Right. Receiving side will convert (if necessary) the code set on the wire into its current locale's code set. This is the model that current ASCII and U.S. EBCDIC automatic conversion is based on.

(m) **SMIR**

Server Makes It Right. The server always converts data into the client's current locale's code set (when conversion is necessary).

(n) **Universal**

A universal code set can be used for exchanging character data outside of **PCS**. Even though there is no assumption for which code set should be used as universal, ISO 10646 is a good candidate.

3. TARGET

The **I18N** enhancements for RPC application development is aimed to be used by application developers who wish to exchange data outside of the **Portable Character Set**.

4. GOALS AND NON-GOALS

The goals of the **I18N** enhancements for RPC application development are:

(a) Provide a way to exchange a variety of **character and code set** values. Automatic code set conversion can be done by using a new ACF feature. But automatic conversion outside of **PCS** is optional and under the application programmer's control.

(b) Provide a mechanism for a client to evaluate a server before it decides to establish a connection with it. Also provide a mechanism to give a client enough information for this evaluation to be done in appropriate fashion.

(c) Provide a mechanism for code set transparency during communication between client and server. Application programs do not need to be aware of which **code set** they are operating on.

(d) Provide necessary tools for global heterogeneous **I14Y** (e.g., *character and code set registry*, and routines to access it).

(e) Provide new management routines to handle attributes in CDS namespace. This new functionality is used for registering, retrieving and removing server's supported **code set** information.

(f) Tags are used to make global **heterogeneous I14Y** possible. These tags are exposed in the interface definition to be consistent with our philosophy of showing all data that is passed between a client and a server.

(g) Allow application interfaces to be written now that could be used with the new ACF feature when it becomes available.

In other words, this functional specification describes the procedure to exchange strings outside of the **PCS**. Even though DCE 1.1 is not available yet, as long as application programs follow this procedure, the future interoperability between DCE 1.1 application programs is guaranteed. To be more concrete, if current application programs set tags appropriately, they can interoperate with DCE 1.1 programs. The current application programs cannot take advantage of automatic code set conversion and other new features like evaluation of character and code set interoperability, however, these features can be added to the application programs when DCE 1.1 is available.

(h) No protocol will be changed. The `idl_char` definition will not be changed in anyway.

(i) Keep backward compatibility between the previous DCE versions and DCE 1.1.

The non-goal is:

(a) Code set tags are part of user data, and they cannot be hidden in internal RPC mechanism. There will be no attempt to hide tags from interface definition.

5. REQUIREMENTS

This enhancement should enable exchanging characters outside of the **PCS** in DCE applications. It includes the *code set conversion mechanism*, which converts from one code set to the other, for cases when server and client both understand the same character set (e.g., the same language), but use different code sets (e.g., the way each character is represented).

In the case where a server only understands character sets which are incompatible with a client's character sets (e.g., Korean and French) establishing a connection can be meaningless. In that case, a client can decide what to do.

There are many *models* for code set conversion. The current automatic character conversion between ASCII and U.S. EBCDIC within the **PCS** domain only supports the **RMIR** (Receiver Makes It Right) model. For the new feature to be flexible enough for world wide requirements, OSF cannot only assume **RMIR** model is enough. That means the architecture of this new feature should be designed in a way that an application can decide what is the right model.

Automatic data conversion outside of the **PCS** is optional. As long as tag(s) are correctly set for data which flow over the wire, world wide inter-operability is guaranteed.

OSF provides a mechanism to give enough information to application programmers to decide the inter-operability between client and server. For this requirement, new *NSI management routines* are used to export code set information to CDS namespace.

OSF will provide several default *character and code set compatibility evaluation routines* and *code set conversion routines*. When these default routines are not enough to support an application program's needs, application developers can provide the customized routines. OSF develops the framework which is used to set these customized routines.

In addition, DCE Shell's *nsi-entry* class will include the method necessary to manipulate code set attributes. OSF provides the API to advertise the server's supported code sets information to CDS entry. In addition, the DCE Shell command will be used to add/delete/view the code set attribute in CDS namespace alternatively.

6. I18N APPLICATION MODELS AND I18N FEATURES IN DCE 1.1

Since there are many new concepts involved in the RPC runtime support for I18N characters, this section is devoted to explaining the various models and features for DCE I18N applications.

There are three major sub-sections here. The first sub-section describes an I18N application as a whole. I18N application models are explained in both the execution model and the development model.

The second sub-section describes Functional Models. In this sub-section, a conceptual model of each feature is discussed without referring to any implementation details. The models discussed here can be implemented in other technologies outside of DCE.

The last sub-section describes Implementation Models. Here we investigate how the conceptual model can be implemented in the DCE architecture.

The intent here is that before we actually see the FUNCTIONAL DEFINITION section, which illustrates the actual coding of each feature, we understand the conceptual background for our design.

6.1. I18N Applications in the Distributed Environment

6.1.1. Overview of the application execution model

In the new application execution model:

(a) A server advertises its supported character and code sets into the namespace.

(b) A client imports this information and evaluates the server. If the client and server are using the same **code set**, there will be no data loss. If they are using different **code sets**, the client can check to determine whether the server's **code set** is compatible with the client's. If it is, data loss (if any) will be minimal. **Code sets** are considered compatible if they encode the same **character set**, or language. For example, a client and server that both support Japanese, but use different **code sets**, are considered compatible. However, a client that supports French and a server that supports Chinese are considered incompatible because conversion between the two would almost certainly result in massive data loss.

(c) A compatible server is returned to the client application in the process of binding.

(d) A client stub is called with data represented in the local form.

(e) These data are converted to an appropriate **code set** representation if necessary, and sent to the server with **code set** identifying information.

(f) A server stub receives the data and converts it into its local representation based on the **code set** identifying information, if necessary.

(g) The manager routine returns the data to the server stub. This data may be converted from local representation to network representation, based on a format from the client about its **code set** support. Server stub sends data together with **code set** identifier for network representation of the data.

(h) A client stub receives the data, converts it to a local representation, if necessary, based on the **code set** identifying information from the server.

(i) Client receives the data from a client stub. At this point, data is represented in client's local **code set**.

There is no conversion involved when both server and client use the same **code set**.

6.1.2. Overview of the application development model

(a) The application programmer writes an IDL file specifying code set identifiers and character data as parameters to operations.

[Implementation note: The interface definition identifies character data that may be automatically converted by creating a particular typedef that may be flagged in the associated attribute configuration file.]

(b) The application programmer writes a matching ACF file which requests that automatic conversion takes place on character data in parameters having a certain data type. This ACF file attribute allows the programmer to name the local data type. Character data will be stored in this data type. ACF also names a routine that will be called from within the stubs that will set the code set identifiers properly.

All names will have appropriate defaults and all default routines will be provided in the DCE stub support library. If the programmer selects routine names other than the default, the programmer must supply these routines.

6.1.3. Server development

(a) The server application programmer calls routines that determine the local **code set** and **code set converter** support available in this process [DCE library routines are provided which request this information from the local operating system].

(b) The server application programmer writes application code which calls NSI routines which place local **code set** and **converter** info into the server entry[ies] for this process. This information is assumed to coexist with the binding info for this server. [This step can be considered to be an extension of the `rpc_ns_binding_export()` step.]

(c) The server application program listens for requests.

6.1.4. Client development

(a) Client developer writes code that calls `rpc_ns_import_begin()` in the normal manner to begin the process of selecting a *compatible* server.

(b) The developer invokes a new server evaluation mechanism which places additional checking on the *acceptable* bindings. In the original import algorithm, only protocol and interface information was used to determine client/server compatibility. This new capability dispatches a user-supplied routine(s) to judge whether the server is acceptable. The routine is passed the name of the current namespace entry, and all information from `rpc_ns_binding_import_begin()` to determine compatibility. In the context of code set compatibility, the evaluation routine will read code set attributes from the namespace and compare it to local code set support to determine whether there is a possibility of significant data loss when conversion is done, or if conversion is needed at all.

OSF will supply several default evaluation routines and examples for describing how to use these features. The default evaluation routines will implement several connection models (between client and server), like **dynamic**, **universal**, **RMIR**, **CMIR**, and **SMIR**. Examples include how these models will be used based on the default code set compatibility evaluation logic. See Functional Model and Implementation Model sections for the default logic.

(We cannot guard against minor data loss (except if an application requires the client and server to use the same code set), but we can prevent massive loss with our evaluation functionality.)

(c) The developer supplies evaluation routines if defaults are not sufficient.

(d) The developer calls the import next routine which evaluates the various server entries (through dispatching of evaluation routines) and returns a compatible server binding.

(e) Client makes a call to the client stub passing data in the representation identified in the API in the ACF.

If the default routines provided with the DCE are not sufficient to handle the application's needs, the client can supply any of the following:

(a) The routine to set tags.

(b) The routine to calculate buffer size.

(c) The routine to convert code sets.

(d) The routine to evaluate the code sets compatibility.

6.2. Functional Model of the New Features

This section describes the model of the functionalities for the new RPC runtime support. The idea behind of each feature is explained in detail without referring to the implementation details.

6.2.1. NSI management

For the server to advertise supported character and code sets, and for a client to obtain this information, a new NSI management mechanism is required. This is because the current NSI management functionality in RPC runtime only accepts pre-architected attributes, like protocol tower, uuid, ns group, or ns profile.

The new NSI management mechanism takes character and code sets information and puts it into the namespace. It also performs a retrieval and removal of the information from the namespace.

6.2.2. Evaluation

The process of selecting a compatible server from a client is called *evaluation*. In this evaluation process, client checks the server's supported character and code sets that the server advertises, and determines if a binding to the server is compatible, based on its own supported character and code sets.

In the binding selection process, we have to make sure that client and server are inter-operable for character and code set compatibility. In other words, we need to make sure that data will be communicated between the client and the server without resulting in a lot of data loss. For example, sending Chinese data to a Greek server would not be communication, since much of the data would be lost. In addition, even if both client and server understand the same character set, e.g., Japanese, if a client uses SJIS code set and a server uses AJEC (EUC variant) code set for processing, we want to know which code set converters are available in client and server, so we can perform the necessary code set conversion.

The problem that needs to be solved is really a two step process:

(a) The character set compatibility issue (i.e., assessing whether the client and the server both support the characters required to be communicated).

(b) The code set conversion problem (i.e., given that the characters can be represented at both client and server, how the data must be transmitted and converted to allow clients and servers to manipulate the data in their own local representation).

A result of the evaluation is whether the client and server can communicate characters adequately. In addition, the client will determine how data will be communicated, i.e., where (client side, or server side, or both sides) the conversion will take place.

For a client to get the information about server's supported character and code sets, a new NSI management mechanism is used. This new mechanism is used by a server to advertise its supporting character and code sets, too.

Also to make its supported character and code sets available for both client and server, an interface to the local operating system is provided. This mechanism queries a local operating system about its supported character and code sets, since different OSs support different groups of code sets. Information about supported character and code sets are mapped to globally unique identifiers via the OSF character and code set registry interface.

The flow of the evaluation process is:

(a) Server gets its supported character and code sets information from local operation system.

(b) Server advertises that information.

(c) Client gets this advertised information.

(d) Client gets its supporting character and code sets information from local operating system.

(e) Client determines and selects the appropriate server.

6.2.3. Interface to the local operating system

For both server and client, the information about supported character and code set is required to determine inter-operability. Since this information belongs to the local operating system domain, a new mechanism is needed to provide the interface to the local operating system.

In addition, since DCE 1.1 will rely on a local code set conversion mechanism to perform code set conversion, we need the following two interfaces to the local operating system resources:

(a) A routine that calls local code set converters.

(b) A routine that queries supported character and code sets.

The first routine is used for code set conversion. The second routine is the new functionality.

6.2.4. OSF character and code set registry

OSF character and code set registry provides a mechanism to uniquely identify code sets. OSF is proposing *OSF Character and Code Set Registry* [RFC 40.0] to provide application programs the ability to uniquely identify every code set, including vendor specific ones. DCE 1.1 will include the routines to access the OSF character and code set registry, so application programs can map platform dependent code sets to the unique identifiers.

The OSF character and code set registry contains information about **character set** and **code set** with **organization identification**. Registered values are required to make inter-operability possible, since different organizations (different vendors) can have different encodings (vendor-specific encodings) as well as different names for the same character and code set (ISO 8859-1 can be named ISO8859-1, Latin-1, 88591, iso8859-1, etc.).

6.2.5. Code set conversion

When a client sends a data to a server, this data should be understood by both the client and the server. This understanding should be achieved in two levels:

(a) Both client and server understand a character set.

(b) Both client and server understand the way the character set is expressed.

Code set conversion provides a way for a character set to be represented in a form both client and server can understand.

The examples of a character set are **PCS**, German, Chinese, or Japanese. Code set conversion cannot take care of character set conversion, since mapping German characters to Chinese characters will probably not make sense in most cases.

However, if both server and client use the same character set, say Japanese, it will make sense to exchange data even though their code set is different, like SJIS and AJEC (EUC variant).

The IDL enhancement can generate a hook to code set conversion mechanism that is invoked from stubs, so from either or both client and server stubs, the code set conversion mechanism can be invoked to perform an automatic conversion of different code sets, which translates a language from one representation to the other.

6.2.6. Buffer sizing

When converting data from one code set to another, there is always a possibility that the converted string can be larger than the original string. This is because one character can be encoded in a variable number of bytes in different code sets.

To support an automatic conversion operation, a buffer sizing mechanism needs to be supplied to calculate the actual size of a buffer, so stubs can allocate a buffer of sufficient size to hold the converted string.

This mechanism is not meant to be used by application programmers, but it is designed to be used for marshalling and unmarshalling of characters outside of **PCS** range.

6.3. Implementation Model of the New Features

This section describes how each feature is actually implemented. Unlike the previous section, which explains the model of functionalities, how each model can be implemented in DCE 1.1 is discussed here.

6.3.1. NSI management

The current RPC runtime library only accepts pre-architected attributes, like NS profile, object uuid, NS tower, or NS group. In other words, only these attributes are assigned **OID**s (object identifiers) in a **OID** domain, which is allocated to OSF. (OSF does not use **OID**s above "1.3.22.1.1.4".)

The new attribute **code sets** will be assigned **OID** "1.3.22.1.1.5", and the new routines will be added to the RPC runtime library. Since there are already routines to handle NSI attributes in the RPC runtime library, the new routines will be similar to the existing routines to add, read, and delete the attribute in/from the CDS namespace.

Since the attribute **code sets** is expressed as an integer, a new IDL compiler feature (*IDL Encoding Services*), will be used to encode the **code sets** values into an endian-safe form before it is stored in CDS namespace. This *IDL Encoding Services* feature is provided with DCE 1.1 IDL compiler.

6.3.2. Evaluation

The evaluation process can be divided into two steps. One is to determine if a server is inter-operable, and the other is to select the way clients and servers inter-operate.

For step one, to determine if a server is inter-operable, the client will set a pointer to an evaluation routine into a context handle structure as a part of the binding selection process. Application developers can provide this evaluation routine, or they can use OSF-supplied routines.

When an evaluation routine is set, a function pointer to an evaluation routine will be a part of context handle to the NS binding selection routine (namely, `rpc_ns_binding_import_next()`). This mechanism can be used not only for character and code set compatibility evaluation, but also for other future functionality, therefore it is designed as stack structure (that includes a list of stacks) which hold function pointer(s).

When the NS binding selection routine is called, it examines a context handle to see if this stack information is included. When the stack contains a function pointer to an evaluation routine, it is dereferenced and the function is executed.

For step two, to select the way the client and server inter-operate, an evaluation routine will choose the model of client and server communication. Based on the model selected, a tag(s) set routine can decide the actual tag value for a data. This tag(s) set routine is called from rpc stubs before the marshalling of data.

6.4. Logic of Evaluation

A client and a server compatibility is determined by three steps: The first is the code set compatibility evaluation. When code set is not compatible, character set(s) is evaluated for compatibility. This character set compatibility checking can be deferred when an application wants to do so. The last is the selection of the connection model between a client and a server, based on the code sets used.

The OSF supplied default evaluation routines will check if the same code set is used by a client and a server. When it is, the evaluation is done. They are compatible.

Next, character sets used by a client and a server is checked. When both a client and a server are using the same character set, it will be safe to exchange coded character values regardless of the encodings used to represent the character set. When a character set is determined compatible, then code sets may be used to select the connection model, such as **RMIR** or **Dynamic**.

The somewhat complicated issue here is that a code set can contain several character sets. For example, some Asian code sets contain the Latin-1 character set in their character domain. Does this mean an ISO 8859-1 client is compatible with an AJEC server if both of them have ISO 8859-1 to/from AJEC converters? They could be compatible as long as only the client sends data, or if the server either returns no data or returns only Latin-1 data to the client. However, in most cases, they are not compatible because the repertoire of characters in AJEC is so much larger than ISO 8859-1, and there is no simple way to ensure that clients and servers limit their data exchange to characters that are in both of their repertoires. Therefore, by default, the OSF evaluation routines will consider ISO 8859-1 and AJEC to be *incompatible*.

Application developers can provide their own the evaluation logic to change the default behavior.

When client's and server's code sets only consist of one character set, the evaluation logic will be simple. If their character sets match, then select the way they can be connected.

But when their code sets consist of more than one character sets, then the character set evaluation will be done based on the logic that

If more than one character sets match between client and server, then their character sets are compatible, otherwise they are not.

6.4.1. Interface to the local operating system

New routines are provided to perform the following two functions:

(a) Routine which calls the local code set converters.

This is a routine to call the appropriate iconv converters for code set conversions. This is a wrapper routine, and will be newly written.

(b) Routine which returns the supported character and code sets.

This is a routine to access local operating system resources, and to produce the information about the supported character and code sets. For example, in the case of an OSF/1.1 system, this routine will access the `/usr/lib/nls/` directory (`LOCPATH` in OSF/1) and will see which converters are available. This routine will be newly written.

6.4.2. OSF character and code set registry

The *OSF Character and Code Set Registry* is a table that maintains

(a) **character set**

and

(b) **code set**.

Each registry entry has this information along with an integer value that represent the entry. This integer value is assigned by OSF. This is a 32-bit integer; the upper 16 bits identify the organization and the lower 16 bits identify the code set. In addition, each entry has one or more character set ID, which is 16-bit integer. [RFC 40.0]

Two new functions are implemented: One takes a string code set name, and returns the integer code set value. The other takes an integer code set value and returns the string code set name.

6.4.3. Code set conversion

DCE 1.1 will provide a wrapper routine, which is explained in *local operating system access*, to invoke the XPG4 code set conversion facility.

In a reference implementation, code set conversion is done by the XPG4 code set conversion facility, namely **iconv**. **iconv** is provided by a local operating system. This is a set of library routines to take one code set and map the value into another code set. (XPG4 also defines a command called `iconv`, to perform code set conversions. But what we will use for DCE 1.1 are the library functions.)

iconv facility consists of the following routines:

(a) `iconv_open()` — Code conversion allocation function.

(b) `iconv()` — Code conversion function.

(c) `iconv_close()` — Code conversion deallocation function.

`iconv_open()` returns a conversion descriptor that describes a conversion from the code set specified by the string pointed to by one argument to the code set specified by the string pointed to by the other argument.

`iconv()` converts the sequence of characters from one code set into a sequence of corresponding characters in another code set.

`iconv_close()` deallocates the conversion descriptor and all associated resources allocated by the iconv_open() function.

Since `iconv_open()` only takes the *code set specified by the string* and there are no standard defined code set names so far, we need to have another mechanism to uniquely identify code sets to be used. Also in some cases for DCE internal operation, using an integer value to distinguish each code set is more appropriate than using string value.

For this purpose, the *OSF Code Set Registry* values are used to access the uniquely identified code sets.

6.4.4. Buffer sizing

Since there can be a need to calculate a buffer size for code set conversion, a routine to determine the storage requirements for converting the data to transmissible form will be provided.

This new routine(s) will be a part of RPC runtime library, and be called from RPC stubs before marshalling and unmarshalling of data.

The logic for determining the new buffer size will be:

```
number-of-bytes * MB_LEN_MAX
```

number-of-bytes is the value that stubs receive from the network. (marshalling always requires the number of bytes be passed over the network).

MB_LEN_MAX is a platform dependent value that is defined as the maximum number of bytes for a character, for any supported locale.

6.4.5. Tags usage

A code set identifier is called *tag*. Tags are necessary to correctly identify data that flows on the wire in case the data is character data outside of the **PCS** range.

A maximum of three tags will be used when sending any character data outside of the PCS over the wire. They are:

(a) The sending tag, indicates the code set the client sends over the wire.

(b) The desired reception tag, indicates the code set the client prefers to receive.

(c) The reception tag, indicates the code set which the client actually received.

When a client sends the data over the wire, the reception tag is not used. The reception tag is set by a server when a data is marshalled in the server stub.

When a client only needs to send a data and no output is expected, only the sending tag is used. When a client needs to receive data from server, both the desired reception tag and the reception tag are used. When client wants to send and receive data, all three tags are used.

These tags can be set directly by an application, or a new ACF feature can be used to set the tags automatically. These tags need to be listed in the IDL file explicitly. This is because:

(a) When data flows on the wire, it needs to be explicitly defined in an interface definition. This is the philosophy OSF is following.

(b) Tags are part of application data. They do not relate to DCE/RPC internal operation in any way.

The IDL description of an operation with a parameter which is a conformant varying of characters from a non-PCS set could be written as:

```
typedef byte my_byte;

void a_op(
        [in] unsigned long stag,
        [in] unsigned long drtag,
        [out] unsigned long *p_rtag,
        [in] long s,
```

```
            [in, out] long *p_l,
            [in, out, size_is(s), length_is(*p_l)] my_byte a[]
    );
```

7. FUNCTIONAL DEFINITION

7.1. NSI Management of New Attribute in CDS Namespace

New attribute *code sets* is to be assigned a new object **OID** within OSF **OID** range.

code sets is defined as a structure to hold a list of integer values. Each integer value is one of the values from the OSF Code Set Registry, and at top of the list is the current **code set** which is used by a process (either client or server).

This *code sets* is converted into an endian-safe format before it is exported to RPC entry within the CDS namespace. The *Incremental Buffer Encoding* and *Incremental Decoding* methods are used to supply buffers for the encoding and the decoding operations. Appropriate interface definition and ACF definition will be written to define these methods.

A routine is written to take this *code sets* and to export it into the CDS namespace. It takes a structure *code sets* and converts it into an endian-safe format before the attribute is exported into CDS namespace. This routine is a part of RPC runtime library.

A set of routines will be written to read this exported attribute from the CDS namespace. The first routine sets up an inquiry context, similar to the `rpc_ns_entry_object_inq_begin()` routine. The second routine reads an attribute based on the inquiry context, similar to the `rpc_ns_entry_object_inq_next()` routine. This routine reads a string from CDS, and converts it into a *code set* structure. The last routine finishes the operation, similar to the `rpc_ns_entry_object_inq_done()` routine. These routines are a part of the RPC runtime library.

For the convenience of reading exported **code set** value, OSF provides a routine to be tailored for the **code set** attribute retrieval. This routine will call the set of routines internally.

A routine will be written to remove a *code sets* attribute from the CDS namespace. This routine operates in a similar to the `rpc_ns_mgmt_binding_unexport()` routine. This routine is a part of the RPC runtime library.

New RPC NSI attributes, *rpc_c_ns_codesets* is used to indicate the new attribute.

7.2. Interface to the Local Operating System

A routine will be written to access local operating system resources and get the information about code set converters. This routine will check which code set converters are available, and fills the values into a *code sets* structure. The first entry of a list of supported code sets should be the current code set under which a process is operating on.

7.3. Compatibility Evaluation Between Client and Server

A stack structure which holds function pointer(s) to an evaluation routine will be defined. The contents of a stack will be poped off and function pointer is dereferenced at the binding selection process in a client.

This stack structure has fields indicating:

(a) Function type.

(b) Function pointer.

(c) Preserved import context.

(d) Pointer to the next element in the stack.

A routine to push a function pointer into this stack structure will be provided. This routine takes:

(a) Import context, which is allocated by `rpc_ns_binding_import_begin()` routine.

(b) Function pointer to an evaluation routine.

(c) Function type.

(d) Status.

When this routine is called, a stack structure is allocated and appended to the import context. One of the fields defined in the current import context, which is not used until the routine `rpc_ns_binding_import_next()` performs the actual binding, will be used as a flag to indicate whether this stack structure is appended to the import context or not.

This flag field is `inq_cntx`, and it is a part of the `rpc_lkup_rep_t` structure (which is defined in nsp.h in the RPC runtime library).

The functional pointer will be dereferenced at the beginning of `rpc_ns_binding_import_next()`. This existing routine will be modified to check if the stack structure is appended to a import context (by looking at the flag field), and execute the evaluation routine. The signature of `rpc_ns_binding_import_next()` routine will not be changed. This is possible since the argument, `import_context`, is `rpc_ns_handle_t` data type which is an opaque pointer. Within the routine, this pointer is cast to the actual data type, so depending on the value of the flag field, this opaque pointer will be cast to the right structure, and the dereference of an evaluation routine becomes possible.

The evaluation routine can evaluate anything, and the use of it is not restricted to code set evaluation. However, the main use of this evaluation routine for DCE 1.1 is *code set evaluation*. A default code set evaluation routine returns the model of code set conversion, which is one of the **No Conversion**, **RMIR**, **CMIR**, **SMIR**, or **UNIVERSAL**. When this evaluation routine cannot find an inter-operable model between client and server, it returns *IMPOSSIBLE* (or some value indicating the connection will result in massive data loss). If the model value is anything except *IMPOSSIBLE*, this model value is appended to RPC binding handle. This RPC binding handle will be passed to stubs, through client RPC call.

7.4. Logic of Compatibility Evaluation

The default OSF evaluation routines will use the following algorithm to determine the compatibility between a client and a server:

```
IF (client and server are using the same code set) {
    Great.  They are compatible;
    return (NO_CONVERSION_NECESSARY);
}

ELSE IF ("no character set comparison" is defined)
{
        Apply "connection model" evaluation;
        return (the result of "connection model" evaluation);
}

ELSE
{
    IF (client's code set has only 1 character set &&
        server's code set has only 1 character set)
    {
        IF (their character sets match)
        {
                Apply "connection model" evaluation;
                return (the result of "connection model"
                        evaluation);
        }
        ELSE
                return (IMPOSSIBLE);
    }
    ELSE
```

```
    {
        IF (more than 1 character sets match between client &
            server)
        {
                Apply "connection model" evaluation;
                return (the result of "connection model"
                        evaluation);
        }
        ELSE
                return (IMPOSSIBLE);
    }
}
```

7.5. Code Set Conversion

The actual code set conversion is done by the XPG4 iconv routines, so this code set conversion routine is a jacket routine to call iconv routine internally based on the tags value.

The tag value is expressed as an integer value, and since the XPG4 iconv routines are defined to take a string code set name, this routine calls the OSF Character and Code Set Registry routine to convert an integer code set to a string code set name.

There are four OSF supplied code set conversion (jacket) routines. For each data type, like `byte` or `wchar_t`, two routines are required to perform code set conversion. One routine converts the local data type for network representation, and is used for marshalling. The other routine converts the network representation of the data to the local data type, and is used for unmarshalling.

The signature for these routines are (`local_data_type` is a place holder):

```
void "local_data_type"_to_netcs (
        /* in */   rpc_binding_handle_t         h,
        /* in */   unsigned32                   tag,
        /* in */   "local data type"            *ldata,
        /* in */   unsigned32                   l_data_len,
        /* out */  "network data type"          *wdata,
        /* out */  unsigned32                   *p_w_data_len,
        /* out */  error_status_t               *p_st
);
void "local_data_type"_from_netcs (
        /* in */   rpc_binding_handle_t         h,
        /* in */   unsigned32                   tag,
        /* in */   "network data type"          *wdata,
        /* in */   unsigned32                   w_data_len,
        /* in */   unsigned32                   l_storage_len,
        /* out */  "local data type"            *ldata,
        /* out */  unsigned32                   *p_l_data_len,
        /* out */  error_status_t               *p_st
);
```

These signatures were defined by DEC. [DEC 1]

(a) `byte_to_netcs()`

This routine takes a (multi-)`byte` value, and converts it to the code set which is defined by a tag, when necessary. The result string is in byte form, and is sent over the wire. This routine is used for marshalling.

(b) `byte_from_netcs()`

This routine receives a `byte` stream from the wire, and converts it to the code set which is used by a server, when necessary. The result string is in `byte` form, and is passed to a manager routine. This

routine is used for unmarshalling.

(c) `wchar_t_to_netcs()`

This routine takes a process code (`wchar_t`) value, and first converts it to (multi-)`byte` string, then converts the string to the code set which is defined by a tag, when necessary. The result string is in `byte` form, and is sent over the wire. This routine is used for marshalling.

(d) `wchar_t_from_netcs()`

This routine receives a `byte` stream from the wire, and converts it to the code set which is used by a server, when necessary. Then this string is converted to process code (`wchar_t`) values of server process. This process code is passed to a manager routine. This routine is used for unmarshalling.

7.6. Character and Code Set Registry Access

The *OSF Character and Code Set Registry* includes a table in which each entry has a 32-bit integer value, a code set name and other supplementary information.

This table is a *code_set_table* file. Each Operating System provides its specific version of a *code_set_table* file, since supported code sets will vary from platform to platform.

In the case of OSF/1, this file is located in:

```
/usr/lib/nls/loc/code_set_table
```

The OSF/1 version of a *code_set_table* would look like the following. The items listed (from the left) are: code set name, 32-bit integer value, and character set ID(s). '#' identifies a comment line.

```
# ISO 8859-1:1987; Latin Alphabet No. 1
 ISO8859-1       0x00010001      0x0011

# ISO 8859-1:1987; Latin Alphabet No. 2
 ISO8859-1       0x00010002      0x0012
  ........

# AJEC:1991, Advanced Japanese EUC Code
 AJEC            0x00020010      0x0011:0x0080:0x0081:0x0082
  ........
```

Two routines are provided to access this table. One takes a string code set name and returns an integer value which represents its corresponding entry in the OSF Character and Code Set Registry. The other takes an integer value and returns a string code set name.

These code set registry routines basically perform a table search, the key being either an integer value or a code set name.

7.7. Setting Tags

The routine `rpc_cs_get_tags()` is called from rpc stubs before the code set conversion routine is called. This routine will take a binding handle, which contains the model information of how the client and server communicate, and sets the tags. When the binding model information is not set, the `rpc_cs_get_tags()` routine will determine a binding model, and will select the tags.

When this routine is called from a client stub, the *sending* tag and the *desired reception* tag are set, and the *reception* tag is ignored. When this routine is called from a server stub, the *reception* tag is set, and the *sending* tag is ignored. *Desired reception* tag is used as input to determine the most appropriate *reception* tag value.

To indicate from which stub it is called, the boolean value `server_side` is used in the signature of this routine. This signature was defined by DEC. [DEC 1]

```
void
rpc_cs_get_tags (h, server_side, p_stag, p_drtag, p_rtag, p_st)
        /* in */     rpc_binding_handle_t h;
```

```
        /* in */     boolean              server_side;
        /* out */    unsigned32           *p_stag;
        /* in/out */ unsigned32           *p_drtag;
        /* out */    unsigned32           *p_rtag;
        /* out */    error_status_t       *p_st;
```

7.8. Buffer Sizing for Marshalling and Unmarshalling

A buffer sizing routines are used within stubs before marshalling and unmarshalling of data, since when code set conversion is performed, there is always a chance that the conversion buffer needs to be enlarged to hold converted data.

The role of this buffer sizing routine is to determine the *conversion type*, and calculate the size of necessary buffer (when a new buffer is required) so the rpc stub can allocate the necessary buffer before it calls a conversion routine.

Conversion type will be either:

(a) **No conversion required.**

(b) **In place conversion.** — The code set conversion can be done in a single storage area; no buffer allocation is required.

(c) **New buffer conversion.** — The converted data must be written to a new storage area; the new buffer needs to be allocated.

These types are defined in `idl_cs_convert_t` enumeration type.

When *new buffer conversion* is selected as the conversion model, the following rule is used to calculate the new buffer size.

```
number-of-bytes * MB_LEN_MAX
```

number-of-bytes is passed to the stub from the network. *MB_LEN_MAX* is defined as a sytem dependent global variable, and can be accessed by including `<limits.h>` header file.

There are four OSF supplied memory management routines. For each data type, like `byte` or `wchar_t`, two routines are required to select the conversion type and calculate the buffer size. One routine is used for marshalling, and the other routine is used for unmarshalling.

Only when *new buffer conversion* type is selected, the buffer size will be different from the *storage length* of a local data.

The signature of these routines are:

```
void "local_data_type"_net_size (
        /* in */   rpc_binding_handle_t h,
        /* in */   unsigned32           tag,
        /* in */   unsigned32           l_storage_len,
        /* out */  idl_cs_convert_t     *p_convert_type,
        /* out */  unsigned32           *p_w_storage_len,
        /* out */  error_status_t       *p_st
);

void "local_data_type"_local_size (
        /* in */   rpc_binding_handle_t h,
        /* in */   unsigned32           tag,
        /* in */   unsigned32           w_storage_len,
        /* out */  idl_cs_convert_t     *p_convert_type,
        /* out */  unsigned32           *p_l_storage_len,
        /* out */  error_status_t       *p_st
);
```

These signatures were defined by DEC. [DEC 1]

(a) `byte_rpc_net_size()`

This routine is called before data marshalling.

It takes a binding handle, a tag that identifies the code set that will be used on the wire, and the local storage size in bytes.

In case of *new buffer conversion* type is required, it calculates the size of the new buffer in bytes.

(b) `byte_rpc_local_size()`

This routine is called before data unmarshalling.

It takes a binding handle, a tag that identifies the code set used on the wire for data being received, and the storage size of the on-the-wire data.

In case of *new buffer conversion* type is required, it calculates the size of the new buffer in bytes.

(c) `wchar_t_rpc_net_size()`

This routine is called before data marshalling.

It takes a binding handle, a tag that identifies the code set that will be used on the wire, and the local storage size in `wchar_t` data type.

In case of *new buffer conversion* type is required, it calculates the size of the new buffer in `wchar_t` data type.

(d) `wchar_t_rpc_local_size()`

This routine is called before data unmarshalling.

It takes a binding handle, a tag that identifies the code set used on the wire for data being received, and the storage size of the on-the-wire data.

In case of *new buffer conversion* type is required, it calculates the size of the new buffer in wchar_t data type.

8. DATA STRUCTURES

8.1. Data Structures Related to Code Set Conversion

```
/* This data enables a user routine to indicate whether I-char data
 * needs conversion from on-the-wire form to local form, and if so,
 * whether this conversion can be done in place
 *
 * This data type was defined by DEC.  [DEC 1]
 */

typedef enum {
     idl_cs_no_convert,          /* No code set conversion required */
     idl_cs_in_place_convert,    /* Code set conversion can be done
                                    in a single storage area */
     idl_cs_new_buffer_convert   /* The converted data must be
                                    written to a new storage area */
} idl_cs_convert_t;

/* These data contain supported code sets for either server or
 * client.
 *  - The first element in codeset[] is a local code set of the
 *    process.
 *  - Other code sets are code sets supported by the process, and
```

```
 *    they are converted to the local code set by iconv converters.
 *  - Comformant array is used, since a number of code sets used in a
 *    host will vary.
 */

typedef struct rpc_codeset_mgmt_s_t {
     unsigned32   version;      /* version of this structure */
     long size;                 /* number of code sets defined */
     [size_is(size)] long *codesets;
} rpc_codeset_mgmt_t, *rpc_codeset_mgmt_p_t;
```

8.2. Data Structures Related to Interoperability Evaluation

```
/* Appending communicating model information to a binding handle.
 *  - This is a modification to the current binding handle which will
 *    be returned from rpc_ns_binding_import_next().  This code set
 *    information is used by rpc_cs_get_tags() routine to set tags.
 *  - When rpc_ns_import_ctx_push_eval() is called by an application
 *    rpc_ns_binding_import_next() will execute a new logic to
 *    determine code set compatibility, and allocate this data
 *    structure to hold desired code set information.
 */

typedef struct {
        rpc_binding_handle_t    bind;   /* binding handle */
        unsigned32              model;  /* communicating model */
        rpc_codesets_t          server; /* server's code sets */
        rpc_codesets_t          client; /* client's code sets */
} rpc_binding_eval_t, *rpc_binding_eval_p_t;

/* Appending evaluation function pointer(s) to a import context
 * handle.
 *  - This is a modification to the current import context which
 *    is returned from rpc_ns_binding_import_begin().
 *  - New stacks structure will be allocated by new
 *    rpc_ns_import_ctx_push_eval() routine, and is appended to
 *    a context handle, which is passed to
 *    rpc_ns_binding_import_next() routine.
 *  - At the beginning to rpc_ns_binding_import_next() routine,
 *    evaluation function will be pop-ed up from the stacks, and be
 *    executed.
 *  - The signature of an evaluation function is
 *      void eval_func (
 *              /* [in] */  void        *args;
 *              /* [out] */ unsigned32, *model;
 *              /* [out] */ unsigned32  *status;
 *      )
 *  - argument of a function is a void pointer.
 *    In the case of an evaluation function, this is a null pointer.
 */

typedef struct rpc_eval_stacks {
   unsigned32   type;                   /* function type */
   void         (*func)(void *args);    /* function pointer */
   void         *args;                  /* arguments of a function */
   void         *cntx;                  /* for preserving context */
```

```
   rpc_cs_eval_stack_p    next;         /* points to next stack
                                           item */
} rpc_cs_eval_stack, *rpc_cs_eval_stack_p;

typedef struct {
        rpc_import_rep_t        c;      /* import context */
        unsigned32              n;      /* number of items in a
                                           stack */
        rpc_cs_eval_stack_p     s;      /* points to a stack */
} rpc_import_func_t, *rpc_import_func_p_t;
```

8.3. Data Structures Related to NSI Attributes Management

```
/* NS Attribute string constants
 *  - There are currently five attributes defined.  So at the end of
 *    these attributes, code sets attribute and arbitrary attribute
 *    are appended.
 * (nsp.h)
 */

#define rpc_c_ns_dna_towers     ((unsigned_char_p_t) "1.3.22.1.3.30")
#define rpc_c_ns_class_version  ((unsigned_char_p_t) "1.3.22.1.1.1")
#define rpc_c_ns_object_uuids   ((unsigned_char_p_t) "1.3.22.1.1.2")
#define rpc_c_ns_group          ((unsigned_char_p_t) "1.3.22.1.1.3")
#define rpc_c_ns_profile        ((unsigned_char_p_t) "1.3.22.1.1.4")
#define rpc_c_ns_codesets       ((unsigned_char_p_t) "1.3.22.1.1.5")

#define rpc_c_attr_class_version    0
#define rpc_c_attr_dna_towers       1
#define rpc_c_attr_object_uuids     2
#define rpc_c_attr_group            3
#define rpc_c_attr_profile          4
#define rpc_c_attr_codesets         5
#define rpc_c_attr_max              6

/* This is a new structure to hold code set attribute.
 * code set attribute is encoded by IDL Encoding Services before
 * it is exported to CDS namespace.
 */

typedef struct {
   unsigned32              value_length; /* <= MAXATTRIBUTE (4000) */
   unsigned_char_p_t       *value;
} rpc_codesets_value_t, *rpc_codesets_value_p_t;

/* Addition to 'Name Service Attributes String constants'
 * - these names are converted to opaque string internally.
 * (nsp.h)
 */

GLOBAL rpc_ns_attributes_t rpc_g_attr_table[rpc_c_ns_max]
                                                [rpc_c_attr_max] =
{
    {
        {rpc_c_ns_class_version},
        {rpc_c_ns_dna_towers},
```

```
        {rpc_c_ns_object_uuids},
        {rpc_c_ns_group},
        {rpc_c_ns_profile},
        {rpc_c_ns_codesets}
    }
}

/* Additional types of RPC memory objects
 * (rpcmem.h)
 */

#define rpc_c_mem_avail         0
/*......*/
#define rpc_c_mem_codesets  83      /* rpc_codesets_value_t */
#define rpc_c_mem_func      84      /* rpc_import_func_t    */
#define rpc_c_mem_eval      85      /* rpc_binding_eval_t   */
#define rpc_c_mem_stack     86      /* stack                */

/* can only use up to "rpc_c_mem_maxtypes - 1" without upping it */
#define rpc_c_mem_max_types    90       /* i.e., 0..(max_types-1) */
```

9. USER INTERFACES

DCE Shell's `nsi-entry` class will include a method to manipulate a code set attribute in CDS namespace. The details of the method are TBD.

10. API'S

10.1. Interface to the Local Operating System

```
/* This function will allocate rpc_codesets_p_t data, and fills the
 * supported code sets values.  It uses nl_langinfo() to determine
 * the current code set, and access /usr/lib/nls/loc (LOCPATH in
 * OSF/1) directory to determine which iconv converters are available
 * in the system.  Then access character and code set registry to get
 * the integer value of each code set.
 */

void rpc_rgy_get_codesets (
   /* [OUT] */  rpc_codesets_p_t        *codesets_p,
   /* [OUT] */  unsigned32              *status
);
```

10.2. NSI Management

```
/* Management routines to register (code sets) attributes to CDS.
 * rpc_ns_mgmt_set_attribute()
 *  - This routine takes attribute type and a pointer to the value,
 *    and add it to CDS entry.  pointer to the value is declared as
 *    'void', means this routine is designed not only for code sets
 *    attributes, but also for other opaque user supplied values.
 *
 * Reading these attributes:
 *  rpc_ns_mgmt_read_attr_begin()
 *   - This routine is called to setup inquiry context.
 *  rpc_ns_mgmt_read_attr_next()
 *   - This routine reads attributes based on the inquiry context.
```

```
 *  rpc_ns_mgmt_read_attr_done() routines are used.
 *   - This routine finishes the operation.
 *
 *  rpc_ns_mgmt_read_codeset()
 *   - A convenience function for reading 'code set' attribute.
 *     This routine calls the above three routines internally.
 *
 * rpc_ns_mgmt_remove_attribute()
 *  - This routine is similar to rpc_ns_mgmt_binding_unexport()
 *    It takes entry name & syntax, and attribute type, then
 *    remove the value of that attribute.
 */

void rpc_ns_mgmt_set_attribute (
   /* [IN] */   unsigned32              entry_name_syntax,
   /* [IN] */   unsigned_char_p_t       entry_name,
   /* [IN] */   unsigned32              attr_type,
   /* [IN] */   void                    *attr_val,
   /* [OUT] */  unsigned32              *status
);
/* In case of code sets value are exported, attr_val will point to
 * rpc_codesets_t data structure.  This routine will convert
 * rpc_codeset_t data structure to an endian-safe format by using the
 * IDL Encoding Services before it actually export the value.
 */

/* This is very similar to rpc_ns_entry_object_inq_begin except it
 * takes attr_type as a signature.  attr_type is rpc_c_attr_codesets.
 * Other types will result rpc_s_mgmt_op_disallowed error.
 */

void rpc_ns_mgmt_read_attr_begin (
   /* [IN] */   unsigned32              entry_name_syntax,
   /* [IN] */   unsigned_char_p_t       entry_name,
   /* [IN] */   unsigned32              attr_type,
   /* [OUT] */  rpc_ns_handle_t         *inquiry_context,
   /* [OUT] */  unsigned32              *status
);

/* This is very similar to rpc_ns_entry_object_inq_next but it
 * is designed to handle arbitrary string.
 * attr_bytes will be rpc_c_attr_codesets.
 * Other types will result rpc_s_mgmt_op_disallowed error.
 * When value is returned, it is cast-ed to (void *), but it will be
 * rpc_codesets_p_t
 */

void rpc_ns_mgmt_read_attr_next (
   /* [IN] */   rpc_ns_handle_t         inquiry_context,
   /* [IN] */   unsigned32              attr_type,
   /* [OUT] */  void                    *value,
   /* [OUT] */  unsigned32              *length,
   /* [OUT] */  unsigned32              *status
);
```

```
/* This is very similar to rpc_ns_entry_object_inq_done.
 * Only difference is that it will checkrpc_e_codesets_member
 * as context usage.
 */

void rpc_ns_mgmt_read_attr_done (
   /* [IN] */   rpc_ns_handle_p_t       inquiry_context,
   /* [OUT] */  unsigned32              *status
);

/* This is a convenience routine for reading the 'code set' attribute
 * value.
 */

void rpc_ns_mgmt_read_codeset (
   /* [IN] */   unsigned32              entry_name_syntax,
   /* [IN] */   unsigned_char_p_t       entry_name,
   /* [OUT] */  rpc_codesets_p_t        codeset_val,
   /* [OUT] */  unsigned32              *status
);

void rpc_ns_mgmt_remove_attribute (
   /* [IN] */   unsigned32              entry_name_syntax,
   /* [IN] */   unsigned_char_p_t       entry_name,
   /* [IN] */   unsigned32              attr_type,
   /* [OUT] */  unsigned32              *status
);
```

10.3. Evaluation

```
/* This function will allocate rpc_import_func_t data and stack data.
 * Then it sets code set compatibility evaluation function pointer in
 * a stack structure.
 * Application programmer is responsible for providing desired
 * code set compatibility evaluation function, or routines which OSF
 * provided can be used.
 * This function is called with import context which is allocated
 * by rpc_ns_binding_import_begin().
 * func_type is CODE_EVAL_FUNC only for now, but it can be extended
 * in the future.
 */

void rpc_ns_import_ctx_push_eval (
        /* [IN/OUT] */ rpc_ns_handle_p_t        import_ctxt,
        /* [IN] */ void                         (*eval_func),
        /* [IN] */ void                         *args,
        /* [IN] */ unsigned32                   func_type,
        /* [OUT] */ unsigned32                  *status
);

/* Sample code set compatibility evaluation functions.
 * OSF provides these functions for reference.  Application
 * developers are free to write up their own evaluation functions, as
 * long as they don't change the signature.
 */
```

```
void rpc_cs_eval_dynamic(
        /* [IN] */  void                *args,
        /* [OUT] */ unsigned32          *model,
        /* [OUT] */ rpc_codesets_p_t    server,
        /* [OUT] */ rpc_codesets_p_t    client,
        /* [OUT] */ unsigned32          *status
);

void rpc_cs_eval_cmir(
        /* [IN] */  void                *args,
        /* [OUT] */ unsigned32          *model,
        /* [OUT] */ rpc_codesets_p_t    server,
        /* [OUT] */ rpc_codesets_p_t    client,
        /* [OUT] */ unsigned32          *status
);

void rpc_cs_eval_smir(
        /* [IN] */  void                *args,
        /* [OUT] */ unsigned32          *model,
        /* [OUT] */ rpc_codesets_p_t    server,
        /* [OUT] */ rpc_codesets_p_t    client,
        /* [OUT] */ unsigned32          *status
);

void rpc_cs_eval_rmir(
        /* [IN] */  void                *args,
        /* [OUT] */ unsigned32          *model,
        /* [OUT] */ rpc_codesets_p_t    server,
        /* [OUT] */ rpc_codesets_p_t    client,
        /* [OUT] */ unsigned32          *status
);

void rpc_cs_eval_universal(
        /* [OUT] */ void                *args,
        /* [OUT] */ unsigned32          *model,
        /* [OUT] */ rpc_codesets_p_t    server,
        /* [OUT] */ rpc_codesets_p_t    client,
        /* [OUT] */ unsigned32          *status
);

/* These is no signature change to rpc_ns_binding_import_next(),
 * however, new logic will be added at the beginning of the routine
 * to check if import_context is actually rpc_import_func_t
 * structure.  If it is, the new logic will dereference code set
 * evaluation function pointer to execute it.
 * This code set evaluation function will return a model of client
 * server communication.  A model value will be appended to binding
 * handle to be reference by stubs.
 */

void rpc_ns_binding_import_next (
        /* [IN] */ rpc_ns_handle_t              import_ctxt,
        /* [OUT] */ rpc_binding_handle_t        *binding,
        /* [OUT] */ unsigned32                  *status
);
```

10.4. Tags Set Routine

```
/* This routine takes binding handle, and determines the tags value
 * depending of a model information appended to a binding handle.
 *
 * The signature of this routine is defined by DEC [DEC 1]
 */

void rpc_cs_get_tags (
        /* [IN] */ rpc_binding_handle_t       h,
        /* [IN] */ boolean                    server_side,
        /* [OUT] */ unsigned32                *p_stag,
        /* [IN/OUT] */ unsigned32             *p_drtag,
        /* [OUT] */ unsigned32                *p_rtag,
        /* [OUT] */ error_status_t            *p_st
);

/*
 * p_stag  -> [OUT] client side, ignored server side
 * p_drtag -> [OUT] client side, [IN] server side
 * p_rtag  -> [OUT] server side, ignored client side
 */
```

10.5. Buffer Sizing and Code Set Conversion

```
/* This routine takes binding handle, tag and local storage size, and
 * calculate the necessary buffer size.  This needs to be done before
 * marshalling of data.
 *
 * The signatures of these routines are defined by DEC.
 */

void byte_net_size (
        /* [IN] */ rpc_binding_handle_t         h,
        /* [IN] */ unsigned32                   tag,
        /* [IN] */ unsigned32                   l_storage_len,
        /* [OUT] */ idl_cs_convert_t            *p_convert_type,
        /* [OUT] */ unsigned32                  *p_w_storage_len,
        /* [OUT] */ error_status_t              *p_st
);

void wchar_t_net_size (
        /* [IN] */ rpc_binding_handle_t         h,
        /* [IN] */ unsigned32                   tag,
        /* [IN] */ unsigned32                   l_storage_len,
        /* [OUT] */ idl_cs_convert_t            *p_convert_type,
        /* [OUT] */ unsigned32                  *p_w_storage_len,
        /* [OUT] */ error_status_t              *p_st
);

/* These routines convert data from local format to on-the-wire
 * format. These routines are used from the stubs.
 */

void byte_to_netcs (
        /* [IN] */ rpc_binding_handle_t         h,
```

```
        /* [IN] */ unsigned32                   tag,
        /* [IN] */ idl_byte                     *ldata,
        /* [IN] */ unsigned32                   l_data_len,
        /* [OUT] */ idl_byte                    *wdata,
        /* [OUT] */ unsigned32                  *p_w_data_len,
        /* [OUT] */ error_status                *p_st
);

void wchar_t_to_netcs (
        /* [IN] */ rpc_binding_handle_t         h,
        /* [IN] */ unsigned32                   tag,
        /* [IN] */ wchar_t                      *ldata,
        /* [IN] */ unsigned32                   l_data_len,
        /* [OUT] */ idl_byte                    *wdata,
        /* [OUT] */ unsigned32                  *p_w_data_len,
        /* [OUT] */ error_status                *p_st
);

/* These routine take binding handle, tag and wire storage size, and
 * calculate the necessary buffer size.  This needs to be done before
 * unmarshalling of data.
 */

void byte_local_size (
        /* [IN] */ rpc_binding_handle_t         h,
        /* [IN] */ unsigned32                   tag,
        /* [IN] */ unsigned32                   w_storage_len,
        /* [OUT] */ idl_cs_convert_t            *p_convert_type,
        /* [OUT] */ unsigned32                  *p_l_storage_len,
        /* [OUT] */ error_status_t              *p_st
);

void wchar_t_local_size (
        /* [IN] */ rpc_binding_handle_t         h,
        /* [IN] */ unsigned32                   tag,
        /* [IN] */ unsigned32                   w_storage_len,
        /* [OUT] */ idl_cs_convert_t            *p_convert_type,
        /* [OUT] */ unsigned32                  *p_l_storage_len,
        /* [OUT] */ error_status_t              *p_st
);

/* These routines convert data from on-the-wire format to local
 * format.  This is called from stubs.
 */

void byte_from_netcs (
        /* [IN] */ rpc_binding_handle_t         h,
        /* [IN] */ unsigned32                   tag,
        /* [IN] */ idl_byte                     *wdata,
        /* [IN] */ unsigned32                   w_data_len,
        /* [IN] */ unsigned32                   l_storage_len,
        /* [OUT] */ idl_byte                    *ldata,
        /* [OUT] */ unsigned32                  *p_l_data_len,
        /* [OUT] */ error_status_t              *p_st
);
```

```
void wchar_t_from_netcs (
        /* [IN] */ rpc_binding_handle_t         h;
        /* [IN] */ unsigned32                   tag;
        /* [IN] */ idl_byte                     *wdata;
        /* [IN] */ unsigned32                   w_data_len;
        /* [IN] */ unsigned32                   l_storage_len;
        /* [OUT] */ wchar_t                     *ldata;
        /* [OUT] */ unsigned32                  *p_l_data_len;
        /* [OUT] */ error_status_t              *p_st;
);
```

10.6. Character and Code Set Registry

```
/* These routines are for the OSF Character and Code Set Registry
 * access.  Since iconv only accepts string as code set name, while
 * code set are represented in integer for internal operations, we
 * need string <-> int value conversion routines.  Also this registry
 * ensures code set compatibility between vendors, and in the
 * heterogeneous network environments.
 */

/* Convert string code set name to integer code set value. */

void dce_cs_loc_to_rgy (
        /* [IN] */ const char                   *local_code_set_name,
        /* [OUT] */ unsigned32                  *rgy_code_set_value,
        /* [OUT] */ error_status_t              *status
);

/* Convert integer code set value to string code set name. */

void dce_cs_rgy_to_loc (
        /* [IN] */ const unsigned32             rgy_code_set_value,
        /* [OUT] */ char                        **local_code_set_name,
        /* [OUT] */ error_status_t              *status
);
```

11. REMOTE INTERFACES

The following are the interface definition and ACF for *IDL Encoding Services* routines used in NSI management to encode and decode **code set** attribute.

11.1. IDL File

```
interface rpc_nsi_management
{
       /* Encoding routine */
       void rpc__codesets_to_nscodesets (
               [in] handle_t   h,
               [in] long       num,
               [in, size_is(num)] long       codesets[]
       );

       /* Decoding routine */
       void rpc__nscodesets_to_codesets (
               [in] handle_t   h,
               [out] rpc_codeset_mgmt_p_t data
```

```
        );
}
```

11.2. ACF File

```
interface rpc_nsi_management
{
        [encode] rpc__codesets_to_nscodesets();
        [decode] rpc__nscodesets_to_codesets();
}
```

12. MANAGEMENT INTERFACES

Not applicable.

13. RESTRICTIONS AND LIMITATIONS

(a) `auto_handle` and *Customized handles* cannot be used with the automatic code set conversion feature.

(b) `pipe` type cannot be used with the automatic code set conversion feature.

(c) Automatic code set conversion feature cannot be used when an array has more than one dimension, or any of the attribute `min_is`, `max_is`, `first_is`, `last_is` or `string` has been applied to an array.

(d) Automatic code set conversion feature cannot be used with `transmit_as` or `represent_as` attributes.

(e) Automatic code set conversion feature cannot be applied to a type which is the target of a pointer with the `[size_is]` or `[max_is]` attribute. For example, following IDL fragment is not acceptable:

```
typedef byte my_byte;

typedef struct {
      long s;
      [size_is(s)] my_byte *arrayfield;
} problem_struct;
```

But the same information can be carried by the following:

```
typedef struct {
      long s;
      [size_is(s)] my_byte arrayfield[];
} conf_struct;
```

(f) If an array parameter has a `[cs_char]` base type, it cannot have `[ptr]` or `[unique]` as parameter attributes. This forbids the cases illustrated by the following, in which the arrays are being treated as pointer targets:

```
void wrong_proc (
       [in] long s1,
       [in, unique, size_is(s1)] my_byte ua[],
       [in] long s2,
       [in, ptr, size_is(s2)] my_byte pa[]
);
```

It does not forbid the following, which is the normal form of an arra parameter:

```
void right_proc (
       [in] long s1,
       [in, size_is(s1)] my_byte ra[]
);
```

14. OTHER COMPONENT DEPENDENCIES

(a) *I18N IDL complier extension* will be done at DEC Nashua, so the completion of this work depends on their work.

(b) *IDL Encoding Services* is a part of 1.1 deliverables from DEC. Encoding/decoding a **code set** attribute into/from an endian-safe format depends on this new feature.

(c) *DCE Shell* is a new facility for 1.1 release. A manipulation of code set attributes by DCE Shell command depends on its availability.

15. COMPATIBILITY

(a) DCE 1.0 IDL compilers will not understand new syntax for **I18N** characters.

(b) DCE 1.0 RPC library will not support the new data structures and new APIs.

16. STANDARDS

(a) POSIX 1003.2/XPG4 I18N facilities and functionalities. (locale, iconv, and other library routines).

17. OPEN ISSUES

(a) The DCE Shell code set manipulatin interface (method) is not yet decided. However, OSF will provide a method to perform, at least, add, show, export, and delete operations for the **code set** attribute.

APPENDIX A. SAMPLE IDL AND ACF USING THE NEW FEATURES

A.1. Fixed Array Example

A very simple example, which exchanges fixed arrays of characters as both input and output, will have IDL and ACF files like the following.

Even though the tags are exposed in the RPC signature, actual tags values will be set by `[cs_tag_rtn(set_tags)]` in ACF definition. So application developers do not need to be aware of the actual tag values.

IDL:

```
[
uuid(001bb502-2190-1be8-b507-08002b0f59bb),
version(1.0)
]
interface fixed_string_example
{
   const long SIZE = 100;
   typedef byte my_byte;

   void fixed_string(
       [in] unsigned long stag,
       [in] unsigned long drtag,
       [out] unsigned long *p_rtag,
       [in,out] my_byte my_string[SIZE]
   );
}
```

ACF:

```
[
implicit_handle(handle_t global_binding_h)
]
interface fixed_string_example
```

```
{
        typedef [cs_char(wchar_t)] my_byte;

        [cs_tag_rtn(rpc_cs_get_tags)]
        fixed_string([cs_stag] stag,
                     [cs_drtag] drtag,
                     [cs_rtag] p_rtag );
}
```

A.2. Conformant Varying Array Example

DEC Design Note [DEC 1] explains the detail of the IDL features. So this is the excerpt from the document. This example is for conformant varying array as both `[in]` and `[out]` argument.

Suppose an IDL file contains:

```
typedef byte my_byte;

void a_op(
         [in] unsigned long stag,
         [in] unsigned long drtag,
         [out] unsigned long *p_rtag,
         [in] long s,
         [in, out] long *p_l,
         [in, out, size_is(s), length_is(*p_l)] my_byte a[]
);

void b_op(
         [in] unsigned long stag,
         [in] unsigned long drtag,
         [out] unsigned long *p_rtag,
         [in] long s,
         [in, out] long *p_l,
         [in, out, size_is(s), length_is(*p_l)] my_byte a[]
);
```

and the associated ACF file contains:

```
/* ltype can be wchar_t, char, or application specific type */
typedef [cs_char(ltype)] my_byte;

[cs_tag_rtn(rpc_cs_get_tags)]
                        a_op( [cs_stag] stag,
                              [cs_drtag] drtag,
                              [cs_rtag] p_rtag );

                        b_op( [cs_stag] stag,
                              [cs_drtag] drtag,
                              [cs_rtag] p_rtag );
```

Then the generated header file will contain:

```
typedef byte my_byte;

void a_op(
       /* [in] */      idl_long_int     s,
       /* [in, out] */ idl_long_int     *p_l,
       /* [in, out, size_is(s), length_is(*p_l)] */ ltype a[]
);
```

```
void b_op(
       /* [in] */      idl_ulong_int    stag,
       /* [in] */      idl_ulong_int    drtag,
       /* [out] */     idl_ulong_int    *p_rtag,
       /* [in] */      idl_long_int     s,
       /* [in, out] */ idl_long_int     *p_l,
       /* [in, out, size_is(s), length_is(*p_l)] */ ltype a[]
);
```

The stubs for `a_op` will call `rpc_cs_get_tags` before marshalling data. For `b_op` it is the application code's responsibility to make sure that the tags are set correctly before data is marshalled. [DEC 1]

If application programmers decide to set tags by themselves, they can use:

(a) new routine (`rpc_rgy_get_codesets()`) to get supported code sets information;

(b) new routine (`rpc_ns_mgmt_read_codeset()`) to read the code set information; and

(c) set tags based on supported code set information and CDS code set attribute value. It might be necessary to write code set compatibility evaluation function to set tags value properly.

APPENDIX B. SAMPLE SERVER AND CLIENT CODE

B.1. Server Side

Using this modification to RPC application development, the new server flow will be like the following (> indicates the new steps):

(a) Register the interface.

(b) Create binding information.

(c) Obtain the binding information.

(d) Advertise the server location in the name service database.

> (e) Get server's suppored code sets.

> (f) Register code sets attributes to CDS.

(g) Register the endpoints in the local endpoint map.

(h) Free the set of binding handles.

(i) Listen for remote call.

> (j) Remove the code sets attributes from CDS.

Note, however, when DCE Shell `nsi-entry` method is used to advertise the supported code set information, these extra steps are not necessary. [DCESH]

So simple server code will look as in the following subsection.

B.1.1. Sample server code

```
void main (void)
{
       unsigned32              status;
       rpc_binding_vector_t    *binding_vector;
       unsigned_char_t         *entry_name = "/.:/my_app_entry";

       rpc_codesets_t          codesets;

       setlocale(LC_ALL, "");  /* <-- NOTE */
```

```
rpc_server_register_if (
        myif_v1_0_s_ifspec,             /* if_handle */
        NULL,                           /* mgr_type_uuid */
        NULL,                           /* mgr_epv */
        &status );

rpc_server_use_all_protseqs (
        rpc_c_protseq_max_reqs_default,/* max_calls_request */
        &status );

rpc_server_inq_bindings (
        &binding_vector,
        &status );

rpc_ns_binding_export (
        rpc_c_ns_syntax_default,       /* entry_name_syntax */
        entry_name,
        myif_v1_0_s_ifspec,            /* if_handle */
        binding_vector,
        NULL,                          /* object_uuid_vec */
        &status );

rpc_reg_get_codesets (  /* <-- NOTE */
        &codesets,                     /* [OUT] codesets */
        &status );

rpc_ns_mgmt_set_attribute (  /* <-- NOTE */
        rpc_c_ns_syntax_default,       /* name entry syntax */
        entry_name,                    /* name entry */
        rpc_c_attr_codesets,           /* attribute type */
        (void *)&codesets,             /* attribute_value */
        &status );

rpc_ep_register (
        myif_v1_0_s_ifspec,            /* if_handle */
        binding_vector,
        NULL,                          /* object_uuid_vec */
        NULL,                          /* annotation */
        &status );

rpc_binding_vector_free (
        &binding_vector,
        &status );

rpc_server_listen (
        rpc_c_listen_max_calls_default, /* max_calls_exec */
        &status );

rpc_ns_mgmt_remove_attribute (  /* <-- NOTE */
        rpc_c_ns_syntax_default,
        entry_name,
        rpc_c_attr_codesets,
        &status );
```

B.2. Client Side

When client code uses OSF supplied macro, there is a very few codes application developers should supply. This is the very simple client code for fixed array operation. (Macro name TBD.)

The following example uses *implicit binding* method, but an application can also use *explicit binding* method when it's needed. When *explicit binding* method is used, ICHAR_SET macro needs to be modified. The first subsection below shows that the client programmer calls the ICHAR_SET macro; the second subsection shows what it looks like after the macro is expanded. (Note, this example omits necessary error checkings to highlight the main flow of the procedure.)

B.2.1. Client side code

```
main ( )
{
        wchar_t                 in_out_string[100];

        setlocale(LC_ALL, "");

        /*
         * arguments to the macro are:
         *      1) interface definition specification
         *      2) NSI entry name
         *
         * 1) Interface definition specification is created by IDL
         *    compiler.
         * 2) NSI entry name should match with the name that server
         *    uses.
         */

        ICHAR_SET(myif_v1_0_c_ifspec, "/.:/my_app_entry");

        my_rpc (in_out_string);
}
```

B.2.2. Expanded client code

```
main ( )
{
       wchar_t                 in_out_string[100];

       unsigned32              status;
       rpc_ns_handle_t         import_context;

       rpc_ns_binding_import_begin (
               rpc_c_ns_syntax_default,        /* defined in nsp.h */
               "/.:/my_app_entry",
               myif_v1_0_c_ifspec,             /* if_handle */
               NULL,
               &import_context,
               &status );

       rpc_ns_import_ctx_push_eval (  /* <-- NOTE */
               &import_context,
               rpc_cs_eval_dynamic,
               (void *)"/.:/my_app_entry",
               CODE_EVAL_FUNC,
               &status );
```

```
        while (1) {
                rpc_ns_binding_import_next (
                        import_context,
                        &global_binding_h,      /* declared in ACF */
                        &status );

                if (status == rpc_s_ok)
                        break;

                rpc_binding_free (
                        global_binding_h,
                        &status
                );
        }

        rpc_ns_binding_import_done (
                        &import_context,
                        &status );

        fixed_string (in_out_string);
}
```

REFERENCES

Some of the following references were not cited explicitly in the body of this document, but nevertheless they all have played a factor in the background of this work.

[DCESH] H. Melman ‘‘DCE Shell Functional Specification’’, to appear.

[DEC 1] T. Hinxman, ‘‘Design Note RPCLang 015 International Character Support in IDL’’, February 1993.

[DEC 2] T. Hinxman, ‘‘IDL Encoding Services’’.

[OGURA] T. Ogura, ‘‘DCE 1.1 I18N Workbook’’, October, 1992.

[RFC 2.1] J. Harrow, ‘‘Proposed Enhancements for DCE 1.1 IDL’’, July 1992.

[RFC 13.0] A. Thormodsen, ‘‘DCE 1.1 Internationalization Requirements’’, January, 1993.

[RFC 23.0] R. Mackey, ‘‘DCE 1.1 Internationalization Guide’’, January, 1993.

[RFC 27.0] S. Martin, ‘‘Coded Character Set Conversions and Data Loss: Providing Interoperability While Preventing Loss’’, December 1992.

[RFC 34.0] H. Melman, ‘‘DCE 1.1 Coding Style Guide’’, April 1993.

[RFC 39.0] HP/IBM, ‘‘An Internationalized DCE Character Handling Proposal’’, March 1993.

[RFC 40.0] S. Martin, ‘‘OSF Character and Code Set Registry’’, April 1993.

[SHIRLEY] J. Shirley, ‘‘Guide to Writing DCE Applications’’, O’Reilly publishers, June 1992.

[XOPEN] X. Open, ‘‘Internationalisation Guide’’.

[XPG4] X/Open Portability Guide (Version 4) — 1993, System Interfaces and Headers.

AUTHOR'S ADDRESS

Mariko Mori
Open Software Foundation
11 Cambridge Center
Cambridge, MA 02142
USA
Internet email: mori_m@osf.org
Telephone: +1-617-621-8981

Richard Mackey
Open Software Foundation
11 Cambridge Center
Cambridge, MA 02142
USA
Internet email: dmackey@osf.org
Telephone: +1-617-621-8924