buidos

omunicación en Sistemas Distribuidos

ernando Pérez Costoya — José Mª Peña Sánchez
Mª de los Santos Pérez Hernández

Índice

- Introducción
- Paso de mensajes
  - Comunicación punto a punto
  - Comunicación de grupo
  - Sistemas de colas de mensajes (MOM)
- Llamadas a procedimientos remotos (RPC)
  - Sun/ONC RPC
- Invocación de métodos remotos (RMI)
  - Java RMI y CORBA

Introducción

- Sistema de comunicación: Espina dorsal del SD
- Paradigmas de comunicación
  - ¿Qué API de comunicación ofrece SD a las aplicaciones?
- Memoria compartida (¿en SD?) vs. Paso de mensajes
- Adecuación del paradigma a las distintas arquitecturas
  - cliente/servidor; editor/subscriptor; P2P
- Grado de acoplamiento del paradigma: espacial y temporal
- Comunicación persistente
  - SD almacena información hasta que destinatario(s) lo obtenga(n)
  - Incluso aunque emisor ya no exista → desacoplamiento temporal
- Patrón de comunicación (cardinalidad):
  - unidifusión versus multidifusión

omunicación

istentes (sockets, MPI,...)

rsistentes (ISIS, JGroups,...)
Message-oriented middleware)

otos (RPC) (ONC, DCE,...)
(RMI)
objetos (Java RMI, CORBA,...)
ntados a servicios

ernando Pérez Costoya — José Mª Peña Sánchez
Mª de los Santos Pérez Hernández

Sistemas Distribuidos

Paso de mensajes

Índice

- Introducción
- Paso de mensajes
  - Comunicación punto a punto
  - Comunicación de grupo
  - Sistemas de colas de mensajes (MOM)
- Llamadas a procedimientos remotos (RPC)
  - Sun/ONC RPC
- Invocación de métodos remotos (RMI)
  - Java RMI y CORBA

de mensajes

PI de paso de mensajes
a sobre protocolo de transporte
tos como:

ica origen/destino de comunicación?

ero-copy)

e funcionalidad (sockets, MPI)

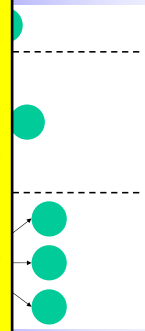
API de paso de mensajes

- MPI
  int MPI_Send(void *buf, int count, MPI_Datatype datatype, int dest, int tag, MPI_Comm comm)
  int MPI_Recv(void *buf, int count, MPI_Datatype, int source, int tag, MPI_Comm comm, MPI_Status *status)
- Sockets datagrama
  ssize_t sendto(int socket, const void *buffer, size_t length, int flags, const struct sockaddr *dest_addr, socklen_t dest_len);
  ssize_t recvfrom(int socket, void * buffer, size_t length, int flags, struct sockaddr *address, socklen_t * address_len);
- Esquemas con conexión
  - Existen además primitivas para conectar y desconectar
  - Operaciones de envío y recepción no incluyen direcciones

Esquemas de direccionamiento

- Usando número de proceso (MPI):
  - En envío: nº proceso destinatario
  - En recepción: nº proceso origen; sólo interacción $1 \rightarrow 1$
    - O cualquiera (*MPI_ANY_SOURCE*): interacción $N \rightarrow 1$
  - Difícil asignar nº proceso único en entorno de propósito general
    - Pero no en aplicación ejecutada en entorno de computación paralela
- Usando puertos: buzón asociado a una máquina (sockets)
  - Comunicación entre puertos
  - Proceso reserva uso de un puerto de su máquina (*bind* de sockets)
  - Envío: desde puerto origen local a puerto destino especificados
  - Recepción: de puerto local; interacción $N \rightarrow 1$
  - Sockets INET: ID puerto = dir. IP + nº puerto + protocolo (TCP|UDP)
- Usando colas: buzón de carácter global; interacción $N \rightarrow N$
  - Sistemas de colas de mensajes; desacoplamiento espacial+temporal

punto-a-punto



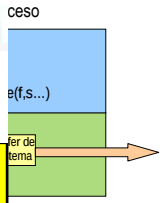
Especificación del mensaje

- Con o sin información de tipos (MPI vs. sockets)
  - Sin ella: aplicación debe gestionar heterogeneidad
  - Con ella: sistema de comunicaciones gestiona heterogeneidad
- Clases de mensajes (etiquetas)
  - Sistema de comunicación puede gestionar clases de mensajes
    - En envío: especifica clase de mensaje enviado
    - Recepción: especifica clase de mensaje que se quiere recibir
      - o usa comodín (*MPI_ANY_TAG*)
  - Múltiples canales sobre una misma comunicación
  - Diversas aplicaciones como por ejemplo:
    - Establecer prioridades entre mensajes
    - En cliente-servidor puede identificar operación a realizar
    - En editor-subscriptor basado en temas como identificador de tema
  - Disponible en MPI como parámetro de primitivas (*tag*)
  - No soportado en sockets, aunque sí mensajes urgentes (OOB)

Zero-Copy

- Reducir al mínimo (≈ a cero) copias entre zonas de memoria
- *Escenario 1*: envío de *N* datos *dispersos* de emisor a receptor
  - N envíos: sobrecarga de llamada de as + fragmentación de mensajes
  - Reserva de *buffer* y 1 envío: sobrecarga de copias
  - Funciones *scatter/gather*: minimizar copias y llamadas
    - UNIX: *readv*, *writev*, *sendmsg*, *recvmsg*
- *Escenario 2*: envío de un fichero
  - Uso de operaciones convencionales de lectura y envío
    - Dos copias de memoria: de *buffer* de sistema a de usuario y viceversa
  - Uso de proyección de ficheros (*mmap*) y envío
    - Una copia de memoria a *buffer* de sistema en envío
  - Uso de operaciones de transferencia directa entre descriptores
    - No requiere copias entres *buffers*; reduce nº llamadas al sistema
    - Linux: *sendfile*, *splice*

copy de fichero



Integridad de los mensajes

- En recepción debe especificarse buffer de tamaño ≥ mensaje
  - Si menor: pérdida de resto del mensaje (MPI y sockets datagrama)
  - Se mantiene integridad de los mensajes
  - Nunca se entregan restos de mensajes ni dos mensajes juntos
- Excepto en comunicación como flujo de bytes (sockets *stream*)
  - Datos de mensaje no leídos se obtienen en próxima recepción
  - Recepción puede devolver datos de fragmentos de mensajes
  - Recepción puede devolver dos mensajes junto
  - Si se requiere no mezclar mensajes de tamaño variable se puede:
    - Enviar longitud
    - Usar un separador
    - Usar 1 conexión/mensaje y hacer *shutdown* de socket de envío

Representación externa de datos

- Emisor y receptor misma interpretación de información
  - Misma cuestión, y soluciones, para lector y escritor de un fichero
- Procesadores, lenguajes, compiladores difieren en:
  - Orden de bytes en tipos numéricos (*endian*)
  - Tamaño de datos numéricos (en C: ¿tamaño de int, long,…?)
  - *Strings* (con longitud vs. carácter terminador (¿qué carácter?))
  - Formatos de texto (ISO-8859-1, UTF-8,…)
  - Organización estructuras datos (compactación, alineamientos,…),…
- Se necesitan "serializar" los datos para enviar/almacenar
  - Asegurando misma interpretación en sistema heterogéneo
  - Eficientemente (en *serialización*, en uso de red/disco,…)
  - Facilitando la programación de la *serialización*
  - Admitiendo cambios incrementales en protocolo
    - P.e. protocolo con nuevo campo opcional pero cliente antiguo sigue OK

ing

ación en emisor
lling) en receptor
er:
(sockets)
s serialización (excepto para int: *htonl*…)
mente
nvierte a formato de receptor
r receptor
onvierte a su formato
alquier emisor
nvierte a formato externo
rno y viceversa
receptor pero ≠ de externo

Formato de *serialización* de datos

- Define cómo se transmiten/almacenan datos (*wire protocol*)
  - Secuencia de bits que representan cada dato
- Alternativas:
  - Formato propio vs. Estándar (mejor)
  - Texto vs. binario: menos compacto pero interpretable por usuarios
  - Información de tipos implícita o explícita:
    - Implícita: emisor y receptor conocen tipos de parámetros
      - no viaja info. de tipos con datos
    - Explícita: disponible información explícita de tipos
      - Viaja mezclada con datos o como referencia a un esquema
    - Explícita más flexible (permite reflexión) pero menos compacto
  - Información de nombres de campos implícita vs. explícita:
    - Explícita: viaja nombre de campo con datos
      - Puede facilitar cambios incrementales de un protocolo
  - Información explícita de campos y tipos
    - Función de *deserialización* genérica

Componentes de un s. de *serialización*

- No sólo define un *wire protocol*, además puede incluir:
  - IDL (*Interface Definition Language*) y API para la *serialización*
- Lenguaje IDL para especificar datos a transmitir/almacenar
  - Permite definir datos con independencia de la plataforma
    - Usando, habitualmente, lenguaje específico (véase sección sobre RPC)
  - *Compilador IDL*: a partir de especificación genera tipos/clases nativos
  - Aplicación usa tipos nativos generados y API para (de)serializar
- API hipotético para *serialización*
  - *Encode(dato, tipo)* → *buffer*
  - *Decode(buffer, tipo)* → *dato*
    - Si información de tipos/campos explícita: *Decode(buffer)*
- Puede estar integrado en entorno de RPC/RMI
  - ¡Buenas noticias!: Programador no realiza *serialización*
  - Código generado usa automáticamente API de *serialización*

de serialización

mplícita campos y tipos
mplícita campos y tipos
, inf. explícita campos y tipos
a XML Schema
pos y tipos
io, no explícita campos y tipos
e cada campo con datos
n protocolo
explícita campos y tipos
con datos; no requiere IDL
Thrift, Apache Avro, BSON,…
ialization formats
m.es/~ssoo/SD.dir/serializacion

XDR

```
struct dato {int id;  string nombre<>; };  // especificación (fichero .x)
```

```
struct dato d;  XDR x;  char buf[TAM];
d.id=1; d.nombre="yo";
xdrmem_create(&x, buf, TAM, XDR_ENCODE);
xdr_dato (&x, &d);            // serializa
write(1, buf, xdr_getpos(&x));
```

```
XDR x; int tam;  char buf[TAM];
struct dato d = {0, NULL};
tam=read(0, buf, TAM);
xdrmem_create(&x, buf, tam, XDR_DECODE);
xdr_dato (&x, &d);            // deserializa
printf("id %d nombre %s\n", d.id, d.nombre);
```

Contenido = 12 bytes: 00 00 00 01 00 00 00 02 'y' 'o' 00 00

JSON (wikipedia)

```
{
    "firstName": "John",
    "lastName": "Smith",
    "age": 25,
    "address": {
        "streetAddress": "21 2nd Street",
        "city": "New York",
        "state": "NY",
        "postalCode": 10021
    },
    "phoneNumber": [
        {
            "type": "home",
            "number": "212 555-1234"
        },
        {
            "type": "fax",
            "number": "646 555-4567"
        }
    ]
}
```


n Javascript)

iza a {"nombre":"yo","tfno":666}

ación genérica

Protocol Buffers (con C++)

```
message Person {
  required int32 id = 1;
  required string name = 2;
  optional string email = 3;
}
```

Especificación (fichero .proto)

```
Person person;
person.set_id(123);
person.set_name("Bob");
person.set_email("bob@example.com");

fstream out("person.pb", ios::out | ios::binary | ios::trunc);
person.SerializeToOstream(&out);
out.close();
```

Serialización a un fichero (C++)

```
Person person;
fstream in("person.pb", ios::in | ios::binary);
if (!person.ParseFromIstream(&in)) {
  cerr << "Failed to parse person.pb." << endl;
  exit(1);
}

cout << "ID: " << person.id() << endl;
cout << "name: " << person.name() << endl;
if (person.has_email()) {
  cout << "e-mail: " << person.email() << endl;
}
```

Deserialización desde un fichero (C++)

Java Serialization

```
public class Dato implements Serializable {
    int id; String nombre; public Dato(int i, String n) {id = i; nombre = n; }}
```

```
class Encode {
    static public void main (String args[]) {
      Dato d = new Dato(1, "yo");
      try {  ObjectOutputStream o = new ObjectOutputStream(System.out);
           /* serialización */ o.writeObject(d); o.close(); }
      catch (java.io.IOException e) { System.err.println("Error serializando");}}}
```

```
class Decode {
    static public void main (String args[]) {
      try {  ObjectInputStream i = new ObjectInputStream(System.in);
           /* deserialización genérica */  Dato d = (Dato) i.readObject(); i.close();
           System.out.println(d.id + " " + d.nombre); }
      catch (Exception e) {  System.err.println("Error deserializando");} }}
```

Contenido = ¡69 B!; *http://www.javaworld.com/article/2072752/the-java-serialization-algorithm-revealed.html*

n entero (a=150)

Contenido

https://developers.google.com/protocol-buffers/docs/encoding

00 00 96 00

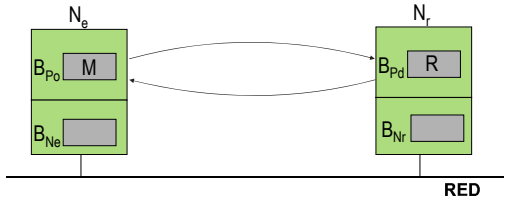
{"a":150}

0B

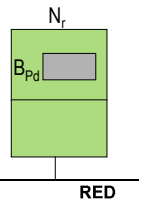
Grado de sincronía y *buffering*

- P_o envía M a P_d: copia entre *buffers* de procesos: $B_{Po} \rightarrow B_{Pd}$
  - Además puede haber *buffers* en nodo emisor B_{Ne} y/o receptor B_{Nr}
    - Minimizar copias entre *buffers* (ideal: *zero copy*)
- De menor a mayor grado de sincronía
  1. Envío devuelve control inmediatamente
     - No requiere B_{Ne} pero P_o no puede reutilizar B_{Po} hasta que sea seguro
       - Fin de operación o mensaje copiado en algún *buffer* (B_{Ne} o B_{Nr})
     - Requiere operación para comprobar si ya se puede reutilizar
  2. Envío devuelve control después de $B_{Po} \rightarrow B_{Ne}$
     - P_o puede reutilizar B_{Po}, pero posible bloqueo si B_{Ne} lleno
  3. Envío devuelve control cuando M llega a nodo receptor (B_{Nr})
     - No requiere B_{Ne}; ACK de N_r a N_e
  4. Envío devuelve control cuando M llega a P_d (B_{Pd})
     - No requiere B_{Ne} ni B_{Nr}; ACK de N_r a N_e
  5. Envío devuelve control cuando P_d tiene respuesta
     - No requiere B_{Ne} ni B_{Nr}: $B_{Po} \leftrightarrow B_{Pd}$; respuesta sirve de ACK

Posibles *buffers* en comunicación

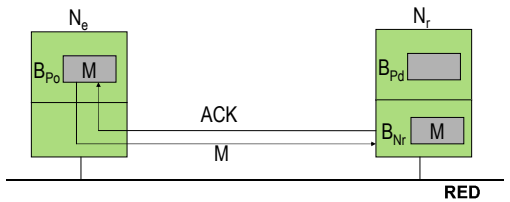


ediato

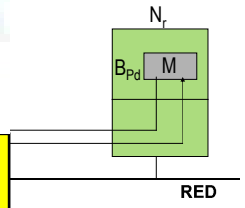


Retorno después de copia local

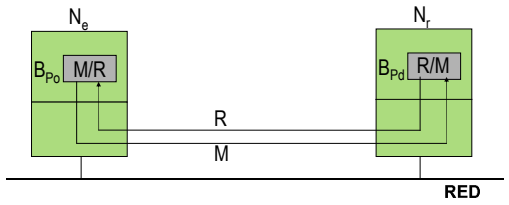
Retorno después de llegada



le recepción



Retorno después de respuesta



Modo de operación en recepción

- Recepción generalmente bloqueante
- Opción no bloqueante: retorna si no hay datos
- Opción asíncrona:
  - Especifica *buffer* donde se almacenará el mensaje y
  - Retorna inmediatamente
  - S. comunicaciones realiza recepción mientras proceso ejecuta
- Espera temporizada: se bloquea un tiempo máximo
- Espera múltiple: espera por varias fuentes de datos

ronía y *buffering*

po 2
con bloqueo si *buffer* local lleno

arse en envío:
scriptor socket: error si *buffer* lleno
robar que envío no bloquea
modo de envío tipo 1
ón bloqueante
arse en recepción:
scriptor socket: error si *buffer* vacío
robar que hay datos que recibir

ediante *select/poll/epoll*

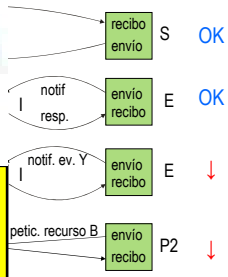
MPI: grado de sincronía y *buffering*

- `MPI_Send` Al retornar emisor puede usar su *buffer* (B_{Po})
  - Modo de operación dependiente de implementación: 2,3 o 4
- `MPI_Bsend` Modo 2 (retorno después de copia local)
  - Aplicación reserva y proporciona a sistema B_{Ne} de tamaño suficiente
- `MPI_Ssend` Modo 4 (retorno en recepción)
- `MPI_Rsend` Emisor sabe que receptor está listo para recibir
- `MPI_SendRecv` Modo 5 (retorno después de respuesta)
- `MPI_I...` Envío devuelve control inmediatamente (Modo 1)
  - Comprobar/esperar *buffer* se puede reutilizar (MPI_TEST|MPI_WAIT)
  - Varias primitivas dependiendo de cuándo se puede reutilizar:
    - `MPI_Isend MPI_Ibsend MPI_Issend MPI_Irsend`
- `MPI_Recv` Recepción bloqueante
- `MPI_Irecv` Recepción asíncrona

Adecuación a arquitecturas del SD

- Paso de mensajes adecuado para cualquier arquitectura
  - Pero cuidado con su asimetría: uno envía y otro recibe
- Cliente/servidor: su asimetría encaja con la del paso mensajes
  - Cliente: envía petición y recibe respuesta
  - Servidor: recibe petición y envía respuesta
- Editor/subscriptor: su asimetría no siempre encaja con p. mens.
  - Si *pull* con intermediario I: buen encaje
    - Su|Ed envían ops. a I y reciben respuestas de I → I siempre pasivo
  - Si *push* con intermediario I: encaje problemático
    - Su envía suscripción(evento X) y espera confirmación de I
    - Pero justo antes I envía notificación de evento Y a Su
    - Soluciones: uso de múltiples puertos y concurrencia en subscriptor
- P2P: arquitectura simétrica
  - ¿Quién envía y quién recibe?

arquitecturas



Índice

- Introducción
- Paso de mensajes
  - Comunicación punto a punto
  - Comunicación de grupo
  - Sistemas de colas de mensajes (MOM)
- Llamadas a procedimientos remotos (RPC)
  - Sun RPC
- Invocación de métodos remotos (RMI)
  - Java RMI y CORBA

Multidifusión: comunicación de grupo

Destino de mensaje → grupo de procesos
- Envío/recepción especifican dirección de grupos de procesos
- Desacoplamiento espacial

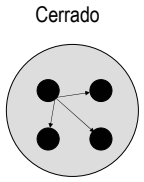
Trabajo seminal: ISIS (posteriores Horus, Ensemble, JGroups)

Adecuación a arquitectura del SD:
- Cliente-servidor replicado
  - Facilita actualizaciones múltiples
- Modelo editor/subscriptor
  - Envío de notificaciones (p.e. 1 grupo/tema)
- Arquitecturas para CD
  - Operaciones colectivas en proc. paralelo (pueden incluir cálculos)

Implementación depende de si red tiene *multicast* (*IP-multicast*)
- Si no, se implementa enviando *N* mensajes

Un proceso puede pertenecer a varios grupos (grupos solapados)

sus cerrado

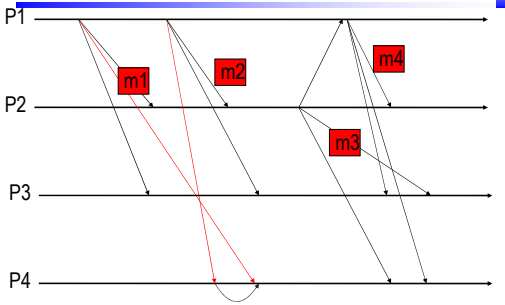


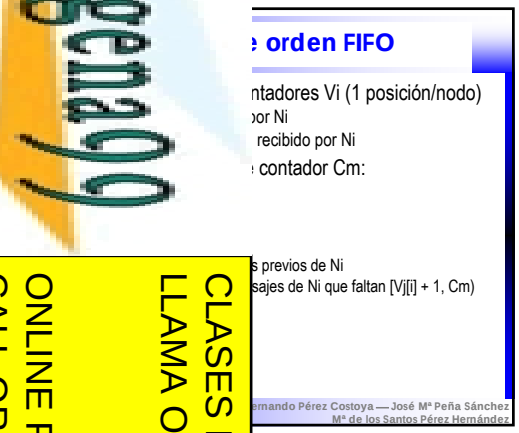
Sólo procesos del grupo pueden mandar mensajes
Suele usarse en procesamiento paralelo

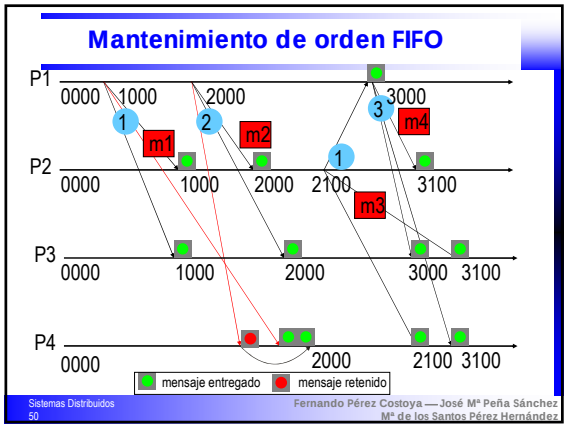
Aspectos de diseño de com. de grupo

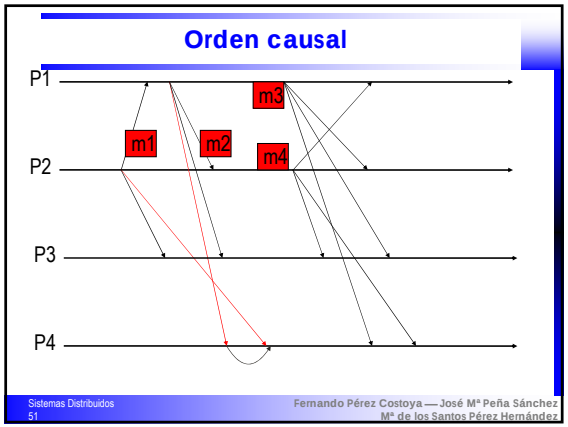
- Atomicidad: o reciben el mensaje o ninguno
  - Con unidifusión fiable (TCP): en medio, se puede caer emisor
  - Con multicast IP: pérdida de mensajes
- Orden de recepción de los mensajes
  - FIFO: mensajes de misma fuente llegan en orden de envío
    - No garantía sobre mensajes de distintos emisores
  - Causal: entrega respeta relación "causa-efecto"
    - Si no hay relación, no garantiza ningún orden de entrega
  - Total: Todos los mensajes recibidos en mismo orden por todos
- El grupo suele tener carácter dinámico
  - Se pueden incorporar y retirar procesos del grupo
  - Gestión de pertenencia debe coordinarse con la comunicación
    - Propiedad denominada "*Virtual Synchrony*"

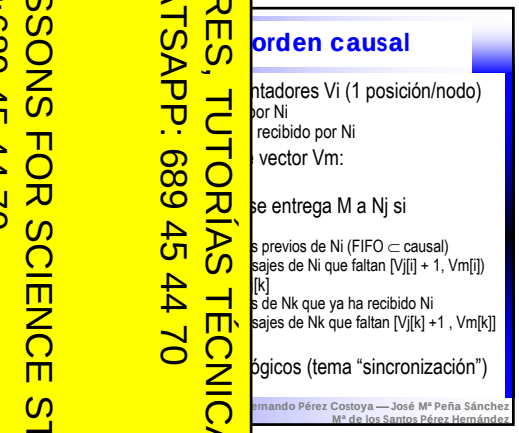
Orden FIFO

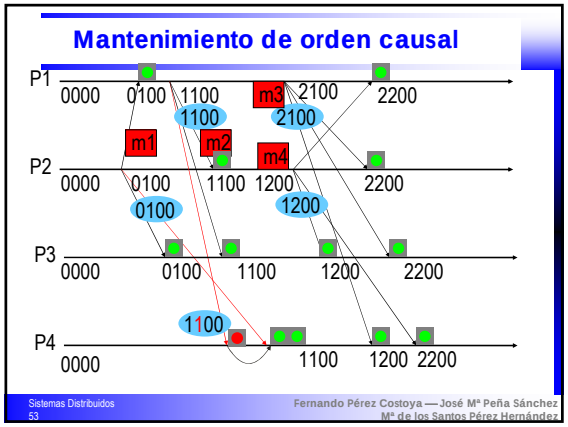


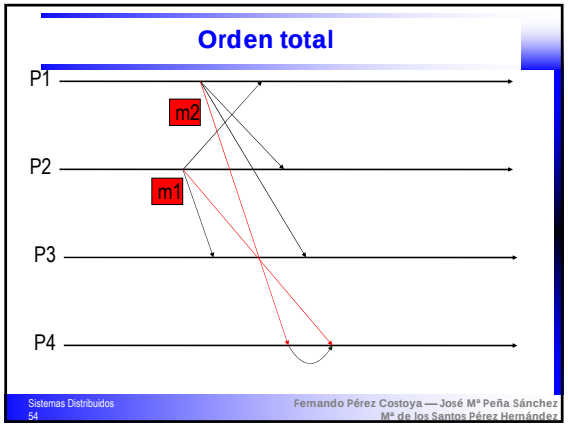










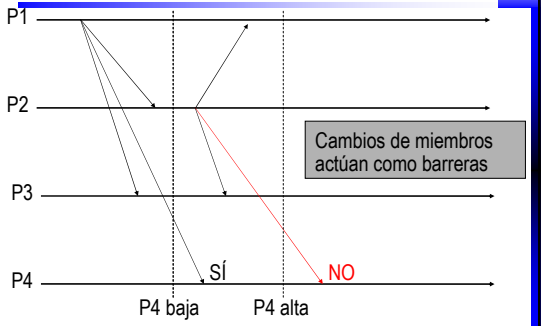


e orden total

asada en secuenciador S
o de fallo
número único a cada mensaje
para cada grupo
todos miembros grupo G + S
ontador; sólo algún tipo de ID de M

pecial Ms = {ID de M, Cs}
mensaje Ms que espera recibir
se retiene
s} en Ni:
y Cs == Ci → entrega de M

Sincronía virtual



Índice

las de mensajes

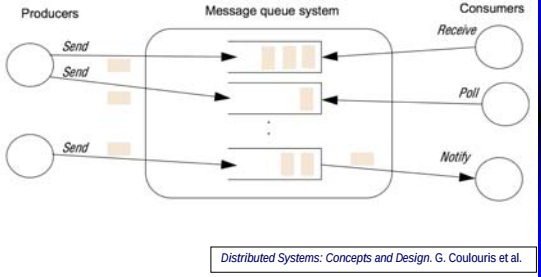
WebSphere MQ, MSMQ, *JMS*
OM
as con comunic. "persistente":

cuando se recibe mensaje

ceptor esté presente
ed de intermediarios) guarda mensaje
espacio y tiempo

a (bloqueante)
o bloqueante)
ado cuando llegue mensaje a cola

Sistema de colas de mensajes

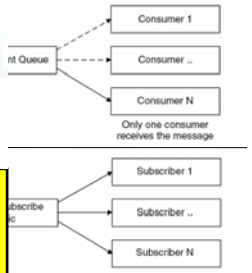


Distributed Systems: Concepts and Design. G. Coulouris et al.

MOM – Sistemas de colas de mensajes

- 2 modelos de comunicación habituales:
  - Basado en colas punto-a-punto
  - Basado en temas editor/subscriptor
- Adecuación a arquitecturas de SD
  - C/S: punto-a-punto → multi-servidor con reparto automático de carga
  - Ed/Su: modelo basado en temas; NOTIFY permite esquema *push*
- Características avanzadas habituales:
  - Filtrado de mensajes: receptor selecciona en cuáles está interesado
    - por propiedades, por contenido,...
    - Puede usarse como filtro por contenido en arquit. Ed./Su.
  - Mensajería con transacciones
  - Transformaciones de mensajes
- Apropiado para integración de aplicaciones de empresa (EAI)

itor/subscriptor



Message-Oriented Middleware. Edward Curry

Sistemas Distribuidos

RPC: Llamada a procedimiento remoto

Índice

- Introducción
- Paso de mensajes
  - Comunicación punto a punto
  - Comunicación de grupo
  - Sistemas de cola de mensajes (MOM)
- Llamadas a procedimientos remotos (RPC)
- Invocación de métodos remotos (RMI)
  - Java RMI y CORBA

s en S. no Dist.

```
.) {
 q, ...);
.
, tb b, ...) {
1;
, ty y, ...) {
2;
```


Provisión de servicios (C/S) en S. Dist.

Servidor

```
int main(...) {
  Mensaje msj,resp;
  .........
  alta(IDservicio,dir_srv);
  while (TRUE) {
    recepción(&msj,&dir_clie);
    switch(msj.op) {
    case OP1:
      resp.r=op1(msj.arg1,...);
    case OP2:
      resp.r=op2(msj.arg1,...);
    .........
    envío(resp,dir_clie);
  }
}
t1 op1(ta a, tb b, ...) {
}
...........
```

Cliente

```
int main(...) {
  Mensaje msj,resp;
  .........
  dir_srv=busca(IDservicio);
  msj.op=OP1;
  msj.arg1=p;
  msj.arg2=q;
  ...........
  envío(msj,dir_srv);
  recepción(&resp, NULL);
  r=resp.r;
  .........
}
```


Cliente-servidor con paso de mensajes

- Provisión de servicios en SD tiene mayor complejidad
- Programador tiene que ocuparse de aspectos como:
  - Operaciones de mensajería
  - *Marshalling/Unmarshalling*
    - Con conversión de formato si sistema de comunicación no lo hace
  - Realizar *binding*
  - Gestión de esquema de concurrencia elegido
    - *Threads*/procesos dinámicos, estáticos o con umbrales
  - Asegurar semántica de comportamiento ante fallos
  - Gestión de conexiones (si se usa esquema con conexión)
    - 1 conexión/petición vs. *N* peticiones de cliente usan misma conexión
  - Si comunicación no fiable, garantizar envío correcto de mensajes
  - Si limitación en tamaño de mensajes, fragmentación y compactación

e las RPC

servicios en SD
tación del cliente y del servidor
rvicio
te a partir de la misma

n sistema no distribuido
as de servicio y aplicaciones
áticamente
e llamadas a procedimiento en SD

Provisión de servicios en SD con RPC

Aplicación

```
int main(...) {
   ....
   r=op1(p, q, ...);
   .........
}
init() {
  dir_srv=busca(IDservicio);
}
t1 op1(ta a, tb b, ...) {
  Mensaje msj,resp;
  msj.op=OP1;
  msj.arg1=a;
  msj.arg2=b;
  envío(msj,dir_srv);
  recepción(&resp, NULL);
  return resp.r;
}
```

Biblioteca

```
int main(...) {
  Mensaje msj,resp;
  .........
  alta(IDservicio,dir_srv);
  while (TRUE) {
    recepción(&msj,&dir_clie);
    switch(msj.op) {
    case OP1:
      resp.r=op1(msj.arg1,...);
    case OP2:
      resp.r=op2(msj.arg1,...);
    .........
    envío(resp,dir_clie);
  }
}
t1 op1(ta a, tb b, ...) {
}
..........
```


Historia y evolución de las RPC

- Primeras ideas: White 1975; White se enrola en Xerox.
- Desarrollo en Xerox: primer sistema de RPC *Courier* (1981)
- Artículo clásico de Birrel y Nelson en 1984.
- Sun: *Open Network Computing* (1985)
  - RPC de Sun/ONC es la base para servicios (NFS o NIS)
- Apollo (competidora de Sun) crea NCA: RPC de NCA
  - HP compra Apollo; HP miembro de Open Group
- Open Group: DCE (*Distributed Computing Environment* 1990)
  - RPC de DCE basada en NCA
  - RPC de Microsoft (sustento de DCOM) basada en RPC de DCE
- RPC de Sun/ONC vs. RPC de DCE
  - RPC de Sun/ONC menos sofisticada pero más extendida

sistema de RPC

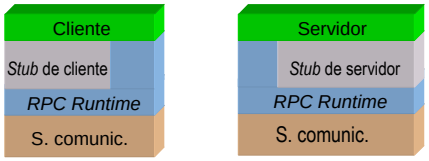
rograma:
Compilador de IDL
→ Resguardos
a múltiples lenguajes
grama:

nte y en servidor
stracción de llamada a procedimiento

que dan soporte a RPC
enticación, comportamiento ante fallos, ...
s pero también por aplicación/biblioteca
alternativas ya presentadas
r. servidor (Sun/ONC RPC)

inder global + *binders* locales (DCE RPC)

Pila de comunicación en RPC



Resguardo del cliente (aka *proxy*)

- Conjunto de funciones enlazadas con la aplicación cliente
  - Ofrece la misma API que el servicio
  - Usa servicios de *runtime* de RPC
- Tareas realizadas por una función del resguardo de cliente:
  - Localiza al servidor usando *binder* (identificador del servicio)
    - Si se usa BG + BL → doble consulta
  - Control de conexiones, si se usa comunicación que las requiera
  - Empaqueta/convierte id. función y parámetros (*marshalling*)
  - Envía el mensaje al servidor
  - Espera la recepción del mensaje
  - Extrae/convierte resultados (*unmarshalling*) y los devuelve
    - Si mensaje info. explícita de tipos, conversión independiente del servicio

r (aka *skeleton*)

vidor con biblioteca de servicio

o de servidor:

stro
procesos/*threads* de servicio
ontrol/gestión de las conexiones
ajes

e (*unmarshalling*)
s, conversión independiente del servicio

ng)
iente

Características de las RPC

- Modo de operación con bloqueo hasta respuesta
  - También asíncrono
    - P.ej. Extensión de Microsoft a RPC de DCE
- Normalmente comunicación desde cliente a servidor
  - En algunos sistemas *Callback* RPC: llamadas de servidor a cliente
- Normalmente implican comunicación uno a uno
  - En algunos sistemas *Multicast* y *Broadcast* RPC (p. ej. RPC de Sun)
- Comportamiento ante fallos garantizado por RPC
  - P.ej. RPC de DCE asegura por defecto como mucho una vez
    - Además permite marcar una función de servicio como idempotente
- Si servidor concurrente
  - Funciones de servicio pueden requerir mecanismos de sincronización

Adecuación a las arquitecturas del SD

- Orientadas a modelo cliente-servidor
- Arquitectura editor/subscriptor
  - Adecuadas si modelo *pull* con intermediario I
    - Ed y Su invocan RPCs de intermediario
      - Ed y Su resguardo de cliente; I resguardo de servidor
  - Modelo *push* más problemático
    - Su invoca RPC del servicio de I para suscripción:
      - Su (y Ed) resguardo de cliente; I resguardo de servidor
    - Pero I invoca RPC del servicio de Su para notificación de eventos
      - I también resguardo de cliente; Su también resguardo de servidor
- Solución habitual *Callback* RPC:
  - Intermediario proporciona función de registro a subscriptor
    - Incluye parámetro que identifica servicio de notificación en subscriptor
  - Intermediario almacena lista de ID de servicio de suscriptores
  - Cuando I recibe evento de Ed, invoca *callback* RPC de subscriptores

ón de interfaces

artir de la interfaz de servicio

des tales como:
número de versión asociado
es, tipos de datos, …
de salida o de entrada/salida
o, si una función es idempotente)
ntación

pecífico vs. uno convencional
faces
ad de lenguajes convencionales

(por ejemplo para C)
do cliente + resguardo servidor

Transparencia de las RPC

- ¿Una invocación remota se comporta igual que una local?
- Existen varias diferencias:
  - Puede producir un error si no puede comunicarse con servidor.
    - ¿Cómo notificarlo a la aplicación?
    - Generalmente, usando excepciones
  - No puede haber variables globales en el servicio accesibles al cliente
  - Los parámetros no pueden pasarse por referencia
    - Uso de copia y restauración: tratamiento de parámetro de entrada/salida
      - Resguardo de cliente envía copia al resguardo servidor
      - Resguardo servidor lo pasa por referencia a función real, que lo modifica
      - Resguardo de servidor envía a resguardo cliente nuevo valor
      - Resguardo cliente lo establece como nuevo valor
  - Problema: parámetro de tipo estructura con punteros internos (listas)
    - Algunos sistemas las prohíben
    - Otros, como RPC de Sun, las permiten
      - el resguardo del emisor las "aplana" y el del receptor las reconstruye

RPC de Sun/ONC

- Estándar (RFC 1831) y disponible en muchas plataformas
- Junto con XDR (*External Data Representation*, RFC 1832)
  - Define IDL y formato de representación externo (binario e implícito)
- Independiente de nivel de transporte subyacente (TCP o UDP)
- Distintos tipos de autenticación de cliente
  - Sin autenticación
  - Modelo UNIX (uid:gid)
  - Basada en cifrado (DES o Kerberos)
    - Secure RPC (base de Secure NFS)
- Comportamiento ante fallos:
  - Si se implementa sobre TCP: como mucho una vez
  - Sobre UDP, *runtime* de RPC retransmite petición si no respuesta
    - Permite activar una caché de respuestas en servidor

Consultar guía de programación RPC en página de la asignatura

RPC de Sun/ONC

rograma (32 bits) + versión
r Sun
lobal (ID + dir. máquina)
áquina en puerto conocido (111)

alta en *binder* local:
ocolo + puerto (efímero)
de programa + versión + protocolo
n *binder* de máquina
ocolo → puerto
ntactar con *binder* para:
e alta, ejecutar el procedimiento
ar si servidor arrancado, etc.

IDL (XDR) de RPC de Sun/ONC

- Extensión de C. Permite definir:
  - Un programa (interfaz) y su nº de versión
  - Constantes, tipos y funciones (hay que asignarles un número)
    - Restricción: función con un 1 solo argumento y 1 solo valor de retorno
- Compilador *rpcgen*:
  - Además de .h y resguardos, genera fichero de *marshalling* (.xdr)
- Algunos tipos específicos de XDR (además de los de C):
  - Vectores de tamaño variable:
    - tipo_elem vector<>; tipo_elem vector<tam_max>;
  - *String*
  - *Union* distinta de la de C: registro variante

```
union nombre_tipo switch (int discriminante) {
    case valor1: declaración1;
    case valor2: declaración2;
    default:  declaraciónN; }
```


Sistemas Distribuidos

RMI: Invocación de método remoto

- **Java RMI**
- **CORBA**

otos (RPC)
(RMI)

Modelo de objetos en sis. distribuidos

- Sistemas distribuidos.
  - Aplicaciones inherentemente distribuidas.
  - Se caracterizan por su complejidad.
- Sistemas orientados a objetos.
  - Más cercanos al lenguaje natural.
  - Facilitan el diseño y la programación.

Objetos-Distribuidos

Características:
- Uso de un *Middleware*: Nivel de abstracción para la comunicación de los objetos distribuidos. Oculta:
  - Localización de objetos.
  - Protocolos de comunicación.
  - Hardware de computadora.
  - Sistemas Operativos.
- Modelo de objetos distribuidos: Describe los aspectos del paradigma de objetos que es aceptado por la tecnología: Herencia, Interfaces, Excepciones, Polimorfismo, ...
- Recogida de basura (*Garbage Collection*): Determina los objetos que no están siendo usados para liberar recursos.

aso de mensajes

adecuado para aplicaciones muy
número de peticiones y respuestas

los

le las operaciones
dario

Ventajas respecto a RPC

- Ventajas derivadas al uso de programación orientada a objetos:
  - Encapsulación
  - Reutilización
  - Modularidad
  - Dinamismo

Objetos distribuidos

- Minimizar las diferencias de programación entre las invocaciones de métodos remotos y las llamadas a métodos locales
- Ocultar las diferencias existentes:
  - Tratamiento del empaquetamiento de los datos (*marshalling*)
  - Sincronización de los eventos
  - Las diferencias deben quedar ocultas en la arquitectura

Distribuidos

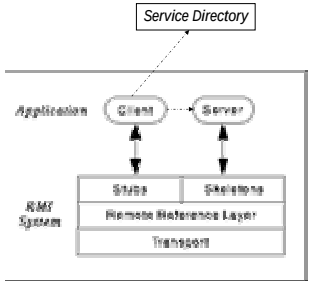
Java RMI

Java RMI

Java: *Write Once, Run Anywhere*

Java RMI extiende el modelo Java para la filosofía *"Run Everywhere"*

Arquitectura de Java RMI



Java RMI

roxy/skeleton) que se encarga
de los parámetros
liente realiza una invocación
ocación de un método del resguardo

ciones de los clientes, realiza la
los resultados.

emotas: interpreta y gestiona
servicio remoto
le las comunicaciones y de
sarias. Basada en TCP

Servicio de directorios

Se pueden utilizar diferentes servicios de directorios para registrar un objeto distribuido
- Ejemplo: JNDI (Java Naming and Directory Interface)

El registro RMI, *rmiregistry*, es un servicio de directorios sencillo proporcionado por SDK (Java Software Development Kit)

Java RMI

El soporte para RMI en Java está basado en las interfaces y clases definidas en los paquetes:
- *java.rmi*
- *java.rmi.server*

Características de Java RMI:
- Se basa en una interfaz remota Java (hereda de la clase Java *Remote*).
- Es necesario tratar mayor número de excepciones que en el caso de invocación de métodos locales
  - Errores en la comunicación entre los procesos (fallos de acceso, fallos de conexión)
  - Problemas asociados a la invocación de métodos remotos (no encontrar el objeto, el resguardo o el esqueleto)
  - Los métodos deben especificar la excepción *RemoteException*.

remota Java

```
zEj extends

l()
oteException;
loat param)
oteException;
```

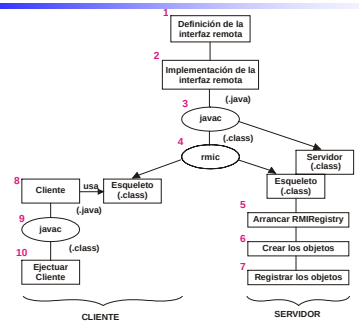

Java RMI

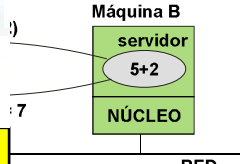
Localización de objetos remotos:
- Servidor de nombres: *java.rmi.Naming*

Ejemplo:

```
Cuenta cnt = new CuentaImpl();
String url = "rmi://java.Sun.COM/cuenta";
// enlazamos una url a un objeto remoto
java.rmi.Naming.bind(url, cnt);
     ....
// búsqueda de la cuenta
cnt=(Cuenta)java.rmi.Naming.lookup(url);
```


Desarrollo de Aplicaciones RMI





Modelización de la interfaz remota (Sumador)

```
public interface Sumador extends java.rmi.Remote
{

  public int sumar(int a, int b)
       throws java.rmi.RemoteException;

}
```


Clase que implementa la interfaz (SumadorImpl)

```
import java.rmi.*;
import java.rmi.server.UnicastRemoteObject;
public class SumadorImpl extends UnicastRemoteObject
  implements Sumador {
   public SumadorImpl(String name) throws RemoteException {
       super();
      try {
       System.out.println("Rebind Object " + name);
           Naming.rebind(name, this);
      } catch (Exception e){
       System.out.println("Exception: " + e.getMessage());
                e.printStackTrace();
      }
  }
   public int sumar (int a, int b) throws RemoteException {
          return a + b;
   }
```


servidor (rServer)

```
tring args[]) {

a = new
("rmi://localhost:1099"
r");
 e) {
n("System exception" +
```


Registro del servicio

Antes de arrancar el cliente y el servidor, se debe arrancar el programa *rmiregistry* en el servidor para el servicio de nombres. El puerto que utiliza el *rmiregistry* por defecto es el 1099.

```
rmiregistry [port_number]
```

El método *rebind* es utilizado normalmente en lugar del método *bind*, porque garantiza que si un objeto remoto se registró previamente con dicho nombre, el nuevo objeto reemplazará al antiguo.

El método *rebind* almacena en el registro una referencia al objeto con un URL de la forma:

```
rmi://<nombre máquina>:<número puerto>/<nombre referencia>
```


Código en el cliente (SumadorClient)

```
import java.rmi.registry.*;
import java.rmi.server.*;
import java.rmi.*;
public class SumadorClient {
   public static void main(String args[]){
  int res = 0;
  try {
       System.out.println("Buscando Objeto ");
       Sumador misuma = (Sumador)Naming.lookup("rmi://" +
  args[0] +                              "/" +"MiSumador");
        res = misuma.sumar(5, 2);
       System.out.println("5 + 2 = " + res);
   }
   catch(Exception e){
            System.err.println(" System exception");    }
  System.exit(0);
   }
```

da

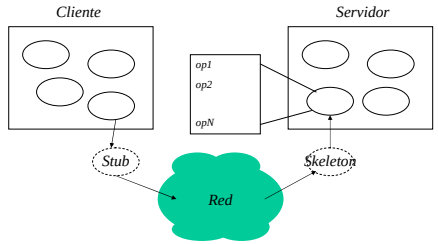
stanciar un objeto remoto debe
uiente forma:
ming.lookup("rmi://" + args[0] +

ferencia remota a un objeto
ota.
miregistry.

Pasos

Java RMI:
- Enlace a un nombre: bind(), rebind()
- Encontrar un objeto y obtener su referencia: lookup()
- refObj.nombre_met()

Cuadro general



ecuta?

RMI

099

Callback de cliente

Hay escenarios en los que los servidores deben notificar a los clientes la ocurrencia de algún evento. Ejemplo: chat.
- Problema: llamada a método remoto es unidireccional

Posibles soluciones:
- Sondeo (*polling*): Cada cliente realiza un sondeo al servidor, invocando repetidas veces un método, hasta que éste devuelva un valor *true*.
  - Problema: Técnica muy costosa (recursos del sistema)
- *Callback*: Cada cliente interesado en la ocurrencia de un evento se registra a sí mismo con el servidor, de modo que el servidor inicia una invocación de un método remoto del cliente cuando ocurra dicho evento.
  - Las invocaciones de los métodos remotos se convierten en bidireccionales

Extensión de la parte cliente para *callback* de cliente

El cliente debe proporcionar una interfaz remota: Interfaz remota de cliente

```
public interface CallbackClient extends java.rmi.Remote {
    public String notificame (String mensaje) throws
   java.rmi.RemoteException;
}
```

Es necesario implementar la interfaz remota de cliente, de forma análoga a la interfaz de servidor (CallbackClientImpl)

En la clase cliente se debe añadir código para que instancie un objeto de la implementación de la interfaz remota de cliente y que se registre para callback (método implementado por el servidor):

```
CallbackServer cs = (CallbackServer) Naming.lookup(URLregistro);
CallbackClient objCallback = new CallbackClientImpl();
cs.registrarCallback(objCallback);
```

dora para *callback*
te

el cliente se registre para

```
ck (CallbackClient
ava.rmi.RemoteException;
eliminarRegistroCallback
```

tructura común (por ejemplo,
referencias a los *callbacks* de
nchronized para acceder a la

Descarga dinámica de resguardo

- Mecanismo que permite que los clientes obtengan dinámicamente los resguardos necesarios
  - Elimina la necesidad de copia de la clase del resguardo en la máquina cliente
  - Se transmite bajo demanda desde un servidor web a la máquina cliente (Similar a la descarga de los applets)
- El servidor exporta un objeto a través del registro RMI (registro de una referencia remota al objeto mediante nombre simbólico) e indica el URL donde se almacena la clase resguardo.
- La clase resguardo descargada no es persistente
  - Se libera cuando la sesión del cliente finaliza

Comparativa RMI vs Sockets

- Los sockets están más cercanos al sistema operativo, lo que implica una menor sobrecarga de ejecución.
- RMI proporciona mayor abstracción, lo que facilita el desarrollo de software. RMI es un buen candidato para el desarrollo de prototipos.
- Los sockets suelen ser independientes de plataforma y lenguaje.

buidos

Entornos de Objetos Distribuidos

OMG

(Object Management Group)

Conjunto de organizaciones que cooperan en la definición de estándares para la interoperabilidad en entornos heteregéneos.

Fundado en 1989, en la actualidad lo componen más de 700 empresas y otros organismos.

OMA

(Object Management Architecture)

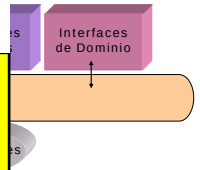
Arquitectura de referencia sobre cual se pueden definir aplicaciones distribuidas sobre un entorno heteregéneo.

CORBA es la tecnología asociada a esta arquitectura genérica.

Formalmente esta dividida en una serie de modelos:

- Modelo de Objetos
- Modelo de Interacción
- ...

A está compuesta por una serie
an entre sí. Estos objetos se
:



OMA

Servicios:

- Proporcionan funciones elementales necesarias para cualquier tipo de entorno distribuido, independientemente del entorno de aplicación.
- Los tipos de funciones proporcionados son cuestiones tales como la resolución de nombres, la notificación asíncrona de eventos o la creación y migración de objetos.
- Concede un valor añadido sobre otras tecnologías (por ejemplo RMI).
- Están pensados para grandes sistemas.

OMA

Aplicaciones:

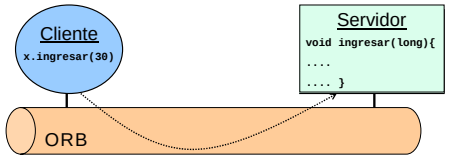
- El resto de funciones requeridas por una aplicación en concreto. Es el único grupo de objetos que OMG no define, pues está compuesto por los objetos propios de cada caso concreto.
- Estos son los objetos que un sistema concreto tiene que desarrollar. El resto (servicios, facilidades) pueden venir dentro del entorno de desarrollo.

ctura. Proporciona las
entre los objetos distribuidos
e aplicación) que forman una

n entre objetos.

ORB

Para posibilitar la comunicación entre dos objetos cualesquiera de una aplicación se realiza por medio del ORB. El escenario de aplicación elemental es:



IDL de CORBA

(*Interface Definition Language*)

Es el lenguaje mediante el cual se describen los métodos que un determinado objeto del entorno proporciona al resto de elementos del mismo.

```
interface Cuenta
{
  void ingresar(in long cantidad);
  void retirar(in long cantidad);
  long balance();
};
```

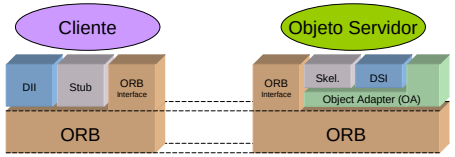
RBA

uaje de programación como:

s de código denominados *Stub* y
go de cliente y servidor

Componentes de un ORB

Arquitectura completa de comunicaciones de CORBA:



Componentes de un ORB

Stub:

Código cliente asociado al objeto remoto con el que se desea interactuar. Simula para el cliente los métodos del objeto remoto, asociando a cada uno de los métodos una serie de funciones que realizan la comunicación con el objeto servidor.

Skeleton:

Código servidor asociado al objeto. Representa el elemento análogo al stub del cliente. Se encarga de simular la petición remota del cliente como una petición local a la implementación real del objeto.

le un ORB

que permite que un cliente solicite
erfaces se desconocían en fase de

tons estáticos definidos en tiempo
do por servidores que durante su
ntes objetos que pueden ser
el servidor.

Componentes de un ORB

ORB/Interface ORB:

Elemento encargado de (entre otras) las tareas asociadas a la interconexión entre la computadora cliente y servidor, de forma independiente de las arquitecturas hardware y SSOO.

Debido a que tanto clientes como servidores pueden requerir de ciertas funcionalidades del ORB, ambos son capaces de acceder a las mismas por medio de una interfaz.

Las dos principales responsabilidades del ORB son:

Localización de objetos: El cliente desconoce la computadora donde se encuentra el objeto remoto.

Comunicación entre cliente y servidor: De forma independiente de protocolos de comunicación o características de implementación (lenguaje, sistema operativo, ...)

Componentes de un ORB

Adaptador de Objetos:

En este elemento se registran todos los objetos que sirven en un determinado nodo. Es el encargado de mantener todas las referencias de los objetos que sirven en una determinada computadora de forma que cuando llega una petición a un método es capaz de redirigirla al código del skeleton adecuado.

Existen dos tipos de Adaptadores de Objetos especificados por OMG:

BOA: (Basic Object Adapter).

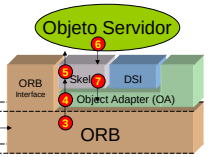
POA: (Portable Object Adapter).

ría CORBA

iado en el stub que realiza el proceso

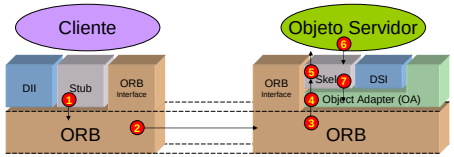
misión de la petición hacia el objeto

ción y la transmite el Adaptador de
sólo hay uno. (ORB servidor)



Comunicación vía CORBA

4- El Adaptador de Objetos resuelve cuál es el objeto invocado, y dentro de dicho objeto cuál es el método solicitado (Adaptador de Objetos)
5- El skeleton del servidor realiza el proceso de de-marshalling de los argumentos e invoca a la implementación del objeto. (Skeleton servidor)
6- La implementación del objeto se ejecuta y los resultados y/o parámetros de salida se retornan al skeleton. (Implementación del objeto)
7- El proceso de retorno de los resultados es análogo.



Localización de Objetos

Los objetos de servicio de una aplicación CORBA se encuentran identificados por medio de una referencia única (Identificador de Objeto).

Esta referencia es generada al activar un objeto en el Adaptador de Objetos.

Por medio de esta referencia el ORB es capaz de localizar la computadora y el Adaptador de Objetos donde se encuentra, y éste último es capaz de identificar el objeto concreto dentro del Adaptador.

del Servidor

a implementación del objeto se iseña como una subclase de la lase generada por el compilador (el *keleton*) de IDL en base a la efinición.

i el lenguaje usado para la mplementación no soporta objetos el ecanismo es diferente.

Servicios CORBA

Conjunto de objetos o grupos de objetos, que proporcionan una serie de funcionalidades elementales. Estos objetos están definidos de forma estándar (interfaces IDL concretas).

Sus especificaciones se encuentran recogidas en los documentos COSS (Common Object Services Specifications).

Los servicios definidos son:

- Servicio de Nombres
- Servicio de Eventos
- Servicio de Ciclo de Vida
- Servicio de Objetos Persistentes
- Servicio de Transacciones
- Servicio de Control de Concurrencia
- Servicio de Relación
- Servicio de Externalización
- Servicio de Consulta
- Servicio de Licencias
- Servicio de Propiedad
- Servicio de Tiempo
- Servicio de Seguridad
- Servicio de Negociación
- Servicio de Colección de Objetos

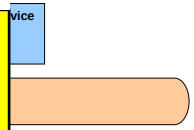
Uso del Servicio de Nombres

Permite asociar un nombre a una referencia de objeto. De esta forma los objetos al activarse pueden darse de alta en el servidor, permitiendo que otros objetos los obtengan su referencia en base a dicho nombre.

Los nombres de los objetos se encuentran organizados en una jerarquía de *contextos de nombre*.

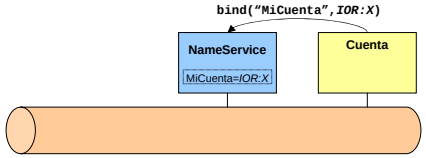
ombres

ue todo objeto del sistema se



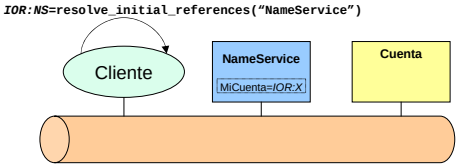
Servicio de Nombres

Un nuevo objeto se arranca y se registra en el servidor de nombres:



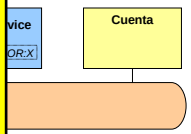
Servicio de Nombres

Un cliente localiza al servidor de nombres. Suele existir una función interna del ORB para localizar al servidor de nombres (**`resolve_initial_references`**) .



ombres

mbres que resuelva el nombre.
buscado.



Servicio de Negociación

Este servicio también permite obtener referencias a objetos usando otra información:

- Los objetos se exportan en el servidor con una serie de características del servicio que proporcionan.
- Los clientes consultan con el servidor cuáles son los objetos ofertados con una serie de características.

Un cliente que buscase un servicio de impresión podría construir una consulta del tipo:

```
Service=Printer AND
PrinterType=HP AND
OS!=WinNT
```

- A lo cual el servidor le indicará los objetos que existen con dichas características.

Comparativa CORBA vs DCOM

CORBA es un estándar abierto y no propietario.
CORBA proporciona soporte para diversos SO.
CORBA es más completo y flexible.
CORBA da una salida a los *legacy systems*

DCOM esta integrado con la tecnología *MicroSoft*.
DCOM ha tenido una fuerte penetración en el mercado.