Uma Arquitetura Baseada em SCTP e SIP para Suporte a Aplicações VoIP Móveis e a Especificação Formal do seu Módulo de Controle

Daniel Gouveia Costa

Natal-RN
Maio de 2006

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

CENTRO DE TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Uma Arquitetura Baseada em SCTP e SIP para Suporte a Aplicações VoIP Móveis e a Especificação Formal do seu Módulo de Controle

Daniel Gouveia Costa

Orientador: Prof. D. Sc. Sergio Vianna Fialho.

Documento apresentado como parte dos requisitos para Defesa de Dissertação do curso de Mestrado em Engenharia Elétrica pelo Programa de Pós-graduação em Engenharia Elétrica da Universidade Federal do Rio Grande do Norte.

Natal-RN
Maio de 2006

Divisão de Serviços Técnicos
Catalogação da Publicação na Fonte / Biblioteca Central Zila Mamede

Costa, Daniel Gouveia.
Uma arquitetura baseada em SCTP e SIP para suporte e aplicações VoIP móveis e a especificação formal do seu módulo de controle / Daniel Gouveia Costa. – Natal, 2006.
... p. : il.

Orientador: Sergio Vianna Fialho.
Dissertação (Mestrado) – Universidade Federal do Rio Grande do Norte. Centro de Tecnologia. Programa de Pós-Graduação em Engenharia Elétrica.

1. Redes de computação (Protocolos) – Dissertação. 2. Comunicação de dados – Dissertação. 3. SCTP (Protocolo de rede de computação) – Dissertação 4. SIP (Protocolo de rede de computação) – Dissertação. 5. VoIP móvel - (Protocolo de rede de computação) – Dissertação. I. Fialho, Sergio Viana. II. Universidade Federal do Rio Grande do Norte. III. Título.

RN/UF/BCZM CDU 004.057.4

Many people spend more time and energy in going around problem than in trying to solve them.

Henry Ford

Agradecimentos

Gostaria primeiramente de agradecer a Deus, por todas as oportunidades que me foram oferecidas. Em seguida, queria agradecer à minha família por todo o apoio dado para o desenvolvimento desse projeto. Um agradecimento mais que especial é destinado a minha esposa, companheira e amiga de todas as horas.

Gostaria de agradecer também ao PoP-RN e a todos os seus membros, bem como os professores e funcionários do PPgEE/UFRN. Especialmente, um agradecimento é oferecido ao meu amigo e orientador, professor D. Sc. Sergio Vianna Fialho.

Por fim, gostaria de agradecer o apoio de todos os meus amigos e colegas.

Resumo

Novas versões do protocolo SCTP permitem sua utilização para implementação de mecanismos de *handover* em nível de transporte, bem como o fornecimento de um serviço de transmissão de dados parcialmente confiável. Integrando o SCTP com o protocolo de iniciação de sessões, SIP, além de utilizar adicionalmente serviços de outros protocolos auxiliares, uma arquitetura de comunicação pôde ser proposta, a fim de atender às aplicações de voz sobre IP com requisitos de mobilidade.

São especificados ainda os procedimentos de localização de usuário em nível de aplicação, utilizando o protocolo SIP, como alternativa aos mecanismos empregados por protocolos tradicionais que suportam mobilidade na camada de rede.

A linguagem de especificação formal SDL é utilizada para especificar o funcionamento de um Módulo de Controle, relacionado à operação coordenada dos protocolos que compõe a arquitetura. Pretende-se assim evitar ambigüidades e inconsistências na definição desse módulo, o que pode auxiliar em implementações corretas de elementos dessa arquitetura.

Abstract

New versions of SCTP protocol allow the implementation of handover procedures in the transport layer, as well as the supply of a partially reliable communication service. A communication architecture is proposed herein, integrating SCTP with the session initiation protocol, SIP, besides additional protocols. This architecture is intended to handle voice applications over IP networks with mobility requirements.

User localization procedures are specified in the application layer as well, using SIP, as an alternative mean to the mechanisms used by traditional protocols, that support mobility in the network layer.

The SDL formal specification language is used to specify the operation of a control module, which coordinates the operation of the system component protocols. This formal specification is intended to prevent ambiguities and inconsistencies in the definition of this module, assisting in the correct implementation of the elements of this architecture.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Acrônimos

CAD – Computer Aided Development

CN – Correspondent Node

CoA – Care of Address

DHCP – Dynamic Host Configuration Protocol

DNS – Domain Name System

FA – Foreign Address

HA – Home Agent

HIP – Host Identity Protocol

HoA – Home Address

IP – Internet Protocol

IPSec – IP Security

IPv4 – Internet Protocol version 4

IPv6 – Internet Protocol version 6

LAN – Local Area Network

MAN – Metropolitan Area Network

MGCP – Media Gateway Control Protocol

MIP – Mobile IP

MIPv4 – Mobile IPv4

MIPv6 – Mobile IPv6

MN – Mobile Node

MSCTP – Mobile Stream Control Transmission Protocol

NAT – Network Address Translator

PR-SCTP – Parcial Reliability Stream Control Transmission Protocol

QoS – Quality of Service

RTCP – Real Time Control Protocol

RTP – Real Time Protocol

SCTP – Stream Control Transmission Protocol

SDL – Specification and Definition Language

GR-SDL – Graphical Representation - SDL

SDP – Session Description Protocol

SIP – Session Initiation Protocol

SRTCP – Secure Real Time Control Protocol

SRTP – Secure Real Time Protocol

SS7 – Signaling System Number 7

TCP – Transport Control Protocol

TLS – Transport Layer Security

UDP – User Datagram Protocol

VoIP – Voice over Internet Protocol

WAN – Wide Area Network

WLAN – Wireless LAN

WMAN – Wireless MAN

WWAN – Wireless WAN

1 Introdução

A necessidade de integração no transporte das mídias de voz, vídeo e dados numa mesma rede para atender às exigências de algumas novas aplicações, tais como videoconferência e telefonia IP, impulsionou o surgimento de padrões de comunicação multimídia na Internet, a exemplo do H.323 [Hersent 2002], SIP (Session Initiation Protocol) [Rosemberg 2002], MGCP (Media Gateway Control Protocol) [Hersent 2002] e outros. Paralelamente, nesses últimos tempos, a questão da mobilidade nas redes de computadores vem se impondo e exigindo novas soluções robustas, tanto para as aplicações quanto para o subsistema de comunicação subjacente, sobretudo quando os ambientes usados para telefonia e comunicação de dados tendem a convergir também na área de comunicação móvel. Nesse cenário, a demanda por aplicações de Voz sobre IP (VoIP) [Hersent 2002], quando essas utilizam enlaces de conexão sem fio, encoraja a especificação de arquiteturas de comunicação móvel adaptadas a novos requisitos operacionais.

O SCTP (Stream Control Transmission Protocol) [Costa 2005a] é um protocolo de transporte, inicialmente definido para permitir o envio de sinalização telefônica através da Internet, mas que, além disso, pode prover uma série de serviços atraentes a diversas aplicações dessa rede. Com algumas adaptações no protocolo original, arquiteturas de comunicação mais robustas podem ser construídas. Uma dessas adaptações, que se beneficia da característica de *multihoming* desse protocolo, aliada a mecanismos de configuração dinâmica de endereços IP [IETF 1981], torna o SCTP capaz de realizar algumas operações necessárias ao suporte de mobilidade na Internet. Dessa forma, mecanismos antes executados na camada de rede podem ser migrados para processamento na camada de transporte dessa arquitetura de comunicação. Aliada a essa característica, o SCTP provê também um serviço opcional de transmissão parcialmente confiável, o que garante a coexistência de dados com características de confiabilidade diversas numa mesma comunicação.

O protocolo SIP é utilizado para criar, modificar e encerrar sessões multimídia entre um ou mais participantes. Em conjunto com protocolos de descrição de capacidades de processamento de mídia [Handley 1998] e de transporte de dados com

características de tempo real [Schulzrinne 1996], o SIP integra uma das mais utilizadas arquiteturas de comunicação multimídia na Internet.

Considerando um ambiente de mobilidade SCTP empregando padrões de sinalização e de transporte de dados sensíveis ao tempo, neste trabalho é descrita uma nova arquitetura de comunicação que utiliza o SCTP em conjunto com o protocolo SIP. Nessa arquitetura, os esquemas de localização de usuário necessários a um ambiente móvel serão executados em nível de aplicação, utilizando mecanismos disponibilizados pelo protocolo SIP. Além disso, essa arquitetura pretende simplificar as implementações desse protocolo referentes a tratamentos de *handover*, eliminado a necessidade desse tipo de operação em nível de aplicação. Os mecanismos de *handover* inerentes a um ambiente móvel de comunicação deverão ser executados em nível de transporte pelo protocolo SCTP.

Na arquitetura de comunicação considerada nesse trabalho, protocolos auxiliares são utilizados para oferecer serviços opcionais a seus usuários. Em relação à segurança de dados, os serviços de autenticação e criptografia são opcionalmente oferecidos ao fluxo de voz pelo protocolo SRTP (Secure Real Time Protocol) [Baugher 2004]. Além disso, o envio de pequenas mensagens de texto entre os usuários da arquitetura estará disponível, sendo implementado através da transmissão de textos redundantes, carregados em um *payload* específico [Hellstrom 2000] do RTP (Real Time Protocol) [Schulzrinne 1996].

Trabalhos anteriores já abordaram a utilização do SIP sobre o SCTP [Marco 2004], mecanismos de mobilidade em nível de aplicação pelo SIP [Wedlung 2004] e a utilização do SCTP em conjunto com padrões de mobilidade em nível de rede [Nooman 2004]. Entretanto, em [Marco 2004] o SCTP é encarregado apenas de realizar o transporte de sinalização SIP entre terminais ou *proxies* comunicantes. Em [Wedlung 2004], operações de suporte a mobilidade são disponibilizadas apenas em nível de aplicação pelo SIP, utilizando elementos operacionais já especificados por esse protocolo. For fim, [Nooman 2004] aborda a integração do SCTP com padrões de mobilidade em nível de rede para oferecer suporte às aplicações móveis. Por outro lado, esse trabalho propõe uma arquitetura que utiliza os protocolos SCTP e SIP em conjunto, para oferecer serviços de mobilidade e de transmissão de dados com requisitos de tempo real, eliminando a necessidade de suporte à mobilidade em nível de rede, como em

[Nooman 2004], e em nível de aplicação, como em [Wedlung 2004]. Além disso, esse trabalho especifica a utilização do SCTP para transporte de sinalização SIP e de dados com características de tempo real, oferecendo um serviço adicional àquele presente em [Marco 2004].

Com o objetivo de descrever mais precisamente a operação dos elementos que compõem a arquitetura, garantindo a ausência de ambigüidades e inconsistências, a linguagem de especificação formal SDL (Specification and Description Language) [SDL 2005] será utilizada. Utilizando essa linguagem, o funcionamento de um Módulo de Controle, responsável pela coordenação da operação conjunta dos protocolos constituintes da arquitetura, será especificado. Essa especificação pretende auxiliar em implementações de elementos da arquitetura.

A estrutura do trabalho está organizada da seguinte forma: no capítulo dois são apresentados os conceitos relativos aos principais protocolos e padrões da arquitetura de comunicação considerada e que são relevantes para a solução proposta, bem como as definições envolvidas com o processo de especificação formal de sistemas distribuídos. No capítulo três é descrita a arquitetura de comunicação a ser trabalhada, baseada na utilização dos protocolos SCTP, SIP e em protocolos auxiliares, cujo propósito é atender às aplicações VoIP com requisitos de mobilidade. No capítulo quatro, a especificação em SDL de um Módulo de Controle da arquitetura é abordada. Por fim, são apresentadas a conclusão do trabalho e as referências bibliográficas.

2 *Tecnologias de referência*

Esse capítulo apresenta os protocolos e tecnologias relacionadas à operação da arquitetura descrita no capítulo três, bem como os procedimentos de especificação formal de sistemas distribuídos, utilizados nas especificações contidas no capítulo quatro.

2.1. SCTP

O SCTP é um protocolo de transporte da pilha TCP/IP que opera num nível equivalente ao TCP [DARPA 1981] e ao UDP [IETF 1980]. A Figura 2.1 apresenta a localização lógica do SCTP nessa pilha de protocolos.

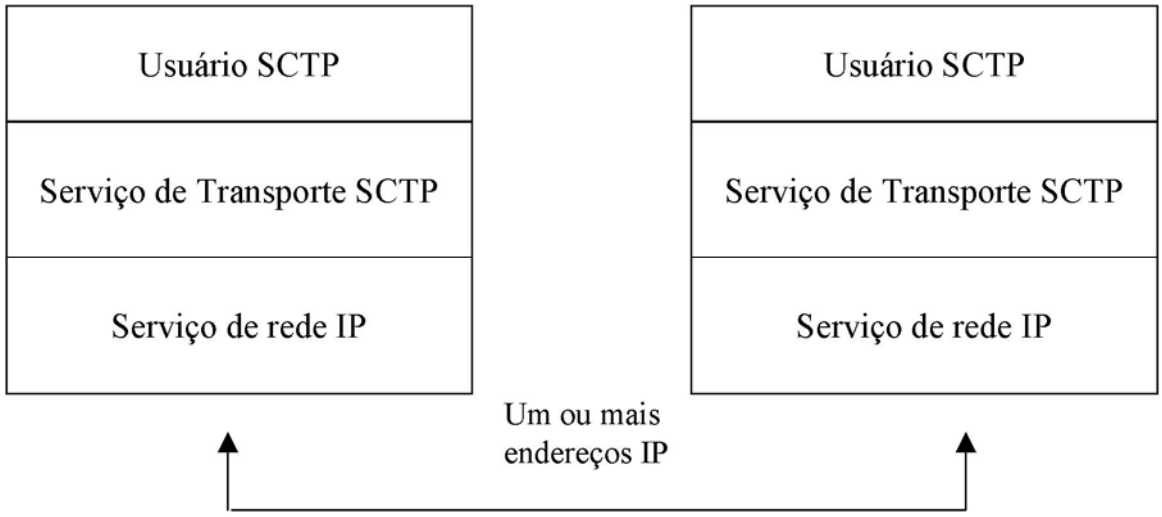


Figura 2.1. Localização lógica do SCTP na pilha TCP/IP.

Especificado originalmente para transportar sinalização telefônica pela Internet, como mensagens SS7 (Signalling System Number 7) [SS7 2005], a aplicabilidade do SCTP foi logo reconhecida como vantajosa em outros escopos de operação, como aqueles ligados às aplicações tradicionais dessa rede. A grande vantagem do SCTP é que ele fornece um número considerável de funções de comunicação robustas e

flexíveis a diversas aplicações, funções essas não presentes nos protocolos TCP ou UDP.

Assim como o TCP, o SCTP fornece entrega confirmada de dados, livre de erros e não duplicados, sendo caracterizado, portanto, como um protocolo de transporte confiável. Adicionalmente, as seguintes funcionalidades são oferecidas pelo protocolo SCTP padrão:

- Entrega seqüencial de dados de usuário em fluxos lógicos;
- Verificação seletiva de erros independente para cada fluxo da comunicação;
- Empacotamento opcional de mensagens num único pacote SCTP;
- Tolerância à falhas de rede através do suporte a caminhos (endereços) múltiplos.

O SCTP é, então, um protocolo orientado a fluxo, onde os dados de usuário devem ser transportados em fluxos *simplex* distintos. Outra característica marcante desse protocolo é o suporte a múltiplos endereços IP numa "conexão" SCTP, permitindo redundância de caminhos em caso de falhas na rede. Essa característica é conhecida como *multihoming* [Stewart 2000].

O protocolo SCTP é orientado a conexão, estabelecendo uma associação SCTP com um número arbitrário de fluxos *simplex* de transmissão e recepção. O escopo de ordenação das mensagens de usuário é restrito a cada fluxo lógico particular, garantindo assim uma maior eficiência em possíveis retransmissões de dados: mensagens pertencentes a fluxos onde não há perdas não são atrasadas ou retransmitidas devido a retransmissões em outros fluxos da mesma associação. Para oferecer um serviço de "mensagem expressa", como no TCP, o SCTP define mensagens urgentes, que não são associadas a qualquer fluxo de comunicação. Sendo assim, nenhum escopo de ordenação é aplicado a essas mensagens.

Diferentemente do TCP, que é orientado a *octetos*, o SCTP é um protocolo orientado a mensagens. Nesse esquema, as informações transmitidas em uma comunicação são transportadas em blocos de dados (*chunks*), que podem ser de usuários (dados da aplicação) ou de controle do protocolo. Mensagens de usuário e de controle podem estar contidas em um mesmo pacote SCTP. Os blocos de controle desempenham diversas tarefas necessárias à operação desse protocolo, como o reconhecimento de

mensagens de usuários recebidas e o estabelecimento de associações. Esse último procedimento é realizado pela troca de quatro mensagens de controle entre cliente e servidor, processo esse conhecido como *four-way handshake*. O fato de utilizar um esquema de mensagens de controle garante ao SCTP uma maior flexibilidade a adições em sua estrutura operacional. Em contrapartida, o TCP possui as funções de controle embutidas em sua estrutura de segmento, o que embora traga benefícios de eficiência computacional de processamento, inibe muitas tentativas de adaptação e evolução do protocolo.

Por fim, o protocolo SCTP é *rate adaptive*, o que garante adaptação dinâmica às variações e problemas ocorridos na rede, de forma a reduzir os impactos negativos causados à aplicação comunicante e à infra-estrutura de comunicação correspondente. Devido à eficiência comprovada em anos de operação, os algoritmos de adaptação a congestionamentos do TCP foram aproveitados no SCTP, com pequenas variações, pelos órgãos padronizadores.

2.2. PR-SCTP

O PR-SCTP (Partial Reliability – Stream Control Transmission Protocol) [Stewart 2004] é uma extensão que define um serviço de transporte parcialmente confiável para ser utilizado com o SCTP. Para tanto, adições ao protocolo são feitas em termos de mensagens e parâmetros de controle, além de adaptação do comportamento padrão do SCTP em relação a algumas situações de comunicação. Devido à estrutura do SCTP, contudo, essas adaptações não causam qualquer comprometimento à operação padrão do protocolo, permitindo que o novo serviço especificado seja opcionalmente selecionado no início de uma associação.

Em outras palavras, essa extensão do SCTP permite ao usuário especificar quão persistente o serviço de transporte deve ser, na tentativa de enviar mensagens através da rede. Assim, além de fornecer transmissão de dados de forma não ordenada e não confiável, como o UDP, o PR-SCTP pode prover ao SCTP a possibilidade de

transmissão de dados de forma ordenada e não confiável, disponibilizando um serviço não encontrado no TCP ou UDP.

Como os pacotes SCTP são formados por mensagens independentes entre si, transmissões confiáveis e não confiáveis podem estar presentes numa mesma associação SCTP que utilize a extensão apresentada. Dessa forma, o número de datagramas IP, bem como o *overhead* de comunicação na rede, são diminuídos, pois outro protocolo de transporte, como o UDP, por exemplo, não seria empregado para transmissão dos dados com características de transmissão não confiáveis. Além disso, o número de portas de transporte necessárias é também reduzido.

Dados parcialmente confiáveis transportados pelo PR-SCTP terão a mesma capacidade de detecção de falha de comunicação e as mesmas capacidades de proteção de tráfego de dados confiáveis do SCTP. Isso inclui as habilidades de rapidamente detectar a falha de um endereço destino, alternar de endereço IP em caso de problemas na rede (*multihoming*) e ser notificado se um receptor de dados se tornar inalcançável [Costa 2005a].

2.3. Protocolos de rede para ambientes móveis – MIPv4 e MIPv6

Mobilidade no contexto das redes de computadores consiste na possibilidade de um determinado *host* poder trafegar entre diversas redes logicamente distintas, sem encerrar ou prejudicar qualquer comunicação que esteja em andamento: esse é o conceito de "mobilidade de terminal". Além disso, a idéia de mobilidade sugere que um *host* móvel possa ser encontrado, para fins de entrega de pacotes, em qualquer que seja sua localização atual.

Genericamente, a mobilidade na Internet está dividida em gerenciamento de localização e operações de *handover*. Gerenciamento de localização é utilizado para identificar a localização atual de um nó móvel e manter o "rastro" das mudanças de rede desse nó enquanto ele se move. *Handover* consiste na mudança dinâmica entre redes

diferentes sem alterar as comunicações atuais do *host* que se move, sendo essa operação necessária para se ter a mobilidade de terminal [Martins 2004].

A mobilidade em redes TCP/IP é naturalmente prejudicada, tendo em vista que os endereços IP identificam, em última análise, pontos de acesso à rede e não *host*s. Um mecanismo adicional muito adotado como solução para permitir mobilidade em redes TCP/IP é descrito na especificação MIP (Mobile IP), que pode estar relacionada à versão quatro [Perkins 2002] ou seis [Johnson 2004] do protocolo IP.

Na especificação MIP, um nó móvel (Mobile Node - MN) é um *host* capaz de alterar seu ponto de acesso à rede sem prejudicar as comunicações em andamento. Um MN está ligado a um endereço IP inicial (Home Address - HoA) no escopo da rede do seu *home agent* (HA), que é um *host* com funções especiais num ambiente MIP. O *home agent* mantém o "rastro" da mobilidade dos MN, criando túneis de comunicação em nível de rede para permitir um *handover* transparente dos nós móveis. Mecanismos de localização de terminais também são implementados nos HoA.

Na especificação MIP para IPv4 (MIPv4), um MN que está fora da sua *home network* deve se *registrar*, opcionalmente, junto a um *foreign agent* (FA), recebendo um endereço que pertence à rede desse elemento. Esse endereço IP é conhecido como *care-of address* (CoA) e muda a cada nova rede que o nó móvel se encontra, juntamente com o *foreign agent* onde o nó deve-se, opcionalmente, registrar. O não registro junto a um FA ainda permite a utilização dos serviços de mobilidade dessa especificação, contudo o nó móvel passa a ser o destino do túnel IP formado a partir do HA correspondente.

Um CN (Correspondent Node) é um nó da arquitetura MIP com o qual os MN realizam comunicações, podendo esse nó ter ou não suporte a mobilidade. Os mecanismos especificados em MIP visam atender, em nível de rede, às necessidades de comunicação entre CN e MN, no que se refere a mecanismos de *handover* e localização de *host*s. A Figura 2.2 apresenta um ambiente MIPv4 típico.

A especificação MIP para IPv6 (MIPv6) possui a mesma abordagem do MIPv4, com algumas modificações operacionais. No MIPv6, os *host* têm um papel mais importante na mobilidade, o que garante a não utilização de *foreign agents* nessa arquitetura, que é mais simples que a MIPv4. Os nós IPv6 comunicantes possuem uma tabela de informações de endereços de *host*s que é atualizada dinamicamente. Numa

comunicação, quando uma estação móvel informa seu endereço atual para o nó correspondente IPv6, a transmissão de pacotes pode ocorrer diretamente, não havendo necessidade de mecanismos de tunelamento IP nesse cenário [Johnson 2004]. Um estudo mais detalhado sobre mobilidade em nível de rede utilizando MIPv6 pode ser encontrado em [Schmidt 2005].

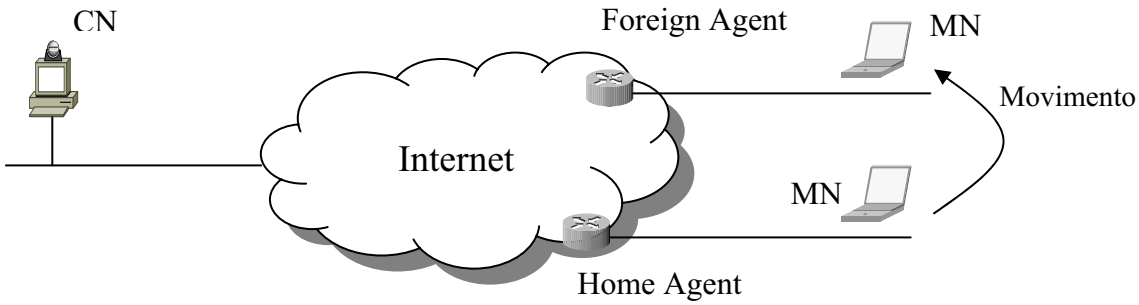


Figura 2.2. Ambiente MIPv4 típico.

2.4. Mobile SCTP

O Mobile SCTP (Mobile Stream Control Transmission Protocol) [Schmidt 2005], ou simplesmente MSCTP, é uma extensão que define o uso do SCTP em conjunto com mecanismos de configuração dinâmica de endereços IP numa associação [Xie 2005]. A idéia é explorar a característica de *multihoming* do SCTP, juntamente com mecanismos de configuração dinâmica de endereços, para permitir operações de *handover* numa arquitetura com demanda por comunicação móvel, garantindo assim "mobilidade de terminal" em nível de transporte. Dessa forma, utilizando o MSCTP em conjunto com a especificação MIP, por exemplo, os serviços de mobilidade dessa última restringem-se a mecanismos de localização de *host*s móveis.

No contexto do MSCTP, dois ambientes operacionais podem ser visualizados. O primeiro deles corresponde às comunicações iniciadas dos MN para os CN. Nesse cenário, serviços de localização de *host*s móveis, como os presentes em MIP, não são necessários e os *handovers* do nó móvel serão tratados de forma nativa pelo MSCTP. Já o segundo ambiente operacional representa as comunicações iniciadas em sentido contrário, onde um mecanismo de localização deve ser empregado. O MSCTP não

possui qualquer mecanismo para localização de *host*s móveis, ao contrário da especificação MIP, que possui elementos e mensagens específicas para o gerenciamento de localização de usuários.

Não há qualquer especificação do MSCTP de como um endereço IP é atribuído a um *host* móvel, nem como ocorre o *roaming* entre redes diferentes (tratado em nível de enlace). Numa arquitetura móvel, uma alternativa para atribuição de endereços é a utilização do protocolo DHCP (*Dynamic Host Configuration Protocol*) [Droms 1997].

É de se notar que as arquiteturas de mobilidade IP podem ser bastante simplificadas com a utilização do MSCTP. O *handover* do MSCTP é realizado ao nível de transporte, enquanto o *handover* MIPv4, por exemplo, é realizado com tunelamento IP, muito mais custoso à rede. O MIPv6 oferece serviços de mobilidade em nível de rede mais robustos e flexíveis que o MIPv4, porém apenas pode ser empregado numa infra-estrutura IPv6.

Outra solução para ambientes móveis, não tão adotada quanto às apresentadas anteriormente, baseia-se na utilização do protocolo HIP (*Host Identity Protocol*) [Moskowitz 2005]. Esse protocolo estabelece uma camada de processamento entre os níveis de rede e de transporte, com o intuito de intermediar a associação entre variáveis desses dois níveis lógicos, facilitando assim operações necessárias a arquiteturas móveis.

Uma comparação das vantagens e desvantagens das soluções de mobilidade mais promissoras na Internet, a saber MIPv6, MSCTP e HIP, pode ser encontrada em [Ratola 2005]. Nesse e em outros estudos, o MSCTP é apontado como a solução mais promissora de mobilidade para a Internet.

Os serviços de mobilidade oferecidos pelo MSCTP podem ser combinados com protocolos da pilha TCP/IP, como o proposto por esse trabalho, para simplificar as implementações de mobilidade em nível de rede e em outros níveis lógicos da arquitetura Internet.

2.5. Arquitetura de comunicação SIP

A arquitetura SIP, destinada às comunicações multimídia em tempo real na Internet, é formada por inúmeros protocolos com objetivos específicos. Combinando as funcionalidades desses protocolos, essa arquitetura de comunicação é capaz de oferecer um serviço flexível e ao mesmo tempo robusto para uma série de aplicações multimídia de escopos distintos. Os principais protocolos constituintes dessa arquitetura de comunicação são o SIP, o SDP (*Session Description Protocol*) [Handley 1998] e o RTP, conforme mostrado na Figura 2.3.

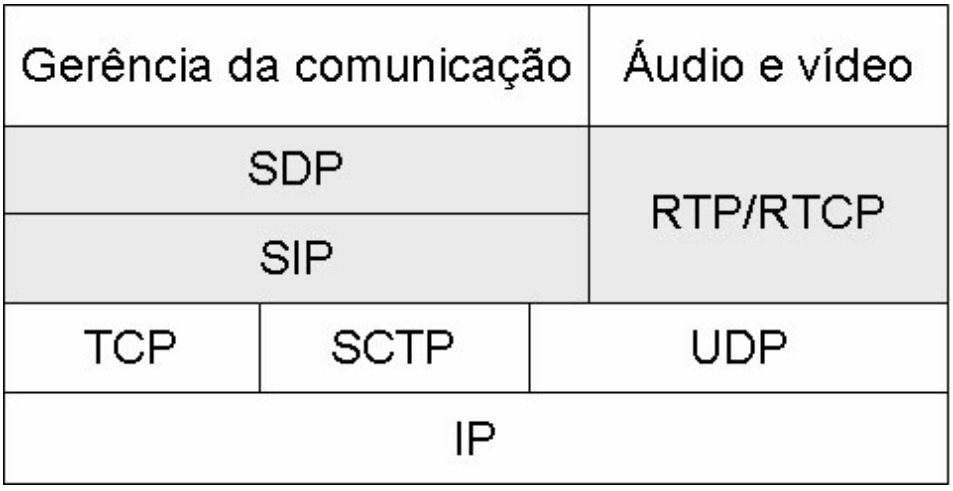


Figura 2.3. Arquitetura de comunicação SIP.

O protocolo SIP [Rosemberg 2002] é destinado à abertura, controle e encerramento de chamadas em comunicações multimídia. Além disso, o SIP possui também funcionalidades para localização de usuários, independente do terminal de comunicação sendo utilizado.

O SIP, por ser um protocolo de aplicação, deve utilizar serviços de algum protocolo de transporte: a utilização do TCP, SCTP ou UDP é possível, sendo a escolha feita de acordo com as necessidades da aplicação. Todas as mensagens SIP são codificadas em formatos textuais.

Uma arquitetura SIP típica é formada por terminais e por servidores *proxies*, *registrars* e de redirecionamento, além de agentes de notificação e *gateways*. Os

terminais SIP, formados por agentes clientes e servidores, são os produtores e consumidores do tráfego multimídia em tempo real, como voz e vídeo. Os SIP *proxies* encaminham mensagens entre terminais, realizando também serviços auxiliares como consulta a registros de localização de usuários. Esse elemento pode ser utilizado para oferecer mobilidade de terminal em nível de aplicação. Já os servidores *registrars* fazem registro dos endereços de rede dos dispositivos que a cada momento estão sendo efetivamente utilizados pelos usuários da arquitetura SIP. Esses servidores estão diretamente ligados aos procedimentos de localização utilizados nessa arquitetura.

Os servidores SIP de redirecionamento não encaminham mensagens, contudo respondem a requisições de usuários informando, por exemplo, onde localizar determinado recurso. Os agentes de notificação informam sobre determinado evento requisitado pelo usuário. Por fim, *gateways* compatibilizam comunicações entre arquiteturas diferentes, como H.323 e SIP. Para a arquitetura SIP padrão, o uso de qualquer desses elementos não é obrigatório, com exceção dos terminais.

2.5.1. RTP e RTCP

O protocolo RTP foi especificado para ser utilizado no transporte de dados sensíveis ao atraso na Internet. Sua utilização é necessária tendo em vista as características dos protocolos de transporte empregados nessa rede: o TCP e o SCTP são protocolos confirmados enquanto o UDP e o PR-SCTP, embora não confirmados, ou parcialmente confirmado, como esse último, não possuem qualquer informação de tempo. As aplicações de tempo real são prejudicadas por retransmissão de pacotes perdidos e por atrasos na transmissão, necessitando, portanto, de marcas de tempo para reconstruir as informações da mídia recebida.

O protocolo RTP, portanto, provê o transporte de dados fim a fim necessário a aplicações multimídia na Internet, possuindo informações de tempo e seqüenciamento necessárias a essas aplicações. Seu objetivo é fornecer um mecanismo para levar dados sensíveis ao atraso, como áudio e vídeo, de um ponto a outro da rede, garantindo a recuperação original da informação transmitida. As aplicações típicas utilizam o RTP

sobre o UDP, embora a utilização do RTP sobre o PR-SCTP seja possível. Em ambos os casos, mecanismos para assegurar a entrega de dados em ordem ou com atraso constante não são esperados do RTP.

O RTCP (*Real Time Control Protocol*) [Schulzrinne 1996] foi especificado para ser um protocolo de controle que auxiliasse o RTP na sua tarefa de transmissão de dados em tempo real. Seu objetivo primário é a disponibilização de *feedback* sobre QoS (Qualidade de Serviço) [Tanenbaum 2003] para que os participantes de uma conferência multimídia possam se adaptar dinamicamente a problemas na rede.

Os dados de mídia codificada são transmitidos em *sessões* RTP distintas, tendo cada uma dessas sessões um fluxo RTP e um RTCP. É a partir de informações contidas em relatórios RTCP que mídias de sessões distintas, como áudio e vídeo, podem ser associadas, operação essa necessária em conferências multimídia.

Uma alternativa para prover autenticidade, integridade e confidencialidade em fluxos de mídia em tempo real é utilizar o SRTP (*Secure Real Time Protocol*) e o SRTCP (*Secure Real Time Control Protocol*) [Baugher 2004]. Sendo perfis para serem utilizados opcionalmente com seus protocolos correspondentes, essas extensões podem ser facilmente implementadas: apenas informações adicionais são inseridas no campo de extensão do cabeçalho desses protocolos.

2.5.2. SDP

O protocolo SDP [Handley 1998] é utilizado para descrever sessões multimídia. Utilizando esse protocolo, aplicações participantes de comunicações com suporte multimídia podem trocar informações sobre suas capacidades de processamento de mídia, permitindo assim compatibilizar os *codecs* (*enCOder/DECoder*) [Hersent 2002] a serem empregados nas comunicações.

Algumas das informações presentes nas mensagens SDP são endereços IP, portas UDP, TCP ou SCTP utilizadas, tipos de mídias suportadas, título e assunto da

sessão multimídia, informações de contatos, entre outras. Todas essas informações são disponibilizadas textualmente seguindo a especificação desse protocolo.

2.6. Aspectos de Segurança no SCTP

No desenvolvimento do SCTP, problemas de segurança comuns ao protocolo de transporte TCP foram considerados. Como exemplo, um dos pontos fracos do TCP, sua alta sensibilidade a ataques do tipo negação de serviço [Cheswick 2005], foi combatida no SCTP. Esse problema ocorre porque o TCP não autentica as partes comunicantes durante o estabelecimento de uma conexão, alocando recursos para qualquer pedido recebido.

O SCTP utiliza um esquema de abertura de conexão com uma chave especial (*cookie*) que deve ser trocada entre as partes comunicantes. Os recursos são apenas alocados quando ocorre essa troca, o que inibe a existência de conexões "meio abertas", comuns ao TCP.

Há ainda outros problemas de segurança que vem sendo tratados no SCTP, como o seqüestro de associação, a autenticação dos usuários e a criptografia da comunicação. Para esses, além dos mecanismos nativos do próprio protocolo, alternativas foram desenvolvidas como extensões ao protocolo.

Uma primeira solução para melhorar alguns aspectos de segurança do SCTP foi criar uma extensão para utilizar o protocolo TLS sobre o protocolo SCTP [Jumgmaier 2002]. Com esse mecanismo, um serviço de entrega seqüencial de dados confiável é disponibilizado. Contudo, nessa especificação, mensagens entregues fora de ordem não podem ser protegidas, proibindo a utilização do protocolo PR-SCTP com a solução especificada.

Outra solução é a utilização da extensão que define o uso do SCTP sobre conexões IPSec [Bellovin 2003]. Nesse caso, as conexões IPSec, em nível de rede, garantem às camadas superiores serviços de integridade e confidencialidade. O grande problema dessa solução é o suporte ao *multihoming* SCTP, principalmente quando a

extensão MSCTP é empregada. Essa solução de segurança causa problemas potenciais nesse cenário de operação.

Outras soluções estão sendo desenvolvidas para oferecer mecanismos de segurança a outros problemas potenciais. A utilização do SCTP por um maior número de usuários impulsionará ainda mais esse desenvolvimento, uma vez que a demanda por soluções seguras com esse protocolo será maior.

Para a arquitetura considerada nesse trabalho, a solução empregada para oferecer autenticidade, integridade e confidencialidade opcional será o SRTP, que, quando utilizado, protegerá os dados de mídia da comunicação. As extensões que utilizam o TLS e o IPsec não serão utilizadas na arquitetura proposta uma vez que inibem o uso, respectivamente, do PR-SCTP e o MSCTP, ambos necessários à sua operação.

2.7. Voz sobre IP

Voz sobre IP (VoIP) [Hersent 2002] consiste na digitalização da voz, empregando codecs apropriados, para transmissão nos backbones Internet, incluindo as aplicações usuárias desse serviço e as técnicas de comunicações desenvolvidas para auxiliar esse processo. Como benefício à adoção desse serviço, pretende-se utilizar uma mesma infra-estrutura de comunicação (Internet) para transmissão de dados e voz.

O crescente aumento do número de usuários dos serviços VoIP, alicerçado pela promessa de economia no custo das ligações, impulsionou o desenvolvimento de tecnologias para ampliar a qualidade desse serviço em ambientes reais, não restritos apenas as LANs (*Local Area Network*) [Soares 1995] institucionais. A ampliação das comunicações VoIP para redes MAN (*Metropolitan Area Network*) e WAN (*Wide Area Network*) [Soares 1995], com novos serviços públicos de comunicação, como as redes celulares de terceira e quarta geração [Wireless 2005], tende a criar novos escopos operacionais onde o IP se tornará condutor básico para diversas aplicações de comunicação.

Nesse contexto, arquiteturas de rede, como a proposta no próximo capítulo, desempenham um papel crucial na evolução das soluções de comunicação móvel e multimídia sobre IP.

2.8. Procedimentos para o desenvolvimento de sistemas distribuídos

Para o desenvolvimento de sistemas distribuídos, uma série de procedimentos deve ser adotada a fim de evitar inconsistências operacionais indesejáveis. A complexidade inerente desses sistemas requer que técnicas de especificação sejam utilizadas para minimizar as chances de implementações incorretas.

A especificação de sistemas distribuídos exige, na maior parte dos casos, que técnicas de especificação formal sejam utilizadas. Isso evita que detalhes de operação sejam mal interpretados em implementações, o que pode ocorrer quando técnicas de especificação informais ou mesmo semi-formais são empregadas.

2.8.1. Especificação Formal

A construção de sistemas computacionais adequados aos diversos requisitos de implementação, e que atendam também necessidades cada vez mais freqüentes, como eficiência e robustez, requerem mecanismos de desenvolvimento mais rápidos e confiáveis. Para suprir essa necessidade, podem-se utilizar técnicas de especificação formal para o desenvolvimento de sistemas. Para sistemas distribuídos, onde os requisitos supracitados são ainda mais relevantes, a utilização de técnicas formais de especificação traz benefícios significativos.

Uma especificação formal utiliza conceitos bem definidos e base matemática sólida para modelar um sistema. Essa natureza garante características como corretude, completude, consistência, concisão e clareza, tão importantes para a especificação de

sistemas distribuídos, porém difíceis de conseguir a partir da utilização de técnicas de especificação informais ou mesmo semi-formais.

O uso de formalismos durante a especificação de sistemas distribuídos permite a eliminação de ambigüidades e inconsistências as quais apenas seriam detectadas mais tarde, durante as fases de implementação e testes do ciclo de desenvolvimento de um sistema. A ocorrência de erros operacionais em fases tardias do processo de desenvolvimento aumenta significativamente o custo total de desenvolvimento de um sistema, o que pode comprometer projetos de implementação de sistemas distribuídos.

Para especificar formalmente um sistema, uma linguagem de especificação formal deve ser utilizada. Em adição, uma especificação que utilize uma linguagem dessa natureza é consideravelmente beneficiada com o uso de ferramentas CAD (*Computer Aided Development*). As subseções 2.8.1.1 e 2.8.1.2 descrevem em detalhes esses dois tópicos, apresentando as soluções adotadas para a especificação descrita no capítulo quatro.

2.8.1.1. Linguagem SDL

A linguagem formal SDL foi criada pelo órgão internacional ITU [ITU 2005]. O desenvolvimento dessa linguagem teve início em 1972, quando seu objetivo era apenas especificar e descrever sistemas de telecomunicações. Atualmente, a SDL é utilizada para especificar sistemas complexos dirigidos a eventos, em tempo real e com atividades concorrentes. O padrão ITU que define a linguagem SDL é o Z.100. Nesse padrão, a versão da linguagem utilizada para a especificação da arquitetura de comunicação descrita nos capítulos três e quatro será a SDL 2000.

As características da SDL fazem com que essa linguagem seja muito utilizada para especificar sistemas de comunicação, como aqueles utilizados em redes de computadores. A operação dos protocolos, bem como a interação entre eles, é especificada satisfatoriamente pela SDL, o que garante o sucesso dessa linguagem formal nessa área.

A SDL é utilizada tanto para representar o comportamento do sistema, como sua estrutura. Uma das características fortes dessa linguagem é que, diferente da maior parte das linguagens de especificação formal, que apresentam apenas uma representação textual, geralmente difícil de utilizar, a SDL possui também uma representação gráfica. Essa representação gráfica do sistema é chamada de GR (*Graphical Representation*). A especificação GR-SDL é baseada em símbolos, o que, além de facilitar especificações, provê maior clareza quando a especificação é lida. Além da representação gráfica, existe também uma representação textual, chamada de PR (*Plain Representation*). Há ferramentas de auxílio à especificação em SDL, que transformam representações de um formato em outro. Assim, é comum que especificações sejam feitas apenas em um formato, sendo convertidas no outro formato possível quando necessário.

Para facilitar a tarefa de especificar a arquitetura, bem como para tornar a especificação mais claramente compreensível pelas pessoas que a leiam, o formato gráfico da linguagem SDL é aqui utilizado. Assim, o tópico seguinte aborda alguns conceitos relacionados a essa representação, que ajudarão à compreensão da especificação desenvolvida.

GR-SDL

Há três grupos de componentes utilizados numa especificação SDL, independente do formato de representação adotado: o estrutural, que define a estrutura estática do sistema sendo modelado, em termos das suas componentes; os sinais, que representam as mensagens trocadas entre os componentes; e os canais, que especificam entre quais componentes os sinais são trocados.

O grupo estrutural contém componentes hierarquicamente segmentados. A Tabela 4.1 apresenta os componentes estruturais, suas finalidades e suas representações gráficas. Os primeiros componentes da Tabela, de cima para baixo, possuem maior hierarquia que os componentes abaixo. Assim, Blocos estão contidos em Sistemas, Processos estão contidos em Blocos, e assim por diante.

Sistemas são unidades estruturais que representam um escopo consideravelmente complexo numa especificação. De certa forma, o que é um sistema depende de como é especificado o mundo sendo modelado. Um sistema pode ser, por exemplo, um protocolo, um conjunto de protocolos ou um conjunto de *hosts* numa rede. Na definição de um sistema, os blocos auxiliam na delimitação dos pontos de processamento. Assim, por exemplo, num sistema que representa um conjunto de protocolos, como o especificado nesse trabalho, cada protocolo pode ser representado por um bloco.

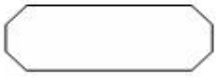
Componente	Descrição	Representação
Sistema	Representa um sistema específico	
Bloco	Reduz complexidade hierárquica	
Processo	Comportamento concorrente do bloco	
Procedimento	Funciona como uma subrotina	

Tabela 2.1. Componentes estruturais da linguagem SDL.

Processamentos concorrentes são representados num bloco pelos seus processos. O comportamento de cada processo é definido como uma máquina de estados finita. Esses elementos estão diretamente relacionados com a definição da dinâmica de um sistema. A Tabela 2.2 apresenta alguns dos elementos de especificação utilizados na definição de um processo.

Processos podem ser criados ou encerrados durante a execução de um sistema. Num processo, dois números entre parênteses e separados por vírgulas costumam ser indicados após o nome do processo. O primeiro número indica o número de instâncias de Processos iniciais, enquanto o segundo número diz a quantidade máxima de processos que podem estar em execução simultaneamente.

Os sinais são as informações trocadas entre os elementos de uma especificação. Os sinais podem ser enviados e recebidos dentro de um elemento estrutural, como num

sistema, ou podem ser enviados para fora ou para dentro de um elemento. Nesse último caso, o sinal trafega entre um elemento externo e um interno.

Ações em um Processo	Representação
Receber sinal externo	
Enviar sinal	
Comentário	
Representar um estado	
Iniciar o processamento	
Texto	
Estabelecer uma decisão	
Parar o processamento	
Definir uma tarefa	

Tabela 2.2. Alguns componentes utilizados na especificação de um processo.

Os sinais podem representar, numa implementação, mensagens de protocolos, mensagens entre objetos, primitivas de serviço, entre outras coisas. O que é feito com

um sinal, como, por exemplo, os parâmetros e opções que eles podem carregar, são detalhes de implementação não definidos pela especificação. A representação gráfica de um sinal é um texto escrito entre os símbolos [e].

A comunicação entre processos, blocos e sistemas é feita através do envio de sinais, que trafegam utilizando elementos chamados canais. Um canal indica um trajeto que o sinal irá seguir. Os canais podem ser unidirecionais ou bidirecionais. A representação gráfica de um canal é uma seta, que pode ser →, ← ou ↔. Os sinais devem estar associados a um ou mais canais.

2.8.1.2. Ferramenta CAD SanDriLa

Para o desenvolvimento da especificação em linguagem SDL, que será apresentada no capítulo quatro, uma ferramenta de suporte foi utilizada. A utilização de uma ferramenta é necessária não apenas para a construção da representação gráfica da arquitetura em SDL, mas também para a verificação da ausência de erros dessa especificação.

A partir da análise das ferramentas de especificação SDL disponíveis, optou-se por utilizar a ferramenta proprietária SanDriLa [Sandrila 2005]. Essa escolha é devida tanto em função da dificuldade de utilização, descontinuidade e falta de suporte de muitas ferramentas de código aberto, quanto em função de políticas de incentivo à pesquisa que algumas empresas produtoras de ferramentas proprietárias possuem. Através do contato direto com os fornecedores do SanDriLa, uma cópia completamente licenciada foi fornecida gratuitamente para a realização da especificação da arquitetura.

SanDriLa é uma ferramenta para a criação de diagramas SDL, MSC (*Message Senquence Chart*), UML 2.0 [UML 2005], entre outros. Para oferecer essas funcionalidades, contudo, uma outra ferramenta é necessária, o Microsoft Visio [Visio 2005]: o SanDriLa é construído como um *plugin* a ser adicionado a essa ferramenta, também proprietária. Também há formas de utilização gratuita do Visio para fins acadêmicos.

Para a especificação da arquitetura, foi utilizado o Microsoft Visio 2000. Assim, apenas sistemas operacionais Microsoft Windows 2000 ou posteriores puderam ser utilizados. A Figura 2.4 apresenta o ambiente de especificação do MS Visio com o SanDriLa instalado.

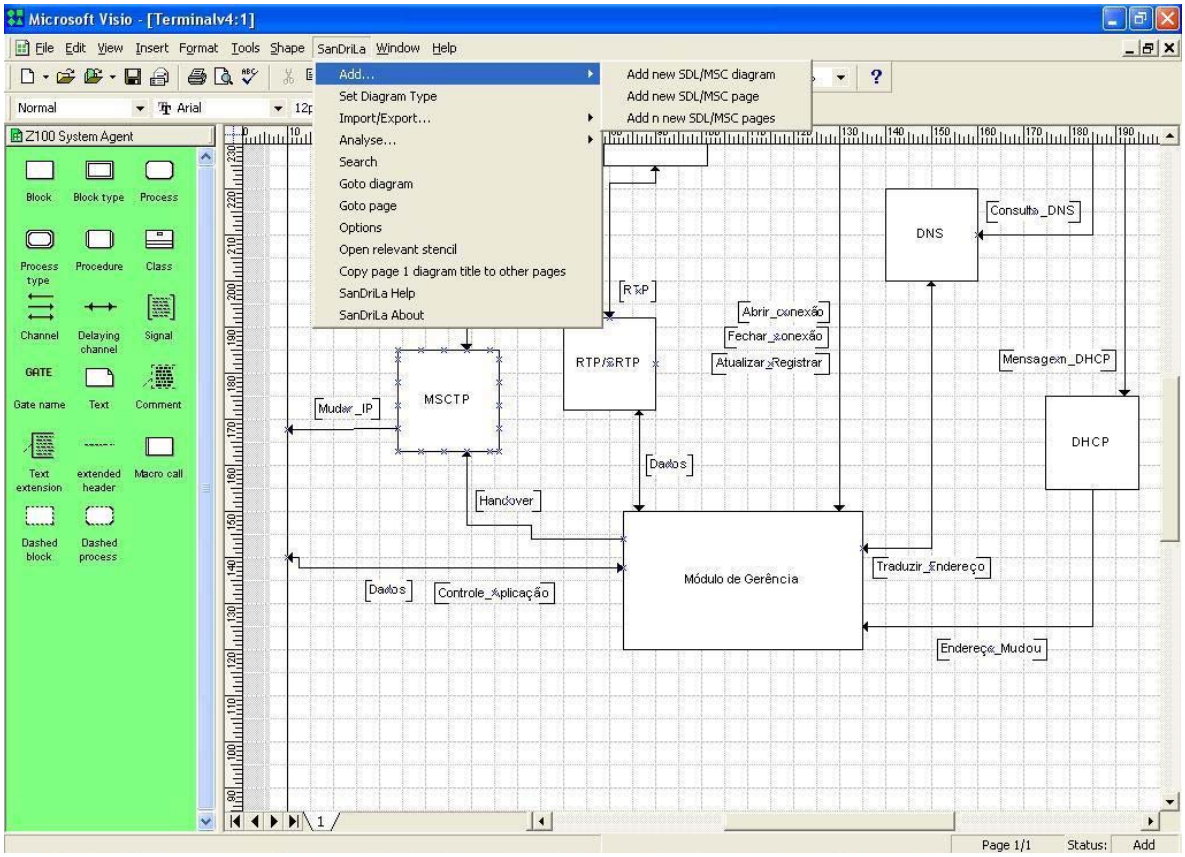


Figura 2.4. Ferramenta CAD SanDriLa.

Algumas funcionalidades que o SanDriLa oferece permitem integração com outras ferramentas de especificação SDL. O módulo de importação permite a abertura de especificações geradas em outras ferramentas. Além disso, fazendo exportação dos diagramas para formatos GR ou PR, é possível utilizar verificadores adicionais para realizar validações extras em especificações.

2.8.2. Verificação Formal

A verificação formal é o processo de validar determinada especificação que utilizou alguma técnica formal para sua definição. Esse processo de validação visa garantir que uma especificação formal está corretamente definida lexicamente, sintaticamente e semanticamente. A garantia de alguma ou de todas essas três características pode variar de acordo com os objetivos da verificação formal desejada.

A verificação léxica visa garantir que os elementos utilizados numa especificação são válidos para a linguagem de especificação formal adotada. Já a verificação sintática está ligada à garantia de que a interação entre os elementos da linguagem utilizados na especificação possui um sentido único e definido pela linguagem. Essa é a verificação estrutural da especificação.

As duas verificações supracitadas garantem que determinada linguagem de especificação formal foi utilizada de forma correta. Contudo, para provar que uma especificação define corretamente um sistema, deve-se empregar a verificação semântica.

A verificação semântica utiliza os formalismos matemáticos em que se baseiam as técnicas formais de especificação para avaliar se uma especificação descreve o sistema da forma esperada. Ela também garante que nenhuma inconsistência operacional oriunda de má especificação, como *deadlocks* e/ou *livelocks*, esteja presente. Diz-se então que a verificação semântica analisa o significado de uma especificação.

A forma e completude que verificações semânticas são realizadas dependem do ferramental de verificação utilizado, o que geralmente é disponibilizado pelas ferramentas CAD. Como os recursos de verificação disponibilizados por tais ferramentas podem ser inferiores ao esperado, uma alternativa para a verificação semântica de uma especificação é a simulação matemática do sistema sendo especificado.

3 *Especificação Funcional da Arquitetura de Comunicação*

Esse capítulo especifica funcionalmente uma arquitetura de comunicação alternativa para alguns cenários de comunicação multimídia na Internet. Essa especificação, de caráter informal, é complementada pela especificação formal descrita no capítulo quatro.

A arquitetura aqui especificada visa atender as aplicações da Internet com capacidades de comunicação de voz e com requisitos de mobilidade, embora a mobilidade não seja obrigatória. Além disso, essas aplicações não devem necessitar de comunicação multiponto, como em audioconferências. Apesar de aparentemente restritiva, essa arquitetura atende a uma grande demanda real de comunicação, embasada tanto pela melhoria dos backbones de rede quanto pela proliferação de dispositivos móveis.

As aplicações usuárias dessa arquitetura podem estar alocadas em diversos equipamentos, como *notebooks*, *palmtops* e telefones IP móveis, só para citar alguns exemplos. Assim, a arquitetura proposta deve englobar diversos ambientes de operação, uma vez que pode ser utilizada sobre enlaces WLAN (*Wireless LAN*), como em redes IEEE 802.11 [Gast 2005], WMAN (*Wireless MAN*), como em redes WiMax [WiMax 2005], e WWAN (*Wireless WAN*), como em redes GSM/GPRS e as tecnologias de telefonia de terceira e quarta geração [Wireless 2005].

Embora essa arquitetura seja destinada a atender simultaneamente requisitos de comunicação multimídia em tempo real e mobilidade na Internet, ela também pode atender aplicações de voz em redes cabeadas que não possuam requisitos de mobilidade, ou mesmo aplicações desse mesmo tipo em redes sem fio que não desejem utilizar serviços móveis. Assim, a mobilidade deve ser considerada uma característica opcional, não acarretando em prejuízos operacionais às aplicações não móveis usuárias da arquitetura: *host*s inicialmente não móveis podem utilizar os serviços de mobilidade da arquitetura a qualquer momento de uma comunicação, sem necessitar encerrar as

transmissões correntes, desde que possuam suporte das camadas física e enlace para isso.

Compondo serviços do protocolo SCTP e de algumas de suas extensões com o protocolo SIP, e incluindo ainda alguns elementos auxiliares presentes em algumas arquiteturas multimídia, como o RTP e o SDP, a arquitetura de comunicação especificada oferece serviços de sinalização de chamada, transferências de dados multimídia em tempo real e envio de mensagens de texto, além de suporte a mobilidade de terminal e gerenciamento de localização, às aplicações e cenários anteriormente citados. Aliada a essas características, recomenda-se utilizar o protocolo SRTP para oferecer serviços opcionais de integridade, autenticidade e confidencialidade aos dados da comunicação.

Adicionalmente, os serviços oferecidos pelos protocolos DHCP e DNS (*Domain Name System*) [Mockapetris 1987] são necessários à arquitetura em questão, porém podem ser oferecidos por outros protocolos com funcionalidades equivalentes, sem prejuízos a mesma. Contudo, estes protocolos são incluídos na especificação da arquitetura, estando estes presentes nos exemplos de operação e na especificação formal apresentados nesse trabalho.

Nenhuma política de qualidade de serviço é especificada pela arquitetura, embora a utilização de políticas de priorização de tráfego [Tanenbaum 2003] tenda a melhorar a qualidade final das comunicações de voz, quando backbones congestionados sejam intermediários dessas comunicações. A Figura 3.1 apresenta a pilha de protocolos utilizada na solução de comunicação considerada.

Na Figura 3.1, o Módulo de Controle corresponde a um código especificamente utilizado para coordenar o funcionamento ordenado dos protocolos que compõem o serviço de comunicação da arquitetura. Além disso, a interação desse serviço com as aplicações usuárias é responsabilidade também desse código. De fato, o serviço de comunicação da arquitetura terá um comportamento semelhante ao serviço oferecido por protocolos individuais, quando consideramos a sua utilização por aplicações. O capítulo quatro trata da especificação formal do Módulo de Controle.

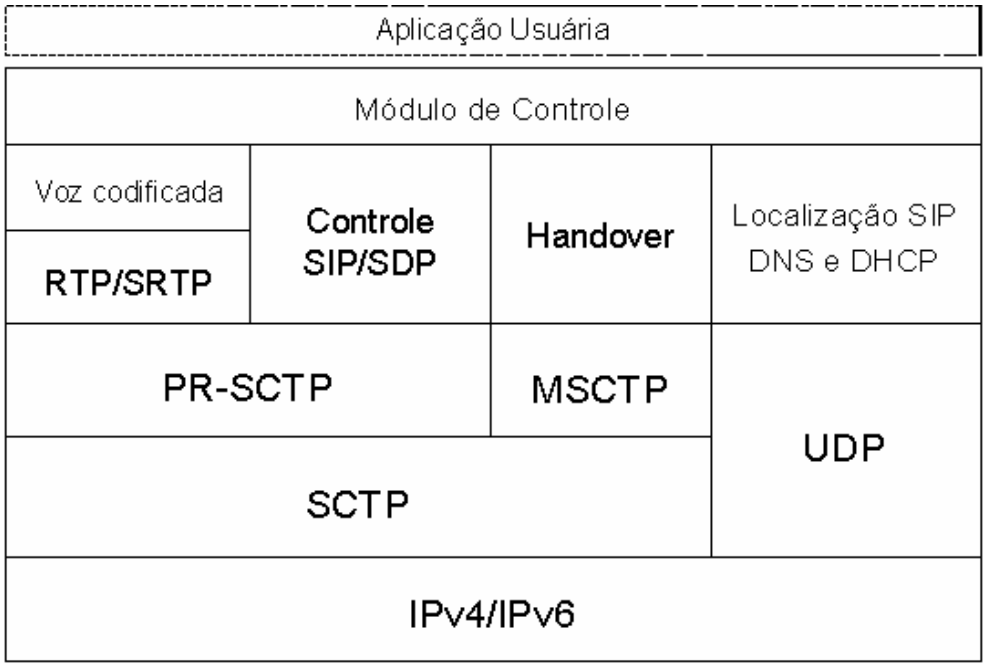


Figura 3.1. Pilha de protocolos da arquitetura.

3.1. Ambiente de comunicações considerado

O ambiente de comunicação considerado é composto por redes sem fio e *host*s móveis que se comunicam entre si e eventualmente com *host*s não móveis em redes cabeadas. Essas redes possuem um esquema de endereçamento que utiliza endereços IP publicamente roteáveis, IPv4 ou IPv6, sendo o protocolo de rede transparente à arquitetura proposta: os protocolos SIP, SDP e SCTP suportam ambas as versões do IP. Com o IPv6, contudo, a grande disponibilidade de endereços IP facilita a utilização de um maior número de *host* comunicantes, requisito esse necessário, por exemplo, na telefonia de terceira e quarta geração.

Além dos *host*s com suporte a voz e telefones IP, que podem ser considerados simplesmente como terminais de comunicação da arquitetura, o ambiente considerado possui servidores SIP *registrar*, que mantém o "rastro" dos terminais de voz, enquanto estes se movem entre as redes sem fio. Os servidores *registrar* também podem ser utilizados para realizar autenticação de usuários, na medida em que realiza a validação de pedidos de localização de terminais. A descrição desse serviço de autenticação, entretanto, está fora do escopo desse trabalho.

A Figura 3.2 apresenta os elementos do ambiente de comunicação considerado, de forma simplificada. Nessa figura, a nuvem WWAN compreende não apenas as tecnologias celulares 2.5G, 3G, 3.5 G e 4G, mas também as tecnologias WMANs, como o WiMax. Além disso, os elementos auxiliares à arquitetura, como os servidores DNS e DHCP, podem ser considerados como implantados nas mesmas máquinas onde se encontram os servidores *registrar*, a título de simplificação.

A nuvem PSTN apresentada na Figura 3.2 compreende o sistema público de telefonia. A partir de *gateways* de conversão entre os padrões utilizados na Internet e na rede telefônica, é possível realizar chamadas entre terminais dessas duas arquiteturas de comunicação. Apesar de completamente possível, a arquitetura de comunicação proposta não oferece qualquer serviço auxiliar a essas comunicações, devendo-se contar com mecanismos adicionais para interação entre essas infra-estruturas.

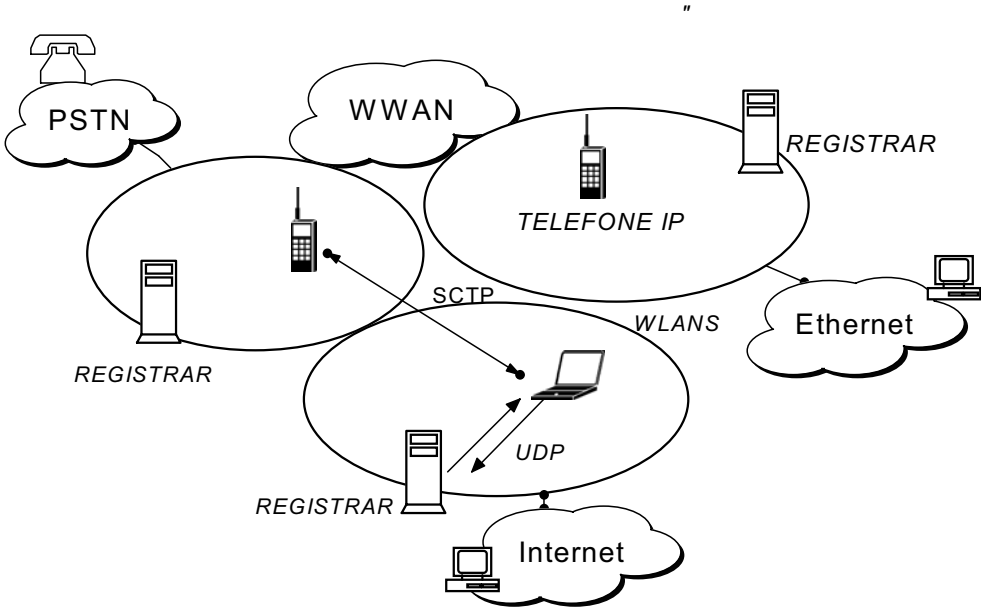


Figura 3.2. Ambiente de comunicação considerado.

Ainda na Figura 3.2, duas classes de comunicações são ressaltadas: as que operam sobre SCTP e as que operam sobre UDP. As que utilizam os serviços desse primeiro protocolo de transporte, como já mencionadas, são as comunicações das sessões multimídia SIP entre os terminais usuários da arquitetura proposta, compreendendo aí tanto os dados de controle como as informações de voz e, opcionalmente, texto da comunicação. Por outro lado, as comunicações que operam

sobre UDP concernem aos registros e consultas a servidores *registrar*, bem como as trocas de mensagens com servidores DNS e DHCP.

Nessa primeira especificação da arquitetura de comunicação, não são utilizados elementos para realizar contabilidade de recursos. Essa restrição visa garantir conexões SCTP fim-a-fim, sem a utilização de elementos intermediários de comunicação, como *proxies*, de forma a explicitar o mecanismo de comunicação idealizado. Entretanto, é possível, desde já, indicar uma solução que inclua outras funcionalidades desejadas: uma forma de garantir a contabilidade de recursos seria através do uso de refletores, que encaminhem as comunicações de controle e de dados isócronos (sensíveis ao tempo) de e para os usuários da arquitetura; estabelece-se assim pelo menos duas associações SCTP, cada uma entre um terminal comunicante e um refletor, que pode ser, por exemplo, um *proxy* SIP em uma rede *wireless* ou um elemento de controle numa rede GSM/GPRS. Apesar de não especificada no ambiente proposto, essa solução manteria todos os requisitos de operação já mencionados, garantindo poucos impactos na rede.

Outros mecanismos de contabilização de recursos são possíveis, porém não serão especificados nessa arquitetura.

Da mesma forma, os elementos auxiliares da arquitetura SIP não foram incluídos, a exceção dos *registrars*, uma vez que os serviços oferecidos por esses elementos não são empregados nas comunicações envolvendo a arquitetura aqui especificada. A utilização de endereços IP publicamente roteáveis, IPv4 ou, preferencialmente, IPv6, não requer o emprego de mecanismos de NAT (*Network Address Translator*) [Comer 1998]. Aliada a essa característica, a necessidade de manutenção de associações SCTP fim-a-fim, diretamente entre os *host*s usuários da arquitetura de comunicação, inibe a adoção de SIP *proxies*: nenhum mecanismo para oferecer compatibilidade com esses elementos, como opções especiais em mensagens DHCP para uso com esses servidores [Schulzrinne 2002], é também necessário. A adoção de mecanismos de mobilidade de terminal em nível de transporte também desencoraja a utilização de servidores SIP de redirecionamento. Por fim, agentes SIP de notificação e *gateways* de tradução não serão adotados na especificação inicial da arquitetura descrita nesse trabalho.

Esquemas híbridos com a utilização do protocolo UDP para transporte de dados isócronos, ou mesmo a utilização do MIP em ambientes onde associações SCTP fim-a-fim não possam ser estabelecidas, como em comunicações onde um dos *host*s está em uma rede privada administrada por NAT, estão fora do escopo da arquitetura proposta, ficando a análise desses casos restrita a especificações futuras.

3.2. Mobilidade em nível de transporte – mecanismos de *handover*

A arquitetura descrita nesse trabalho especifica a utilização dos mecanismos de atualização dinâmica de endereços IP do protocolo SCTP para implementar operações de *handover* em nível de transporte. Esse procedimento depende de três pontos básicos: primeiro, o *host* deve descobrir que mudou de rede lógica, para assim poder iniciar os procedimentos de atualização de endereço. Em seguida, caso descubra que a rede lógica em que o *host* encontra-se é diferente da rede anterior a seu deslocamento, um procedimento para aquisição de um novo endereço IP deve ser adotado. Por fim, os mecanismos de configuração dinâmica de endereços do SCTP devem ser empregados. As seções seguintes tratam desses aspectos operacionais da arquitetura.

3.2.1. Detecção de mudança de rede lógica

Considerando os enlaces sem fio no ambiente de comunicação idealizado, a troca de informações entre os elementos constituintes desses enlaces, a nível físico, ocorre por ondas eletromagnéticas. Dessa forma, uma estação móvel, a partir do nível de potência dos sinais recebidos de pontos concentradores, como, por exemplo, *Access Point* em redes 802.11 [Gast 2005], consegue identificar quando deve ocorrer mudança entre redes físicas distintas, sendo esse procedimento gerenciado pelo nível de enlace. Deve-se confiar nesse nível lógico para que a mudança de rede priorize a estabilidade da rede escolhida, caso haja sinais recebidos de diversos pontos concentradores.

A mudança de rede física pode acarretar em mudança de rede lógica IP. Embora a configuração de redes lógicas distintas em diferentes células *wireless* seja a configuração predominante, uma mesma rede lógica IP pode estar configurada em diferentes células de redes sem fio. Esse procedimento é comumente adotado para aumentar a área de abrangência de determinada rede lógica, sobretudo quando o raio de alcance das células *wireless* é limitado.

Quando não há mudança de rede lógica, não é necessário que haja alteração de endereço IP. O roteamento feito pelo protocolo IP em *host*s utiliza uma máscara de subrede para decidir se pacotes devem ser roteados por um caminho pré-estabelecido (geralmente um *gateway* IP) ou se devem ser encaminhados na própria rede. Como esse protocolo encaminha diretamente pacotes que estejam em uma mesma rede lógica, não é necessária que haja mudança de endereço IP e de máscara de subrede caso a rede lógica não seja alterada, embora a mudança de endereço seja possível. Para a primeira versão dessa arquitetura, o procedimento adotado é manter o mesmo endereço IP nesse tipo de cenário.

Como não há mudança de endereço de rede nesse caso, em nível de transporte SCTP não haverá também necessidade de utilizar mecanismos para atualizar o endereço IP do *host* móvel participante da associação corrente.

Com a mudança de rede lógica, contudo, o endereço IP do *host* móvel deve ser alterado; em algumas situações, a mudança de endereço IP pode implicar também na alteração da máscara de subrede. É necessário que essa mudança de rede lógica seja identificada de alguma maneira pela camada de rede ou por um nível superior, já que, em nível de enlace, não é utilizada a abstração de endereço IP. Existem várias maneiras de se notificar a mudança de rede lógica IP: em nível de rede, podem ser empregadas mensagens de anúncio de agentes da especificação MIP, enquanto em nível de aplicação, pode-se utilizar o protocolo DHCP para informar essa mudança. Nessa última solução, utilizada na especificação inicial da arquitetura, a identificação de mudança de rede lógica ocorre a partir do endereço IP e da máscara de subrede presente em mensagens desse protocolo. Utilizando esses dados, é possível calcular o endereço de rede (*netid*) para identificar se houve ou não mudança de rede lógica: um valor de *netid* diferente do atual indica que houve essa mudança. A partir daí, com a mudança de rede

lógica, pode-se iniciar o procedimento de obtenção de um endereço IP pertencente à rede visitada.

3.2.2 Atribuição de endereço IP

Quando o *host* móvel migra para uma rede lógica diferente da rede de origem, um outro endereço IP deve ser a ele atribuído, bem como uma máscara de subrede (que pode ser a mesma) e endereços do default gateway e servidor DNS. Uma forma de realizar esse procedimento é através do protocolo DHCP.

Nessa alternativa, cada servidor DHCP mantém uma série de endereços a serem atribuídos diretamente aos *host*s visitantes de sua rede lógica de influência, sendo esses endereços utilizados durante toda a permanência do *host* na rede visitada. Na arquitetura especificada, nenhum mecanismo de *timeout* será empregado nos endereços atribuídos pelo DHCP.

3.2.3 Atualização dinâmica de endereço na associação

Quando uma estação móvel migra para uma rede lógica diferente da original, com a subseqüente obtenção de um endereço IP (e demais configurações), o processo de atualização dinâmica de endereço tem início em nível de transporte. Para tanto, mensagens SCTP de atualização de endereço são trocadas entre os *host*s participantes da associação. Essas mensagens são ASCONF (*Address Configuration*) e ASCONF-ACK (*Address Configuration ACK*) [Costa 2005a], que podem ser trocadas a qualquer momento de uma associação SCTP.

Após a atribuição do novo endereço, uma mensagem ASCONF é enviada para informar ao par da comunicação o novo endereço IP a ser utilizado, substituindo o endereço anterior. Com uma mensagem ASCONF-ACK, a requisição de inserção do novo endereço na associação é confirmada. Esse procedimento de reendereçamento é

mais rápido do que os presentes nas soluções MIPv6 e HIP, sendo a melhor das três soluções para telefonia celular de terceira e quarta geração [Ratola 2005]: os constantes *handovers* dos usuários dessa rede necessitam de mecanismos rápidos de reendereçamento. Embora os mecanismos de *handover* adotados pelo MSCTP possam causar eventuais atrasos e perdas de pacotes na comunicação, o impacto final sentido pelo usuário é pequeno, devido, sobretudo, às características da mídia de voz, que é menos sensível a ruídos e perdas de informação que outras mídias como vídeo e música.

Por fim, deve-se notar que em nível de rede não é necessário qualquer adaptação adicional ao protocolo IP, considerando a solução proposta.

Apenas como comentário final, o MSCTP opera eficientemente quando ocorre *handover* de apenas um *host* móvel, participante de uma associação, em determinado instante de tempo. Contudo, caso ambos os *host*s móveis da associação realizem operações de *handover* simultaneamente, a comunicação, bem como a associação SCTP, pode ser encerrada. Na solução atual, especificada nesse trabalho, o procedimento adotado é tentar uma nova chamada automática ao endereço atual do *host* destino. Caso os dois terminais tentem iniciar a chamada simultaneamente, apenas uma delas deve ser estabelecida.

Embora pareça ineficiente a solução inicialmente adotada para tratar *handovers* simultâneos, a ocorrência desses eventos será comum apenas em algumas comunicações telefônicas de terceira e quarta geração, onde a taxa provável de operações de *handover* é alta em comparação a outros cenários de comunicação. Uma solução alternativa, que utiliza redundância para tratar o problema de *handovers* simultâneos com o MSCTP, pode ser encontrada em [Dreibholz 2003].

3.3 Mobilidade em nível de aplicação – gerenciamento de localização

A especificação MSCTP sugere a utilização de *home agents* [Perkins 2002] para localização de usuários. Contudo, em vez de utilizar serviços de localização em nível de rede, como aqueles especificados em MIP, podem-se empregar outros mecanismos de localização, como servidores DNS dinâmicos [Vixie 1997], realizando assim

localização de usuários em nível de aplicação. Usando um enfoque semelhante, pode-se ainda utilizar mecanismos implementados por elementos de determinadas arquiteturas de comunicação multimídia, que operam também em nível de aplicação. Essa última solução é adotada na arquitetura especificada nesse trabalho, onde o gerenciamento de localização de usuários é realizado pelo protocolo SIP e por certos elementos de sua arquitetura. Como essa arquitetura é voltada a aplicações de voz sobre backbones Internet, a utilização do SIP para localização de usuários não causa *overhead* adicional considerável, uma vez que esse protocolo também poderá estar responsável pelas sinalizações necessárias às comunicações de voz.

Sendo os mecanismos de *handover* executados em nível de transporte pelo protocolo MSCTP, as funções de mobilidade da arquitetura aqui descrita estão localizadas nas camadas lógicas de transporte e de aplicação, que, por serem fim-a-fim, garantem flexibilidade ao serviço proposto. O benefício final esperado pela utilização de uma solução fim-a-fim, que não depende da infra-estrutura de comunicação, já é um fator promissor à adoção dessa arquitetura.

3.3.1 Localização de usuários

As comunicações entre usuários em um ambiente móvel, como já mencionado, podem ser classificadas em dois tipos: a) as iniciadas a partir de nós móveis para nós não móveis, em redes cabeadas ou sem fio, e b) as iniciadas por nós, móveis ou não, para nós móveis. As comunicações explicitadas em (a) não necessitam de qualquer mecanismo de localização de usuário além do convencional, bastando apenas o emprego do protocolo MSCTP para prover serviços de *handover*. Contudo, as comunicações mencionadas em (b) ainda necessitam de um mecanismo de localização de usuários especial, como aquele oferecido pelos *home agents* da especificação MIP, por exemplo.

O protocolo SIP transparentemente suporta “mobilidade pessoal”, que é a possibilidade de um usuário poder utilizar terminais diferentes, porém sendo identificado por apenas um endereço lógico “globalmente” único. Com isso, o usuário pode ser localizado independentemente do ponto de acesso (*host*) à rede utilizado por

ele em determinado momento. Esse esquema de endereçamento é baseado em endereços de e-mail, que, no escopo da Internet, possui significado global.

Para suportar a mobilidade de terminal, como aquela fornecida pela arquitetura MIP e pelo protocolo MSCTP, o SIP emprega servidores *proxies* e de redirecionamento, bem como mecanismos de atualização de sessões multimídia por reenvio de mensagens de sinalização. Os procedimentos de *handover* implementados pelo SIP operam em nível de aplicação e não são empregados na arquitetura especificada nesse trabalho, uma vez que essa função é esperada do protocolo MSCTP.

Na arquitetura considerada, portanto, o SIP é utilizado para realizar sinalização de chamadas, necessária às comunicações multimídia na Internet, e gerenciamento de localização, esse último executado pelo SIP através de servidores *registrars*. Os mesmos mecanismos auxiliares de localização empregados pelo SIP também podem ser utilizados na arquitetura proposta, como o uso de servidores DNS para auxiliar na localização de *registrars* e, opcionalmente, de usuários, embora as informações dos usuários devam estar obrigatoriamente armazenadas pelo menos nos *registrars* para atender à especificação da arquitetura.

3.3.2 Servidores SIP *registrar*

Para manter o "rastro" dos terminais móveis, enquanto estes trafegam pelas redes que implementem o serviço proposto, esses devem registrar-se junto ao seu *home registrar* correspondente, que é, a princípio, uma estação não móvel. Esse procedimento é conhecido como "SIP registration". Cada usuário dos terminais da arquitetura, móveis ou não, deve possuir um endereço SIP globalmente único, tendo esse endereço mesmo domínio (escopo) que o endereço do seu *home registrar* associado, ou mesmo um domínio que seja associado ao *home registrar* através de algum sistema de nomes de domínio, como o DNS. Dessa forma, mantêm-se um escopo de endereçamento voltado aos usuários, onde a localização ocorrerá independente do *host* utilizado no momento: os *host*s são meros pontos de acesso ao serviço de comunicação disponibilizado.

Como já mencionado, na arquitetura considerada, cada endereço SIP está associado a um *home registrar*, e cada usuário possui pelo menos um endereço desse tipo. Através de mensagens REGISTER (transmitidas sobre UDP), os usuários podem criar, consultar ou eliminar informações de localização nos seus servidores (*home*) *registrar*. Os registros nesses servidores fazem uma associação entre endereços SIP e endereços IP, tendo essa associação um "tempo de vida" determinado e configurável. É de se notar que a inclusão de mais de um endereço IP em um mesmo registro é possível na especificação do protocolo SIP, porém não será adotada inicialmente pela arquitetura.

A cada mudança de endereço IP de um *host* móvel no ambiente considerado, deve ocorrer uma atualização de seu registro no *home registrar*, onde será informado o endereço atual desse terminal. Nesse procedimento, o endereço previamente associado ao terminal, no registro, será substituído pelo seu novo endereço. Operações de registro por terceiros não são possíveis nessas atualizações, por questões de segurança, embora sejam possíveis na operação padrão do protocolo SIP.

Os endereços SIP estão escritos na forma "*sip:usuario@dominio*". A parte *dominio* do endereço SIP é utilizada para descobrir o *registrar* que mantém o registro do endereço IP, sendo essa descoberta feita empregando-se algum serviço auxiliar de localização, como consulta a servidores DNS. Essa consulta ao serviço de domínio é realizada diretamente pelos terminais SIP da arquitetura: para aumentar a eficiência em consultas seqüenciais a servidores *registrar*, os terminais devem manter em cache o resultado das últimas consultas realizadas, por um período dependente dos recursos disponíveis de cada terminal e por alguma política de atualização eficiente. Em seguida, empregando a parte *usuario* do endereço SIP e de posse da localização do *registrar* a ser contatado, os terminais da arquitetura podem realizar uma consulta ou atualização de endereço nesse servidor. Como todo usuário está registrado em seu (*home*) *registrar*, uma consulta a determinado servidor sempre encontrará um registro referente a um usuário a ele associado, mesmo que não haja qualquer endereço IP incluído no registro (registro expirado, por exemplo).

O encerramento de uma comunicação corrente de forma abrupta (ABORT SCTP), ou mesmo o encerramento imediato da associação SCTP sem notificação por mensagem, sugere que os terminais participantes da comunicação encerrada tentem

novas consultas a *registrars* e novas chamadas aos terminais envolvidos na associação recém encerrada, de acordo com algum procedimento estabelecido pelos usuários. Caso mensagens de redirecionamento ou de erro sejam recebidas de servidores *registrar*, deve-se adotar o comportamento padrão do protocolo SIP.

É uma exigência para o uso da solução proposta, a inclusão obrigatória de um registro de recurso nos servidores DNS dos domínios envolvidos na mobilidade dos terminais, descrevendo o *home registrar*, a exemplo de: "IN <endereço IPv4> A registrar" ou "IN <endereço IPv6> AAAA registrar". Assim, para localizar o servidor *registrar* desejado, a consulta deve ser feita ao recurso *registrar.dominio* do servidor DNS correspondente ao domínio buscado.

Por fim, para melhorar a eficiência das atualizações de localização e consultas feitas nos *registrars* pelos *host*s móveis, é recomendada que a atribuição dos endereços SIP aos usuários da arquitetura leve em conta as prováveis redes de maior freqüência de acesso desses usuários, a fim de diminuir o tempo de propagação das mensagens de e para esses servidores.

3.3.3. Uma versão simplificada do SIP para a arquitetura

Os procedimentos de abertura, modificação e encerramento de sessões multimídia, executados pelo protocolo SIP, são empregados na arquitetura para o controle das comunicações de voz e, opcionalmente, de texto. Contudo, um aspecto importante dessa utilização deve ser mencionado. Quando um *host* realiza uma operação de *handover*, efetuada em nível de transporte pelo MSCTP, não deve ser transmitida qualquer mensagem SIP INVITE ou UPDATE ao par da comunicação correspondente, por não ser necessário controle de *handover* por parte do SIP. Dessa forma, simplificam-se as implementações do SIP com uma operação mais leve do protocolo.

Entretanto, caso a rede visitada por um *host* móvel não possua recursos suficientes para atender aos requisitos do codec de voz empregado na comunicação corrente, o *host* móvel visitante poderá informar essa condição a outra ponta da comunicação reenviando uma mensagem INVITE indicando um codec alternativo para

consumir menos recursos da nova rede. Essa indicação é feita através do protocolo SDP, que envia tais informações em mensagens do tipo previamente mencionado.

Nesse procedimento de atualização com mensagens SIP, contudo, não é necessário que os *host*s comunicantes informem seus endereços IP uns aos outros, uma vez que essa notificação de mudança de endereço já ocorre em nível de transporte na arquitetura de comunicação especificada.

3.4 Transporte de dados isócronos

Fazendo uso do serviço parcialmente confiável fornecido pelo PR-SCTP, bem como do serviço de transmissão confiável do SCTP, o envio de dados isócronos e de informações de controle, entre *host*s de determinado padrão de comunicação multimídia, pode ser realizado numa mesma associação. Os dados referentes à voz digitalizada podem ser transmitidos de forma não confiável e não ordenada, enquanto as informações de controle seguem num fluxo ordenado e confiável. Dessa forma, o overhead da comunicação é reduzido, uma vez que o número de datagramas IP, o número de portas de transporte necessárias e o processamento na máquina destino são diminuídos. Essa é mais uma vantagem na adoção da solução idealizada.

No contexto da arquitetura, os dados isócronos são transmitidos de forma não confiável e não ordenada pelo protocolo PR-SCTP, utilizando para tanto o encapsulamento do protocolo RTP como intermediário. A transmissão dos dados de tempo real não necessita ser ordenada ao nível de transporte SCTP, uma vez que o RTP realiza ordenação utilizando o "número de seqüência" contido em seu cabeçalho. Indica-se o RTP para transmissão de dados isócronos, porque ele carrega informações necessárias a mídias com características de tempo real, que não estão presentes nos protocolos de rede e transporte da pilha TCP/IP: marcas de tempo no RTP permitem a sincronização da reprodução da mídia decodificada no usuário destino e o número de seqüência capacita à realização de reordenação de pacotes, caso esse serviço não esteja disponível, como no UDP. Por fim, na operação da arquitetura, a sinalização SIP segue em dois fluxos (*simplex*) SCTP confiáveis e ordenados, um para cada direção da comunicação.

Embora o RTP tenha sido adotado como protocolo de transporte de dados isócronos na arquitetura de comunicação considerada, a ser operado utilizando os serviços do PR-SCTP, o RTCP não será inicialmente utilizado. Tendo em vista o ambiente de comunicação adotado, os relatórios RTCP com função de reportar dados de QoS não são necessários a princípio, por simplificação da arquitetura. Também não é necessário que o RTCP carregue informações gerais de determinado participante de uma sessão RTP, como telefone e endereço de e-mail, uma vez que os protocolos SIP e SDP podem se encarregar dessa tarefa. E, finalmente, o CNAME, presente em pacotes SDES RTCP [Schulzrinne 1996], não é necessário, tendo em vista que apenas as mídias de voz e de texto são utilizadas na arquitetura idealizada, não sendo, portanto, preciso qualquer mecanismo de associação entre essas mídias, como ocorre com as mídias de vídeo e voz em videoconferências.

Deve-se notar que as aplicações multimídia que sigam determinado padrão, como o H.323 ou o SIP, não podem utilizar transmissão multicast com o PR-SCTP, uma vez que esse protocolo é "orientado a conexão". Caso sejam necessárias conferências multiponto, algum componente concentrador deve ser utilizado, como o MCU (*Multipoint Control Unit*) do padrão H.323 [Hersent 2002], ou então deve-se reverter ao uso do protocolo UDP para transmissão dos dados. Como já mencionado, a arquitetura de comunicação multimídia aqui especificada se enquadra em aplicações VoIP que não possuam requisitos de conferências multiponto, sendo todas as comunicações realizadas em modo unicast.

Como as comunicações que trafegam por enlaces sem fio tendem a possuir uma maior taxa de erro, e esse será o cenário de comunicação mais comum da arquitetura, deve-se priorizar a adoção de codecs com maior resistência a perdas de pacote, a exemplo do iLBC [Andersen 2004].

3.5 Mensagens de texto

A arquitetura de comunicação oferece o serviço opcional de transmissão de mensagens de texto entre seus terminais. Esse serviço é utilizado para envio de

mensagens instantâneas aos usuários da arquitetura, podendo ser utilizado também como um serviço de chat. A transmissão dos dados de texto é feita utilizando *payloads* redundantes RTP [Hellstrom 2000], podendo essa transmissão usufruir de serviços de criptografia e autenticação opcionais do SRTP.

Para transmissão de mensagens instantâneas, deve-se iniciar uma comunicação com um usuário da arquitetura, enviar a mensagem requerida e encerrar a associação estabelecida. Em versões posteriores da arquitetura, elementos concentradores poderão ser utilizados para captação e envio posterior de mensagens, garantindo a entrega mesmo com a ausência do usuário destino da mensagem em determinado momento. De qualquer maneira, a intenção de enviar mídias de texto na associação SCTP deve ser informada pelo protocolo SDP, enviado em mensagens INVITE, para poder preparar a aplicação destino à recepção das informações textuais.

O tipo de formatação do texto a ser utilizado na comunicação exposta acima está fora do escopo da arquitetura proposta, uma vez que pode variar de acordo com a implementação.

3.6 Aspectos de segurança da arquitetura

A arquitetura proposta utiliza um serviço de segurança opcional para o transporte de dados isócronos e mídias de texto, disponibilizado pelo SRTP. Contudo, esse é apenas um serviço inicial, uma vez que apenas os dados carregados por esse protocolo seriam protegidos. As mensagens de controle SIP e, em nível de transporte, as mensagens de controle SCTP, não possuiriam inicialmente qualquer mecanismo de segurança, assim como as mensagens SIP carregadas sobre UDP para atualizações de registros em servidores *registrar* estariam também desprotegidas.

Um serviço de segurança mais abrangente, englobando todas as mensagens envolvidas nas comunicações da arquitetura, deve ser especificado em versões futuras.

3.7 Seqüência de comunicação típica entre terminais

Inicialmente, para efetuar uma chamada a um usuário qualquer, é necessário que o usuário interessado conheça o endereço de rede atual do terminal do usuário a ser chamado. Se esse endereço for conhecido pelo *host* chamador, devido algum mecanismo de cache, por exemplo, a fase de localização pode ser ignorada ou postergada e um pedido de abertura de comunicação pode ser enviado. Contudo, caso o endereço contido em cache não seja mais válido (expirado) ou mesmo não esteja presente, deve-se realizar o procedimento de localização especificado na arquitetura.

Para descobrir o endereço de rede atual do usuário que participará da comunicação, deve-se conhecer de antemão o seu endereço SIP. A partir desse endereço é possível descobrir a localização do *home registrar* do usuário a ser chamado, uma vez que a parte *dominio* desse endereço permite o descobrindo do endereço IP desse servidor (utilizando consultas a servidores DNS, por exemplo). Em seguida, com a parte *usuario* do endereço SIP, é possível obter o endereço IP atual do terminal usado pelo usuário que será chamado, realizando uma busca ao registro correspondente presente no *registrar* consultado. Essa consulta retornará um endereço de rede que é independente do tipo de terminal utilizado pelo usuário sendo chamado.

A troca de mensagens, executada sobre o protocolo UDP, é apresentada na Figura 3.3. As consultas feitas a servidores DNS utilizam mensagens do protocolo DNS, enquanto as comunicações entre cliente e servidor *registrar* utilizam mensagens SIP específicas: a mensagem SIP REGISTER, utilizada para operações de consulta, não informa qualquer endereço IP, mas apenas a parte *usuario* do endereço SIP; a mensagem 200 OK contém o endereço de rede consultado.

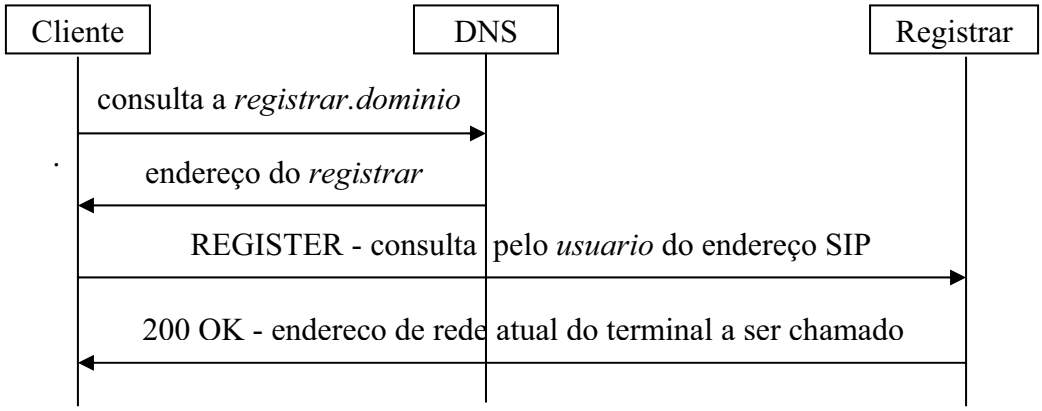


Figura 3.3. Procedimento para localização de usuário.

Com o endereço de rede do terminal destino, o terminal chamador (agente cliente SIP) inicia o estabelecimento de uma associação SCTP, com um único fluxo *simplex* em cada direção. O insucesso no estabelecimento dessa associação pode indicar, entre outras coisas, a indisponibilidade de comunicação do usuário destino naquele momento. A Figura 3.4 apresenta a troca de mensagens para o estabelecimento da associação SCTP após o sucesso na localização do usuário. Essa troca de mensagem corresponde ao mecanismo de *four-way handshake* necessário à abertura de qualquer associação SCTP.

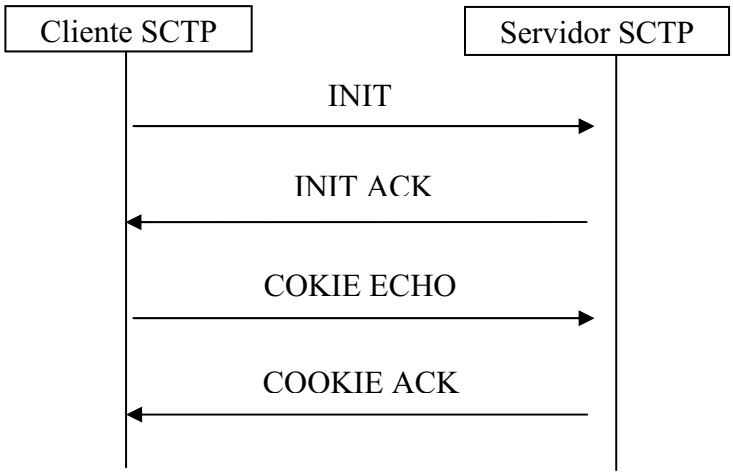


Figura 3.4. Estabelecimento de associação SCTP.

No par de fluxos *simplex* pertencente à associação recém estabelecida ocorre a troca de mensagens SIP, iniciada pelo envio de uma mensagem INVITE ao terminal chamado (agente servidor). Em seguida, uma mensagem "180 Ringing" pode ser opcionalmente enviada para indicar que o terminal chamado está "tocando". Logo após, para prosseguir com a fase de abertura da comunicação em nível de aplicação, o terminal chamado envia a mensagem SIP 200 OK, indicando que a chamada foi aceita. Por fim, completando o processo de *handshake* SIP, de três vias, uma mensagem ACK é enviada no mesmo sentido de envio do INVITE. A Figura 3.5 apresenta a troca de mensagens desse procedimento.

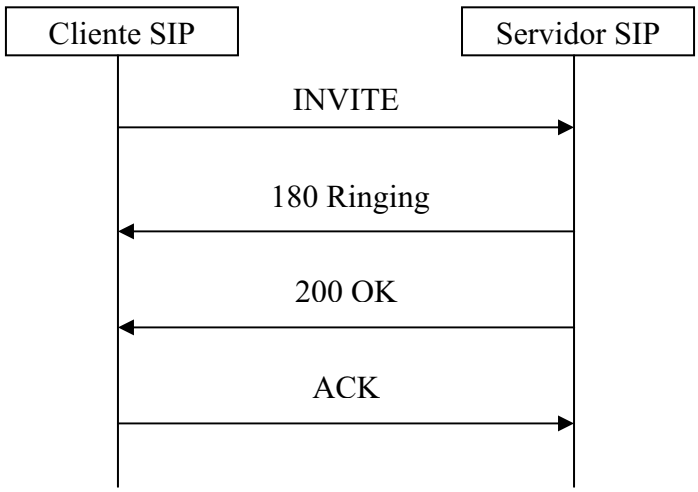


Figura 3.5. Estabelecimento de conexão SIP.

A troca de mensagens SIP apresentada anteriormente é suficiente para estabelecer o canal onde irá fluir a sinalização pretendida. A partir de então, a troca de dados isócronos ocorre utilizando o protocolo RTP, de tal maneira que os dados podem ser enviados de forma desordenada (não associado a um fluxo) e não confiável através da associação SCTP. Nesse caso, os dados carregados sobre RTP, que podem ser voz codificada ou mensagens de texto, são transportados utilizando os serviços da extensão PR-SCTP.

Enquanto um terminal SIP move-se entre redes lógicas distintas, dois procedimentos devem ser adotados, além da aquisição do novo endereço IP. O primeiro procedimento diz respeito à troca de mensagens ASCONF e a ASCONF-ACK pelo SCTP [Costa 2005a], para substituição dos endereços IP da associação corrente. Esse

mecanismo de troca de mensagens é apresento na Figura 3.6. Nessa Figura, o terminal que adquiriu o novo endereço inicia o procedimento de atualização.

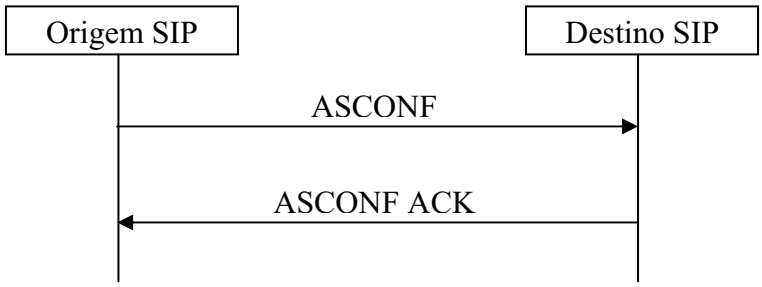


Figura 3.6. Atualização dinâmica de endereço na associação.

Em seguida, no segundo procedimento, o terminal emissor do ASCONF deve enviar uma mensagem de atualização ao seu *home registrar*, informando sua localização atual após a mudança de endereço IP. Essa medida tem a finalidade de tornar o terminal móvel alcançável por outros terminais SIP da arquitetura. Essa operação é apresentada na Figura 3.7.

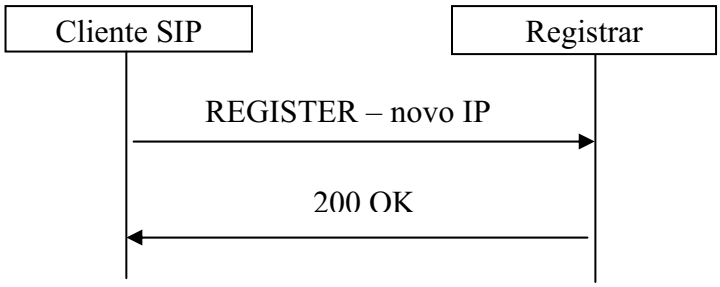


Figura 3.7. Atualização de registro em um SIP registrar.

A ordem de processamento apresentada nas duas trocas de mensagens anteriores deve ser mantida, tendo em vista que a atualização dos endereços IP da associação SCTP deve ocorrer o mais rápido possível, evitando assim eventuais perdas de pacotes e atrasos na transmissão dos dados na comunicação corrente. A atualização do registro no *home registrar* pode ocorrer em seguida, sem maiores problemas para a operação eficiente dos terminais comunicantes.

Para encerrar a comunicação de voz, a sessão multimídia SIP deve ser desfeita, com a indicação em nível de aplicação através de mensagens SIP BYE. Em seguida, a associação SCTP é também encerrada, com o *three-way handshake* de mensagens SCTP SHUTDOWN, como apresentado na Figura 3.8. Após esses procedimentos de finalização da comunicação, nenhuma mensagem de atualização deve ser enviada ao *home registrar* das estações comunicantes.

Seguindo esses procedimentos, os usuários tornam-se aptos a participarem de novas comunicações de voz, seguindo as especificações de comunicação da arquitetura considerada.

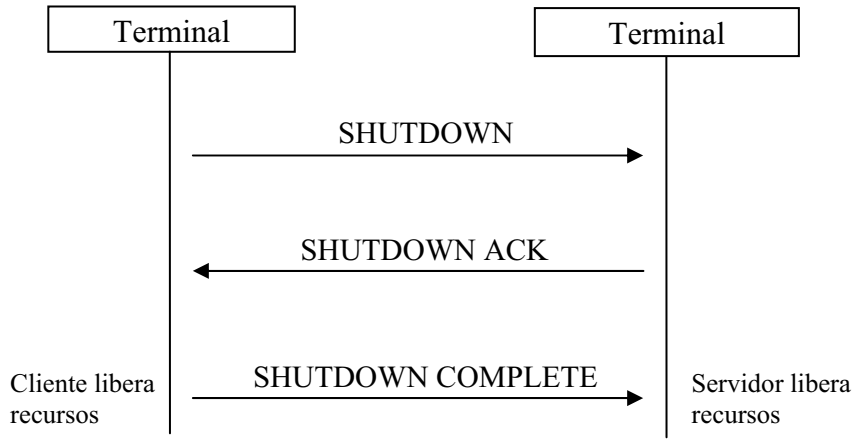


Figura 3.8. Encerramento coordenado da associação SCTP.

4 Especificação do Módulo de Controle

A Figura 3.1 apresentou a pilha de protocolos da arquitetura de comunicação considerada. Nessa Figura, um módulo específico, chamado de Módulo de Controle, é utilizado para coordenar a operação conjunta dos protocolos que compõe essa arquitetura.

O Módulo de Controle é um elemento crucial para a operação da arquitetura, não apenas pela coordenação da utilização dos protocolos envolvidos, mas também por ser esse módulo responsável pela interação entre as aplicações usuárias e a arquitetura propriamente dita.

Para evitar ambigüidades e interpretações errôneas das particularidades de operação desse módulo, pode-se empregar um formalismo para a sua especificação. Assim, eventuais implementações da arquitetura poderiam utilizar, além das descrições textuais presentes no capítulo anterior, um conjunto de descrições relacionadas ao funcionamento do Módulo de Controle da arquitetura.

O capítulo anterior apresenta os detalhes de operação da arquitetura e suas características. Os elementos operacionais da arquitetura, bem como as mensagens de controle trocadas entre os protocolos constituintes, foram também apresentados. Nesse capítulo, contudo, o foco é dirigido para a especificação do Módulo de Controle, enquanto o funcionamento da arquitetura como um todo é especificado apenas de forma superficial. Como os protocolos que compõe a arquitetura já são especificados por órgãos de padronização, não se faz necessária a especificação desses componentes. O Módulo de Controle, entretanto, é específico da arquitetura considerada, sendo bastante relevante a sua especificação.

Como já ressaltado, esse trabalho especifica a arquitetura de comunicação considerada em nível de seus componentes, especificamente do Módulo de Controle, devido ao cenário atual de utilização de alguns desses componentes. As implementações do protocolo SCTP e de suas extensões, por exemplo, não estão ainda completamente disponíveis, principalmente quando consideramos extensões mais recentes como o

MSCTP. O estado de maturação que muitas das implementações ainda estão passando desencoraja, no momento, eventuais esforços de codificação da arquitetura considerada.

Numa esfera paralela, uma especificação pode vir a auxiliar também em tarefas de simulação da operação da arquitetura. Atualmente, as ferramentas de simulação de protocolos comumente usadas em cenários acadêmicos não ofereceram suporte adequado ao protocolo SCTP e as suas extensões. O NS-2 [NS-2 2005], por exemplo, possui limitações no seu projeto funcional que impede a utilização de *host*s *multihomed* que não sejam roteadores. Além disso, para implementar operações de mobilidade com o SCTP nesse simulador, são necessárias adaptações que ocultariam o desempenho real da arquitetura de comunicação considerada.

Assim, optou-se pela especificação do Módulo de Controle, de forma que a descrição desse componente fosse formalmente definida. Essa especificação utilizou a linguagem de especificação formal SDL e a ferramenta CAD SanDriLa.

4.1 Sistemas da arquitetura de comunicação

A arquitetura considerada utiliza diversos protocolos para obter o serviço de comunicação desejado, que envolve comunicações em tempo real e mobilidade. Cada um dos protocolos utilizados possui especificações próprias, sendo o funcionamento deles já conhecido. Essa arquitetura define como alguns protocolos existentes no ambiente Internet operam em conjunto, a fim de obter um serviço de comunicação melhor que o oferecido por soluções comuns. Assim, o objetivo da arquitetura descrita no capítulo anterior é modelar como os protocolos se comunicam, e não especificar o funcionamento de elementos já definidos. Esse comportamento de operação conjunta de protocolos é modelado através do Módulo de Controle. Esse módulo recebe requisições de aplicações usuárias e utiliza os protocolos da arquitetura de forma correta.

Na especificação do Módulo de Controle utilizando a linguagem SDL, a comunicação entre os diversos protocolos é especificada. A operação de cada protocolo individual não é definida, uma vez que versões padrões dos protocolos são utilizadas. No caso do protocolo SIP, que teve sua operação simplificada com a não utilização de

handover por mensagens do próprio protocolo, uma especificação adicional não é feita: as diversas maneiras através das quais o protocolo SIP pode ser adaptado estão fora do escopo da especificação do Módulo de Controle.

O Módulo de Controle é especificado como membro do sistema Terminal. Esse sistema está relacionado ao ambiente de comunicação típico da arquitetura de comunicação, que envolve, além dos terminais de voz, servidores *registrar*, DNS e DHCP. Na especificação em SDL, apenas os terminais de voz, que contém o Módulo de Controle, precisam ser considerados em detalhes: os servidores SIP *registrar*, DNS e DHCP operam de forma padrão (de acordo com suas RFCs), de maneira que são considerados apenas como sistemas externos na especificação. Já os terminais de voz, onde o Módulo de Controle atua, utilizam protocolos de comunicação que devem operar ordenadamente entre si. Assim, os terminais de voz serão especificados, sendo a ênfase dessa especificação o Módulo de Controle.

As seções seguintes apresentam os artefatos referentes à especificação do sistema Terminal, onde o Módulo de Controle está contido. Deve-se notar que a especificação contempla apenas as características estruturais e comportamentais desse sistema SDL, não sendo incluídos detalhes de implementação.

Sistema Terminal

O sistema Terminal é um sistema SDL composto por um conjunto de blocos que se comunicam entre si. Cada bloco representa um protocolo ou o Módulo de Controle envolvido com a operação dos terminais da arquitetura. A Tabela 4.1 descreve os blocos que compõe o sistema Terminal.

O Módulo de Controle é responsável por ordenar a comunicação entre os protocolos no tocante às comunicações em tempo real e à mobilidade. Para o usuário, o serviço todo é visto como apenas um protocolo de comunicação, estando o Módulo de Controle responsável pelo fornecimento de pontos de acesso para a aplicação usuária utilizar o serviço oferecido pelo terminal de voz. Detalhes como qual o algoritmo de codificação de voz está sendo empregado, por exemplo, dependem da aplicação que está

utilizando esse serviço de comunicação, que pode ser considerado como um sistema externo na especificação. O terminal de voz, no contexto da arquitetura, corresponde apenas ao elemento que oferece o serviço de comunicação.

Bloco	Descrição
Módulo de Controle	Coordena a operação conjunta dos outros blocos. Esse bloco recebe e envia dados e requisições da aplicação usuária
SIP	Corresponde ao protocolo SIP simplificado
RTP	Corresponde ao protocolo RTP
DHCP	Corresponde ao protocolo DHCP
DNS (*resolver*)	Corresponde ao módulo resolver do protocolo DNS
UDP	Corresponde ao protocolo UDP
PR-SCTP	Corresponde à extensão PR-SCTP e implementa funcionalidades específicas para a comunicação parcialmente confiável
MSCTP	Corresponde à extensão MSCTP e implementa funcionalidades específicas para a mobilidade de terminal
SCTP	Corresponde ao protocolo SCTP padrão
Protocolo IP	Corresponde ao protocolo IP. É o ponto de interação do sistema Terminal com os sistemas "externos"

Tabela 4.1. Blocos do sistema Terminal.

Há dois pontos de interação do sistema Terminal com os sistemas externos: o Módulo de Controle, que faz a comunicação com a aplicação usuária, e o bloco Protocolo IP, que faz a comunicação com os sistemas *registrar*, DNS, DHCP e Terminal, que numa implementação típica estariam ligados em rede.

A Figura 4.1 apresenta o sistema Terminal com todos os seus blocos constituintes. Nessa Figura, podemos ver um conjunto variado de sinais e canais. A Tabela 4.2 apresenta os sinais que trafegam no sistema Terminal. Na próxima seção são apresentados os sinais que correspondem às primitivas do serviço de comunicação oferecido, trocados entre a aplicação usuária e o Módulo de Controle.

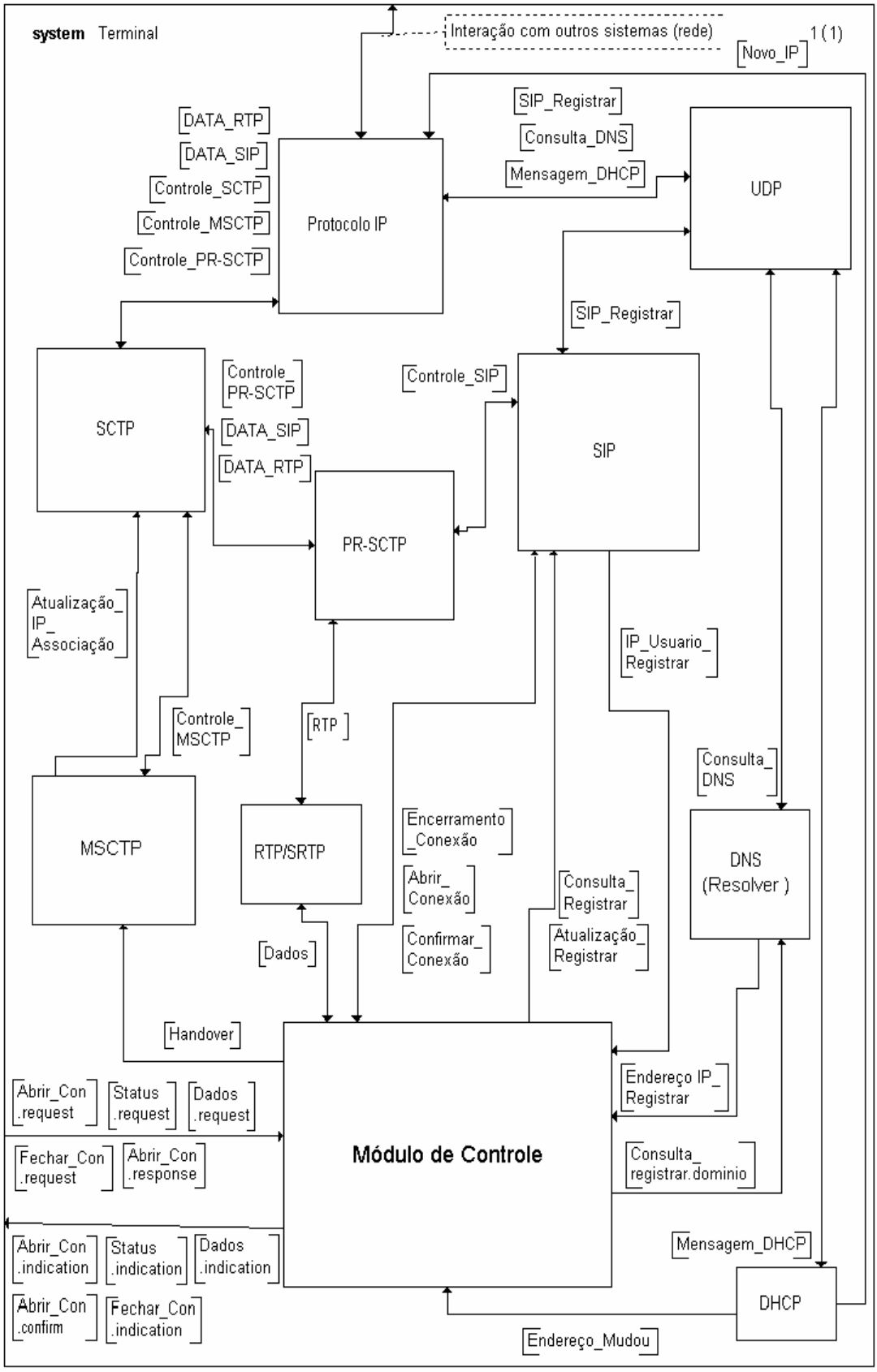


Figura 4.1. Sistema Terminal.

Sinal	Canal	Descrição
DATA_SIP	SCTP ↔ PR-SCTP	Mensagens SIP carregadas em mensagens SCTP DATA
Controle_SIP	PR-SCTP ↔ SIP	Mensagens de controle SIP para o controle da comunicação: INVITE, ACK, 200 OK, 180 RINGING e BYE.
DATA_RTP	SCTP ↔ PR-SCTP	Pacotes RTP transmitidos em mensagens SCTP DATA
RTP	RTP ↔ PR-SCTP	Representa os pacotes RTP
Dados	Sistema Usuario ↔ Módulo de Controle RTP ↔ Módulo de Controle	Dados da comunicação (mídia codificada ou texto)
Controle_SCTP	SCTP ↔ Protocolo IP	Mensagens de controle SCTP: INIT, INIT ACK, SACK, etc
Controle_MSCTP	MSCTP ↔ SCTP	Mensagens de controle MSCTP: ASCONF e ASCONF ACK
Controle_PR-SCTP	SCTP ↔ PR-SCTP	Mensagem de controle FWD-TSN
Handover	Modulo de Controle → MSCTP	Solicitação para enviar uma mensagem ASCONF
Novo_IP	DHCP → Protocolo IP	Indicação à camada de rede para alterar o endereço IP
Consulta_registrar. dominio	Modulo de Controle → DNS	Consulta ao registro “registrar.dominio”, que irá retornar o IP do home registrar do usuário que será chamado
Endereço_IP_Registrar	DNS → Modulo de Controle	Resultado da consulta pelo IP do home registrar do usuário que será chamado
Consulta_DNS	DNS ↔ UDP	Representa as mensagens DNS envolvidas em consultas. Trocados entre o sistema Terminal e o sistema DNS
Mensagem_DHCP	DHCP ↔ UDP UDP ↔ Protocolo IP	Mensagens de operação do DHCP. Trocados entre o sistema Terminal e o sistema DHCP
Endereço_Mudou	DHCP → Modulo de Controle	Indicação de mudança de rede lógica (endereço IP mudou)
Abrir_Conexão	Modulo de Controle ↔ SIP	Pedido para abertura de conexão (SIP/SCTP)
Confirmar_Conexão	Modulo de Controle ↔ SIP	Confirmação (sim/não) do pedido de abertura de conexão

Sinal	Canal	Descrição
Encerramento_Conexão	Modulo de Controle ↔ SIP	Pedido para encerramento de conexão
Atualização_Registrar	Modulo de Controle → SIP	Pedido para atualizar registro do terminal no home registrar
Consulta_Registrar	Modulo de Controle → SIP	Consulta ao registro do usuário a ser chamado no seu home registrar
IP_Usuario_Registrar	SIP → Modulo de Controle	Resultado da consulta pelo IP do usuário a ser chamado
SIP_Registrar	SIP ↔ UDP	Mensagem SIP REGISTER. Trocados entre o sistema Terminal e o sistema Registrar
Atualização_IP_ Associação	MSCTP → SCTP	Atualização de endereço da associação ao receber um ASCONF

Tabela 4.2. Sinais do sistema Terminal.

4.2. Serviço de comunicação do Módulo de Controle

O Módulo de Controle é responsável pelo fornecimento de um serviço de comunicação às aplicações usuárias de um Terminal. Através de primitivas de serviço, esse bloco recebe pedidos da camada usuária e os processa, ativando os protocolos da arquitetura necessários à operação requerida. Da mesma forma, situações específicas de estado da arquitetura são indicadas à aplicação.

A Tabela 4.3 apresenta os serviços de comunicação providos pelo Módulo de Controle e as primitivas de serviço relacionadas à utilização de um Terminal da arquitetura.

Serviço	Primitiva de Serviço	Parâmetros
Abrir Conexão	Abrir_Con.request	Endereço SIP destino
	Abrir_Con.indication	Endereço SIP origem
	Abrir_Con.response	Confirmação (sim/não) do pedido de abertura de conexão

	Abrir_Con.confirm	Confirmação (sim/não) do pedido de abertura de conexão
Encerrar Conexão	Fechar_Con.request	Endereço SIP destino
	Fechar_Con.indication	Endereço SIP origem
Enviar Dados	Dados.request	Dados (voz/texto)
	Dados.indication	Dados (voz/texto)
Estado do Terminal	Status.request	Parâmetros requeridos
	Status.indication	Parâmetros consultados

Tabela 4.3. Primitivas de serviço.

4.3. Processos do Módulo de Controle

O bloco Módulo de Controle é responsável pela interação das aplicações usuárias do serviço de comunicação com os protocolos que compõem a arquitetura. Esse bloco possui relevância crucial na operação da arquitetura, pois é ele o responsável por ordenar a troca de mensagens entre os protocolos relacionados ao serviço de comunicação oferecido, além, é claro, de oferecer primitivas de acesso a esse serviço de comunicação.

O bloco Módulo de Controle é formado por três processos distintos. A Figura 4.2 apresenta os processos que compõem esse bloco: os processos Controlar Chamadas, Controlar Mobilidade e Controlar Dados, que são descritos nas três subseções seguintes.

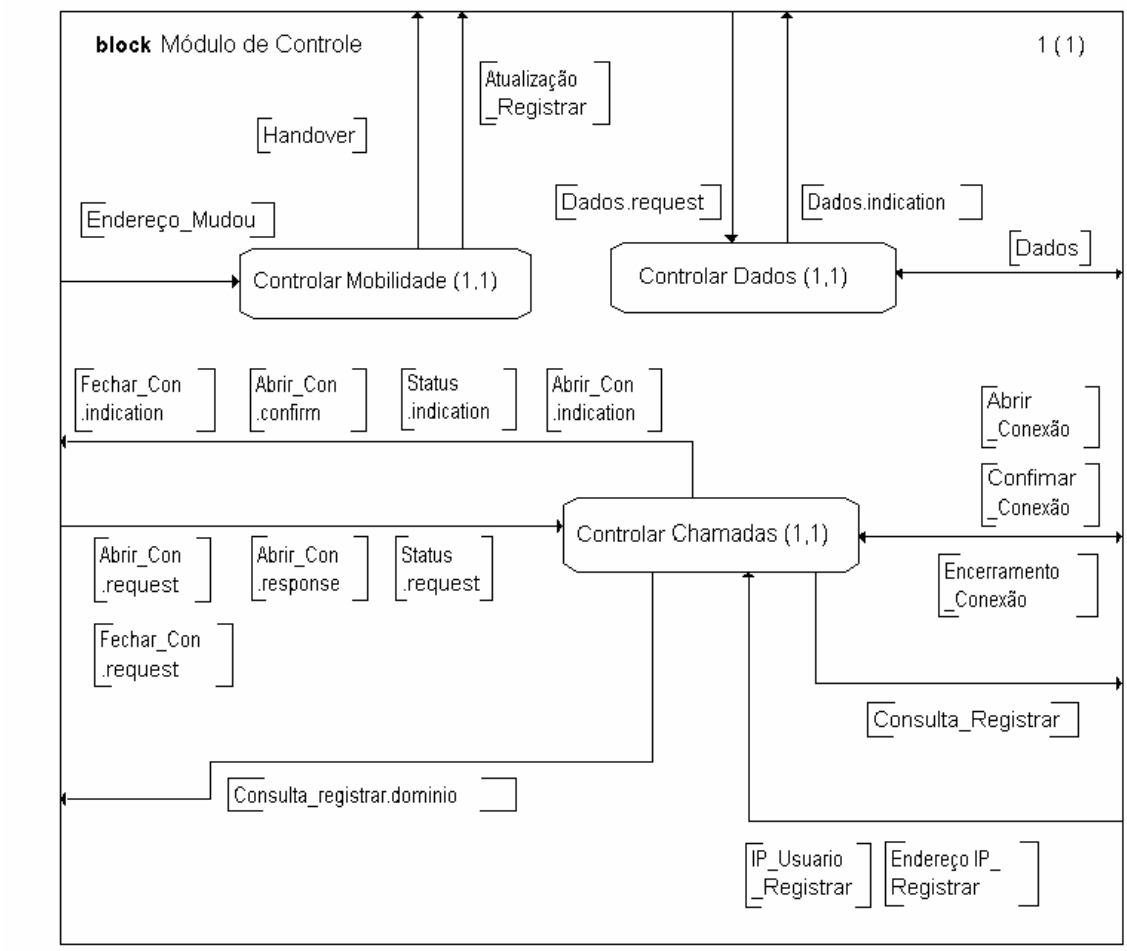


Figura 4.2. Bloco Módulo de Controle.

4.3.1 Processo Controlar Chamadas

O processo Controlar Chamadas é responsável por receber pedidos de abertura e encerramento de conexões por parte da aplicação usuária, e acionar os protocolos envolvidos com a comunicação de voz em tempo real e localização de usuário. Além disso, esse processo também recebe pedidos de abertura e encerramento de conexão de um terminal remoto, através do bloco SIP, e os repassa para a aplicação usuária, que pode aceitar ou não o pedido, quando esse for uma solicitação de abertura de conexão.

Para permitir o acompanhamento da situação de um terminal da arquitetura, esse processo também recebe primitivas de consulta ao estado atual da comunicação.

Quando o processo quer abrir uma conexão, uma indicação é enviada ao bloco SIP. Esse bloco, então, solicita a abertura de uma conexão SCTP com o terminal chamado, para realizar a comunicação de voz em tempo real. As Figura 4.3 e 4.4 apresentam a especificação do processo Controlar Chamadas.

Um aspecto interessante em relação a definições de processos é que tarefas de conteúdo "genérico" podem ser definidas. Na Figura 4.3, a tarefa "Alocar recursos", por exemplo, possui significado subjetivo no contexto dessa especificação, na medida em que os detalhes de como os recursos serão processados podem variar entre implementações. O mesmo ocorre também com a tarefa "Desalocar recursos", apresentada na Figura 4.4.

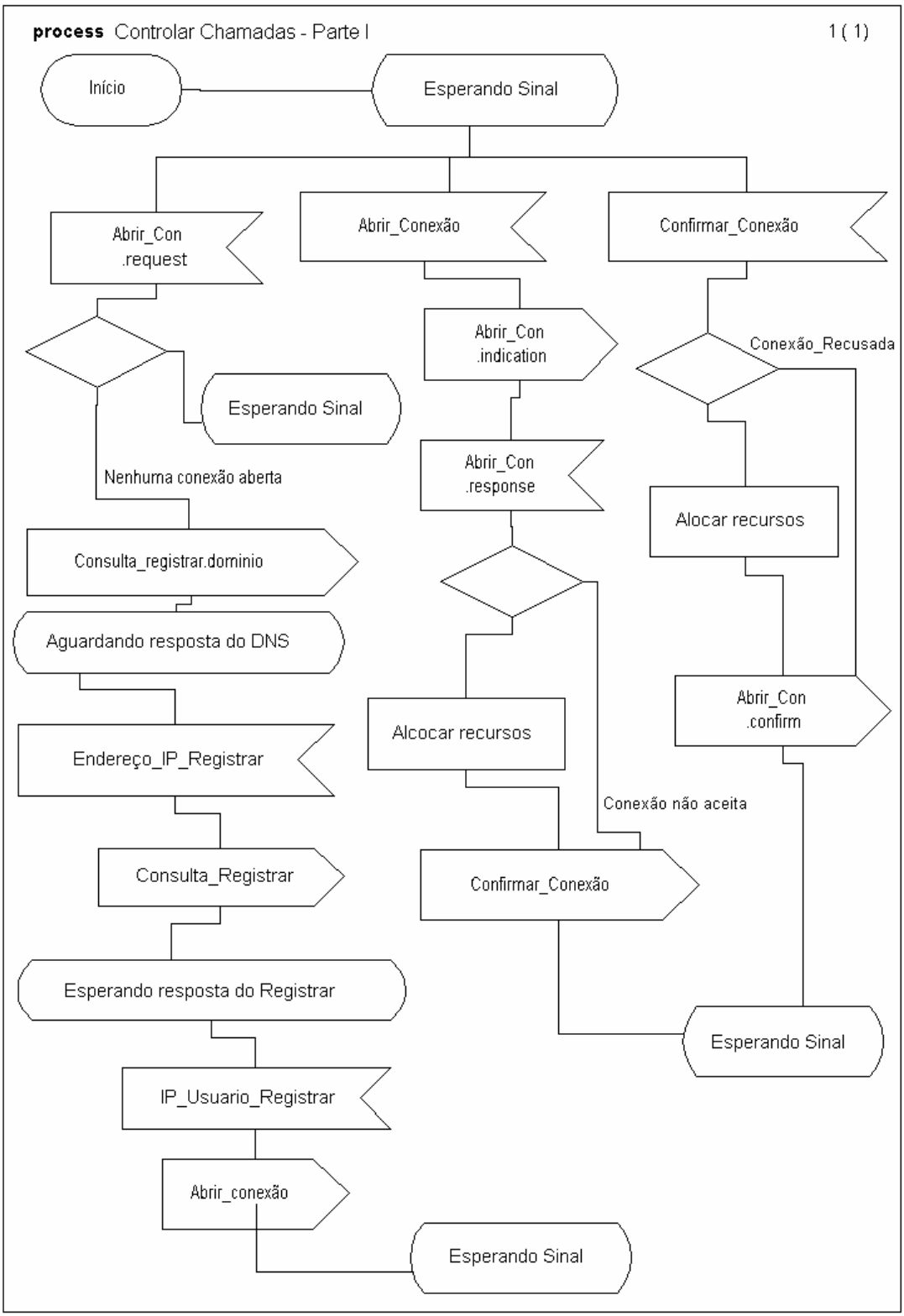


Figura 4.3. Processo Controlar Chamadas - Parte I.

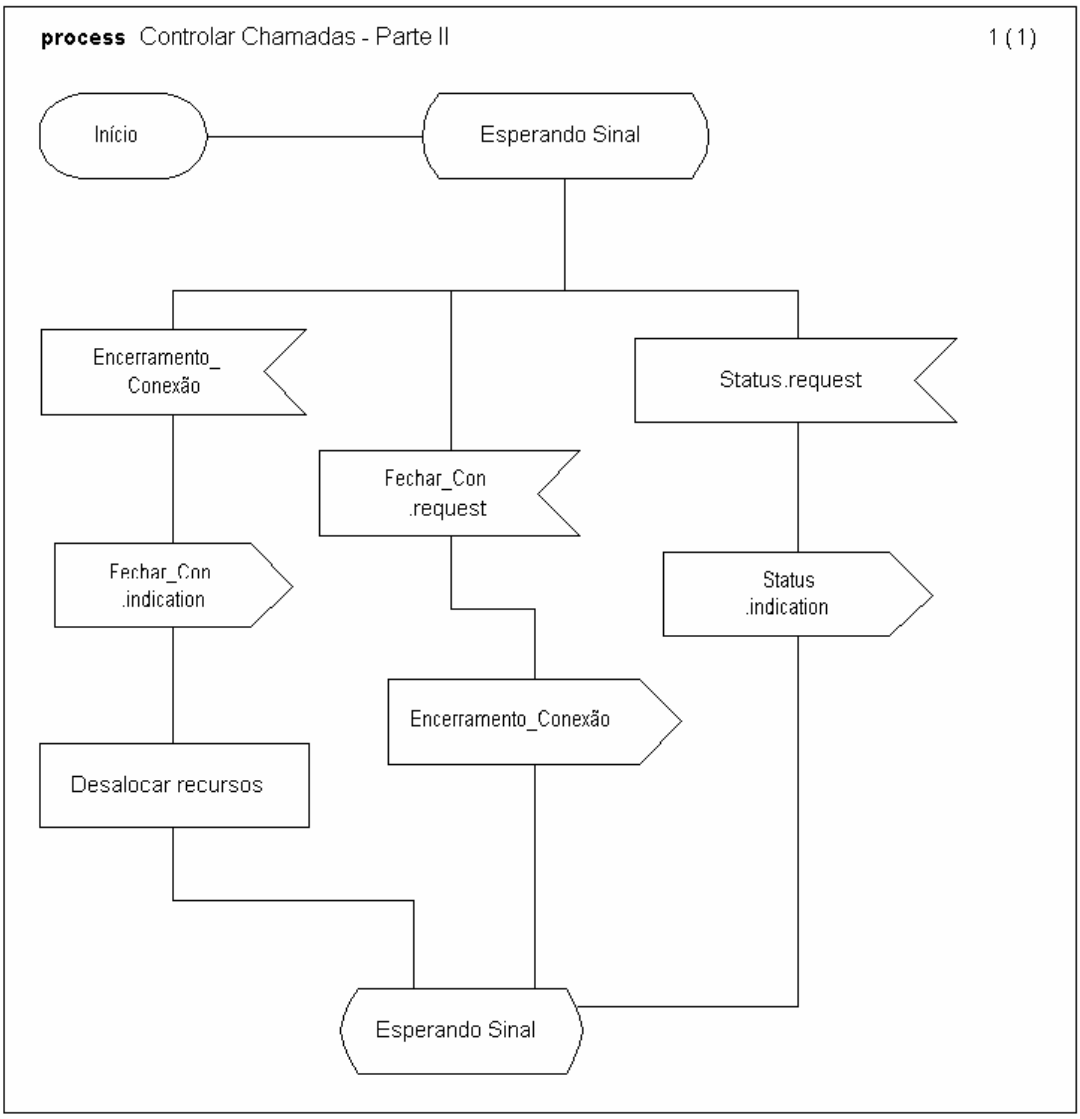


Figura 4.4. Processo Controlar Chamadas - Parte II.

4.3.2 Processo Controlar Mobilidade

O processo Controlar Mobilidade está relacionado às atualizações dinâmicas de endereço que devem ocorrer nas comunicações que seguem a arquitetura, quando um *host* móvel altera seu endereço IP. Essas atualizações estão divididas em *handover*, que ocorre em nível de transporte (MSCTP) e atualização do registro do terminal no seu *home registrar*, que ocorre na camada de aplicação (SIP).

A Figura 4.5 apresenta o processo Controlar Mobilidade.

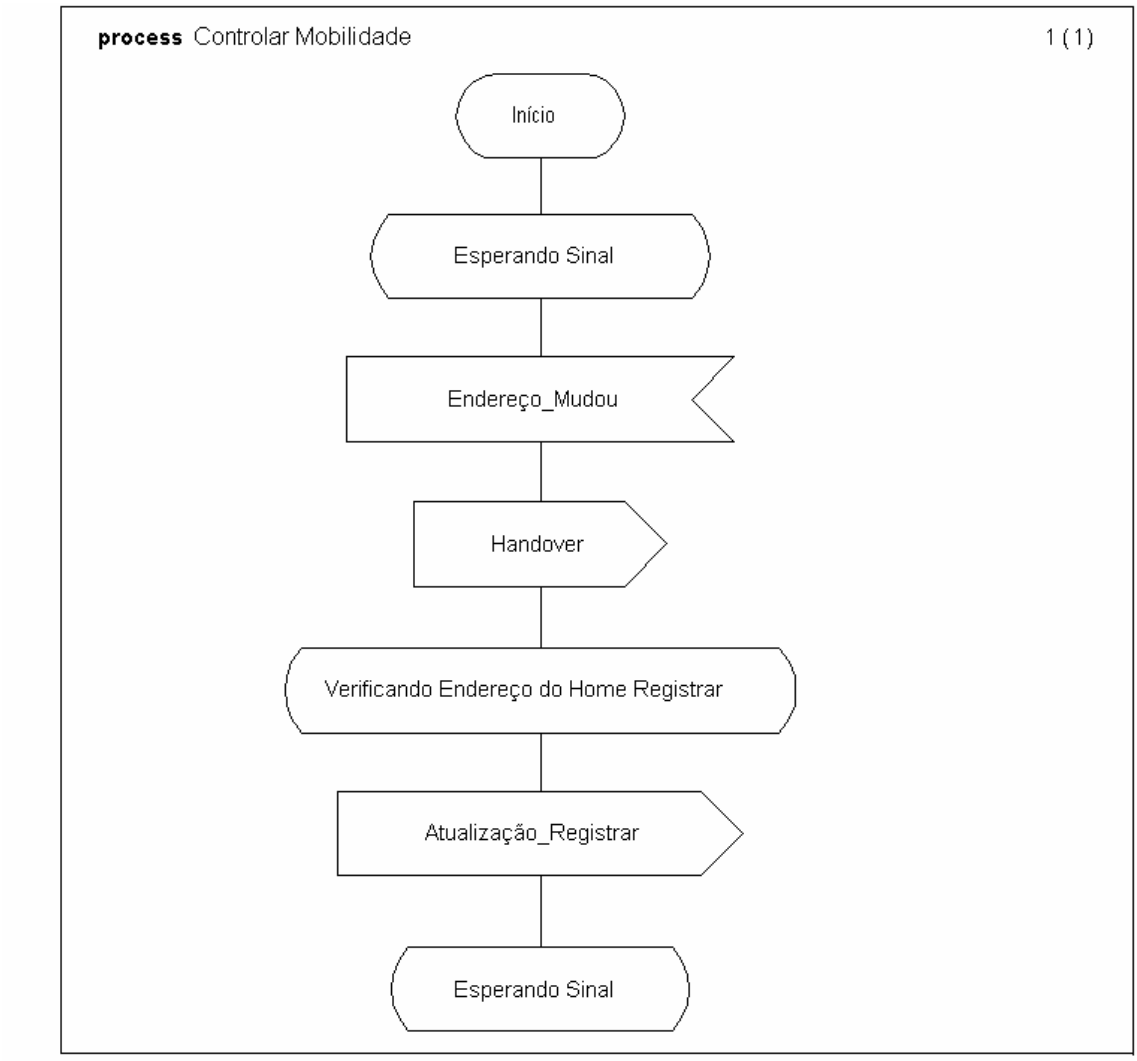


Figura 4.5. Processo Controlar Mobilidade.

Detalhes como a obtenção do endereço do *home registrar*, indicado pelo estado "Verificando Endereço do Home Registrar", estão diretamente ligados à implementação, não sendo, portanto, considerados na especificação da arquitetura. Nesse caso, uma implementação típica poderia utilizaria consultas DNS ao registro *registrar.dominio* para descobrir esse endereço, onde esse domínio poderia ser obtido a partir do endereço SIP do usuário.

4.3.3 Processo Controlar Dados

O processo Controlar Dados é responsável por enviar dados (voz ou texto), provenientes da aplicação usuária, para o terminal remoto, e, inversamente, do terminal remoto para a aplicação usuária. Esse processo possui uma operação bastante simples, uma vez que sua finalidade é apenas construir uma "interface" entre o sistema Terminal e a aplicação usuária. Pode-se pensar nesse processo como sendo um "canal", onde os dados de voz e texto fluem nos dois sentidos.

Para a transmissão dos dados pelas redes de comunicação, o protocolo RTP é utilizado. O processo Controlar Dados faz o encapsulamento e desencapsulamento dos dados contidos nas unidades de transmissão desse protocolo.

Detalhes de implementação relacionados ao processamento das mídias presentes na arquitetura de comunicação considerada não são cobertos por esse processo. Assim, se esses dados estão codificados por algum codec ou não, por exemplo, depende da implementação da aplicação usuária do terminal de voz.

A Figura 4.6 apresenta o processo Controlar Dados.

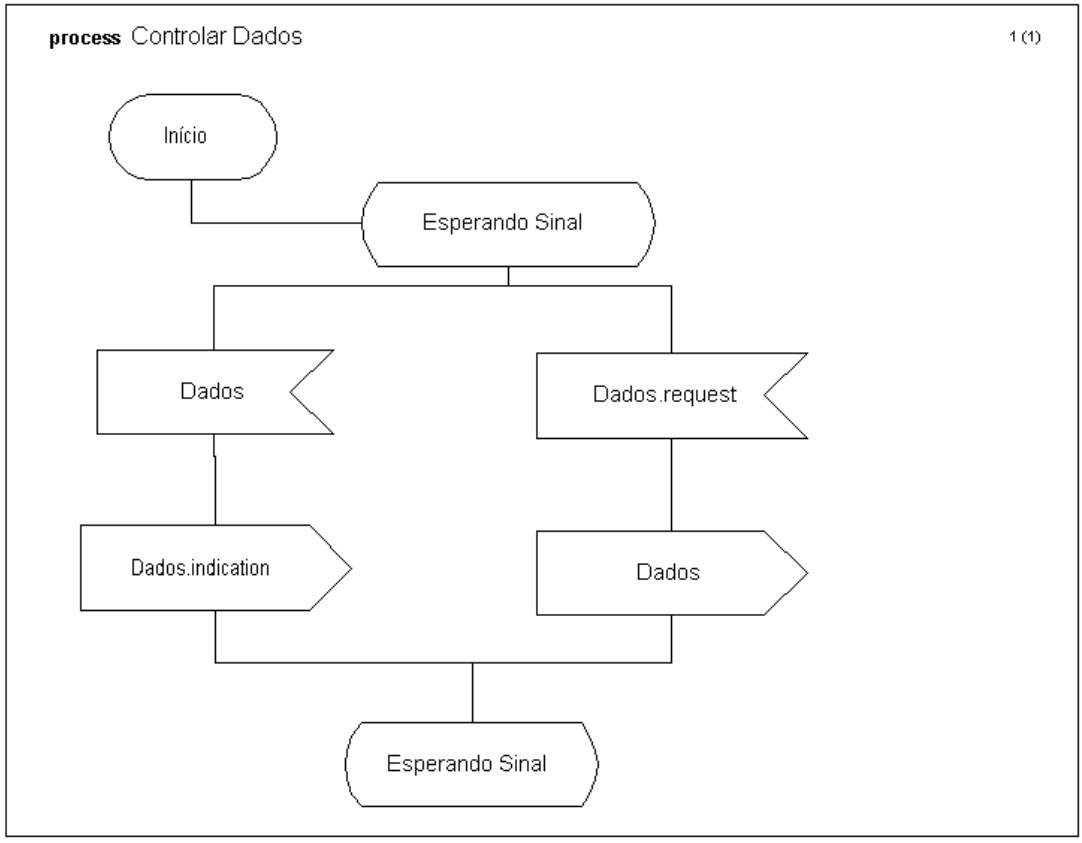


Figura 4.6. Processo Controlar Dados.

4.4 Verificação da especificação

A especificação desenvolvida teve como objetivo definir as funcionalidades do Módulo de Controle, apresentando adicionalmente a troca de mensagens entre os blocos que constituem um Terminal da arquitetura proposta. Contudo, o foco da especificação está voltado à operação dos processos que compõem esse bloco. Assim, a verificação da especificação deve tentar assegurar a não existência de erros na definição do Módulo de Controle.

A verificação léxica ocorreu de forma "automática", uma vez que apenas os elementos da linguagem SDL são fornecidos pelo ambiente de especificação. Assim, não há como utilizar elementos não padronizados na especificação. Portanto, lexicamente, a especificação não contém erros.

A verificação sintática é necessária para assegurar que a linguagem SDL foi utilizada corretamente, ou seja, que os elementos estão ligados da forma esperada e que os diagramas estão de acordo com as regras de utilização dessa linguagem. Através do ambiente de especificação utilizado, é possível realizar uma verificação sintática de cada diagrama. Realizando essa verificação, foi possível constatar a ausência de erros dessa natureza na especificação.

A Figura 4.7 apresenta uma tela de verificação sintática do diagrama Sistema Terminal, apresentado na Figura 4.1.

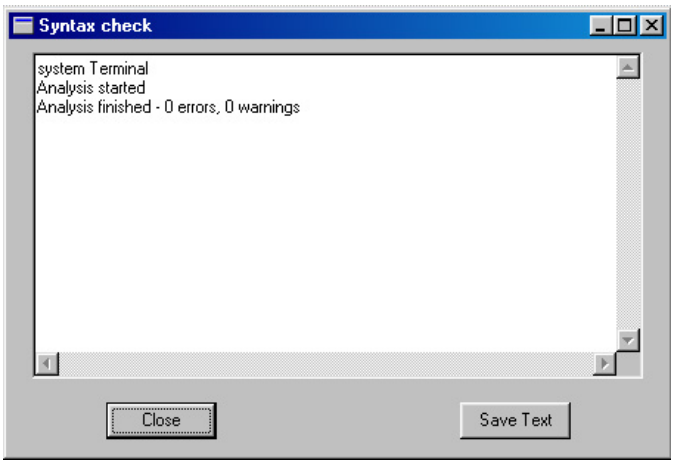


Figura 4.7. Tela de verificação sintática.

Outra verificação realizada analisa se os diagramas estão relacionados de forma correta. A Figura 4.8 apresenta as ligações semânticas entre os diagramas.

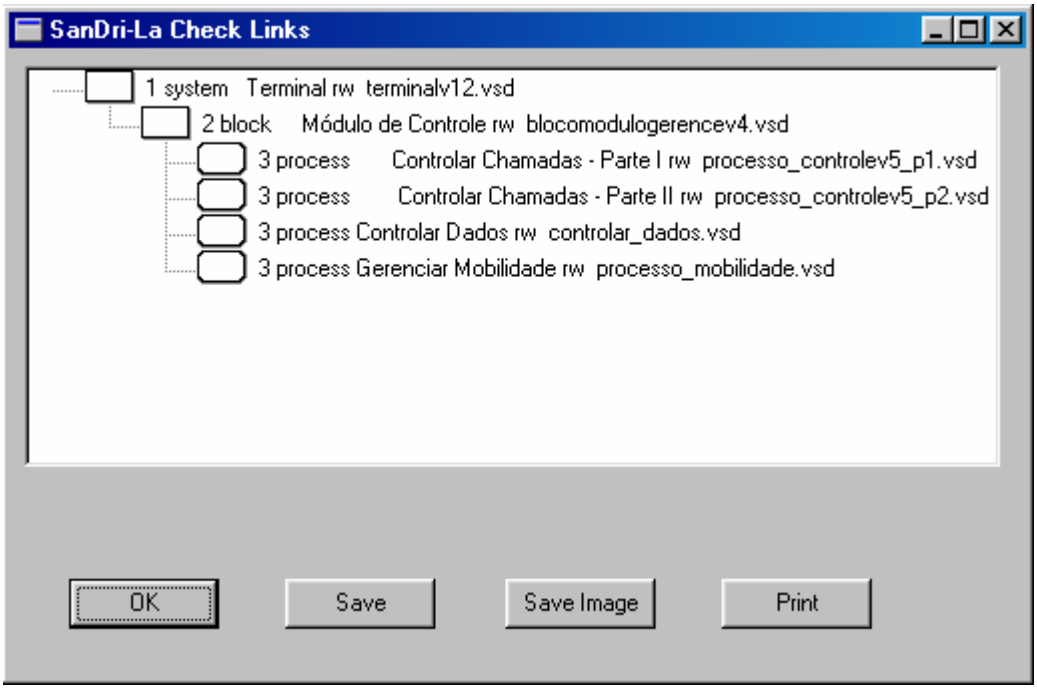


Figura 4.8. Relacionamento entre os diagramas.

Apesar das verificações realizadas, não se pode afirmar que a especificação desenvolvida corresponde ao planejamento do funcionamento do Módulo de Controle. A falta de verificação semântica não permite afirmar, por exemplo, a ausência de *deadlocks* e *livelocks* no funcionamento desse módulo. Verificações desse tipo não estão disponíveis no SanDriLa. Para sanar esse problema, seria necessário se ter acesso a ferramentas que permitam fazer análises semânticas, ou mesmo simulações de operação, para poder garantir a ausência de inconsistências operacionais na especificação desse módulo. Como esse não foi o caso, somente verificações manuais da operação do módulo foram realizadas, não tendo sido encontrados erros de especificação.

Apesar dessa limitação, a especificação ainda pode ser utilizada para guiar implementações de Terminais da arquitetura, uma vez que a deliberada ausência de detalhes de implementação na especificação diminui, porém não extingue, prováveis inconsistências operacionais no funcionamento dos Terminais.

5 *Conclusão*

O protocolo SCTP vem se apresentando à comunidade Internet como alternativa real a aplicações que demandam serviços adicionais da camada de transporte, não disponibilizados pelos protocolos tradicionais UDP e TCP. Por ainda não estar consolidado como protocolo de transporte de ampla utilização na Internet, o SCTP permite maior flexibilidade à adoção de novos mecanismos operacionais, como aqueles referentes ao suporte à mobilidade e à transmissão parcialmente confiável de dados de usuário, o que deve culminar numa maturação mais eficiente desse protocolo.

Em relação às arquiteturas de comunicação multimídia em tempo real, o SIP está se tornando padrão de mercado para aplicações que necessitem de um protocolo leve e flexível. Na área de telefonia IP, esse protocolo é um dos mais fortes candidatos à adoção em arquiteturas de comunicação em tempo real.

A arquitetura aqui descrita, integrando a eficiência desses dois protocolos de comunicação em seus escopos operacionais, apresenta-se como uma boa solução para as aplicações de voz sobre IP, sobretudo àquelas com requisitos de mobilidade. Além das vantagens apresentadas nesse trabalho, a facilidade de implantação e utilização dessa arquitetura é apontada como ponto positivo para consideração desse projeto. A característica de comunicação fim-a-fim, não diretamente dependente dos *backbones* de comunicação, como ocorre com MIP, facilita sobremaneira a adoção dessa arquitetura na Internet.

Para promover a integração adequada desses protocolos, tornou-se necessário especificar também um Módulo de Controle, com o objetivo de coordenar a operação dos vários módulos existentes envolvidos. Esse módulo foi totalmente especificado em SDL, utilizando-se a ferramenta SanDriLa, e a especificação foi verificada tão extensamente quanto essa ferramenta o permitiu.

Os resultados obtidos com o desenvolvimento do trabalho são satisfatórios. Embora não tenha sido implementada, os esforços desprendidos com as definições e especificações da arquitetura devem facilitar trabalhos adicionais envolvendo a arquitetura e possíveis variações.

Os esforços de pesquisas em soluções de mobilidade prometem ainda aquecer o campo das discussões sobre a melhor solução a ser adotada. Ainda mais, esse cenário de pesquisa é alimentado pela demanda por telefonia de terceira e quarta geração, onde os ambientes de micro e macro mobilidade devem estar integrados. Soluções como Mobile IP, Mobile SCTP e HIP são alternativas já estudadas no meio acadêmico para atender essa demanda; contudo, outras arquiteturas de mobilidade, frutos de extensões dessas soluções ou oriundos de novos paradigmas, devem ser ainda consideradas.

Como propostas de trabalhos futuros, relacionadas ao projeto realizado, novas especificações da arquitetura deverão abordar, de forma mais detalhada, alguns aspectos adicionais de comunicação. Dentro dessa abordagem mais detalhada da arquitetura, podem-se incluir: aspectos de segurança, com a adoção de mecanismos que englobem todas as comunicações; aspectos de contabilidade de recursos consumidos da rede, para algum eventual procedimento de controle ou cobrança; e, ainda, aspectos relacionados ao uso de políticas de qualidade de serviço, que podem ser recomendados em função do ambiente onde ocorrem as comunicações de voz. Contudo, um ponto mais importante, que deve dominar os esforços de desenvolvimento de uma nova especificação da arquitetura, refere-se a serviços de comunicação WWAN, como os oferecidos pela telefonia de terceira e quarta geração. Nesse contexto, técnicas mais robustas de *handover* deverão ser pesquisadas, a fim de permitir um desempenho satisfatório da arquitetura nesse novo ambiente de comunicação.

Paralelamente, esforços voltados à implementação e simulação dos terminais e das comunicações entre os elementos da arquitetura devem guiar futuros trabalhos relacionados ao amadurecimento e evolução dessa arquitetura.

Referências bibliográficas

[Andersen 2004] ANDERSEN, S. DURIC, A. ASTROM, H. RFC 3951: Internet Low Bit Rate Codec (iLBC). dez. 2004. Disponível em: <http://www.ietf.org/rfc/rfc3951.txt>.

[Baugher 2004] BAUGHER, M., MCGREW, D. RFC 3711: The Secure Real-time Transport Protocol (SRTP). mar. 2004. Disponível em: <http://www.ietf.org/rfc/rfc3711.txt>.

[Bellovin 2003] BELLOVIN. S., IOANNIDIS, J. RFC 3554: On the Use of Stream Control Transmission Protocol (SCTP) with IPsec. jul. 2003. Disponível em: <http://www.ietf.org/rfc/rfc3554.txt>.

[Cheswick 2005] CHESWICK, W. R., BELLOVIN, S. M., RUBIN, A. D. Firewalls e Segurança na Internet. 2ª ed. Porto Alegre: Editora Bookman. 2005. 400p.

[Comer 1998] COMER, D. E. Interligação em Redes com TCP/IP, Volume 1. 2ªed. Rio de Janeiro: Editora Campus. 1998. 672p.

[Costa 2005a] COSTA, Daniel Gouveia. SCTP: Uma Alternativa aos Tradicionais Protocolos de Transporte da Internet. 1ª ed. Rio de Janeiro: Editora Ciência Moderna. 2005. 120p.

[Costa 2005b] COSTA, D. G., FIALHO, S. V. Uma Arquitetura Compondo SCTP, SIP e Protocolos Auxiliares para Suporte a Aplicações VoIP Móveis. In: Simpósio Brasileiro de Telecomunicações 2005. Campinas, SP, Set. 2005.

[DARPA 1981] IETF DARPA. RFC 793: Transmission Control Protocol. sep. 1981. Disponível em: <http:// www.ietf.org/rfc/rfc0793.txt>.

[Dreibholz 2003] DREIBHOLZ, T., TÜXEN, M. A new Scheme for IP-based Internet Mobility. out. 2003. Disponível em: <http://tdrwww.exp-math.uni-essen.de/inhalt/forschung/sctp_fb/lcn_2003_a_new_scheme_for_ip_based_internet_mobility_10_2003.pdf >.

[Droms 1997] DROMS, R. RFC 2131: Dynamic Host Configuration Protocol. mar. 1997. Disponível em: <http://www.ietf.org/rfc/rfc2131.txt>.

[Gast 2005] GAST, M. S. 802.11 Wireless Networks: The Definitive Guide. 2ª ed. O´reilly. 632p.

[Handley 1998] HANDLEY, M., JACOBSON, V. RFC 2327: Session Description Protocol. abr. 1998. Disponível em: <http://www.ietf.org/rfc/rfc2327.txt>.

[Hellstrom 2000] HELLSTROM, G., OMNITOR, A. B. RFC 2793: RTP Payload for Text Conversation. mai. 2000. Disponível em: <http://www.ietf.org/rfc/rfc2793.txt>.

[Hersent 2002] HERSENT, O., GUIDE, D., PETIT, J-P. Telefonia IP: Comunicação Multimídia Baseada em Pacotes. São Paulo: Editora Prentice Hall. 2002.

[IETF 1980] IETF DARPA. RFC 768: User Datagram Protocol. ago. 1980. Disponível em: <http:// www.ietf.org/rfc/rfc0768.txt>.

[IETF 1981] IETF DARPA. RFC 791: Internet Protocol. sep. 1981. Disponível em: <http:// www.ietf.org/rfc/rfc0791.txt>.

[Johnson 2004] JOHNSON, D., PERKINS, C. RFC 3775: Mobility Support in IPv6. jun. 2004. Disponível em: <http://www.ietf.org/rfc/rfc3775.txt>.

[Johnston 2004] JOHNSTON, Alan B. SIP: Understanding the Session Initiation Protocol. 2ª ed. Artech House, 2004.

[Jumgmaier 2002] JUMGMAIER, A., RESCORLA, E. RFC 3436: Transport Layer Security over Stream Control Transmission Protocol. dez. 2002. Disponível em: <http://www.ietf.org/rfc/rfc3436.txt>.

[Marco 2004] MARCO, G., VITO, D., LONGO, M., LORETO, S. SCTP as a transport for SIP: a case study. fev. 2005. Disponível em: <http://www.coritel.it/publications/IP_download/SCTP%20as%20a%20transport%20for %20SIP.pdf>.

[Martins 2004] MARTINS, L., BARBOSA, A., BARBOSA, M. Redes móveis baseadas nos protocolos IPv4 e IPv6 - uma visão geral do MIPv4 e MIPv6. nov. 2004. Disponível em: <http://www.rnp.br/newsgen/0301/mip.html>.

[Mockapetris 1987] MOCKAPETRIS, P. RFC 1034: Domain names - concepts and facilities. nov. 1987. Disponível em: <http://www.ietf.org/rfc/rfc1034.txt>.

[Moskowitz 2005] MOSKOWITZ, R., NIKANDER, P. Internet Draft: Host Identity Protocol. out. 2005. Disponível em: < http://www.ietf.org/internet-drafts/draft-ietf-hip-base-04.txt>.

[Nooman 2004] NOOMAN, J., PERRY, P., MURPHY, J. A study of SCTP services in a Mobile-IP network. fev. 2005. Disponível em: <http://www.eeng.dcu.ie/ ~jnoonan/publications/a-study-of-sctp.pdf>.

[NS-2 2005] The Network Simulator – NS-2. out. 2005. Disponível em: <http://www.isi. edu/ nsnam/ns>.

[Perkins 2002] PERKINS, C. RFC 3344: IP Mobility Support for IPv4. ago. 2002. Disponível em: <http://www.ietf.org/rfc/rfc3344.txt>.

[Ratola 2005] RATOLA, M. Which Layer for Mobility? – Comparing Mobile IPv6, HIP and SCTP. fev. 2005. Disponível em: <http://www.tml.hut.fi/Studies/T-110.551/2004/ papers/Ratola.pdf>.

[Riegel 2005] RIEGEL, M., TUXEN, M. Internet draft: Mobile SCTP. jul. 2005. Disponível em: <http://www.ietf.org/internet-drafts/draft-riegel-tuexen-mobile-sctp-05.txt>.

[Rosemberg 2002] ROSEMBERG, J. SCHULZRINNE, R. RFC 3261: Session Initiation Protocol. jun. 2002. Disponível em: <http://www.ietf.org/rfc/rfc3261.txt>.

[Sandrila 2005] SanDriLa. nov. 2005. Disponível em: <http://www.sandrila.co.uk>.

[Schmidt 2005] SCHMIDT, T C., WÄHLISCH, M. Roaming Real-Time Applications – Mobility Services in IPv6 Networks. fev. 2005. Disponível em: <http:// arxiv.org/ftp/cs/papers/0408/0408002.pdf>.

[Schulzrinne 1996] SCHULZRINNE, R., CASNER, S. RFC 1889: RTP: A Transport Protocol for Real-Time Applications. jan. 1996. Disponível em: <http://www.ietf.org/rfc/rfc1889.txt>.

[Schulzrinne 2002] SCHULZRINNE, H. RFC 3361: Dynamic Host Configuration Protocol (DHCP-for-IPv4) Option for Session Initiation Protocol (SIP) Servers. ago. 2002. Disponível em: <http://www.ietf.org/rfc/rfc3361.txt>.

[SDL 2005] SDL Forum Society. fev. 2005. Disponível em: < http://www.sdl-forum.org/>.

[Soares 1995] SOARES, L. F. G., LEMOS, G., COLCHER, S. Redes de Computadores: das LANs, MANs e WANs às redes ATM. Rio de Janeiro: Campus. 1995.

[SS7 2004] SS7 News Portal. fev. 2005. Disponível em: <http://www.ss7.com>.

[Stewart 2000] STEWART, R., XIE, Q. RFC 2960: Stream Control Transmission Protocol. out. 2000. Disponível em: <http://www.ietf.org/rfc/rfc2960.txt>.

[Stewart 2004] STEWART, R., XIE, Q. RFC 3758: Stream Control Transmission Protocol Partial Reliability Extension. mai. 2004. Disponível em: <http://www.ietf.org/rfc/ rfc3758.txt>.

[Tanembaum 2003] TANEMBAUM, A. S. Redes de Computadores. 4. ed. Rio de Janeiro: Editora Campus. 2003, 946p.

[UML 2005] Unified Modeling Language – Object Management Group. dec. 2005. Disponível em: <http://www.uml.org/>.

[Visio 2005] Site do Microsoft Visio. nov. 2005. Disponível em <http://www.microsoft.com/brasil/office/visio/default.asp>.

[Vixie 1997] VIXIE, P., THOMSON, S. RFC 2139: Dynamic Updates in The Domain Name System (DNS UPDATE). abr.1997. Disponível em: <http://www.ietf.org/rfc/rfc2139.txt>.

[Wedlung 2004] WEDLUNG, E., SCHULZRINNE, H. Mobility Support Using SIP. fev. 2005. Disponível em: <http://www.cs.columbia.edu/~hgs/papers/Wedl9908_Mobility.pdf>.

[WiMax 2005] WiMax Forum. fev. 2005. Disponível em: <http://www.wimaxforum.org /home>.

[Wireless 2005] Portal Independente de Telecomunicações. fev. 2005. Disponível em: <http://www.wirelessbrasil.org>.

[Xie 2005] STEWART, R., XIE, Q. Internet Draft: Stream Control Transmission Protocol Dynamic Address Reconfiguration. jun. 2005. Disponível em: <http://www.ietf.org/internet-drafts/draft-ietf-tsvwg-addip-sctp-12.txt>.