

Apectos do Sistema de Arquivamento Virtual no Contexto da OSI

ESTHER DE CASTRO PACITTI

Núcleo de Computação Eletrônica — UFRJ

Caixa Postal 2324 - Cep 20001 Rio de Janeiro - RJ - Brasil

E-mail: ncd10123@ufrj.bitnet

ANA CLAUDIA DE MEDEIROS ROCHA VAZ

Cobra Computadores e Sistemas Brasileiros

Av. Comandante Guarany, 447 - Cep 22785 Jacarépagua

Rio de Janeiro - RJ - Brasil

RESUMO

Este artigo tem como objetivo analisar os principais aspectos relativos a implementação do sistema de arquivamento virtual no contexto do FTAM/OSI. Serão levadas em consideração as diversas estruturas de arquivos existentes e os modelos virtuais propostos pelas normas referentes ao assunto. Baseado nestes fatos será feita uma análise de quais os mapeamentos possíveis para cada caso, e os problemas decorrentes desse mapeamento.

1. INTRODUÇÃO

A ISO (International Organization for Standardization) vem se concentrando num projeto, denominado OSI (Open Systems Interconnection), que visa a padronização de vários aspectos relativos à comunicação de dados. O modelo de referência OSI foi criado para prover uma estrutura de trabalho na padronização de protocolos necessários para a interconexão de máquinas heterogêneas. O objetivo deste modelo é encorajar o desenvolvimento de interconexões de modo independente em máquinas de fabricantes diferentes que, ao final, consigam se comunicar obedecendo a uma arquitetura pré-estabelecida.

O ambiente de Transferência, Acesso e Gerenciamento de Arquivos para arquitetura OSI é coletivamente chamado de FTAM (File Transfer Access and Management). Esse ambiente permite que arquivos de entidades computacionais distintas sejam compartilhados.

Para que as operações requisitadas pela aplicação sejam realizadas, é necessário resolver as incompatibilidades dos equipamentos envolvidos e, em seguida, viabilizar a comunicação entre os sistemas cooperantes para que eles sejam informados das operações a serem efetuadas e ajam de modo adequado.

O modo de se armazenar e manipular arquivos varia muito nos diferentes sistemas. Existem inúmeros estilos de descrição e acesso a arquivos. Para que um protocolo de transferência de arquivo seja viável é necessário, primeiro, estabelecer um modelo comum de arquivo e seus atributos. Este modelo é denominado de arquivo virtual. O arquivo virtual permite que todas as incompatibilidades sejam mascaradas através de uma função de mapeamento, permitindo que sistemas distintos troquem arquivos sem se preocuparem com os aspectos locais de implementação. O protocolo de FTAM permite, no progresso da troca de mensagens, definir qual o modelo de arquivo virtual a ser adotado numa determinada operação. Isto permite que sistemas mais simples interajam com máquinas mais sofisticadas, pois toda complexidade será mascarada por modelos virtuais adequados.

Este artigo tem como objetivo analisar os principais aspectos relativos a implementação do Sistema de Arquivamento Virtual, levando em consideração as diversas estruturas de arquivos existentes e, baseado em experiências de implementação das autoras. Inicialmente será dada uma visão geral do FTAM, em seguida serão abordados os diversos modelos de arquivos virtuais propostos para o FTAM junto com as estruturas de arquivos mais comuns. Com isso será feita uma análise de quais os mapeamentos possíveis para cada caso e os problemas decorrentes. É importante ressaltar que o escopo deste artigo não aborda o protocolo de arquivo pois este se encontra bem definido nas normas referentes ao assunto.

2. VISÃO GERAL DO FTAM

O ambiente de FTAM provê serviços que permitem que sistemas incompatíveis cooperem eficazmente para a realização de operações de Transferência, Acesso e Gerenciamento de arquivos num contexto de interconexão de redes. Quanto aos serviços, o de Transferência consiste na troca de arquivos, o de Acesso diz respeito ao acesso a partes do arquivo (registros), e o Gerenciamento consta na leitura e modificação de atributos de um arquivo.

A interação entre duas entidades de aplicação FTAM se dá por meio do Protocolo de Arquivo Virtual (PAV). Este protocolo possui duas características fundamentais. A primeira se refere ao número de entidades envolvidas numa atividade de serviço FTAM. Por definição, apenas duas E.A.'s (Entidade de Aplicação [8]) poderão estar interagindo para a efetuação de algum serviço FTAM (através do PAV). A segunda característica diz

respeito a assimetria dessa interação, isto é, uma das E.A. será a mestre ou iniciadora e a outra será escrava ou respondente. Contudo, em qualquer momento, uma E.A. pode estar envolvida em mais de uma instância de serviços FTAM. Cada instância está associada a requisitos de serviços independentes.

A atividade é iniciada por um usuário (iniciador), que possui algum objetivo a alcançar. A entidade escrava (respondente) simplesmente reage às iniciativas de modo passivo quanto a solicitação de serviços, porém ativo no que diz respeito a filtragem dos pedidos efetuados pelo mestre. Na transferência de dados, uma das entidades é a emissora e a outra é a receptora. Em qualquer instante existe uma direção preferida para o fluxo de dados.

Os serviços oferecidos pelo FTAM são definidos sob unidades funcionais. Uma classe de serviço consiste na combinação de um grupo de unidades funcionais para um determinado propósito (Transferência, Acesso, Gerenciamento, Transferência e Gerenciamento). O nível de serviço define se o protocolo é confiável ou não, isto é, se o protocolo provê mecanismos de recuperação de erros que são transparentes ao usuário (External File Service - EFS) ou não (Internal File Service - IFS). Deste modo, em cada inicialização de uma associação de aplicação [8] é necessário negociar a classe de serviço, as unidades funcionais opcionais, o nível de serviço e os modelos de arquivos virtuais suportados.

A operação do protocolo de FTAM cria um contexto para realização das atividades desejadas pela entidade iniciadora. A criação desse contexto se dá em estágios. O período de tempo em que alguma informação contextual permanece válida é chamado de regime. São definidos quatro regimes: FTAM, Seleção, Arquivo Aberto e Transferência de Dados. O regime de FTAM visa estabelecer uma associação de aplicação correta, e negociar as classes de serviço e unidades funcionais que estarão disponíveis. O regime de Seleção tem como objetivo identificar uma única entidade de armazenamento (arquivo) e o escopo sob o qual as operações requisitadas serão efetuadas. Já no regime de Arquivo Aberto é feita a modelagem virtual do arquivo para posterior transferência no regime de Transferência de Dados. Esses regimes são aninhados e a conclusão de um regime encerra todos os outros que o sucederam, sendo que em cada associação de aplicação só pode ocorrer uma instância de cada regime. Um regime concluído pode ser reiniciado quando necessário ou desejado.

O SASE (Specific Application Service Element) FTAM utiliza obrigatoriamente os serviços comuns de aplicação ACSE (Association Control Service Element) e opcionalmente o CCR (Commitment Concurrency and Recovery), dependendo das unidades funcionais negociadas na inicialização. A utilização dos serviços de CCR torna algumas primitivas de serviço atômicas, proporcionando maior confiabilidade no protocolo.

Existem dois tipos de protocolos especificados pela norma 8571/4: o protocolo básico

e o de recuperação. O protocolo básico suporta o nível de serviço corrigível pelo usuário. O protocolo de recuperação suporta o nível confiável (EFS) utilizando o IFS, ou seja, ele é uma extensão do protocolo (IFS). A utilização destes protocolos também depende das unidades funcionais negociadas na inicialização.

Como se pode notar o protocolo de FTAM é bem fechado e não oferece grande complexidade. O problema reside na questão de mapeamento dos arquivos virtuais para o formato real, e vice-versa, pois cada ambiente de implementação pode oferecer um estilo de armazenamento de arquivos diferente, se pensarmos em uma rede heterogênea. Além disto, é necessário escolher um modelo virtual adequado para mapear o arquivo real sem que a semântica do arquivo seja perdida. Para melhor abordarmos o assunto, serão introduzidos os principais conceitos relativos ao Sistema de Arquivamento Virtual.

3. SISTEMA DE ARQUIVAMENTO VIRTUAL

Denomina-se Sistema de Arquivamento Virtual (S.A.V.), o gerenciador que controla o mapeamento dos modelos virtuais para os possíveis formatos reais, e vice-versa, interagindo com o protocolo de FTAM (Figura 1).

Podemos fazer uma analogia do S.A.V. com o Sistema de Arquivos Convencional (S.A.C.) de uma determinada máquina. O S.A.C. virtualiza, para seus usuários, a forma como o arquivo está armazenado fisicamente. Podemos até pensar que ele age como um algoritmo que recupera/armazena informações de unidades físicas de discos, fitas, etc. As informações são apresentadas num formato organizado para que o usuário tenha a impressão de que o arquivo esteja realmente armazenado no modelo apresentado. Do mesmo modo o S.A.V. provê uma visão padronizada dos possíveis arquivamentos convencionais, através dos modelos virtuais, de tal forma que as diversas E.A.'s possam trocar e compartilhar arquivos de uma forma pré-estabelecida (Figura 2).

MODELOS VIRTUAIS

Os modelos virtuais possuem as propriedades de uma estrutura hierárquica. São elas:

- a) A estrutura de acesso ao arquivo é uma árvore ordenada;
- b) Cada nó possui associado zero ou uma unidade de dado (D.U.). A D.U. é a menor unidade capaz de ser acessada. Para propósitos de manipulação a D.U. pode ser sub-dividida em partes menores denominadas elementos de dados (data elements);
- c) Cada nó na estrutura dá acesso a uma sub-árvore, que é denominado FADU (FILE ACCESS DATA UNIT), que é compreendida pelos nós da sub-árvore com suas

unidades de dados. Neste âmbito o nó raiz da árvore dá acesso ao arquivo todo;

- d) Opcionalmente cada nó possui um nome associado;
- e) O número de níveis, o comprimento de arco, e o número de arcos originários de cada nó não são restringidos.

Analisando de outra forma, os modelos virtuais utilizam as FADU's para descrever a posição de cada nó dentro da estrutura hierárquica, denotando seus irmãos, filhos e pai, definir se possuem unidade de dado associadas e revelar sua identidade dentro da estrutura. A estes modelos virtuais adicionado às ações válidas sobre o arquivo dá-se o nome de conjunto restrito. Os conjuntos restritos definidos pelas normas são *unstructured*, *flat* e *hierárquico*. Em seguida será feita uma descrição sucinta de cada um deles:

- (i) **UNSTRUCTURED**: Consiste num modelo no qual existe um nó raiz associado a uma única unidade de dado (Figura 3). Esta organização é um caso especial do modelo hierárquico. Aplica-se a arquivos passíveis de serem transferidos ou acessados como um todo.
- (ii) **FLAT**: Consiste de um nó raiz e uma série de nós folhas conectados a raiz por arcos de comprimento um. Cada nó folha possui associado uma unidade de dado (Figura 4). O acesso a este tipo de modelo pode ser feito numa base individual.
- (iii) **HIERÁRQUICO**: Já descrito no início deste tópico (Figura 5).

ARQUIVOS CONVENCIONAIS

Os métodos de acesso utilizam, nos arquivos convencionais, registros para apresentarem seus dados de maneira organizada. A esses registros podem estar associados números de seqüência, índices e chaves de acesso, revelando assim o conteúdo arquivo. Para cada organização estão associadas ações válidas no arquivo. Neste contexto denomina-se Sistema de Arquivos o gerenciador que efetua o mapeamento dos dados físicos para o formato real, e vice-versa, interagindo com o usuário.

Os métodos de acesso mais difundidos são *stream*, *sequencial*, *relativo* e *sequencial indexado*. Os tres últimos, seguindo padrão ANSI definido para a entrada/saída padrão da linguagem COBOL. Em seguida será feita uma descrição sucinta de cada um deles:

- (i) **STREAM**: Aplica-se a arquivos que se apresentam como uma seqüência de octetos. O acesso aos dados se faz sequencialmente segundo tamanho desejado pelo usuário.
- (ii) **SEQUENCIAL**: Aplica-se a arquivos que admitem registros de tamanho fixo ou variável, cujo acesso se dá na ordem estabelecida pela gravação dos mesmos. Não existe acesso direto a um registro, só é permitido a recuperação do registro sucessor ao último acessado, a gravação do registro no final do arquivo, e a atualização do

registro corrente.

- (iii) *RELATIVO*: Aplica-se a arquivos que admitem registros de tamanho fixo ou variável, limitado ao tamanho de célula, cujo acesso se dá na ordem sequencial e direta dos registros. O acesso direto se dá pelo número da cédula, (posição) relativa do registro em relação ao início do arquivo.
- (iv) *SEQUENCIAL INDEXADO*: Aplica-se a arquivos que admitem registros de tamanho fixo ou variável, cujo acesso se dá pela ordem sequencial e direta dos registros. No acesso sequencial pode-se recuperar/atualizar um registro segundo a ordem física ou pela ordem de suas respectivas chaves dentro do arquivo. O acesso direto se dá através da chave (sequência de caracteres) ou via um identificador (índice) que determina unicamente a posição do registro dentro do arquivo.

4. SOLUÇÕES PARA OS PROBLEMAS DE MAPEAMENTO

Como já mencionado, o gerenciamento de arquivos no contexto do FTAM é uma questão local de cada sistema, cada implementador fornecendo a sua solução. Com base na experiência das autoras foi concluído a não necessidade de manter um sistema virtual de arquivos em coexistência com o inerente a máquina em questão (real), pois a sobrecarga para manter o real compatível com o virtual é muito grande se pensarmos em um sistema multi-usuário. Isto foi constatado também, pela forma com que as normas abordam a questão. Isto é, muitas questões não são fechadas e alguns pontos mais complexos são deixados como opcionais propiciando a oportunidade ao projetista especificar e viabilizar desde um modelo simples até um de maior complexidade, dependendo do Sistema Operacional a que se aplica e ao objetivo da implementação.

Como já definido o respondente apenas reage as solicitações do usuário iniciador. Devido ao fato destas solicitações envolverem interações com arquivos, entendeu-se que o S.A.V, na realidade, é o elemento usuário da E.A. respondente. Deste modo ele filtra as ações que um usuário remoto possa solicitar, apesar do serviço ser oferecido, e age como interface entre o sistema de arquivos e os modelos virtuais, resolvendo a questão de mapeamento.

Em seguida serão apresentados algumas propostas de mapeamento baseado nos conceitos apresentados nos itens anteriores. Cabe ressaltar que um arquivo que se encontra originalmente sob uma organização inerente a uma máquina local pode ser transferido de acordo com um modelo virtual qualquer, previamente negociado, e ao ser recebido, na máquina remota, ser armazenado em outra organização qualquer.

Caso o arquivo, no formato transferido e/ou recebido, não possua um significado coerente com o S.O. da máquina receptora, caberá ao usuário desenvolver um conversor

de arquivos, de tal modo que se recupere a semântica do arquivo original. Deste modo, uma organização complexa pode ser transferida num modelo virtual simples, cabendo ao receptor entender seu conteúdo e remontá-lo de acordo com seu objetivo. Este tipo de operação não é um atributo do S.A.V.

Devido ao fato das organizações de arquivos encontrarem uma correspondência visível com os modelos virtuais, optou-se por utilizar, com algumas restrições, as potencialidades dos modelos virtuais.

Em seguida serão apresentadas algumas propostas de mapeamento:

(i) STREAM

Neste caso, o mapeamento é realizado diretamente sob o modelo *unstructured*, onde cada octeto do stream é mapeado como um elemento de dado e um stream completo consiste na D.U. do modelo *unstructured*.

(ii) SEQUENCIAL

Mapeia-se esta estrutura no modelo *flat*, onde cada registro consiste de uma D.U., de tamanho variável ou não. Neste caso as FADU's associadas as sub-árvores não possuem identificação. Apenas o nó raiz possui uma identificação, o nome do arquivo.

Durante o processo de análise de como mapear os arquivos relativos e sequencial indexado nos modelos virtuais propostos, verificou-se que os modelos virtuais não ofereciam suporte suficiente para abrigar informações necessárias a manutenção da semântica do arquivo. Para exemplificar, como seriam passadas informações como tamanho máximo da cédula de um arquivo relativo? Como seriam as chaves e índices de um arquivo sequencial indexado? Inicialmente levantou-se a possibilidade de passar como uma primeira FADU do arquivo (em sua D.U.) tais informações. Mas tal solução obrigaria o receptor do arquivo não só saber de antemão qual seria o modelo virtual escolhido e o formato original do arquivo, para que ele pudesse interpretar o conteúdo da D.U. apropriadamente, como também a adulteraria o conceito de conteúdo do arquivo, adicionando uma D.U. que realmente não existe.

A segunda solução levantada foi a de passar o arquivo sem estas informações, assumindo que o usuário que solicitou o arquivo já conhece sua organização a priori. Tal solução baseou-se numa analogia feita no fato de que, um programador de sistemas, ao ler um arquivo de uma unidade de fita por exemplo, já conhece toda a especificação a do arquivo, isto é, tais informações não fazem parte do conteúdo do arquivo. Do mesmo modo um usuário dos serviços FTAM ao requisitar a transferência de um arquivo sequencial indexado, provavelmente utilizará tal arquivo em uma determinada aplicação, por exemplo banco de dados, e saberá como pesquisar e organizar o arquivo (quais as chaves e índices, tamanho de registro, etc.) de modo que este seja útil para sua aplicação.

A terceira solução assemelha-se a primeira, a diferença reside no fato de que, ao invés das informações de controle serem enviadas como uma primeira FADU, definiria-se um contexto de apresentação (FTAM-FILE-CONTROL) no qual todas as informações de controle, se houvessem, seriam enviadas para que a E.A. receptora pudesse entender a semântica do arquivo que seria recebido. Para que este procedimento possa ser efetuado, uma sintaxe abstrata baseada, na ASN-1 deve ser elaborada de modo a prover a maior flexibilidade possível na descrição das informações de controle, de forma a abranger o maior número possível de organizações de arquivos. A alteração do contexto de apresentação, de FTAM PCI para FTAM-FILE-CONTROL, se daria após o envio das primitivas de serviço FTAM [3] F-READ e F-WRITE, através da primitiva de serviço da apresentação P-ALTER. Após o envio das informações de controle, haveria uma nova alteração de contexto para o contexto FTAM-FADU.

Esta solução não desobedece as normas pois é permitido a definição de novas sintaxes abstratas e contextos de apresentação. É função da camada de apresentação negociar quais os contextos válidos em uma conexão de apresentação.

Os seguintes mapeamentos serão baseados na terceira proposta apresentada:

(iii) RELATIVO

Mapeia-se esta estrutura no modelo *flat*, onde cada registro consiste de uma D.U. O número da cédula não identifica a FADU. As informações de controle, Control-Info, são passadas no contexto de apresentação FTAM-FILE-CONTROL, de acordo com a sintaxe abstrata descrita abaixo.

(iv) SEQUENCIAL INDEXADO

Mapeia-se esta estrutura no modelo *flat* onde cada registro consiste de uma D.U. O número do índice identifica a FADU, assim como a chave de acesso. As informações de controle são passadas no contexto de apresentação FTAM-FILE-CONTROL, de acordo com a sintaxe abstrata descrita abaixo.

Control-Info ::= [APPLICATION 0] CHOICE {

cell-size [0] IMPLICIT INTEGER,
index [1] IMPLICIT SEQUENCE OF SEQUENCE,
index-number [0] IMPLICIT INTEGER OPTIONAL,
index-type [1] IMPLICIT BOOLEAN {
primary (0),
secondary (1) }
key [2] IMPLICIT SEQUENCE OF SEQUENCE {
initial-position [0] IMPLICIT INTEGER,
size [1] IMPLICIT INTEGER,

ordering [2] IMPLICIT BOOLEAN {

ascending (0),

descending (1) }}}

5. CONCLUSÕES

Observa-se uma grande variedade de propostas de mapeamento. A opção de qual delas adotar irá depender das necessidades de cada aplicação.

O ideal seria que os futuros sistemas definissem as estruturas de seus arquivos seguindo ao máximo a proposta dos modelos virtuais apresentados pela OSI. Desta forma o esforço de interconexão, no âmbito de transferência e acesso a arquivos, seria drasticamente reduzido. Porém, como não se pode deixar de levar em consideração todos os sistemas já existentes com suas definições de arquivos já tão solidificadas, e num momento em que novos esforços de implementação de sistemas procuram não inviabilizar a compatibilidade com os anteriores, esta possibilidade apresenta-se muito teórica, e quando a prática ainda é muito remota.

Uma outra maneira de minimizar este esforço seria enriquecer os modelos virtuais apresentados de modo a englobar o maior número possível de organizações de arquivos nestas estruturas. Nesta ótica a utilização de contextos de apresentação específicos seria reduzida, aumentando a capacidade de interconexão no contexto do FTAM.

6. REFERÊNCIAS

1. Information Processing Systems - Open Systems Interconnection - *File Transfer, Access and Management; Part 1: General Introduction* (1988 - ISO/IS 8571/1).
2. Information Processing Systems - Open Systems Interconnection - *File Transfer, Access and Management; Part 2: Virtual Filestore Definition* (1988 - ISO/IS 8571/2).
3. Information Processing Systems - Open Systems Interconnection - *File Transfer, Access and Management; Part 3: File Service Definition* (1988 - ISO/IS 8571/3).
4. Information Processing Systems - Open Systems Interconnection - *File Transfer, Access and Management; Part 4: File Protocol Specification* (1988 - ISO/IS 8571/4).
5. Keith g. Knightson, Terry Knowles, John Larmouth. *Standards for Open Systems Interconnection*, Mc Graw Hill 1987.
6. Pacitti, Esther de Castro. *Estudos dos Serviços de Transferência, Acesso e Manipulação de Arquivos em uma Rede de Sistemas Abertos*, Relatório Técnico, NCE - 01187, Julho de 1987.

7. Pacitti, Esther de Castro. Análise e Especificação dos Serviços de Transferência Acesso e Manipulação de Arquivos ("FTAM"), *XIII Conferência Latino Americana de Informática*, Novembro de 1987, Bogotá - Colômbia.
8. Carmo, Luiz F.R.C.. *Ambiente OSI de Transferência e Manipulação de Tarefas (JTM) para um Sistema VAX/VMS*, Tese - Universidade Federal do Rio de Janeiro - COPPE, Rio de Janeiro, Setembro 1988.
9. Information Processing Systems- Open Systems Interconnection - *Basic Connection Oriented Presentation Protocol Specification*; 1988 (ISO/IS 8823).
10. Gien, Michel. A File Transfer Protocol (FTP), *Computer Networks 2*, 1978,312-319.
11. Linington, Peter F.. The Virtual Filestore Concept, *Computer Networks 8*, 1984, 13-16.
12. Information processing Systems - Open Systems Interconnection - *Specification of Abstract Notation One (ASN.1)*.
13. Information Systems - Programmer Language COBOL - ASN.1 X3-23-1974.
14. Sauv e, Jacques, *A Organiza o Interna do Unix* - 1985.
15. Sistema de Arquivamento Estendido SOX (SAX) - COBRA (1989).

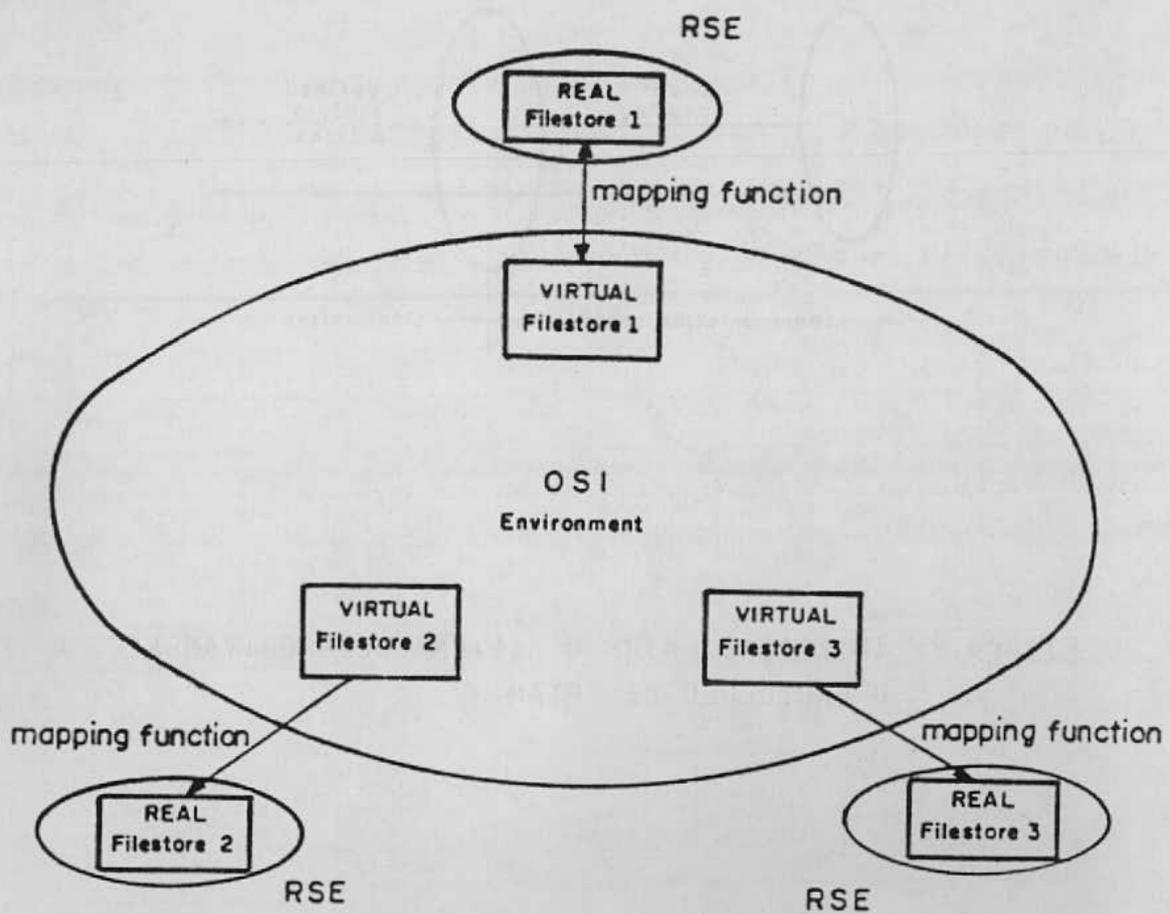


Figura 1: MAPEAMENTO ENTRE ARQUIVOS CONVENCIONAIS E VIRTUAIS

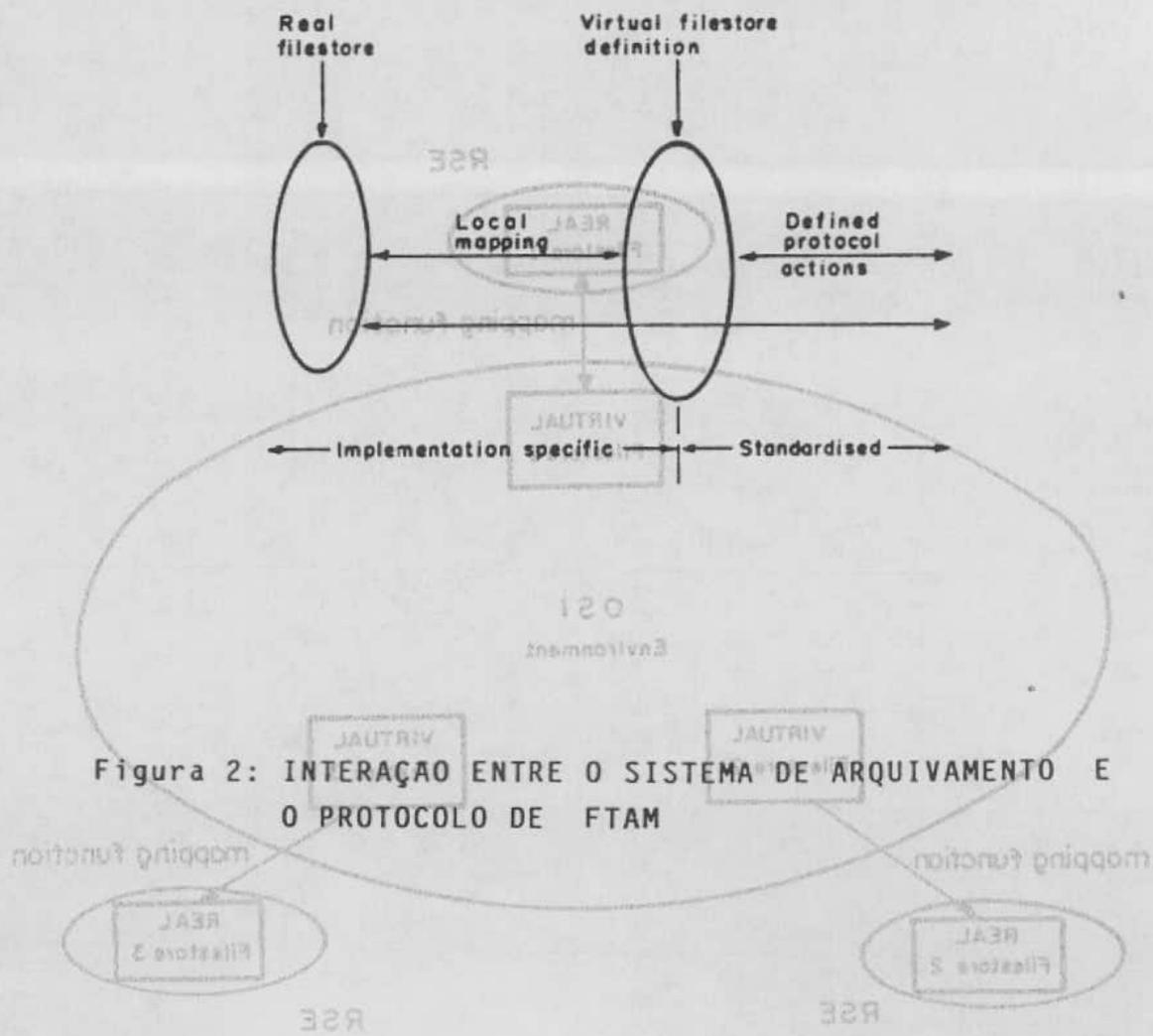


Figura 2: INTERAÇÃO ENTRE O SISTEMA DE ARQUIVAMENTO E O PROTOCOLO DE FTAM

Figura 1: Mapeamento entre arquivos convencionais e virtuais

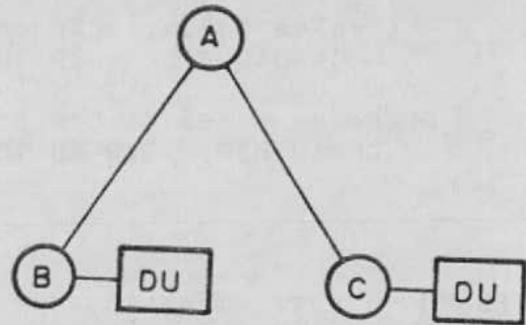
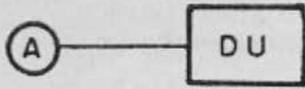


Figura 3: MODELO "UNSTRUCTURED"

Figura 4: MODELO "FLAT"

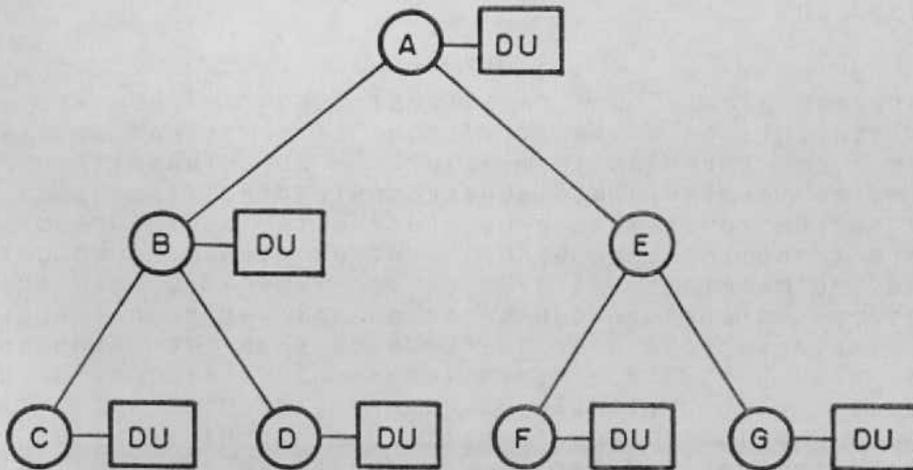


Figura 5: MODELO HIERÁRQUICO