UNIVERSIDADE FEDERAL DO PARANÁ

ANTONIO RODRIGUES BARROS

APLICAÇÃO DE INTEGRAÇÃO DE BANCO DE DADOS VISANDO CONTROLE DE ACESSO CENTRALIZADO: UMA ALTERNATIVA DE *SINGLE SIGN-ON* PARA INSTITUIÇÕES FEDERAIS DE ENSINO SUPERIOR

CURITIBA

2010

ANTONIO RODRIGUES BARROS

APLICAÇÃO DE INTEGRAÇÃO DE BANCO DE DADOS VISANDO CONTROLE DE ACESSO CENTRALIZADO: UMA ALTERNATIVA DE *SINGLE SIGN-ON* PARA INSTITUIÇÕES FEDERAIS DE ENSINO SUPERIOR

Dissertação apresentada ao Programa de Pós-Graduação em Ciência, Gestão e Tecnologia da Informação, Setor de Ciências Sociais Aplicadas, Universidade Federal do Paraná, como requisito parcial á obtenção do grau de Mestre.

Orientador: Prof. Dr. José Simão de Paula Pinto

CURITIBA

2010

Barros, Antonio Rodrigues
Aplicação de integração de banco de dados visando controle de acesso centralizado: uma alternativa de *single sign-on* para instituições federais de ensino superior / Antonio Rodrigues Barros. – Curitiba, 2010.
133 p. : il.

Dissertação (mestrado) – Universidade Federal do Paraná, Setor de Ciências Sociais Aplicadas, Programa de Pós-Graduação em Ciência, Gestão e Tecnologia da Informação.
Orientador: José Simão de Paula Pinto

1. Banco de dados -- Medidas de segurança. I. Pinto, José Simão de Paula. II. Título.

CDD 658.478

TERMO DE APROVAÇÃO

ANTONIO RODRIGUES BARROS

"APLICAÇÃO DE INTEGRAÇÃO DE BANCO DE DADOS VISANDO CONTROLE DE ACESSO CENTRALIZADO: UMA ALTERNATIVA DE *SINGLE SIGN-ON* PARA INSTITUIÇÕES FEDERAIS DE ENSINO SUPERIOR"

DISSERTAÇÃO APROVADA COMO REQUISITO PARCIAL PARA OBTENÇÃO DO GRAU DE MESTRE NO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA, GESTÃO E TECNOLOGIA DA INFORMAÇÃO DA UNIVERSIDADE FEDERAL DO PARANÁ, PELA SEGUINTE BANCA EXAMINADORA:

Prof. Dr. José Simão de Paula Pinto
(Orientador/UFPR)

Prof. Dr. Edelvino Razzolini Filho
(Examinador/UFPR)

Prof. Dr. Mauro José Belli
(Examinador/UFPR)

Prof. Dr. Marcos Sfair Sunyé
(Examinador/UFPR)

Curitiba, 21 de dezembro de 2010

À minha esposa Marcela que esteve
sempre ao meu lado me incentivando
e apoiando incondicionalmente.
Aos meus pais João e Lourdes (*in memoriam*),
por todo carinho e força que recebi,
ao longo da vida.

AGRADECIMENTOS

A Deus pela vida e ajuda, nos momentos de dificuldades.

Ao Professor José Simão de Paula Pinto, pela orientação, apoio, incentivo e principalmente pelo companheirismo.

Aos amigos e colegas do Centro de Computação Eletrônica (CCE), em especial aos da Divisão de Sistemas de Informação.

Ao servidor Eduardo Renan Manika, pela preciosa ajuda nesta pesquisa.

Ao diretor do CCE, Édson Flávio de Souza, pela compreensão e apoio.

RESUMO

O principal propósito de integração de base de dados, bens importantes para qualquer organização, é proporcionar a visão integrada dos mesmos. A integração, aliada aos conceitos de segurança da informação, fornece rico instrumental para gestão de identidade, que se configura como base para os processos de controle de acesso. Com foco nesses aspectos aspectos, a pesquisa é desenvolvida, introduz os principais conceitos e arquiteturas de integração de banco de dados heterogêneos, discorre sobre questões de segurança da informação, controle de acesso e gestão de identidade. Apresenta como proposta um modelo de controle de acesso centralizado, tendo como suporte procedimentos de integração de base de dados, mecanismos de segurança da informação e gestão de identidades. A proposta foi viabilizada por meio da implementação de um protótipo de autenticação centralizada baseado em *single sign-on* e integração de três bases de dados institucionais para compor um diretório de contas de usuários. Para a implementação e testes foi utilizada a infraestrutura de autenticação Shibboleth, possibilitando a simulação de três aplicações distintas, compartilhando o mesmo ambiente seguro de autenticação. Por fim, apresenta uma proposta de gestão de identidades, elaborada com base em dois estudos exploratórios, realizados na Universidade Federal do Paraná, no ano de 2010, um para analisar o perfil dos usuários em relação ao uso de conceitos de segurança da informação, e outro para identificar possíveis falhas nos processos de controle de acesso dos principais sistemas de informação e serviços de comunicação.

Palavras-chave: Integração de banco de dados. Controle de acesso. Segurança da informação. *Single Sign-On*.

ABSTRACT

Database integration, coupled with the concepts of information security, provides rich instrumental in identity management, which is configured as a basis for access control processes. Focusing on these aspects aspects, research is developed, introducing the main concepts and architectures for integrating heterogeneous database, discusses about information security, access control and identity management. Presented as a proposed model of centralized access control, supported by database integration, information security and identity management. To validate the proposal was implemented a prototype based on centralized authentication single sign-on. For the implementation and testing was used to Shibboleth authentication infrastructure, enabling the simulation of three different applications sharing the same secure environment for authentication. Finally, it presents a proposal for identity management, based on two exploratory studies, conducted at the Universidade Federal do Paraná, in 2010, one to analyse the profile of users regarding the use of concepts of information security, and another to identify potential gaps in access control.

Key words: Database integration. Access control. Information security. Single Sign-On.

LISTA DE FIGURAS

LISTA DE QUADROS

LISTA DE ABREVIATURAS

ABNT	- Associação Brasileira de Normas Técnicas
ACM	*- Association for Computing Machinery*
ACS	- Assessoria de Comunicação Social
AOL	- America OnLine
CAFe	- Comunidade Acadêmica Federada
CAS	- *Central Althentication Service*
CCE	- Centro de Computação Eletrônica
CIPEAD	- Coordenadoria de Integração de Políticas de Educação a Distância
DBA	*- Database Administration*
DECIGI	- Departamento de Ciência e Gestão da Informação
EID	*- Export Import Data Tool*
EID2LDAP	*- Export Import Data Tool to LDAP*
GSI	- Gabinete de Segurança Institucional
GSI/PR	- Gabinete de Segurança Institucional da Presidência da República
HTML	*- Hiper Text Mark-Up Language*
ICA	*- Interschema Correspondence Assertion*
IDP	*- Identity Provider*
IEEE	- I*nstitute of Electrical and Electronics Engineers*
JOSSO	*- Java Open Single Sign-On*
LDAP	*- Lightweight Directory Access Protocol*
MEC	- Ministério da Educação e Cultura
NBR	- Norma Brasileira
PPCGI	- Programa de Pós-Graduação em Ciência, Gestão e Tecnologia da Informação
PROPLAN	- Pró-Reitoria de Planejamento, Orçamento e Finanças
PSI	- Política de Segurança da Informação
RFC	*- Request for Comments*

RNP	- Rede Nacional de Ensino e Pesquisa
SAP	- Sistema de Administração de Patrimônio
SGBD	- Sistema Gerenciador de Banco de Dados
SI	- Segurança da Informação
SIBI	- Sistema de Bibliotecas
SIE	- Sistema de Informação para o Ensino
SIGEPE	- Sistema Integrado de Gestão de Pessoal
SLO	- *Single Log-Out*
SP	- *Service Provider*
SQL	- *Structured Query Language*
TCU	- Tribunal de Contas da União
TI	- Tecnologia da Informação
UFBA	- Universidade Federal da Bahia
UFMG	- Universidade Federal de Minas Gerais
UFPR	- Universidade Federal do Paraná
UFRGS	- Universidade Federal do Rio Grande do Sul
VOIP	- *Voice Over Internet Protocol*
XML	- *Extensible Markup Language*

SUMÁRIO

1 INTRODUÇÃO

1.1 PROBLEMÁTICA

Analisando alguns serviços de tecnologia da informação e comunicação, disponibilizados pela Universidade Federal do Paraná (UFPR) aos seus usuários: alunos, docentes, técnicos e terceiros, nota-se que os dados são armazenados em bancos de dados distintos e isolados, ou seja, para cada sistema de informação ou serviço de comunicação há um banco de dados correspondente para armazenamento dos seus respectivos dados. Como consequência dessa distribuição de dados, é possível identificar redundância de dados, uma vez que cada sistema funciona de forma isolada e independente, possuindo mecanismos próprios de cadastro e armazenamento de dados (quadro 1, terceira e quarta coluna) relativos a pessoas, usuários e vínculos.

Nº	Sistema de informação / Serviço	Armazenamento de dados	Qtde de usuários	(%) usuários sem vínculo	Fluxo de autenticação (uso de criptografia)
1	Sistema Acadêmico (SIE)	IBM DB2	86466	45,00%	NÃO
2	Sistema de Gestão de Pessoas (SIGEPE)	PostgreSQL	84	0,00%	SIM
3	Serviço de Correio Eletrônico e Serviço de Proxy	MySQL	9827	18,00%	SIM
4	Serviço de Voz Sobre Internet (VOIP)	LDAP	167	34,00%	NÃO
5	Serviço de Terminal Remoto	Active Directory	***	***	NÃO
6	Sistema de Administração Patrimonial (SAP)	Firebird	535	27,00%	NÃO
7	Sistema de Registro de Pesquisa (THALES)	Oracle	2311	23,00%	NÃO
8	Plataforma de Educação a Distância (MOODLE - UFPR)	PostgreSQL	3491	***	NÃO
9	Plataforma de Educação a Distância (MOODLE - CIPEAD)	PostgreSQL	***	***	NÃO
10	Sistema de Bibliotecas (SOPHIA)	MSSQL Server	***	***	NÃO
*** Não foi possível obter dados referente ao quantitativo de usuário e, portanto, não foi possível estabelecer o percentual de usuários sem vínculo formal					

QUADRO 1 - INFRAESTRUTURA DE SISTEMAS DE INFORMAÇÃO E SERVIÇOS DE COMUNICAÇÃO NA UFPR
FONTE: O autor (2010)

O quadro 1, elaborado com base em pesquisa exploratória, realizada entre os meses de julho a setembro de 2010, tendo como base controles propostos pela

norma NBR ISO/IEC 27002 de 2005, apresenta o cenário da instituição pesquisada. Uma miscelânea de sistemas de informação, serviços de comunicação e bases de dados heterogêneas. Cada repositório de dados (quadro 1, terceira coluna), representa também o respectivo mecanismo de controle de acesso.

De maneira geral, essa independência de mecanismos de controle de acesso acarreta problemas de inconsistência e redundância de dados, visto que, cinco dos dez serviços pesquisados apresentam percentuais (quadro 1, quinta coluna) bastante significativos de usuários com potencial de acesso, mesmo sem vínculo formal com a instituição. Esses percentuais representam ausência ou ineficiência de mecanismos de controle para registro de usuários, o que se configura como problemas de segurança para o ambiente de tecnologia da informação.

Outro aspecto relevante, também relacionado à questão de segurança, diz respeito aos fluxos de autenticação. Um fluxo de autenticação ocorre quando um usuário requisita acesso a uma funcionalidade protegida (de acesso restrito) de um sistema ou serviço de tecnologia da informação, que está disponibilizado em ambiente de rede (figura 1). Dos dez recursos contidos no quadro 1 oito deles não utilizam mecanismos para proteger o fluxo de autenticação (sexta coluna), de modo que os dados sensíveis, nomes e senhas de usuários, são transmitidos de forma desprotegida.

A figura 1 apresenta, de modo simplificado, o fluxo de autenticação para acesso à funcionalidade Alterar.Senha. As setas representam o tráfego de dados entre usuários, representado por um aplicativo instalado no computador local do usuário, e um Sistema disponibilizado na rede.

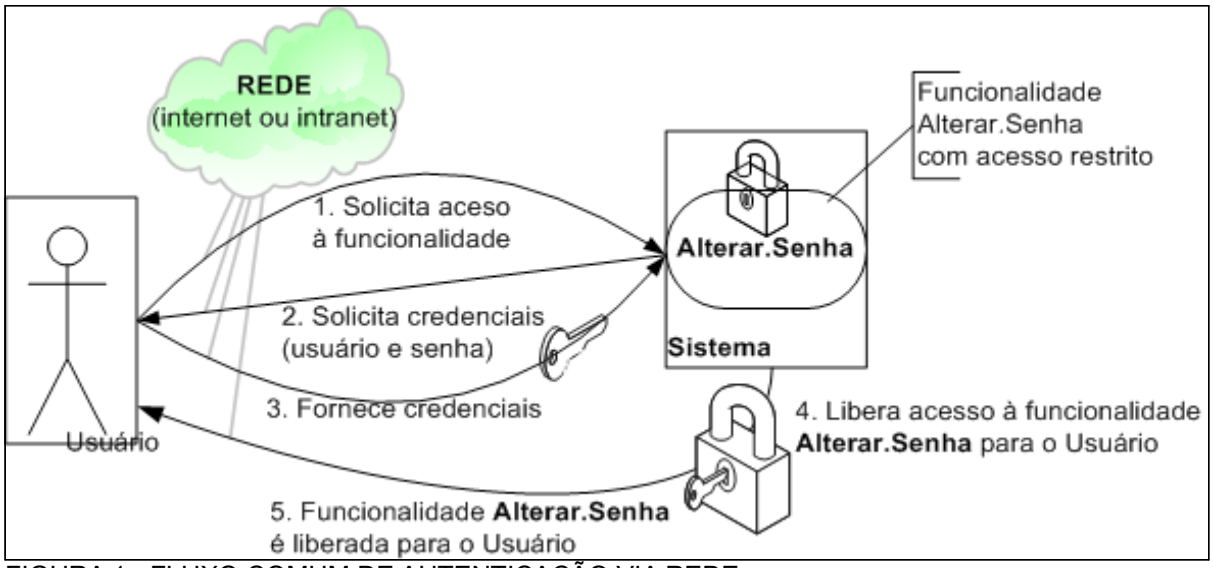

FIGURA 1 - FLUXO COMUM DE AUTENTICAÇÃO VIA REDE
FONTE: O autor (2010)

Conforme apresentado o ambiente, que inclui questões de heterogeneidade de bancos de dados, bancos de dados isolados, controles de acessos isolados e indícios de problemas de segurança da informação. **Como é possível a implementação de controle de acesso centralizado fortemente apoiado em conceitos de segurança da informação?**

1.2 OBJETIVOS

A figura 2 representa um modelo conceitual de bases de dados unificados para prover um meio de controle de acesso centralizado, de forma que o usuário ao requisitar acesso a um provedor de serviços (SP), tal como o portal do aluno, por exemplo, imediatamente é redirecionado ao provedor de identidades (IDP), que efetua a validação verificando na base unificada (Diretório), se este possui acesso ou não. Caso o usuário tenha acesso, então a autorização é concedida para o recurso requisitado.

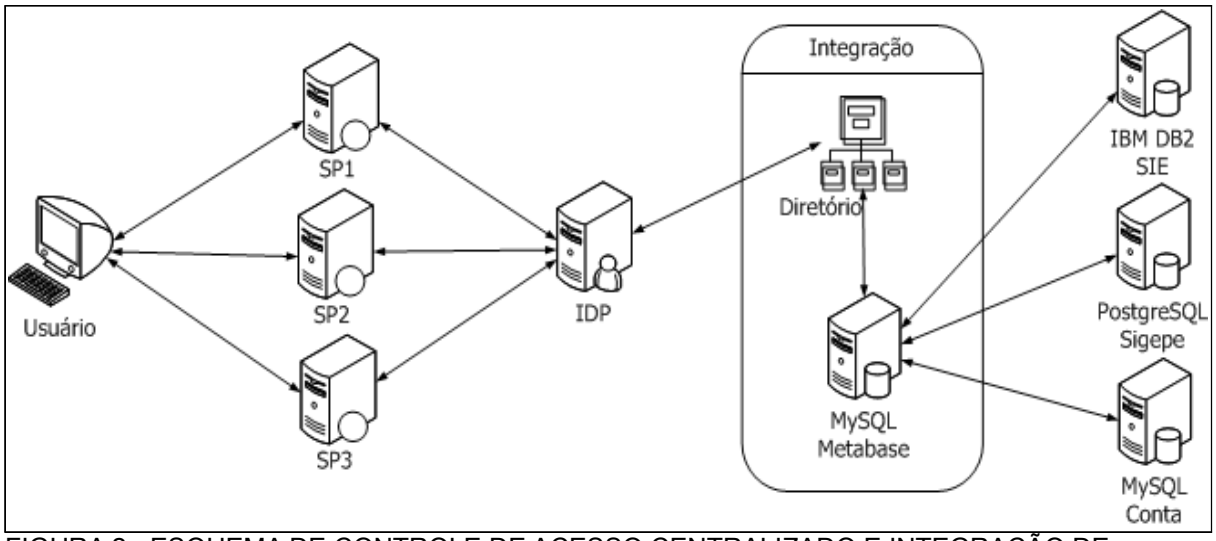

FIGURA 2 - ESQUEMA DE CONTROLE DE ACESSO CENTRALIZADO E INTEGRAÇÃO DE BANCO DE DADOS
FONTE: O autor (2010)

A proposta limita-se à integração de três bancos de dados relacionais, que representam respectivamente dados pessoais de alunos, dados pessoais de docentes, técnicos e dados de contas de correio eletrônico.

Apresentado o contexto, chega-se ao objetivo deste trabalho, que foi desdobrado em objetivo geral e objetivos específicos, a seguir tratados.

1.3 OBJETIVO GERAL

Viabilizar processo de controle de acesso centralizado, baseado em *Single Sign-On* por meio de integração de bases de dados pessoais, relativos à comunidade acadêmica[1] da Universidade Federal do Paraná.

1.3.1 Objetivos específicos

1. levantar os principais sistemas de informação e serviços disponibilizados pela instituição, quanto ao uso de segurança da informação, principalmente na questão de controle de acesso lógico;
2. analisar o perfil dos usuários quanto ao uso de senhas;
3. propor a integração de bases de dados para construir metabase de identidades digitais;
4. implementar projeto piloto de *Single Sign-On,* levando-se em consideração os requisitos de segurança sugeridos pela ABNT NBR ISO/IEC 27002:2005;

1.4 JUSTIFICATIVA

Dentre as principais motivações para desenvolvimento da pesquisa destacam-se: a legislação vigente sobre a questão de segurança da informação e a necessidade de se ter uma base única de usuários.

Segue o detalhamento das justificativas:

1. Regulamentação do Tribunal de Contas da União (TCU) e do Gabinete de Segurança Institucional (GSI), para uso de normas e padrões de gestão de segurança da informação tais como:
   a) Decreto Nº 4.553, de 27 de dezembro de 2002, que trata a salvaguarda dos dados, documentos e materiais sigilosos no âmbito da Administração Pública Federal;
   b) Guia de boas práticas em segurança da informação, elaborado pelo TCU, em 2008, que trata, em linhas gerais, como proceder em questões

[1] Comunidade acadêmica é composta de pessoas que possui vínculo formal com a instituição de ensino pesquisada. Alunos, professores, técnicos administrativos e funcionários terceirizados são exemplos de pessoas que compõem a comunidade acadêmica.

relacionadas ao ambiente informatizado;

c) Instrução Normativa GSI/PR Nº 1, de 13 de junho de 2008, cujo objetivo é disciplinar a Gestão da Segurança da Informação e Comunicações no âmbito da Administração Pública Federal; entre outros documentos oficiais;

2. Gerenciamento de uma base de dados única e centralizada de contas de usuários é mais fácil de administrar e manter do que várias bases;
3. Possibilidade de implementação de normativas mais rigorosas quanto ao uso de senhas de acesso e gerenciamento de contas de usuário, por se tratar de identificação única de usuário;
4. Questões de segurança como: gerenciamento de privilégios, registro de usuários, controle de acesso e auditoria são facilitados em decorrência da integração das bases e autenticação centralizada;
5. Adequação ao projeto Comunidade Acadêmica Federada (CAFe[2]), projeto proposto pela Rede Nacional de Ensino e Pesquisa (RNP), cujo objetivo é integrar instituições de ensino e pesquisa Brasileiras em uma rede de confiança (RNP, 2009);
6. Necessidade de maior gestão dos recursos disponibilizados pela instituição, ou seja, saber quantos e quais são os indivíduos possuem acesso a determinado recurso;
7. Redução e organização do trabalho, visto que por trás da duplicação de cadastros há duplicação de esforços, tendo como consequência a má distribuição dos recursos financeiros, destinados ao pagamento de pessoal e consultorias externas para manutenção dos cadastros duplicados;
8. Outra questão está relacionada ao sigilo das informações depositadas em sites e sistemas de informação, pois nem sempre recebem o tratamento adequado sob o aspecto de segurança da informação, permitindo que haja incidentes de segurança, tais como acessos não autorizados.

---

[2] Comunidade acadêmica é composta de pessoas que possui vínculo formal com a instituição de ensino pesquisada. Alunos, professores, técnicos administrativos e funcionários terceirizados são exemplos de pessoas que compõem a comunidade acadêmica.

1.5 ORGANIZAÇÃO DO TEXTO

No Capítulo 2, a fundamentação teórico conceitual é apresentada, abordando os principais temas relacionados à pesquisa. Inicialmente são apresentados os conceitos de integração de banco de dados heterogêneos, com fundamentação nos principais autores da literatura do assunto. Descreve de maneira breve quais os fatores que influenciam a heterogeneidade das bases de dados, fases de integração. Aborda, ainda, a integração de esquemas e por fim as arquiteturas de integração. Na sequência, são apresentados os conceitos de segurança da informação e legislação vigente no Brasil sobre o tema. Trata também da gestão de risco e conteúdos correlatos. Por fim é apresentado o controle de acesso, uma das medidas mais importantes e difíceis de serem tratadas para a proteção da informação.

No Capítulo 3, discorre-se sobre a metodologia, apresentando os principais subsídios da pesquisa.

No Capítulo 4, são apresentados os resultados e discussões de duas pesquisas exploratórias, uma com foco na identificação do perfil do usuário e sistemas de informação e outra específica para sistemas, com objetivo de identificar as principais discordâncias em relação à norma NBR ISO/IEC 27002. Apresenta um protótipo conceitual de integração de base de dados, baseado no conceito de materialização de visões. Discute gestão de identidade, apresentando como proposta o kit de provisionamento de usuários. Por fim, apresenta a implementação da proposta da pesquisa, um protótipo funcional de *Single Sign-On*, baseado na ferramenta Shibboleth.

Finalmente, no Capítulo 5, são descritas as considerações a respeito das contribuições da pesquisa para a instituição, dificuldades encontradas, perspectivas de trabalhos futuros e conclusão.

2 REFERENCIAL TEÓRICO

2.1 INTEGRAÇÃO DE BANCO DE DADOS

O processo de integração de banco de dados consiste na junção de várias fontes de dados para gerar uma visão integrada dos dados, independente de onde e como estão dispostas as fontes de dados. É uma tarefa complexa e que torna difícil criar um processo totalmente automatizado para realizá-lo em sua totalidade.

Batini, Lenzerini e Navathe (1986) descrevem que os principais problemas a serem considerados são de ordem estrutural e semântica, portanto é necessário considerar a ação humana para resolução de conflitos no projeto de unificação.

Sheth e Larson (1990) introduzem o conceito de unificação de esquemas como uma forma de integrar vários bancos de dados, formando uma federação. Esta abordagem permite que haja unificação e, ao mesmo tempo, permite que as bases de dados permaneçam autônomas.

O principal benefício do processo de integração está no fato do usuário saber qual dado escolher ao invés de concentrar os esforços em como obter. Por outro lado, uma das maiores dificuldades envolvida no processo de integração de base de dados está concentrada na resolução de conflitos. Isso ocorre devido à heterogeneidade das fontes de dados envolvidas no projeto.

2.1.1 Fatores que Influenciam a Integração de Bancos de Dados

Hasselbring (2000) aborda, genericamente, a integração, destacando três dimensões, parâmetros importantes que devem ser observados num processo de integração: autonomia, heterogeneidade e distribuição. Essas dimensões são detalhadas a seguir, tomando como base os trabalhos de Hasselbring (2000) e Tatbul *et al.* (2001).

2.1.1.1 Autonomia

A autonomia é um fator bastante crítico em um processo de integração, devido à individualidade de cada sistema. É bastante comum identificar conflitos entre os requisitos de integração. Quanto à autonomia das bases de dados, pode-se

classificá-las em quatro:

1. Autonomia de projeto – normalmente as fontes de dados são criadas anteriormente a um processo de integração. Portanto podem ocorrer grandes diferenças entre as fontes a serem integradas. Outro fator é a forma de percepção que cada indivíduo possui sobre determinada entidade do mundo real. Na figura 3, é possível observar dois projetos de bases de dados (Base de dados “A” e Base de dados “B), modelados de formas diferentes. No entanto trata-se das mesmas questões do mundo real;

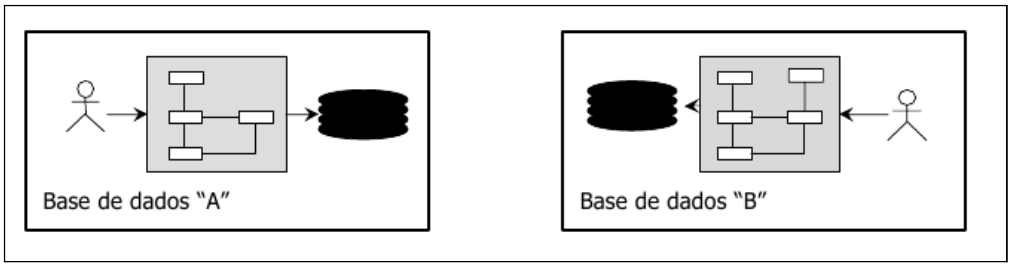

FIGURA 3 - AUTONOMIA DE PROJETO
FONTE: KROTH (2006)

2. Autonomia de comunicação Autonomia de comunicação – as fontes de dados possuem autonomia para responder as requisições no momento que for conveniente. Por exemplo, um banco de dados pode estar configurado para processar requisições apenas das 00h30min às 6h da manhã. Outra situação pode ocorrer, quando o banco de dados estiver configurado para retornar como respostas apenas um formato específico de arquivo (figura 4);

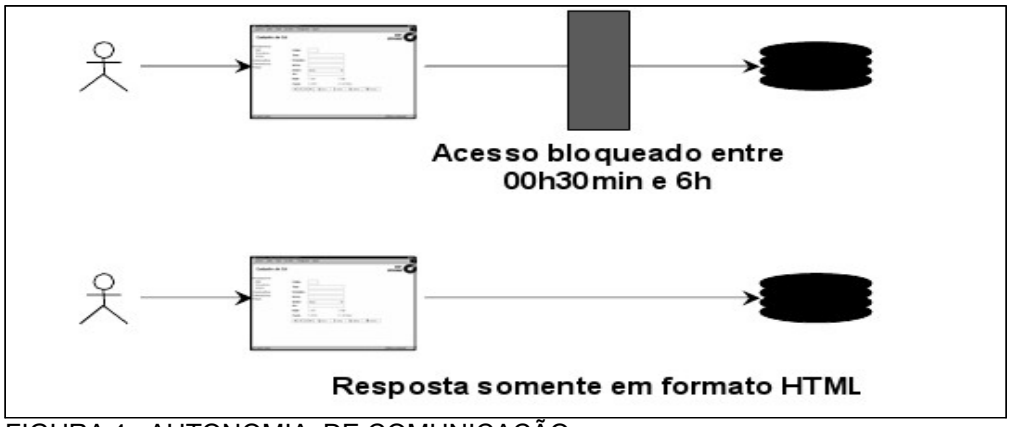

FIGURA 4 - AUTONOMIA DE COMUNICAÇÃO
FONTE: Adaptado de KROTH (2006)

3. Autonomia de execução – independente da ordem das requisições, o banco possui sua ordem interna (figura 5);

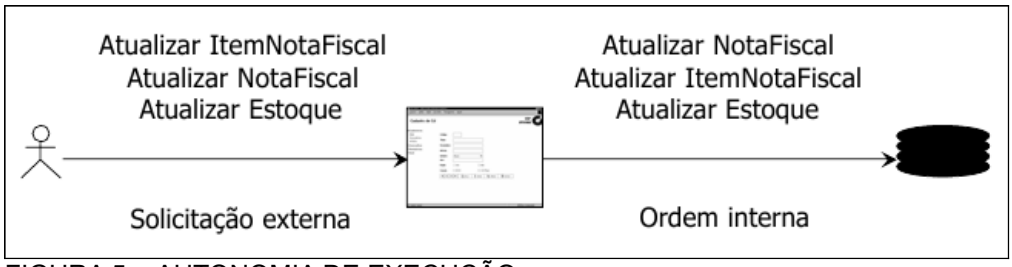

FIGURA 5 - AUTONOMIA DE EXECUÇÃO
FONTE: KROTH (2006)

2.1.1.2 Heterogeneidade

Em banco de dados há diversas formas de heterogeneidade, como, por exemplo, dispositivos e equipamentos, plataformas diferentes, modelos de dados diferentes, linguagens de manipulação de dados diferentes etc. Um dos principais elementos que motivam a heterogeneidade de um sistema de informação ou banco de dados é a independência que este possui em relação aos seus similares. Este assunto é seguramente o que gera maior dificuldade num processo de integração de base de dados. Para melhor entendimento, este assunto é dividido em três níveis principais:

1. Nível semântico – as fontes de dados podem ser projetadas com diferentes visões do que se pretende representar. Por exemplo, em uma modelagem de dados pode-se optar por armazenar os dados de uma pessoa em uma única tabela de banco de dados, enquanto os de outra podem ser armazenados em duas ou mais tabelas distintas;
2. Nível de projeto – as fontes de dados podem ser modeladas usando diferentes paradigmas, de forma que em determinada modelagem é utilizado banco de dados relacional, enquanto em outra se pode utilizar banco de dados orientado a objetos. Ambas possuem o propósito de resolver o mesmo problema, contudo as formas de projeto são diferentes;
3. Nível de implementação – está relacionado aos diferentes tipos de dados utilizados para representar um mesmo atributo. Exemplo deste nível de heterogeneidade é observado em alguns modelos de dados que utilizam caracteres numéricos para armazenar a identificação de um indivíduo, enquanto em outros são utilizados caracteres alfabéticos.

2.1.1.3 Distribuição

Distribuição está relacionada à arquitetura física em que as fontes de dados estão dispostas. Um exemplo é quando existem sistemas legados que passaram por processo de reformulação ou integração, mesmo estando em locais fisicamente distintos encontram-se ligados através de rede e compartilham recursos por meio de procedimentos RPC (*Remote Procedure Call* ou Procedimentos de Acesso Remoto).

O grau de complexidade do processo de integração de base de dados está diretamente relacionado aos fatores citados nos itens 2.1.1.1, 2.1.1.2, 2.1.1.3, ou seja, quanto mais autônomas, heterogêneas e distribuídas, mais complexo e difícil é o processo de integração.

2.1.2 Fases da Integração

As fases descritas pelos autores Batini, Lenzerini e Navathe (1986), cujo trabalho foi um estudo comparativo de metodologias de integração, são: pré-integração, comparação de esquemas, adequação de esquemas e união e reestruturação. Os autores ainda afirmam que as metodologias estudadas por eles seguem individualmente suas particularidades, variando em quantidade e nomenclatura de fases, no entanto todas estão inseridas no contexto das quatro fases.

Zamboni (2004) faz um apanhado acerca das fases propostas por Batini, Lenzerini e Navathe (1986) e Parent e Spaccapietra (1998) para geração de uma visão nova e mais completa e mais adequada ao trabalho de Zamboni.

As fases elencadas por Zamboni são:

1. considerações iniciais – procura-se a uniformidade dos esquemas a serem integrados;
2. pré-integração – escolha e análise dos esquemas a serem integrados e escolha de política eficiente de integração;
3. identificação de correspondências – correspondências entre os esquemas selecionados são identificadas e descritas;
4. identificação de conflitos – análise e comparação entre os esquemas para determinar as correspondências entre conceitos e detectar possíveis conflitos;

5. integração – os conflitos entre os esquemas são resolvidos e os itens correspondentes são unificados;
6. união e reestruturação - geração de esquemas intermediários para análise e, se necessário, reestruturação para que alcance as qualidades desejáveis: a) completude e corretude, o esquema integrado deve apresentar de forma correta todos os dados contidos nos esquemas fontes; b) minimalidade, eliminação de redundância, de forma que, quando há dados repetidos em mais de uma fonte, estes dados apareçam apenas uma vez; c) e compreensão, o esquema final, que representa a visão dos dados integrados, deve ser fácil de entender, tanto pelos administradores quanto pelos usuários finais.

2.1.2.1 Considerações iniciais

É fundamental que haja uma visualização comum das funcionalidades das fontes de dados a serem integradas, tornando-as mais homogêneas para consultas futuras. De forma geral, esta fase busca o entendimento e a transformação dos esquemas a serem integrados para que se tornem sintática e semanticamente homogêneos.

Uma vez definido o CDM (*Common Data Model* ou Modelo Comum de Dados), em geral, definido por um especialista de banco de dados, DBA (*Database Administrator* ou Administrador de Banco de Dados), o próximo passo é a tradução dos esquemas de entrada para o CDM. Há muitas questões envolvidas na escolha de um CDM. Segundo Parent e Spaccapietra (1998), a maioria dos pesquisadores é favorável ao modelo orientado a objetos, visto que contempla todos os conceitos dos outros modelos de dados e seus métodos podem ser utilizados para implementar as regras específicas de mapeamentos. No entanto, deve ser levada em consideração a escolha de um modelo que simplifique a implementação, posto que quanto mais rico o modelo, mais complexo fica o processo de integração.

Uma medida bastante oportuna para simplificar o processo de integração é a escolha de um CDM simples, ou seja, com o mínimo de semântica.

2.1.2.2 Pré-Integração

Nesta fase é feita a delimitação da integração através da análise de cada esquema a ser integrado. A investigação está concentrada em definir quais esquemas serão integrados e em que nível, o qual está diretamente relacionado à proporcionalidade dos esquemas, ou seja, se a integração será parcial ou total.

Na pré-integração são estabelecidas as correspondências entre os esquemas sob a forma de relacionamentos e declarações entre os componentes dos esquemas envolvidos. Nessas correspondências pode ser identificado, por exemplo, que um objeto em um esquema é resultado de algumas operações de um conjunto de outros objetos em outro esquema.

Independente da metodologia utilizada, todas contemplam esta fase, mesmo que implicitamente, considerando a sequência e o agrupamento das fontes de entrada.

Outro ponto que deve ser definido nessa fase é a estratégia de integração, com objetivo de otimizar o processo e ao mesmo tempo reduzir sua complexidade. Em geral, adota-se um processamento sequencial para os componentes dos esquemas.

A figura 6 apresenta quatro estratégias possíveis para processamento da integração das fontes de dados dispostas em forma de árvore. Os nós da folha correspondem aos esquemas da fonte, os nós não folha correspondem aos resultados intermediários da integração e o nó raiz é o resultado final da integração.

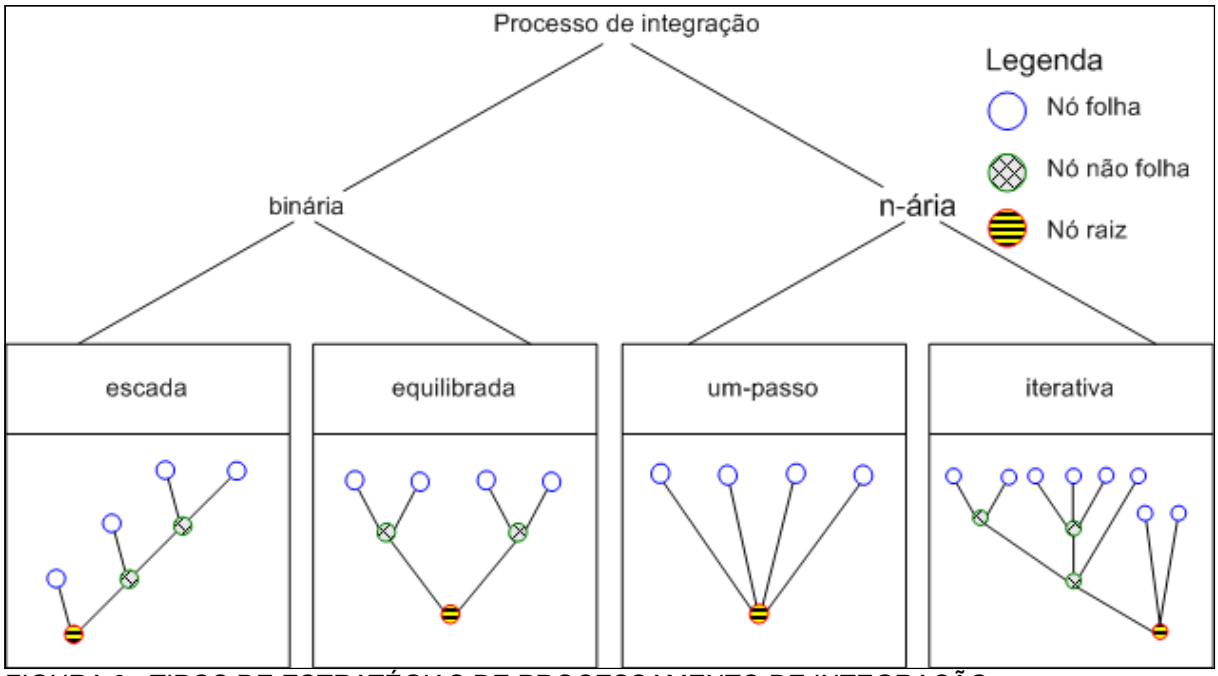

FIGURA 6 - TIPOS DE ESTRATÉGIAS DE PROCESSAMENTO DE INTEGRAÇÃO
FONTE: Adaptado de BATINI, LENZERINI e NAVATHE (1986)

As estratégias (figura 6) podem ser classificadas em binárias e n-árias. As binárias permitem que os esquemas possam ser integrados aos pares, as binárias podem ser em escada (*ladder*) ou equilibrada (*balanced*). A estratégia em forma de escada permite que cada esquema fonte seja integrado juntamente com um estado intermediário, naturalmente partindo de dois esquemas fonte. Já a estratégia em forma dita equilibrada requer que as fontes de dados sejam divididas em pares, para que a integração ocorra simetricamente. As estratégias classificadas como n-árias podem ser um-passo (*one-shot*) ou iterativa (*iterative*) e permitem a integração de n esquemas de uma só vez (sendo n > 2). A estratégia um-passo (*one-shot*) ocorre, quando todos os esquemas a serem integrados ocorrem simultaneamente. Já a iterativa pode ser considerada como uma combinação de todas as outras estratégias, por ser a mais flexível (BATINI, LENZERINI e NAVATHE, 1986).

A maior vantagem das estratégias binárias está na simplificação das atividades de comparação entre os esquemas, dividindo-os em vários passos e em cada passo analisando apenas dois esquemas. Por essa razão as estratégias binárias são mais utilizadas, visto que pesquisas indicam que a complexidade da comparação entre n esquemas opera na ordem de n^2. Por outro lado, a maior desvantagem dessa estratégia está no aumento do número de operações de integração para chegar à análise final.

A estratégia n-ária, que realiza análise de n esquemas de uma só vez, permite o aumento considerável na análise semântica. Esta pode ser analisada, antes da união dos esquemas de forma que evita transformações futuras nos esquemas integrados. Essa é uma das grandes características dessa estratégia, além de diminuir a quantidade de passos para integração, porém é a mais complexa.

2.1.2.3 Identificação de correspondências

A identificação de correspondências vai além do que está sendo representado no modelo a ser integrado, mas como foi representado cada objeto do mundo real.

Essa correspondência é tida como válida, quando dois objetos estão dispostos em dois esquemas distintos, que possuem descrições iguais ou semelhantes e representam o mesmo fato do mundo real. Portanto a identificação de correspondência vai além do que está descrito em um modelo de dados, para uma visão de mais de um significado.

Parent e Spaccapietra (1998) introduzem o conceito de ICA (*Interschema Correspondence Assertion* ou Afirmações de Correspondência entre Esquemas), que indica o grau de correspondência entre esquemas. De modo geral, um ICA descreve quando e como dois objetos são equivalentes.

Para definir completamente um ICA, deve-se levar em consideração quatro passos:

1. identificar o conjunto de elementos de cada esquema fonte que representa o mesmo objeto do mundo real;
2. definir que tipo de relacionamento entre os elementos que representam o mesmo objeto do mundo real. Os relacionamentos podem ser descritos como: equivalência (≡), inclusão (⊇), interseção (∩) ou disjunção (≠);
3. mapear as instâncias correspondentes de cada ICA com a fonte correspondente, de forma que quando solicitado determinado dado o mecanismo saiba exatamente onde buscá-lo;
4. finalmente, identificar os atributos que estão presentes em ambos conjuntos de elementos de cada fonte de dados, com objetivo de evitar redundância no esquema integrado.

2.1.2.4 Identificação de conflitos

A ideia principal desta fase, como o próprio nome sugere, é identificar todos os conflitos contidos nos esquemas a serem integrados.

A principal causa de conflitos em bases de dados reside no fato de que pessoas diferentes possuem percepções diferentes acerca de um determinado problema, de forma que uma pessoa pode representar um objeto do mundo real de forma completamente diferente de outra, assim com descrevê-lo também de maneira diferente.

A classificação que segue leva em consideração as categorizações de Sheth e Kashyap (1993), Kim e Seo (1991), e contribuições dos trabalhos acadêmicos de Rodacki (2000) e Zamboni (2004).

Os conflitos estruturais estão dispostos em duas classes: incompatibilidade de definições de domínio e incompatibilidade de definições de entidades. Os relativos a domínios estão relacionados a problemas de nomes (sinônimos e homônimos), incompatibilidades de tipos, unidades, precisão, valor padrão e regras de integridade de dados. Os relacionados à definição de entidades estão ligados a questões estruturais tais como equivalências de chaves, compatibilidade de uniões, isomorfismo dos esquemas e falta de atributos (SHETH e KASHYAP, 1993). A descrição detalhada dos tipos de conflitos segue:

1. Sinônimos – quando um conjunto de objetos equivalentes é descrito de forma diferente. Por exemplo, uma entidade nomeada como Professor pode ser representada em outro esquema como Docente, no entanto representam o mesmo objeto do mundo real.
2. Homônimos – ao contrário dos conflitos de sinônimos, os homônimos ocorrem quando dois elementos semanticamente diferentes possuem o mesmo nome. Por exemplo, uma entidade nomeada como Estado pode representar uma unidade da Federação, no caso do Brasil, já em outro esquema pode representar o estado de um processo como ativo, inativo, neste caso com sentido de *status*.
3. Tipo de dados – ocorrem quando dois elementos equivalentes possuem tipos diferentes. Por exemplo, um esquema possui a entidade Pessoa, que possui

o atributo matrícula como numérico, já em outro esquema este mesmo atributo pode ser representado por um atributo como caracteres alfa numérico.

4. Unidade – ocorre, quando os atributos armazenam unidades de medidas diferentes. Por exemplo, os esquemas possuem um elemento que armazena valores monetários, em um esquema os valores são representados em Real, moeda Brasileira, já em outro a moeda pode ser em Dólar, moeda Americana.
5. Precisão – ocorre quando elementos semanticamente parecidos são representados por diferentes unidades e medidas. Por exemplo, em um esquema o atributo nota é representado por um conjunto numérico e pode variar de zero a cem, já em outro, o atributo pode ser conceito representado pelo conjunto {A, B, C, D, E}.
6. Valor padrão – ocorre quando são definidos valores de padrão em diferentes esquemas. Por exemplo, em uma entidade Funcionário, o atributo vínculo pode ter seu valor padrão como sendo “efetivo” ou “terceirizado”, já em outro esquema, a mesma entidade pode ter o valor padrão nulo, caso o funcionário seja “efetivo”.
7. Equivalência de chaves – ocorre quando entidades semanticamente parecidas possuem identificadores semanticamente diferentes. Por exemplo, a entidade Funcionário de um esquema possui seu atributo identificador (chave primária) com sendo o CPF, já em outro esquema, a entidade Funcionário possui o atributo matrícula como sendo seu atributo chave.
8. Incompatibilidade união – ocorre quando duas entidades semanticamente parecidas são representadas com número diferente de atributos ou atributos não relacionados, de forma que a união irá resultar em uma entidade diferente e discrepante das fontes originais. Por exemplo, uma entidade Aluno do esquema “a” possui os atributos matrícula, nome e idade; a outra, do esquema “b”, possui os atributos: matrícula, nome e CPF. Ocorrerá um problema, quando aplicada uma operação de união, visto que o atributo idade do esquema “a” nunca existiu no esquema “b”, da mesma forma que ocorre no esquema “b” para “a” com relação ao atributo CPF.
9. Isomorfismo – ocorre quando entidades semanticamente parecidas possuem

número de atributos diferentes para representar conceitos similares. Por exemplo, a entidade Aluno possui os atributos: matrícula e nome, já outra possui matrícula, nome e sobrenome. Ambas possuem o propósito de armazenar o mesmo dado (nome do aluno), o que muda é apenas a estrutura.

10. Falta de atributo – ocorre quando entidades semanticamente semelhantes possuem número de atributos diferentes, ou seja, um atributo que está presente em uma entidade não está presente em outra. Este conflito foi abordado nos conflitos de incompatibilidade de união e isomorfismo.
11. Relações e atributos – ocorre quando uma entidade é modelada como atributo de um esquema e como atributo em outro esquema. Por exemplo, o esquema “a” possui a entidade Aluno que possui os atributos matrícula, nome, idade e orientador; o esquema “b” possui as entidades AlunoPG e Aluno, sendo que Aluno possui os atributos matrícula, nome e idade; e AlunoPG possui os atributos matrícula e orientador. Note que no esquema “a” todos os atributos de Aluno estão em uma única entidade enquanto no esquema “b” os atributos estão dispostos em duas entidades (Aluno e AlunoPG).
12. Atributos e dados – os conflitos entre atributos e dados ocorrem quando um valor de um atributo em um esquema corresponde a um atributo de outro esquema. Por exemplo, a entidade Aluno do esquema “a” possui os atributos matrícula, nome e titulação (titulação refere-se ao grau de escolaridade, graduação, mestrado e doutorado). Já a entidade Aluno do esquema “b” possui os atributos matrícula, nome, grad, mestr e dout, sendo os atributos (grad, mestr e dout) a titulação do aluno. O grau de escolaridade no esquema “a” está representado sob a forma de dados, já em “b” está como atributo.

2.1.2.5 Integração

Esta etapa, normalmente, é efetuada após a descrição das correspondências e resolução de conflitos. É comum, nesta fase, surgirem problemas, sejam relacionados ao mapeamento de correspondências ou a conflitos que não foram resolvidos de forma adequada.

Uma das formas de realização desta etapa é por meio do uso de linguagens procedurais, declarativas ou lógicas e, normalmente, é feito por um profissional da área de banco de dados, DBA, ou de forma automatizada, por meio de ferramentas destinadas para esse fim.

2.1.2.6 União e reestruturação

Neste ponto da integração normalmente já é apresentado um esquema integrado temporário, possivelmente um protótipo de banco de dados. Esta etapa é marcada pela união simples dos esquemas por sobreposição de conceitos comuns, sendo posteriormente realizadas operações de reestruturação no esquema obtido pela união.

Várias propriedades são aplicadas nesta fase, tais como subconjuntos, generalização, agrupamentos, união, categorização, etc, com objetivo de obter alguns quesitos de qualidade como:

1. completude e corretude – a base de dados integrada deve apresentar o mesmo quantitativo de dados das bases de dados fonte e estes dados devem ser isentos de erros;
2. minimalidade – quando um mesmo dado está presente em mais de uma base fonte e é apresentado na base integrada apenas uma vez;
3. clareza – diz respeito à facilidade de entendimento da base integrada.

Completude e corretude são alcançadas analisando-se o esquema integrado para verificar se todos os dados contidos nas fontes estão dispostos de maneira correta e completa no esquema integrado. É comum a completude ser deixada de lado visto que em muitos casos é esperado apenas um subconjunto dos dados integrados.

A minimalidade consiste na descoberta e eliminação de redundâncias, de forma que se um dado estiver presente nas várias fontes de dados integradas, este dado deve aparecer apenas uma vez no esquema integrado. Esse processo também é utilizado nos relacionamentos de forma a obter mais simplicidade e clareza no esquema integrado.

A clareza está diretamente ligada à compreensão que os usuários e administradores possuem acerca do esquema integrado.

2.1.3 Integração de esquemas

Esquema é a descrição de como os dados estão estruturados em um banco de dados. Em um banco de dados relacional, por exemplo, descrevem-se as tabelas, os relacionamentos e os atributos envolvidos, assim com em um banco de dado orientado a objetos, descrevem-se a definição das classes, atributos, métodos etc.

A integração de esquemas é um processo de integração de base de dados, e tem como partida um conjunto de esquemas, gerando ao final do processo uma descrição unificada, também chamada de esquema integrado ou esquema global, e um mapeamento de informações associadas que dão suporte de acesso aos dados existentes nos esquemas integrados (BATINI, LENZERINI e NAVATHE, 1986; PARENT e SPACCAPIETRA, 1998).

A integração de esquemas pode ser uma tarefa relativamente fácil ou extremamente difícil, dependendo das discrepâncias encontradas nas bases de dados a serem integradas. Quanto mais heterogêneas são, mais complexa é a tarefa de integrá-las.

Outro fator que está diretamente relacionado à complexidade é a dimensão total ou parcial da integração. Na integração parcial apenas parte dos objetos do esquema são incluídos no contexto da integração, por outro lado a integração total envolve todo esquema de banco de dados. Dessa forma pode representar um aumento de complexidade que, por vezes, pode inviabilizar o projeto de integração.

A principal vantagem dessa abordagem é que apresenta uma visão uniforme dos dados ao usuário, por outro lado apresenta uma série de problemas (SCOPIM, 2003):

1. o processo de integração requer intervenção do usuário para mapeamento dos esquemas para o esquema global, resolução de conflitos estruturais, semânticos e comportamentais, desta forma pode ser automatizado, parcialmente;
2. a autonomia dos bancos de dados integrantes é sacrificada para resolução de conflitos semânticos durante a criação do esquema global;
3. existem diversos métodos de integração para mais de dois bancos de dados:

combinação de dois a dois, combinação de todos juntos de uma só vez etc. Desta forma pode-se definir esquemas globais completamente diferentes;

4. enrijecimento do modelo, visto que a cada alteração no modelo local é necessário alteração no modelo global.

Scopim (2003) em seu trabalho acadêmico propõe uma ferramenta chamada J-SCHEMA INTEGRATOR, cujo principal objetivo é facilitar o processo de integração de esquemas em bases de dados relacionais. A ferramenta possui as seguintes funcionalidades: a) visualização dos esquemas de banco de dados; b) identificação dos conflitos; c) geração do esquema integrado a partir dos esquemas informados e dos conflitos apontados pelo usuário; e d) geração das informações de mapeamentos entre o esquema gerado e os esquemas iniciais.

Há inúmeros projetos na mesma linha de Scopim (2003), tais como AutoMed e TSIMMIS. AutoMed (Geração Automática de Ferramentas mediadoras para Integração de Bancos de Dados Heterogêneos) da Universidade de Birkbek de Londres em parceria com a Universidade Imperial de Londres (BOID, MCBRIEN e TONG, 2002). TSIMMIS (The Stanford-IBM Manager of Multiple Information Sources), cujo objetivo é integrar de forma rápida e facilitada fontes de dados heterogêneas, podendo ser bases estruturadas ou não estruturadas (CHAWATHE et. al. 1994).

Apesar da quantidade de ferramentas disponíveis para auxiliar a integração de esquemas ainda se faz necessária a intervenção de um profissional da área, normalmente, um DBA, como pode ser observado na figura 7.

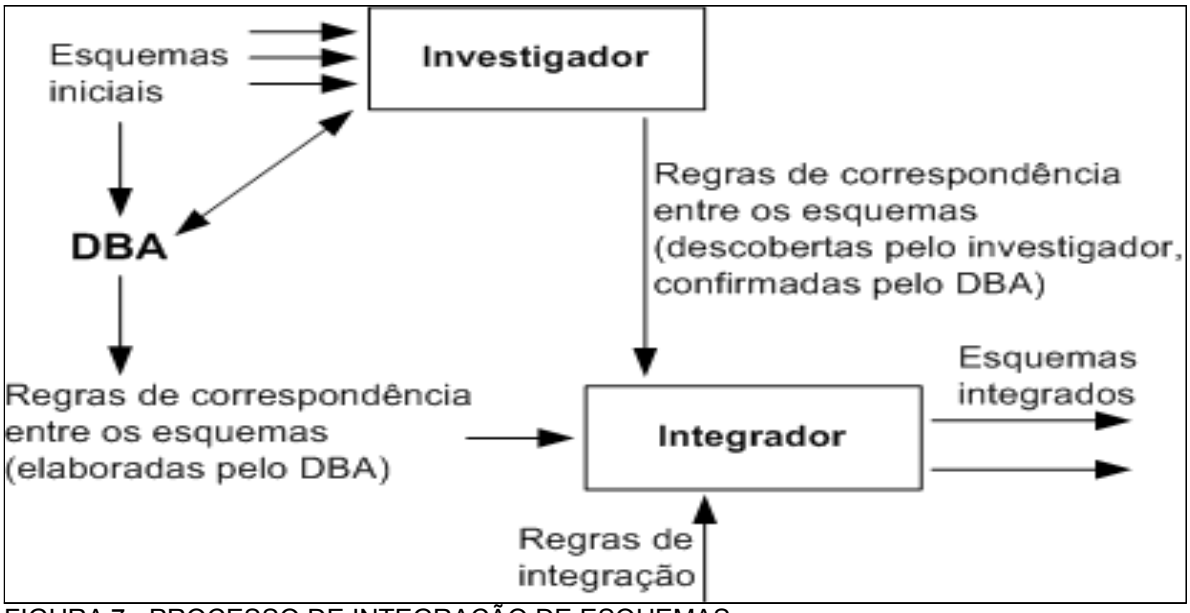

FIGURA 7 - PROCESSO DE INTEGRAÇÃO DE ESQUEMAS
FONTE: DUPONT, PARENT E SPACCAPIETRA (1992)

A ideia geral proposta por Dupont, Parent e Spaccapietra (1992) (figura 7) consiste na definição de um processo de integração de banco de dados, que parte de um conjunto de especificações para geração de um novo esquema de banco de dados distribuído. Os autores apresentam um modelo bifásico de integração que inicia com um processo de investigação e posteriormente à integração propriamente dita.

Na fase de investigação, o objetivo é identificar os pontos em concordância e discordância de cada esquema, etapa manual realizada pelo DBA. Este profissional examina os esquemas de entrada e define um conjunto de correspondências entre os esquemas. Esta fase pode ser semiautomatizada com uso de ferramentas de auxílio, que automatizam a descoberta de correspondências entre os esquemas por meio de similaridade de nomes, estruturas e relacionamentos. Neste processo, o investigador é a ferramenta que identifica as correspondências e as exibe para que o DBA tome decisão de aprovação ou reprovação sobre as correspondências encontradas.

Na segunda fase, a integração é realizada de forma semiautomatizada de acordo com as correspondências entre esquemas e regras de integração. Novamente há necessidade de intervenção do DBA para resolução de conflitos entre os esquemas de entrada. Esses conflitos são em decorrência da forma individual que cada pessoa possui de modelar a resolução de um problema.

2.1.4 Arquiteturas de integração

Atualmente há muitas formas de integração de bancos de dados, contudo Tatbul *et al.* (2001) enumeram três arquiteturas diferentes: a) integração baseada em federação; b) mediadores; c) e data *warehouse*. Estas podem ser divididas em abordagem virtual e materializada.

A diferença entre ambas está na forma de acesso, que pode ocorrer diretamente ou indiretamente. Na abordagem virtual, os dados são acessados indiretamente sob demanda, ou seja, quando o usuário requisita um sistema de informação, inicia-se um subprocesso para a obtenção dos dados, já na abordagem materializada os dados são armazenados em um repositório chamado de *warehouse*. Sendo assim, a requisição é feita diretamente no repositório materializado e, consequentemente, é imediatamente atendida.

2.1.4.1 Abordagem virtual

A abordagem virtual pode ser dividida em duas, integração por mediadores e federação, que pode ser classificada em duas abordagens de implementação, as fracamente acopladas e as fortemente acopladas.

2.1.4.1.1 Mediadores ou integradores

O principal objetivo é prover uma visão unificada dos dados de forma virtualizada. O mediador possui um esquema global ou mediador. Neste esquema o usuário realiza as operações, normalmente, efetuadas na forma de consulta. Neste modelo o usuário não tem a mínima ideia de onde estão armazenadas as fontes de dados, tão pouco que tipo de fonte de dados está consultando, uma vez que as fontes de dados podem ser as mais variadas possíveis, podendo ser bancos de dados, arquivos textos, sistema legado, ou qualquer outro tipo de fonte de dados (TATBUL *et al.* , 2001).

As principais características desta arquitetura são:

1. componentes diferentes de bancos de dados podem ser fontes de dados em um esquema mediado, por exemplo, um componente de serviço legado, uma página de dados da internet, um arquivo XML[3] etc;

[3] XML – *Extensible Markup Language* ou Linguagem de Marcação Estendida

2. em consequência da heterogeneidade das fontes de dados os mediadores podem possuir linguagens próprias, uma vez que possuem, em sua estrutura, mecanismo que decompõe as requisições vindas do usuário e, portanto, deixam as fontes mais autônomas;
3. as operações realizadas pelo usuário, normalmente, são consultas, ou seja, esta arquitetura tem foco na leitura dos dados;
4. as fontes possuem total autonomia, além do fato de que é bastante simples adicionar ou remover fontes de dados do esquema global;

O mediador é um componente ou serviço de software que faz uma ponte entre as requisições do usuário e sua origem (figura 8).

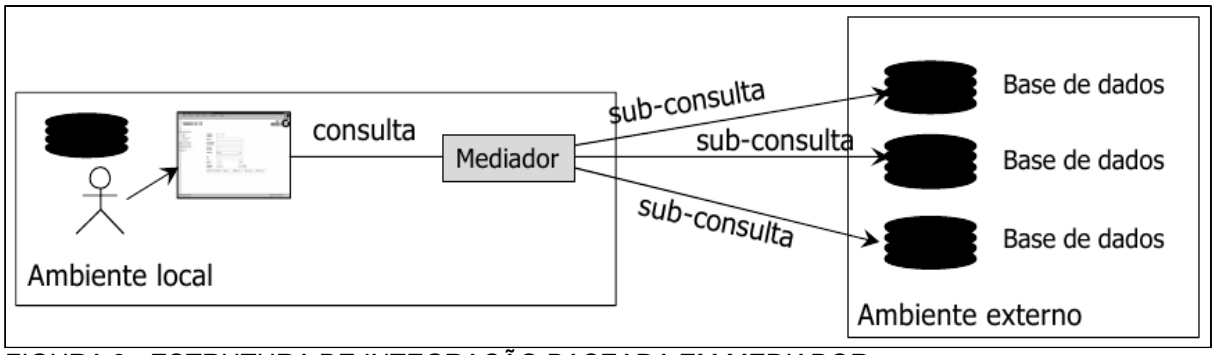

FIGURA 8 - ESTRUTURA DE INTEGRAÇÃO BASEADA EM MEDIADOR
FONTE: KROTH (2006)

A figura 8 mostra a estrutura geral de integração por mediador. O papel do mediador pode ser entendido como um componente de *software* que recebe as requisições do usuário, sob a forma de instruções SQL ou linguagem própria, supondo que a requisição seja uma consulta, o mediador a decompõe em subconsultas, e as envia respectivamente para as fontes correspondentes, cada fonte de dados retorna sua parcela de dados que, ao chegarem no mediador, são convertidos para o modelo do esquema global e então os dados são retornados ao usuário requisitante.

Tatbul *et al.* (2001) destacam dois componentes principais em uma arquitetura de integração por meio de mediador, uma estrutura chamada mediador ou integrador e um componente *wrapper* (serviço de comunicação) para cada fonte de dados. O mediador trata uma consulta em quatro passos:

1.recebe a consulta feita pelo usuário;

2.decompõe a consulta em subconsultas, com base na descrição das fontes de dados;

3.otimiza as subconsultas, com base no plano de execução dessas consultas;

4.envia as subconsultas, para cada *wrapper* individual de cada fonte de dados, que, por sua vez, envia à fonte de dados, após o processamento os dados são retornados e apresentados ao requisitante.

Em resumo, um mediador atua como intermediário em um processo de recuperação de dados. O usuário requisita um determinado dado ao mediador, o mediador conhece as fontes de dados através do mapeamento que possui dessas fontes, realiza as buscas individuais em cada fonte de dados, agrupa os resultados para contemplar a visão do usuário e retorna os dados requisitados.

2.1.4.1.2 Integração baseada em federação

Integração de bancos de dados federados é uma coleção bancos de dados autônomos, cujo objetivo comum é o compartilhamento de todo ou parte dos seus dados, através de exportação de esquemas, de modo a prover interoperabilidade entre bancos de dados fisicamente distribuídos (SHETH e LARSON, 1990).

Segundo Heinbigner e McLeod (1985) este modelo de integração possibilita que cada membro da federação possa interagir cooperativamente entre cada membro da federação e ao mesmo tempo mantém a autonomia local de cada membro envolvido.

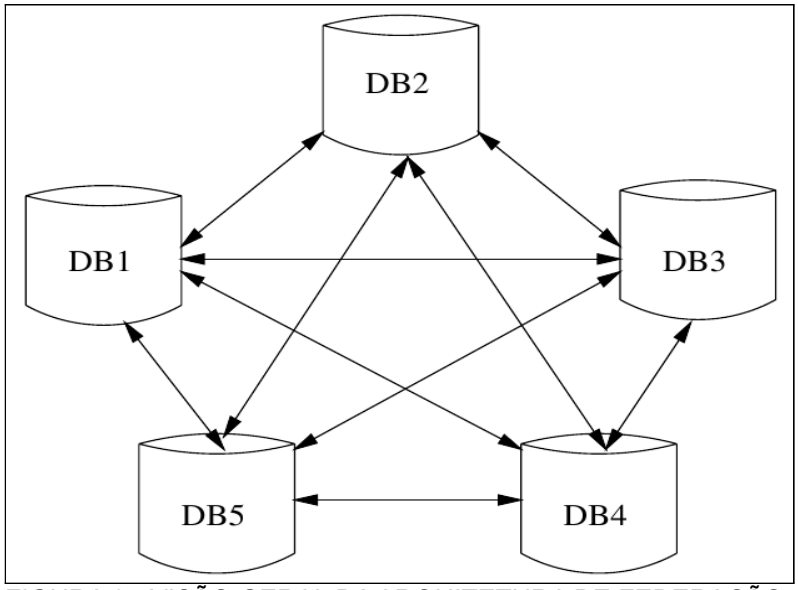

FIGURA 9 - VISÃO GERAL DA ARQUITETURA DE FEDERAÇÃO
FONTE: TATIBUL *et al.* (2001)

A figura 9 representa bem o processo de cooperação apresentado, tanto por Heinbigner e McLeod quanto por Sheth e Larson, em que cada um dos bancos de dados representados por DB1, DB2, DB3, DB4 e DB5, podem comunicar com qualquer outro membro da federação. Essa comunicação, segundo Heinbigner e McLeod (1985) se dá através da comunicação de dados, compartilhamento de transação ou realizar atividades de forma cooperativa.

Sheth e Larson (1990) classificam o banco de dados federados de acordo com o grau de autonomia e tipo de integração existente entre os membros da federação. Os fracamente acoplados e os fortemente acoplados. O nível de acoplamento está diretamente ligado à independência ou dependência de cada membro na federação.

2.1.4.1.3 Fracamente acoplados

Não existe controle central da federação, nos fracamente acoplados, a responsabilidade de criação e manutenção da federação fica a cargo do usuário.

O objetivo principal desta abordagem é facilitar o compartilhamento dos dados entre os membros da federação de forma simplificada e flexível, de forma que a qualquer tempo podem entrar ou sair novos membros sem prejuízo à federação. Daronco (2003) explica que os sistemas fracamente acoplados surgiram para minimizar os problemas do esquema conceitual global, que são fortemente acoplados. Exemplo desse tipo de implementação pode ser observado no trabalho de Soares e Medeiros (1999), que descreve a implementação de integração baseada em federação aplicada em sistemas ligados de informações geográficas.

2.1.4.1.4 Fortemente acoplados

Nos bancos fortemente acoplados existe uma autoridade que centraliza todo o processo que possui o papel de administração da federação. Esta abordagem é bem semelhante à integração de esquemas por meio de esquema global. Esse modelo é também chamado de estático, uma vez que qualquer alteração em qualquer membro da federação deve todo o processo de integração ser revisado. Portanto configura-se como um tipo rígido de difícil implementação.

Esse tipo de abordagem pode ser interessante em casos em que há

necessidade de sincronização de bases de dados, posto que qualquer modificação (inclusão, alteração ou exclusão) em qualquer base de dados deve ser refletida para todos os membros da federação.

Uma diferença bastante significativa entre as duas abordagens pode ser observada sob a percepção do usuário: na fracamente acoplada, o usuário tem a visão de todas as bases, isoladamente, já na fortemente acoplada, o usuário possui a visão de apenas uma base de dados.

2.1.4.2 Abordagem materializada

Nesta arquitetura, assim como nas anteriores, o objetivo é fornecer uma visão integrada dos dados dispostos em fontes de dados distribuídas. A peculiaridade da abordagem materializada, também chamada de *data warehouse*, é que os dados que representam a visão integrada são armazenados, em uma base de dados local. O produto da integração é disposto fisicamente em um local e não como uma visão conceitual. Por conseguinte, as operações, normalmente de consultas, são realizadas diretamente no *data warehouse* (figura 10).

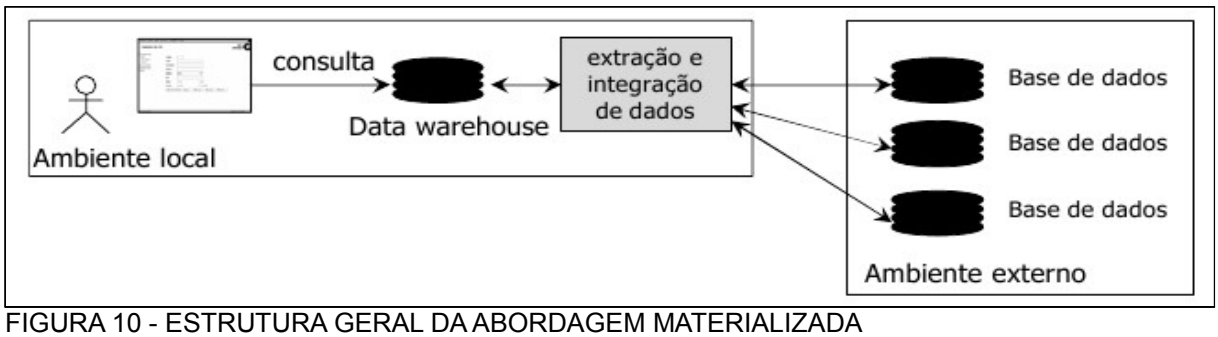

FIGURA 10 - ESTRUTURA GERAL DA ABORDAGEM MATERIALIZADA
FONTE: KROTH (2006)

Alguns aspectos devem ser levados em consideração em um projeto de integração baseado na abordagem materializada, que estão diretamente relacionados à criação e manutenção de um data warehouse: a) modelagem do projeto – quais fontes de dados e quais dados dessas fontes serão selecionados para compor a visão materializada; b) manutenção (atualização) – como será a forma de atualização dessas visões, ou seja, define como será o processo para manter a base física atualizada. A atualização é a principal atividade a ser planejada em um projeto de visões materializadas (TATBUL *et al.*, 2001).

A manutenção da visão materializada é um fator bastante crítico que por

vezes pode inviabilizar um projeto dessa natureza, visto que a sincronização dos dados entre as fontes de dados e o data warehouse é um processo bastante oneroso e depende de alto poder de processamento. O processamento é feito por algoritmos que comparam os dados entre as duas fontes de dados (fontes de dados de origem e fonte de dados de destino) (ZHUGE *et al.* 1995; QUASS e WIDON, 1997).

Esta abordagem possui vantagens e desvantagens. Quando a fonte de dados a ser integrada sofre constantes atualizações, esta pode não ser uma das melhores implementações de integração, por outro lado se a necessidade é performance e não haja necessidade de que os dados integrados sejam constantemente atualizados, esta é uma das soluções mais robustas a serem aplicadas, visto que os dados estão isolados de sua origem e eventualmente dispostos para agilizar um processo de busca.

Um caso real em que pode ser aplicada esta metodologia é, supondo que haja um banco de dados relacional de uma universidade, que possui pessoas, estas pessoas podem assumir diversos papéis dentro dessa instituição, denominados vínculos, que são: aluno, professor, técnico administrativo, terceirizados e visitantes. Esta base de dados possui aproximadamente 300 tabelas interrelacionadas; para conseguir todos os vínculos de uma pessoa cadastrada neste banco há necessidade de unir mais de 50 tabelas. Observa-se que os vínculos, assim como o status das pessoas podem mudar com certa regularidade, ou seja, um professor novo pode ser contratado, um aluno conclui seu curso, um terceiro é demitido etc. Para exemplificar, assume-se que a frequência seja semanal. Neste caso a abordagem de materialização da consulta que busca os vínculos se torna bastante desejável, visto que tira toda a carga de busca do banco de dados para uma visão externa fisicamente distinta da base original.

Existem muitas ferramentas destinadas à criação e manutenção de data warehouse, algumas comerciais como a Oracle Data Integration Suite da Oracle® e DB2 Data Warehouse Edition da IBM® (ORACLE, 2010 e IBM, 2010), além de outras de código livre, tais como Export Import Directory (EID) etc.

EID é uma ferramenta de importação e exportação de dados que pode ser utilizada para criação e manutenção de data warehouse. O EID é desenvolvido e

mantido pela Universidade Federal de Minas Gerais (UFMG), sua principal função é prover suporte para integração de base de dados heterogênea, podendo conectar-se a qualquer base de dados relacional para criar uma base única, chamada de metabase. O uso dessa ferramenta é recomendação do projeto Comunidade Acadêmica Federada (CAFe), que possui objetivo de integrar universidades e centros de pesquisa brasileiros em uma rede de confiança para prover autenticação federada.

2.2 SEGURANÇA DA INFORMAÇÃO

A segurança da informação (SI) remete à proteção dos ativos de informação de uma organização. Um ativo pode ser considerado como qualquer dado, informação ou conhecimento relevante para o negócio da organização. São representantes de ativos: pessoas, documentos impressos ou em formato eletrônico, procedimentos estruturados, sistemas de informação, dispositivos etc. O principal objetivo da SI é proteger os ativos contra as ameaças que possam causar algum prejuízo ou dano à organização. As ameaças podem ser classificadas como físicas ou lógicas. Essas ameaças possuem ação direta sobre as vulnerabilidades de um determinado ativo.

Em linhas gerais, Beal (2008) define que o objetivo da segurança da informação visa à preservação dos ativos de informação, levando em consideração três objetivos fundamentais: confidencialidade, integridade e disponibilidade.

A adoção de um projeto de SI em uma organização, pública ou privada, proporciona a redução de ameaças e consequentemente redução de riscos, que podem comprometer o negócio desta organização, ao mesmo tempo, que aumenta produtividade dos usuários, através de um ambiente mais organizado e controlado. A ABNT (2005) aponta que a SI, além de promover a competitividade, lucratividade, conformidades legais e a imagem da organização junto ao mercado, ainda contribui para a diminuição de fraudes, espionagem, sabotagem, vandalismo, incêndio e inundação. Tanto na iniciativa pública quanto na iniciativa privada possibilita interconectividade entre redes públicas e privadas, permitindo que haja disponibilização de serviços como o governo eletrônico (*e-gov*) e serviços de comércio eletrônico (*e-commerce*).

Para a implementação de um modelo de gestão para SI é necessário o envolvimento de todas as pessoas da organização, principalmente a alta direção, assim como, o estabelecimento de um plano de gestão de SI que contemple o ciclo de vida da informação, desde a criação até o descarte, passando pelo tratamento, uso, armazenamento e distribuição. Este modelo deve ser apoiado em políticas de SI e estas devem ser elaboradas com base em normas e padrões reconhecidos e comprovados e, ao mesmo tempo, estarem alinhadas ao negócio da organização, ou seja, levar em consideração a cultura organizacional, de forma que o processo de SI não se torne obstáculo e sim uma ferramenta de padronização para a organização.

Entre as inúmeras motivações para a implementação de um projeto de SI estão os instrumentos jurídicos, sob a forma de leis, decretos, acórdãos, instruções normativas, portarias, além de manuais técnicos de boas práticas:

1. Lei Nº 9.983, de 14 de julho de 2000. Altera o Decreto-Lei no 2.848, de 7 de dezembro de 1940, – Código Penal e dá outras providências: trata da inserção de dados falsos em sistemas de informação e modificação ou alteração não autorizada de sistemas de informação.
2. Decreto Nº 3.505, de 13 de junho de 2000. Institui a Política de Segurança da Informação nos órgãos e entidades da Administração Pública Federal.
3. Decreto Nº 4.553, de 27 de dezembro de 2002, cujo objetivo é disciplinar a salvaguarda de dados, informações, documentos e materiais sigilosos, bem como das áreas e instalações onde tramitam.
4. Instrução Normativa GSI/PR Nº 1, de 13 de junho de 2008. Normativa elaborada pelo Gabinete de Segurança Institucional, cujo objetivo é disciplinar a Gestão de Segurança da Informação e Comunicações na Administração Pública Federal.
5. Boas Práticas em Segurança da Informação. Manual técnico elaborado pelo Tribunal de Contas da União (TCU) em 2008, cujo objetivo é apresentar boas práticas em SI a qualquer pessoa que interaja de alguma forma com ambientes informatizados. Um dos focos principais deste guia é auxiliar os Gestores da Administração Pública Federal. O manual traz ainda referências de acórdãos e decisões do TCU, sendo que a maioria dos acórdãos tem

como base as seções e controles da norma NBR ISO/IEC 17799:2005 (TCU, 2008).

6. Decreto Nº 7174, de 12 de maio de 2010. Regulamenta a contratação de bens e serviços de informática e automação. Entre outras coisas, o decreto trata da segurança para usuários e instalações como exigências no processo de contratação. Assim como a qualidade e padronização dos itens contratados.

2.2.1 Gestão de risco

Risco de segurança é uma combinação de ameaças, vulnerabilidades e impacto. As ameaças são eventos que exploram as fragilidades (vulnerabilidades) e podem causar danos. O impacto é a consequência de uma vulnerabilidade ter sido explorada (ESR-RNP, 2010).

Gestão de risco é um conjunto de processos que dão condições às organizações para identificar e implementar medidas de proteção, com intuito de diminuir os riscos a que estão sujeitos os seus ativos de informação (BEAL, 2008).

Os procedimentos da gestão de risco não se limitam tão somente à eliminação dos fatores de risco, mas ao tratamento adequado destes, por meio de procedimentos formais alinhados aos objetivos da organização. Segundo a ESR-RNP (2010), este processo de tratamento pode ser simplificado em quatro etapas:

1. Identificação dos riscos: os riscos são identificados para um determinado ativo, levando em consideração as possíveis ameaças para suas respectivas vulnerabilidades. Neste caso é admitido que para cada vulnerabilidade de um ativo existirá uma ameaça associada a esse ativo, que por sua vez possui uma probabilidade de ação dessa ameaça sobre a vulnerabilidade (a ação da ameaça sobre a vulnerabilidade é chamada de incidente). Esta probabilidade normalmente é definida, atribuindo valores ao grau de ocorrência dos incidentes;
2. Análise dos riscos frente aos requisitos organizacionais: nesta etapa é feita uma análise mais detalhada dos riscos alinhada ao negócio da organização e de conformidade com a legislação, normas e cultura organizacional;
3. Avaliação do impacto dos riscos sobre os requisitos de proteção da

informação: mensurar quantitativa ou qualitativamente, ou ambos, o impacto do incidente relacionado a um determinado requisito de proteção. Os requisitos de proteção elencados pela ABNT NBR ISO/IEC 27002:2005 são:

a) confidencialidade – compreende a proteção dos dados transmitidos contra ataques passivos, ou seja, acessos não autorizados;
b) autenticidade – garantir que uma comunicação seja autentica, verdadeira entre a origem e o destino;
c) integridade – garantia contra alteração ou remoção não autorizada, está relacionada à completude de uma comunicação desde a origem até o destino;
d) não repúdio – garantia de aceitação de uma comunicação da origem ao destino;
e) conformidade – garantia de conformidade com os regulamentos internos e externos impostos às atividades da organização;
f) controle de acesso – garantia de controle de acesso físico e lógico aos ativos da organização por meio de identificação, com objetivo de evitar acessos não autorizados;
g) disponibilidade – garantia que os recursos estejam disponíveis para acesso por entidades autorizadas, sempre que solicitados.

   O impacto normalmente é medido, atribuindo-se valores de acordo com o grau de ocorrência de um incidente que afeta diretamente os requisitos de proteção da informação.

4. Seleção da estratégia adequada para tratamento de riscos: as estratégias podem ser classificadas em quatro:
   a) aceitar o risco;
   b) reduzir o risco, neste caso cabe a aplicação de controles e procedimentos de segurança;
   c) evitar o risco, está relacionado à retirada do ativo tratado do processo;
   d) compartilhar ou transferir o risco, compreende a delegação do risco a outra organização. A contratação de um plano de seguros constitui-se um exemplo de transferência de risco;
   e) aplicação de controles para redução de minimização dos riscos: se a

escolha da estratégia de tratamento de riscos for a redução, então cabe a aplicação dos controles descritos na norma ABNT NBR ISO/IEC 27002:2005. O procedimento inicia-se com a chamada declaração de aplicabilidade, onde cada controle é analisado quanto sua aplicabilidade, ou não, para cada vulnerabilidade. Após serem selecionados os controles adequados, parte-se para a definição de políticas, normas e procedimentos de segurança da informação.

2.2.2 Melhores práticas em segurança da informação

Outro objeto de estudo da segurança da informação está relacionado à continuidade do negócio, cujo objetivo principal é a manutenção dos procedimentos relacionados à segurança da informação de forma cíclica, ou seja, sempre que ocorrer alguma mudança, seja nos ativos da organização ou processos, torna-se necessária a reanálise de riscos e impactos que estão diretamente relacionados aos procedimentos de SI.

Segundo Beal (2008) e Fontes (2008), existem diversas referências internacionais criadas para auxiliar a implementação de melhores práticas em gestão da segurança da informação e da tecnologia da informação. Os autores destacam, ainda, que há uma grande variedade de referências importantes, como ITIL (IT Infraestructure Library), COBIT (Control Objectives for Information and Related Technology), normas NBR da série 27000 (27000, 27001, 27002 e 27005) entre outras. Estes padrões, reconhecidos internacionalmente, normalmente não ensinam como fazer, estão limitados a apresentar o que deve ser feito para atingir níveis satisfatórios de segurança.

Algumas normas de renome internacional, homologadas pela Associação Brasileira de Normas Técnicas (ABNT), são utilizadas como guias para criação, implantação e operação da gestão da segurança da informação em uma organização, as normas geralmente possuem abrangência tanto para pequenas, médias ou grandes empresas, independente de serem públicas ou privadas. As principais normas utilizadas são:

ABNT NBR ISO/IEC 27002:2005 – Tecnologia da informação – Técnicas de segurança – Código de prática para gestão de segurança da informação. É um guia

prático para o desenvolvimento e implementação de procedimentos e controles de SI em uma organização;

ABNT NBR ISO/IEC 27001:2006 – Tecnologia da informação – Técnicas de segurança – Sistema de gestão de segurança da informação – Requisitos: esta norma representa um modelo para implementar os requisitos de um Sistema de Gestão de Segurança da Informação (SGSI);

ABNT NBR ISO/IEC 27005:2008 – Tecnologia da informação – Técnicas de segurança – Gestão de riscos de segurança da informação. Tem por objetivo fornecer as diretrizes para o processo de gestão de riscos de segurança da Informação.

A principal norma é a ABNT NBR ISO/IEC 27002:2005, pois apresenta, de maneira abrangente, controles que devem ser utilizados no processo de gestão de segurança, contudo limita-se a recomendação a quais controles devem ser implementados e não como implementá-los. A forma de implementação dos controles pode ser entendida como políticas, normas e procedimentos. É uma atividade desempenhada pelo comitê gestor de segurança da organização. A norma está estruturada em dezesseis capítulos (de 0 a 15), sendo que os primeiros cinco são introdutórios. Os capítulos são chamados de seções, que possuem trinta e nove categorias, que por sua vez possuem controles que somam um total de cento e trinta.

Esses controles estão estruturados em:

- controles – apresenta a descrição e definição do controle;
- diretrizes para implementação – informações auxiliares para implementação do controle;
- informações adicionais – apresenta uma explicação mais detalhada do controle.

O ITIL é um conjunto de documentos desenvolvido pelo governo do Reino Unido para registrar as melhores práticas na área de gestão de serviços de tecnologia da informação. Embora não represente um padrão de SI, o ITIL contempla gestão de incidentes, problemas, configuração, atendimento ao usuário final, nível de serviço e desenvolvimento, implantação e suporte de software. Desta forma colabora tanto para a padronização e melhoria da qualidade dos serviços de

TI, quanto para processos de SI (BEAL, 2008; FONTES, 2008)

O COBIT é um guia de boas práticas dirigido para gestão de tecnologia da informação. Trata-se de um conjunto de diretrizes para gestão e auditoria de processos, práticas e controles de tecnologia da informação. Criado e mantido pelo *Information Systems Audit and Control Association* (ISACA). O COBIT oferece um modelo de maturidade para controle dos processos de TI e abrange quatro domínios, os quais possuem trinta e quatro controles. Os domínios são: a) planejar e organizar; b) adquirir e implementar; c) entregar e dar suporte; e d) monitorar e avaliar. O principal objetivo do COBIT é auxiliar a organização a equilibrar os riscos e retorno de investimento em TI (BEAL, 2008; FONTES, 2008).

As normas e padrões têm por objetivo de sistematizar os processos e serviços, a fim de garantir qualidade, padronização e segurança.

2.2.3 Política de segurança da informação

A política de segurança da informação (PSI) faz parte de um conjunto de ferramentas para proteção dos ativos de informação da organização e inclui-se na Governança. De maneira geral, deve padronizar todo o processo de gestão da informação, desde criação, uso, distribuição até o descarte, contudo estas devem ser do conhecimento de todos os envolvidos no processo de gestão da informação e principalmente ter o apoio da alta administração para que sejam cumpridas. Portanto, as PSIs possuem papel estratégico dento da organização.

“A política de segurança da informação define um conjunto de normas, métodos e procedimentos, utilizados para manutenção da segurança da informação”, segundo Ferreira e Araújo (2006). Os autores ainda acrescentam que esta deve ser formalizada na organização e divulgada amplamente para a todas as pessoas que fazem uso do objeto de abrangência da mesma. Esse mecanismo deve fornecer instrumentos normativos jurídicos, processuais e administrativos e abranger estruturas físicas, tecnológicas e administrativas, para garantir a salvaguarda da informação.

Uma PSI pode ser aplicada com finalidade específica, como pode ser observada em uma política de segurança para uso do correio eletrônico, ou poderá ter maior abrangência, como em uma política para disponibilização de dados

produzidos dentro da organização para serem publicados externamente. Em ambos os casos a política de segurança deve seguir algumas diretrizes, como expõem Ferreira e Araújo (2006): a) conceituação de que as informações são ativos importantes para a organização; b) envolvimento da alta administração no processo de decisão e apoio a política de informação.

Políticas, normas e procedimentos devem ser simples, claros, estar em conformidade com os anseios da alta administração da organização, serem estruturadas de forma que possam ser implantadas em fases, estarem em conformidade com os procedimentos existentes na organização, ou melhor, devem ter comportamento adaptativo, serem orientadas a riscos, flexíveis, possuírem escalas de prioridades, dando mais valor aos ativos de maior importância para a organização e não se concentrarem apenas em ações proibitivas ou punitivas, entretanto estas devem existir (FERREIRA e ARAÚJO, 2006; ABNT NBR ISO/IEC 27002:2005, 2005).

Para melhor exemplificar a abrangência da PSI, Ferreira e Araújo (2006) a subdividem em três blocos (figura 11). Os autores explicam que:

1. diretrizes ocupam papel estratégico, portanto devem expressar a importância que a organização dá a seus ativos, tanto os ativos de informação quanto humanos;
2. normas são para detalhar as situações, ambientes, processos específicos e fornecem orientações para regulamentar o uso dos ativos;
3. procedimentos representam descrições detalhadas sobre como atingir os resultados esperados.

De modo geral, o modelo da figura 11 demonstra o processo lógico da PSI. No nível estratégico é representado pela alta administração da organização. São definidas as diretrizes (o que será padronizado) que devem levar em consideração o negócio, a cultura e os ativos da organização, além de contemplar a legislação e normas vigentes. No nível tático, que pode ser representado por diretores de área, serão definidas quais as normas (como será padronizado) para regulamentar as diretrizes. No nível operacional, que pode ser representado pelo pessoal técnico, serão elaborados os procedimentos e instruções, que representam o detalhamento da implementação das normas.

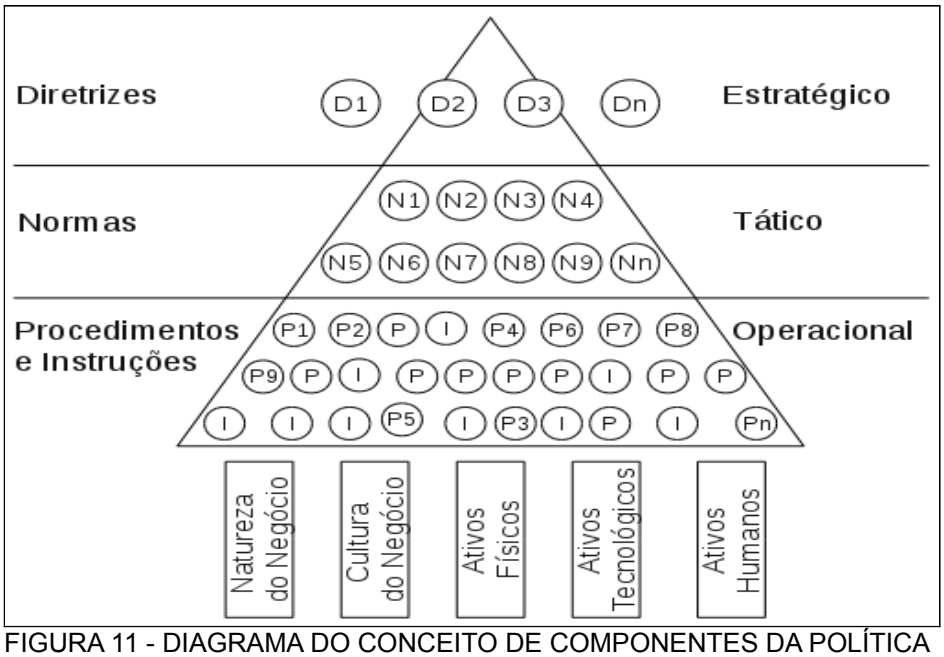

FIGURA 11 - DIAGRAMA DO CONCEITO DE COMPONENTES DA POLÍTICA E SEUS PILARES DE SUSTENTAÇÃO
FONTE: Adaptada de FERREIRA e ARAÚJO (2006)

2.2.4 Segurança da informação com foco no usuário

A pessoa tem papel fundamental no processo de SI, como comenta Fontes (2008), pois esta faz acontecer a proteção da informação, portanto deve ser alvo de cuidados especiais dento da organização para garantia de sucesso no processo de SI. O autor ainda atenta para o comprometimento do usuário em alguns aspectos, tais como: acesso à informação, conscientização da SI, plano de continuidade do negócio, serviço de suporte ao monitoramento do sistema e definição das cópias de segurança. Estes requisitos demonstram a preocupação com o usuário no processo de SI.

Marciano e Lima Marques (2006) atentam para a coerência ao elaborar PSIs, citando que não podem ser evasivas e devem sempre estar em conformidade com os comportamentos dos usuários. Salientam que ao introduzir uma política de informação, deve-se levar em conta o comportamento dos usuários e a cultura da organização, a fim de melhor aproveitamento e cumprimento das regras.

É natural alguns procedimentos de SI cercearem algumas atividades das pessoas e, por essa razão, é tão importante a adesão dessas pessoas ao processo de gestão da segurança da informação. Neste ponto é fundamental o processo de conscientização das pessoas, visto que estes procedimentos possuem uma via de

duas mãos, se por um lado é restritivo, limitando a ação das pessoas, por outro lado tem a função de padronizar e organizar as ações, por meio de documentos formais como uma PSI, por exemplo.

As pessoas devem participar do processo de definição das regras e políticas para que estas tenham sentimento de pertencimento ao processo de SI, e com isso aumentar as chances de sucesso no processo de gestão da SI.

Ao elaborar procedimentos de SI que envolvam a pessoa, deve-se levar em conta que estes procedimentos devem ser descritos utilizando linguagem simples, de fácil compreensão e rápidos de serem executados, contudo devem existir controles punitivos para que não haja repetições de incidentes ocorridos.

Na afirmação "Temos que olhar com o coração de usuário, mas com a mente da proteção", Fontes (2008) retrata bem essa proximidade com o usuário, entretanto não abre mão do objetivo principal que é a segurança.

2.2.5 Controle de acesso

O controle de acesso, segundo Beal (2008) é a medida mais importante de proteção da informação que deve ser adotada em um processo de SI, tem objetivo de impedir acessos não autorizados contra os ativos de informação. Este processo de controle é fundamental para que se tenha o mínimo de segurança dentro da organização, serve como suporte para resguardar os pilares da SI que são: a) confidencialidade – está relacionada a sigilo, privacidade intimidade e seu objetivo é evitar acessos não autorizados a informações sigilosas e desta forma proteger, além de dados pessoais, os segredos de negócio da organização; b) integridade – está relacionada à completude, autenticidade e precisão da informação. Neste caso o controle de acesso tem objetivo de evitar a criação, alteração ou destruição da informação; c) disponibilidade – está relacionada à prontidão da informação, ou seja, a informação estará disponível para usuários legítimos quando necessário.

Um processo de controle de acesso deve ser amplo o suficiente para abranger tanto o acesso em meio físico quanto em meio lógico, de forma a controlar um acesso a uma sala ou uma funcionalidade de um sistema de informação.

A ABNT (2005) aponta para a criação e manutenção de um processo formal de controle de acesso, de forma a estabelecer uma política proibitiva partindo do

princípio onde "tudo é proibido a menos que seja expressamente permitido". A norma apresenta uma forma um tanto quanto radical, quanto às regras de controle de acesso, mas esse entendimento expressa de forma clara a importância que deve ser dada a esta área.

2.2.5.1 Controle de acesso lógico

O controle de acesso em computador é motivado pela necessidade de disponibilização de um recurso ou serviço somente a uma entidade ou grupo de entidades. Entidade, neste contexto é um termo genérico para representar um agente com possibilidade de acesso, pode ser uma pessoa, um computador servidor, um serviço de rede ou qualquer outro mecanismo capaz de acessar um recurso protegido. Os modos de acesso podem ser categorizados em: acesso de leitura ou de gravação, e estão relacionados ao nível de restrição de uma entidade a determinado recurso (BENANTAR, 2006).

O controle de acesso lógico está relacionando com o uso de sistemas informatizados, cujo objetivo é proteger contra acessos não autorizados a sistemas corporativos, sistemas de gerenciamento de bancos de dados, códigos fonte, arquivo de dados, sistema operacional, aplicativos etc. Normalmente, o controle de acesso representa uma barreira lógica, que é criada para dividir uma área de acesso restrito de uma área de acesso público. Por exemplo: um sistema de cadastro e divulgação de notícias deve conter duas áreas bem distintas, uma destinada ao público (leitores) e outra, de acesso restrito, destinada aos editores. Neste caso, é conveniente que haja um mecanismo de autenticação para segregar estas duas áreas, de forma que quando um editor que possui acesso queira gerenciar notícias basta passar pelo processo de autenticação e realizar as operações desejadas.

Para que o controle de acesso seja eficiente é necessário que o acesso do usuário seja gerenciado. As recomendações da ABNT (2005) estão dispostas em sete categorias para esse propósito:

1. Requisitos de negócio para controle de acesso – controlar o acesso à informação;
2. Gerenciamento do acesso do usuário – assegurar acesso de usuários autorizados a fim de prevenir acessos não autorizados a sistemas de

informação;

3. Responsabilidades dos usuários – prevenir o acesso não autorizado dos usuários e evitar o comprometimento ou roubo da informação e recursos informatizados da organização;
4. Controle de acesso à rede – prevenir acessos indevidos aos serviços de rede;
5. Controle de acesso ao sistema operacional – prevenir acessos indevidos aos sistemas operacionais das estações de trabalho;
6. Controle de acesso à aplicação e à informação – prevenir acessos indevidos à informação contida em sistemas de informação;
7. Computação móvel e trabalho remoto – assegurar a segurança da informação quando se utilizam recursos de computação móvel e recursos de acesso remoto.

A norma NBR ISO/IEC 27002:2005 dispõe de vinte e cinco controles para tratamento do controle de acesso lógico, que estão dispostos em sete categorias, conforme citadas nos itens de 1 a 7. Cada controle apresenta diretrizes de implementação.

Tomando como base uma organização com cinco mecanismos diferentes de controle de acesso para implementação dos controles recomendados pela norma NBR ISO/IEC 27002:2005, certamente será um trabalho bastante complexo, por dois motivos principais: a) volume de registros de usuários para se manter atualizado; b) e volume de diretrizes que deverá ser convertido em procedimentos de controles de acesso. A solução coerente nestes casos é buscar ferramentas que minimizem o trabalho, tais como mecanismos de controle de acesso centralizado.

2.2.5.2 Autenticação e autorização

Uma das formas mais comuns de controle de acesso a sistemas de informação e serviços de TI ocorre por meio de um ou mais processos de autenticação e autorização. O uso de par de usuários e senha é a forma mais comum de validação de acesso nos sistemas de informação e serviços de TI. Aparentemente os processos de autenticação e autorização parecem ser relativamente simples. Fontes (2008) se refere ao processo de autenticação de pessoas ou recursos, como uma das ações mais difíceis da segurança da

informação, principalmente por envolver questões de sigilo dos dados, o autor recomenda o uso de algoritmos de criptografia e serviços de certificação digital, a fim de evitar que os dados envolvidos nesse processo sejam expostos.

No processo de autenticação se faz necessário o entendimento claro dos conceitos de autenticação e autorização, pois são distintos e devem ser entendidos como tal. A autenticação se limita à validação das credenciais do usuário, enquanto a autorização possui a função de verificar se o usuário, que foi validado pela autenticação, possui permissão de acesso à determinada funcionalidade (figura 12).

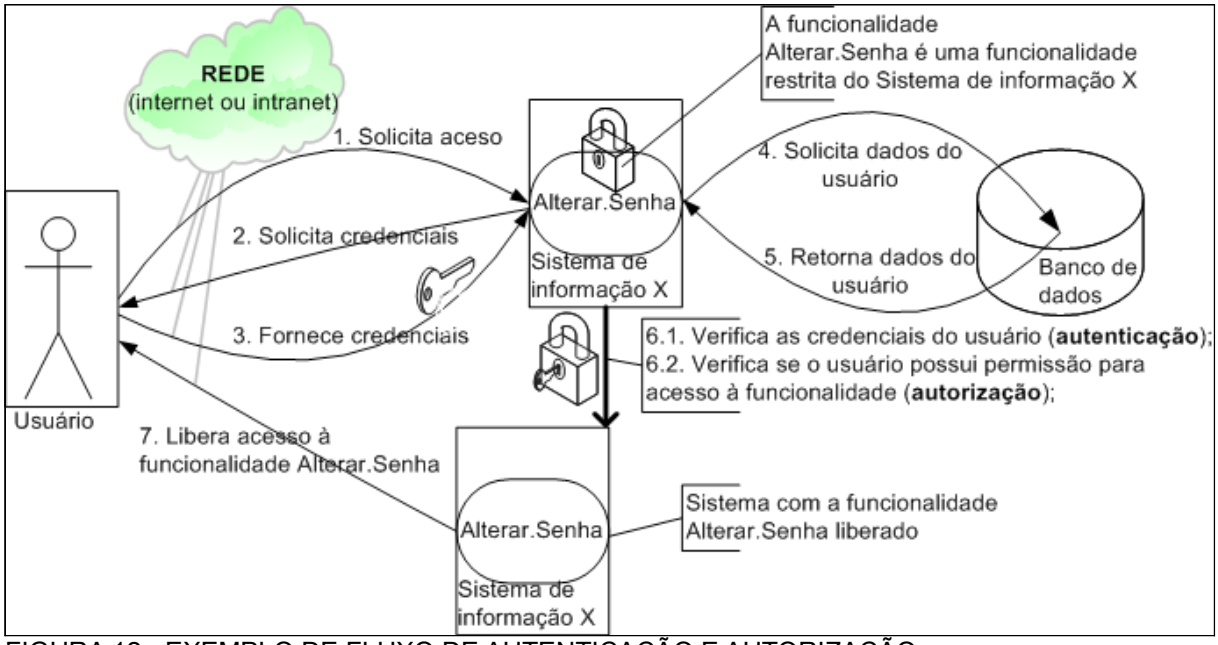

FIGURA 12 - EXEMPLO DE FLUXO DE AUTENTICAÇÃO E AUTORIZAÇÃO
FONTE: O autor (2010)

A figura 12 apresenta a sequência do fluxo de autenticação e autorização utilizado na maioria dos sistemas de informação e serviços de tecnologia da informação. A sequência de passos pode ser melhor entendida da seguinte forma:

1. usuário solicita acesso à funcionalidade Alterar.Senha do sistema de informação X, funcionalidade protegida por controle de acesso;
2. sistema de informação X solicita ao usuário as credenciais de acesso, neste caso o nome de usuário e senha;
3. usuário fornece as credenciais ao sistema de informação X;
4. sistema de informação X solicita dados do usuário ao banco de dados;
5. banco de dados retorna os dados do usuário. Neste caso convencionou-se

que os dados retornados pelo banco de dados contêm uma conta de usuário, que além das credenciais possui, também, dados de perfil de acesso, desta forma elimina um novo acesso ao banco de dados para recuperar dados para autorização;

6. sistema de informação executa duas verificações, autenticação (6.1) e autorização (6.2):
   6.1. verifica, se as credenciais fornecidas pelo usuário conferem com os dados recuperados do banco de dados;
   6.2. verifica, se o usuário, cujas credenciais foram conferidas, possui acesso à funcionalidade Alterar.Senha e libera acesso à funcionalidade;

Na figura 12, foi descrito apenas o fluxo principal de autenticação e autorização, entretanto os sistemas de informação devem implementar, tanto os fluxos alternativos quanto os de exceção.

Apesar da maioria dos sistemas informatizados utilizar mecanismos tradicionais para autenticação, baseados em usuário e senha, a norma ABNT ISO/IEC 27002:2005 orienta para o uso de mecanismos mais fortes de autenticação, principalmente quando houver a necessidade de um processo de autenticação mais robusto. Estes mecanismos podem ser alternativos ao uso tradicional de identificador de usuário e senha ou complementares. A combinação dos mecanismos pode variar de acordo com o nível de segurança desejado e o valor a ser investido na implementação. Estes mecanismos adicionais são: criptografia, cartões inteligentes (*smart cards*[4]), chaves eletrônicas (*tokens*[5]) e características biométricas.

Além do fluxo apresentado, há outras alternativas mais complexas como uso de mais de uma fonte de dados para autenticação e autorização, outras ainda fazem uso de mecanismos que centralizam o processo de autenticação em um único ponto. Estas são chamadas de CAS (*Central Authentication Service* ou Serviço de Autenticação Centralizada). Uma CAS normalmente é uma aplicação responsável pelo processo de autenticação de várias outras aplicações, a comunicação entre a

[4] *Smart card* ou cartão inteligente – é um tipo de cartão, semelhante a um cartão de crédito, a diferença é que possui capacidade de processamento (possui um microprocessador no seu interior), memória para armazenamento de dados e sofisticado mecanismo de segurança.

[5] *Token* ou chave eletrônica – é um dispositivo eletrônico, semelhante a um chaveiro, que serve para gerar senhas. A senha gerada no dispositivo é também gerada no servidor sem que haja conexão entre o servidor e o *token*, as senhas são geradas baseadas em sincronização de tempo.

CAS e as outras aplicações se dá por meio de protocolo de comunicação específico, que garante a segurança e troca de informações entre as aplicações envolvidas. O fluxo principal de autenticação centralizada funciona da seguinte forma: usuário acessa um recurso protegido de um sistema de informação, automaticamente a aplicação redireciona o usuário para autenticação na CAS, uma vez autenticado o usuário é novamente redirecionado à origem com acesso à funcionalidade requisitada.

2.2.5.3 Autenticação e integração de base de dados

A figura 12, apresentada anteriormente, traz um exemplo de fluxo de autenticação e autorização onde é utilizado apenas um banco de dados, ou seja, o mesmo banco de dados armazena dados de contas e perfis de usuários, além de outros dados do sistema. No entanto, há outras alternativas (figura 13).

A figura 13 mostra uma alternativa interessante, onde os processos de autenticação e autorização são realizados separadamente. Esta arquitetura pode se adequar a qualquer tipo de organização, mesmo uma organização que possua apenas um sistema de informação, visto que a base de dados que armazena a identidade (dados de contas de usuário) pode ser acessada por qualquer sistema informatizado para fazer autenticação, tais como sistemas operacionais, sistemas de informação e serviços de redes.

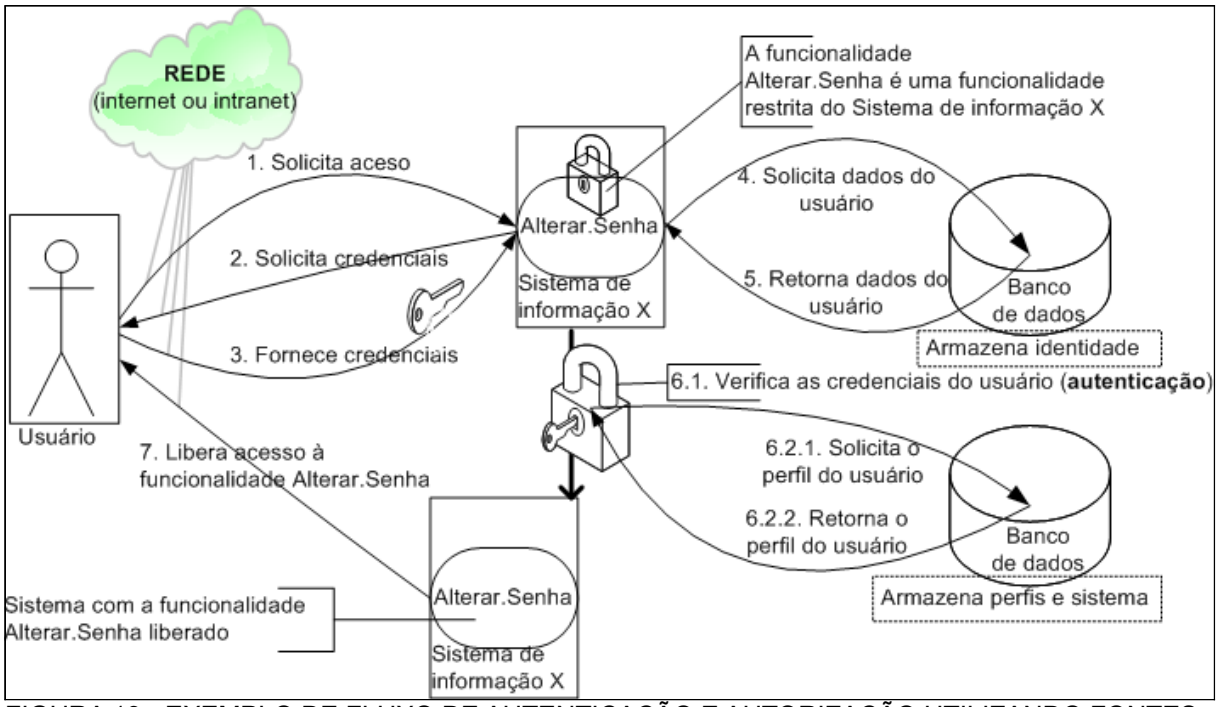

FIGURA 13 - EXEMPLO DE FLUXO DE AUTENTICAÇÃO E AUTORIZAÇÃO UTILIZANDO FONTES DE DADOS DISTINTAS
FONTE: O autor (2010)

O fluxo de autenticação, apresentado na figura 13, é bastante semelhante ao apresentado na figura 12. As diferenças estão no processamento de autenticação e autorização (passos 6.1.1, 6.2.1, 6.2.2 e 6.2.3 a seguir), de forma que utiliza um banco de dados para autenticação e outro para autorização e armazenamento de outros dados do sistema.

Os fluxos de autenticação e autorização são:

6. sistema de informação executa quatro passos para verificação da autenticação e autorização:

6.1. verifica, se as credenciais fornecidas pelo usuário conferem com os dados recuperados do banco de dados de identidade;

6.2. realiza três passos, para verificar se o usuário possui autorização para acesso à funcionalidade requerida;

6.2.1. solicita o perfil do usuário ao banco de dados de perfis e sistema;

6.2.2. banco de dados de perfis e sistema retorna o perfil do usuário;

6.2.3. verifica e libera acesso à funcionalidade Alterar.Senha.

Existem outros modelos que podem ser implementados como, por exemplo, uso de três bancos de dados distintos: um armazena identidades de usuários, outro armazena os perfis de acesso dos usuários e um terceiro armazena os dados do

sistema.

Ao comparar os dois modelos de autenticação e autorização apresentados (figura 12 e 13), observa-se que não há integração de banco de dados e sim distribuição. Esta estratégia de segregação é utilizada para que outros recursos que necessitem de autenticação possam também se beneficiar da mesma base de usuários (identidades), sem necessidade de conhecer as fontes de dados autoritativas .

A estratégia de segregação de base de dados representa um ganho real, principalmente se aliada à integração de base de dados, que pode ser aplicada em organizações que possuem muitos sistemas de informação, serviços de TI e sites, que dependem de autenticação para serem acessados, ou seja, quando há muitos sistemas que possuem a arquitetura de autenticação padrão (figura 12). Neste cenário, a integração de base de dados pode ser uma forte aliada no processo de autenticação, principalmente, em casos de sistemas de bases de dados heterogêneas que possuem suas bases de contas de usuários de forma distribuída. Neste caso, a integração pode ser aplicada para concentrar todas as bases de dados de contas de usuário em uma única, que represente todos os usuários da organização. No entanto o benefício da integração só é possível se a base de dados fonte de credenciais (nome usuário e senha) for única. Um bom exemplo que pode ser utilizado é eleger a base de dados de contas de *e-mail* para ser a fonte base de credenciais.

O isolamento de base de dados de contas de usuário é observado em ABNT (2005), que recomenta o isolamento de sistemas sensíveis, em casos onde seja necessário processamento exclusivo ou quando há necessidade de maior segurança. A norma também recomenda procedimentos bastante rígidos para gerenciamento de acesso do usuário, que inclui o registro de usuários, gerenciamento de privilégios, gerenciamento de senhas do usuário e análise crítica dos direitos dos usuários.

Os benefícios obtidos com o uso de integração de base de dados aplicado ao gerenciamento de contas de usuários, pode ser observado de duas maneiras: a) fica evidente que a redução de trabalho para a manutenção de contas de usuários só ocorrerá em um único ponto e todos os outros sistemas e serviços se

beneficiarão desta para seus processos de autenticação; b) o gerenciamento de contas do usuário é facilitado devido à integração com as fontes autoritativas[6] de dados, de modo que se um funcionário for demitido, mudar de cargo ou função, este perderá automaticamente seus direitos de acesso.

2.2.6 Controle de acesso e criptografia

"Criptografia é a ciência e arte de escrever mensagens em forma cifrada ou em código" (CERT.BR, 2006). Trata da proteção de dados ou informações em trânsito ou armazenadas. Dentre as principais finalidades estão: a) autenticar a identidade de usuários; b) autenticar e proteger o sigilo de comunicações pessoais e de transações comerciais e bancárias (CERT.BR, 2006).

Beal (2008) recomenda o uso de criptografia assimétrica para proteção dos dados em trânsito, ou seja, estabelecer um processo de comunicação segura que garanta a integridade dos dados da origem ao destino.

A criptografia assimétrica utiliza duas chaves distintas, pública e privada, para codificar e decodificar mensagens. As chaves públicas podem ser disponibilizadas publicamente, enquanto que a chave privada deve ser mantida em segredo. As chaves públicas são utilizadas para cifrar e as privadas são utilizadas para decifrar (CERT.BR, 2006).

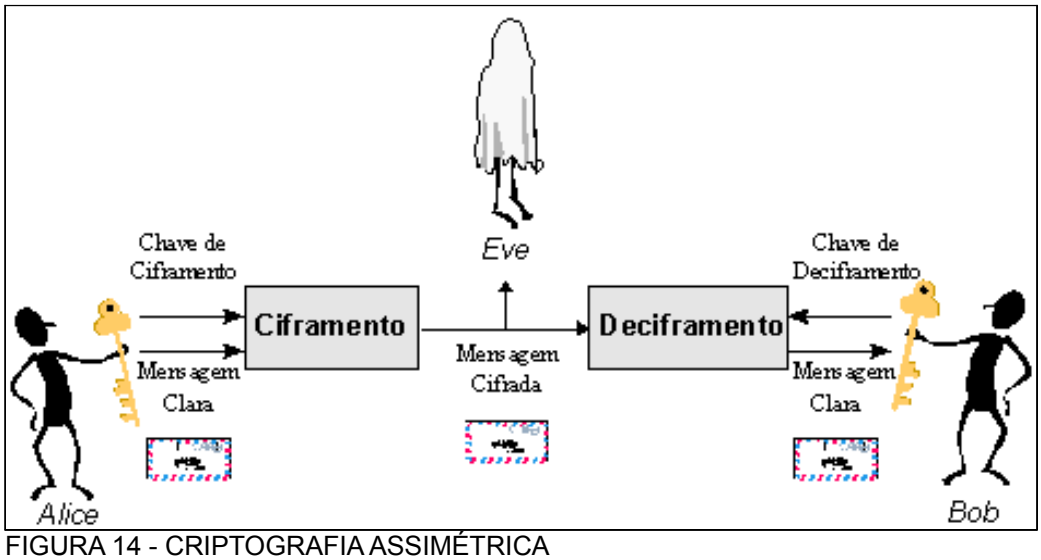

FIGURA 14 - CRIPTOGRAFIA ASSIMÉTRICA
FONTE: MAIA e PAGLIUSI (2010)

[6] Fontes autoritativas – são fontes de dados que armazenam e gerenciam os vínculos de uma pessoa. Base de dados de recursos humanos de uma organização ou de controle acadêmico de uma universidade, são exemplos de fontes autoritativas.

A figura 14 representa um exemplo de comunicação segura na internet entre duas pessoas, Alice e Bob. O processo pode ser dividido em quatro etapas:

1. Alice codifica uma mensagem utilizando a chave pública de Bob, que está disponível para o uso de qualquer pessoa;
2. depois de criptografada, Alice envia a mensagem para Bob, através da Internet;
3. Bob recebe e decodifica a mensagem, utilizando sua chave privada, que é apenas de seu conhecimento;
4. se Bob quiser responder a mensagem, deverá realizar o mesmo procedimento, mas utilizando a chave pública de Alice.

Nesse exemplo, é possível observar que Eve (pessoa desconhecida), mesmo tendo acesso ao fluxo de dados, não consegue acessar as mensagens trocadas por Alice e Bob por se tratar de mensagem cifrada.

O uso de criptografia nos fluxos de autenticação traz alguns benefícios importantes para a segurança do sistema que a utiliza. Ao trafegar dados sensíveis como credenciais, por meio da intranet ou internet, devem se levar em consideração questões de segurança, uma vez que os dados em trânsito estão vulneráveis a vários tipos de ameaças, como interceptação, espionagem, invasão de privacidade entre outros. Portanto, o uso de mecanismos de criptografia é uma solução que pode minimizar os riscos desse tipo de ameaça.

De modo geral a adoção de mecanismos de criptografia garante a confidencialidade dos dados em trânsito. Isto significa que os dados só poderão ser revelados pela fonte emissora e receptora, funcionando como um canal seguro, por isso o uso de criptografia é tão importante para os processos de autenticação. Portanto, a aplicação desse tipo de mecanismo nos processos de controle de acesso, como os apresentados nos itens 2.2.5.2 e 2.2.5.3 (figuras 12 e 13), se torna indispensável.

2.3 GESTÃO DE IDENTIDADE

Identidade pode ser entendida com a representação de uma pessoa em meio digital, que, normalmente, é composta por um identificador único e outros dados (atributos) como nome, cadastro de pessoa física (CPF), matrícula acadêmica, matrícula funcional, *e-mail* etc. Gestão de identidade consiste em um conjunto de processos formais e tecnologias voltadas ao provimento e manutenção de usuários desde seu nascimento, quando cadastrados no sistema de recursos humanos até as aplicações gerenciadas, tais como sistemas operacionais, correio eletrônico, acesso físico, sistemas de informações etc. (SANTOS, 2007).

A gestão de identidade, como afirma FONTES (2008), vai além da simples manutenção de uma identificação e senha única. Defende que alguns posicionamentos estratégicos devem ser tomados para o alcance da gestão de identidade adequada: a) existência de governança em SI – de forma que a gestão de identidades deve fazer parte do projeto de segurança da organização; b) gestão de identidade não é um projeto apenas de TI – o autor aponta que apesar da área de TI ser muito importante para a boa gestão de identidade, o projeto é da organização como um todo e, portanto, deve estar contido no planejamento de segurança da organização; c) necessidade de definições ou outros projetos serão necessários para dar suporte à gestão de identidade – o autor considera importante algumas definições que podem representar fatores de sucesso: identificação de perfis comuns que caracterizam as funções mais comuns das pessoas. Por exemplo: grupo de alunos de graduação, professores etc.; existência formalizada do gestor da informação para autorizar ou não o acesso à informação; gestão adequada dos perfis, de forma que sejam constantemente revisados. Desta forma, quando houver atualização, mudança de cargo, exclusão de função etc., os perfis sejam atualizados para evitar acessos não autorizados. O autor ainda reitera que as medidas só serão efetivas, se a organização estiver consciente e educada em segurança da informação.

Conforme AHN e LAM (2005) gestão de identidade centralizada é um modelo padronizado que envolve um único provedor de identidade (IDP – *Identity Provider*), que interage com provedores de serviços (SP – *Service Provider*), por

meio de um laço de confiança. O IDP centraliza todos os processos de autenticação e comunicação com a base de identidade, de modo que todos os SPs envolvidos possam realizar de maneira fácil seus processos de autenticação. Os autores apontam que o ponto fraco dessa abordagem está na disponibilidade do IDP, caso haja falha todos os SPs são afetados. Esta abordagem oferece comodidade ao usuário por meio do *Single Sign-On* (SSO), que possibilita que o usuário faça o processo de autenticação uma vez e pode acessar todos os SPs sem que haja necessidade de autenticar em cada SP.

Um dos benefícios mais importantes da gestão de identidade é a possibilidade de integração com os sistemas autoritativos, ou seja, os sistemas armazenam o estado do funcionário na organização ou mesmo a situação de um aluno dentro de uma universidade, desta forma qualquer alteração funcional ou acadêmica que o usuário tiver dentro da organização ou universidade pode ser refletida na base de identidades, consequentemente o acesso pode ser revisado de forma automática minimizando assim os riscos de acessos não autorizados.

Outro benefício advindo da gestão de identidade é a possibilidade do uso de CAS, de modo que todos os processos de autenticação podem ser concentrados em um único ponto seja por meio do acesso da base de identidade para autenticação ou utilização de SSO, que possibilita uma única entrada para acesso a todos os sistemas e serviços integrados à estrutura de identidade única.

Damiani, Vimercati e Samarati (2003) indicam três requisitos básicos para a questão de identidade digital:

1. segurança e confiabilidade – a identidade do usuário deve ser protegida contra ataques e falsificação, de forma que seja garantido a segurança das transações entre usuário e recurso;
2. privacidade – o usuário deve ter o controle de quais dados da identidade serão divulgados, de modo a proteger a integridade dos dados do usuário;
3. questões de mobilidade – deve se levar em consideração a estrutura computacional envolvida, tais como questões de desempenho por limitação de conexão, tamanho de tela etc.

Alguns elementos desejáveis para implementação de um projeto de gestão de identidade:

1. metabase ou metadiretório – base de dados unificada de usuários, convenientemente separada das bases de dados institucionais, entretanto deve ser mantida por dados de bases institucionais, normalmente é armazenada em uma base LDAP (*Lightweight Directory Access Protocol*) por questões de performance e flexibilidade;
2. agregação e sincronismo de identidade – integrar as múltiplas identidades existentes em diversos sistemas para uma base unificada;
3. automatizar provisionamento e desaprovisionamento – essa medida deve ser tomada para que qualquer alteração que haja nas fontes de dados sejam refletidas na base de identidade. Dessa forma quando um novo funcionário for registrado, automaticamente, sua identidade é criada, caso seja demitido sua identidade é inativada;
4. provedor de identidade ou *identity provider* (IDP) - serviço que conecta diretamente com a metabase ou metadiretório e processa as requisições de autenticação de identidade quando necessário. Esse provedor deve implementar mecanismos de segurança para proteger os dados, de forma que garanta sigilo nas transações de comunicação. Outro fator desejável em um IDP é a possibilidade de combinar mecanismos de autenticação, como criptografia além do identificador de usuário e senha;
5. provedor de serviço ou *service provider* (SP), é o agente que faz a comunicação de forma segura entre o sistema de informação ou serviço de TI e o IDP;
6. estabelecer procedimentos formais para gerenciamento de senhas;
7. padronização no desenvolvimento de sistemas, para que possa integrar com com o SP;
8. adoção de uma ferramenta de SSO que atenda tanto os requisitos funcionais quanto aos de segurança.

2.3.1 *Single Sign-On* – SSO

SSO é um processo de *logon* único, feito pelo usuário para obter acesso a várias aplicações ou serviços, este *logon* fica válido até a expiração da sessão, desta forma, o usuário pode acessar dois ou mais serviços sem que haja

necessidade de fornecer novamente as credenciais (SANTOS, 2007).

Uma arquitetura de SSO é indicada, quando os serviços de TI (sistemas de informação) dispostos de forma distribuída, vários sistemas de controle de usuários, ou seja, uma estrutura heterogênea e descentralizada. Nesse cenário o processo de SSO tem papel de centralizador, provendo uma estrutura robusta de autenticação, configurando-se como uma fonte provedora de identidade digital. Entretanto é necessário minimamente de uma base centralizada de identidades (metabase ou metadiretório).

Uma boa ferramenta de SSO normalmente oferece as seguintes vantagens:

1. acesso centralizado, de forma que o usuário tenha menos senha para guardar;
2. padronização para processo de controle de acesso em sistemas;
3. autenticação uma única vez, podendo acessar diversos recursos disponíveis sem ter que recorrer a novo processo de autenticação;
4. auditoria das atividades do usuário;
5. autorização de acesso a recursos.

2.3.1.1 Ferramentas de SSO

OpenID é uma ferramenta gratuita de provedor de identidade, mantido pela OpenID Foundation – organização internacional sem fins lucrativos. Este mecanismo pode ser usado por qualquer pessoa que queira ter um OpenID, contudo para disponibilizar o serviço de autenticação pelo provedor de identidade deve pagar uma taxa refente à associação, que pode variar de 25 até 10.000 dólares, dependendo da quantidade de usuários. Entre os principais usuários estão: Google, Facebook, Yahoo!, Microsoft, AOL, MySpace, Sears, Universal Music Group, France Telecom, Novell, Sun, Telecom Italia, entre outras. (OPENID, 2009)

Shibboleth® é uma ferramenta de SSO completa de código aberto. Apresenta em seu pacote de software o IDP, SP, DS (*Discovery Service*), além de outras ferramentas úteis. Esta ferramenta é resultado de um projeto do consórcio americano de redes avançadas (Internet2 - www.internet2.edu). Um dos objetivos da ferramenta é unir as universidades americanas em federações com intuito de compartilhar usuários e serviços entre as instituições (INTERNET2, 2009).

Outro SSO é o Java *Open Single Sign-On* (JOSSO), ferramenta de código aberto, cujo objetivo é fornecer uma solução centralizada de autenticação e autorização em ambiente web, porém pode ser adaptado para sistemas que funcionam em ambiente *desktop*, visto que sua arquitetura é baseada em *Web Service* (arquitetura que permite integração entre sistemas através de protocolo XML). O JOSSO é uma infraestrutura bastante versátil, pois permite integração com os principais servidores de aplicações Web existentes e pode ser integrado com diversas linguagens de desenvolvimento de software. Foi criado e é mantido pela AtricoreTM, empresa Americana que atua nas áreas de treinamento, suporte e consultoria (JOSSO, 2009).

Algumas universidades brasileiras utilizam a tecnologia de SSO, para autenticação de seus usuários. A Universidade Federal de Minas Gerais (UFMG) usa o Shibboleth como ferramenta de SSO, desde 2003. Assim como a UFMG, a Universidade Federal do Rio Grande do Sul (UFRGS) também utiliza tecnologias de SSO, nos processos de autenticação.

3 METODOLOGIA

Com base no contexto ambiental apresentado, nas características de integração de banco de dados, necessários ao processo, na necessidade de segurança da informação e suas características, a metodologia para realização deste trabalho foi definida nas seguintes fases: a) levantamento de informação de usuários e sistemas; b) levantamento bibliográfico de metodologias de integração de banco de dados e resolução de conflitos, segurança da informação e políticas de segurança, controle de acesso e gestão de identidade; c) realização do modelo conceitual de integração; d) prototipação e teste da solução proposta; e) elaboração de documento com sugestões de melhorias no processo de gestão de identidade.

3.1 ESTUDO EXPLORATÓRIO DO AMBIENTE

Para conhecimento do ambiente institucional, foram elaborados dois instrumentos de pesquisa. Um questionário submetido à comunidade acadêmica, cujo objetivo principal foi traçar o perfil dos usuários da Universidade Federal do Paraná (UFPR), alunos, docentes, técnicos e terceiros, em relação ao tema Segurança da Informação (SI), mais precisamente o uso de senhas utilizadas para acesso aos principais serviços disponibilizados pela instituição. Também foi realizado um levantamento, baseado na técnica de observação, nos principais sistemas de informação e serviços de comunicação, disponibilizados pela instituição pesquisada, com objetivo de identificar possíveis falhas de segurança, principalmente as relacionadas ao controle de acesso. Tanto para o questionário, quanto para o levantamento foi utilizada a norma NBR ISO/IEC 27002 de 2005. Para o questionário, além da norma, foi utilizado o guia RFC 2196 - *Site Security Handbook*, que apresenta um conjunto de procedimentos de segurança para serviços dispostos na Internet.

Os controles selecionados abrangem, principalmente, o controle de acesso lógico e estão dispostos, respectivamente, nas seções 10 (gerenciamento das comunicações) e 11 (controle de acesso) da norma NBR ISO/IEC 27002 de 2005. Nestas seções, os assuntos são tratados de maneira abrangente, contudo serão selecionados alguns, mais representativos, relacionados ao controle de acesso

lógico a sistemas e serviços de tecnologia da informação.

A seguir, serão detalhados os controles utilizados nas pesquisas exploratórias:

1. controle **10.6.2 Segurança dos serviços de rede**: um dos principais benefícios deste controle é o emprego de mecanismos de criptografia para prover segurança a serviços interconectados. Esta técnica se faz necessária, principalmente, em serviços que necessitam transmitir dados sigilosos por meio de rede de comunicação, por exemplo, nos processos de autenticação;
2. controle **10.10.1 Registros de auditoria**: convém que seja mantido registro de atividades de usuários (logs ou trilhas de auditoria), de forma que possibilite a rastreabilidade das atividades do usuário. Um registo adequado de auditoria deve conter informações para responder às seguintes perguntas: a) quem? b) quando? c) onde? d) e o que? Assim é possível saber quem, quando, onde e que operação foi realizada, possibilitando que haja rastreabilidade das ações;
3. controle **11.1.1 Política de controle de acesso**: diz respeito à existência de uma política de controle de acessos, ou seja, um documento formal que contenha todas as regras de controle de acesso claramente expressas. Esta política deve contemplar o negócio da organização, levar em consideração a legislação vigente, segregação de função, procedimento formal para obtenção e remoção de acessos e análise periódica do controle de acesso;
4. controle **11.2.1 Registro de usuário**: este controle recomenda que haja um processo formal para registro e cancelamento de usuários, de forma que seja possível atribuir permissão de acesso e retirá-lo de todos os sistemas de informação, quando houver necessidade. A norma ainda recomenda a implementação de algumas diretrizes nos procedimentos de controle de acesso que envolva registro e cancelamento de usuário, que são:
   a) utilizar um identificador (ID de usuário) único para cada usuário, para assegurar a responsabilidade das ações praticadas por este usuário;
   b) verificar, se o usuário possui autorização do proprietário do sistema para uso do sistema de informação ou serviço;
   c) verificar, se o nível de acesso concedido é apropriado ao propósito do

negócio da organização e à função do usuário;

d) informar o usuário quais são seus direitos de acesso;
e) requerer comprometimento do usuário, através de declaração de aceite dos termos do acesso;
f) assegurar que os sistemas de informação e serviços somente serão liberados após todos os procedimentos de registro serem concluídos;
g) manter registro formal de todas as pessoas registradas para usar os sistemas ou serviços;
h) remover ou bloquear, imediatamente, os direitos de acesso de usuários que mudaram de cargo ou funções, ou deixaram a organização;
i) verificar, periodicamente, e remover ou bloquear as contas de usuário com mesmo identificador (ID) a fim de garantir que não haja redundância;
j) assegurar que os identificadores de usuário (ID de usuário) redundantes não sejam atribuídos a outros usuários;

5. controle **11.2.2 Gerenciamento de privilégio**: este controle é complementar ao 11.2.1 (Registro de usuários) e prevê que os privilégios sejam restritos e controlados. A norma ainda recomenda que os recursos sejam categorizados em níveis de acesso, a fim de evitar o uso de usuários administradores (usuários que possuem todos os perfis de acesso);
6. controle **11.2.3 Gerenciamento de senha do usuário**: a norma orienta que as senhas sejam controladas por procedimentos formais, tanto no que diz respeito ao usuário, quanto pelos sistemas de informação e serviços. Para os usuários, a norma indica que seja assinado um termo de confidencialidade das senhas pessoais, para que sejam mantidas exclusivamente pelos usuários. A norma recomenda ainda que:
   a) as senhas fornecidas temporariamente sejam trocadas no primeiro acesso, isso deve ser garantido pelo sistema;
   b) as senhas devem ser armazenadas de forma protegida;
   c) os mecanismos de recuperação e obtenção de senhas devem ser seguros, de modo seja garantida a identidade do usuário antes de fornecer uma nova senha;
   d) as senhas não devem ser transmitidas de forma desprotegidas ou

enviadas por meio de mensagens de correio eletrônico;

e) quando houver geração automática de senhas, devem ser geradas senhas únicas para cada usuário e estas devem ser difíceis de serem adivinhadas;
f) por fim a norma recomenda o uso de outras tecnologias para identificação e autenticação do usuário, tais como biometria, *smart cards* e *tokens*, caso seja possível.

7. controle **11.2.4 Análise crítica dos direitos do usuário**: visa manter o controle efetivo sobre os acessos de dados e serviços de informação. Os direitos de acesso dos usuários devem ser revisados, periodicamente a cada seis meses, ou após qualquer mudança. Por exemplo, mudança de cargo, função, rebaixamentos, promoções ou encerramento de contrato. Em caso de autorização para direitos de acesso privilegiado especial, o período de análise deve ser a cada três meses;
8. controle **11.3.1 Uso de senhas**: este controle diz respeito à conduta do usuário em relação ao uso de senhas, ou seja, recomenda um conjunto de ações que os usuários devem seguir, por conseguinte os sistemas de informação e serviços devem dar suporte para que estas práticas sejam seguidas. As recomendações aos usuários são:
   a) manter a confidencialidade das senhas;
   b) manter as senhas em local seguro, evitando que sejam anotadas em papel ou meio digital de forma desprotegida;
   c) trocar de senha sempre que houver qualquer indício de comprometimento do sistema de informação ou da própria senha;
   d) criar senhas de qualidade: fáceis de lembrar, evitar palavras conhecidas ou dados pessoais, tais como data de nascimento, placa do carro etc., evitar uso de caracteres consecutivos idênticos, sequências numéricas ou alfabética;
   e) modificar as senhas com regularidade;
   f) modificar as senhas temporárias no primeiro acesso;
   g) evitar o uso de mecanismos de memorização de senhas por navegadores de internet ou sistemas de informação;

h) não compartilhar senhas individuais; e
i) não utilizar a mesma senha para finalidades pessoais e profissionais. A norma ainda recomenda o uso de uma senha única quando há muitos serviços na organização, contudo esta senha deve ser de qualidade;

9. controle **11.6.1 Restrição de acesso à informação**: este controle deve estar alinhado à política de segurança da informação. Tem como objetivo evitar acessos não autorizados, de forma que os sistemas de informação devem possuir mecanismos para controlar as ações que cada usuário poderá fazer no sistema. Estas ações devem possuir níveis de acesso de leitura, escrita e exclusão. A norma aponta para o aspecto da minimalidade, ou seja, ao tratar as saídas dos sistemas de informação, principalmente as mais sensíveis, estas devem conter apenas informações relevantes e garantir que as informações sejam disponibilizadas para terminais e locais autorizados. Convém analisar, periodicamente, as saídas, a fim de se evitar duplicidade de disponibilidade de informação.

3.1.1 Questionário

O instrumento de pesquisa teve por objetivo identificar o comportamento dos usuários da Universidade Federal do Paraná (UFPR), alunos, docentes, técnicos e terceiros, em relação ao tema Segurança da Informação (SI), mais precisamente ao uso de senhas utilizadas para acesso aos serviços disponibilizados pela instituição. Dentre os serviços que a UFPR oferece acesso a seus usuários estão: serviço correio eletrônico (*Webmail*), Sistema de Informação para o Ensino (SIE), Sistema de registro de pesquisas (Thales), Sistema de Educação a Distância (Moodle), Sistema Integrado de Gestão de Pessoal (SIGEPE), Sistema de Administração Patrimonial (SAP), Sistema de Voz Sobre IP (VOIP), entre outros.

3.1.1.1 O método

A escolha do instrumento de pesquisa foi o levantamento de dados, o instrumento de coleta desses dados foi o questionário em formato eletrônico autoaplicado via Internet, por se tratar de um método rápido e simples de se obter resultados. A opção pelo uso de questionário desta modalidade de aplicação está

centrada no fato de que todos os que responderam possuem acesso à Internet, seja na instituição ou em casa.

O levantamento é caracterizado pela interrogação direta das pessoas e tem como objetivo conhecer o comportamento dos indivíduos pesquisados. As principais características são: a) conhecimento direto da realidade - os indivíduos são submetidos a questionamentos cujas respostas se referem aos seus comportamentos, crenças e opiniões; b) economia e rapidez - quando há uma equipe treinada; c) quantificação - os dados podem ser tabulados e analisados estatisticamente. Por outro lado este tipo de estudo é caracterizado como superficial, apresenta uma visão estática do problema estudado e possui foco nas percepções das pessoas, que podem apresentar distorções entre o que pensam e o que falam (GIL, 2009).

Dentre as três técnicas de levantamento - questionário, entrevista e formulário, o questionário destaca-se pela forma rápida e barata de obter informações, além de proporcionar o anonimato (GIL, 2009).

A escolha de questionário como instrumento de coleta de dados pode ser ainda mais vantajosa, quando aliada a recursos de tecnologias da informação. De acordo com Santos e Amaral (2004), a aplicação de questionários via Internet pode apresentar maiores vantagens quando comparados aos meios tradicionais:

1. menor tempo – os questionários baseados na web oferecem maior rapidez na resposta que os meios tradicionais;
2. menor custo – os questionários baseados na web são mais baratos, por não envolverem os custos associados com a impressão, papel, envelopes e envio. Os custos associados ao suporte tecnológico acabam sendo diluídos em vários estudos;
3. maior qualidade – vários estudos mostram que a qualidade da resposta dos questionários baseados na web são melhores.

A utilização da Internet na aplicação de questionários permite maior agilidade no processo além de assegurar o anonimato, cuja característica é bastante desejável nesse tipo de estudo, acrescentam Giovinazzo e Fischmann (2001).

As questões foram elaboradas, tendo como base a norma NBR ISO/IEC 27002:2005 (Tecnologia da informação - Técnicas de segurança - Código de prática

para a gestão da segurança da informação), que trata, em linhas gerais, todos os processos de segurança da informação dentro de uma organização e RFC2196 (*Request for Comments*: 2196 – *Site Security Handbook*), um guia para desenvolvimento de políticas e boas práticas de segurança da informação voltado para Internet.

O questionário é composto de doze questões fechadas de múltipla escolha (apêndice A). O instrumento foi realizado em duas etapas. Inicialmente, foi realizado um pré-teste com número de usuários reduzido e, posteriormente, aplicado para todos os usuários da instituição.

A ferramenta utilizada para confecção do questionário de pré-teste foi o Google Forms da empresa Google.com, que disponibiliza gratuitamente a ferramenta de criação de formulários dinâmicos, bem como o armazenamento dos dados inseridos pelos respondentes, no formato de planilha e após as respostas apresenta um resumo das respostas, que facilitou o processo de tabulação e análise dos dados obtidos.

Na segunda etapa foi utilizada a ferramenta Lime Survey, programa de código aberto, que permite aplicação de questionários *on-line* e oferece infraestrutura para coleta, armazenamento e publicação de questionários via internet. A ferramenta ainda exibe graficamente as respostas por meio de métodos estatísticos (LIME SURVEY, 2010).

3.1.1.2 Amostra para o questionário

O questionário de pré-teste foi envidado para uma amostra de trinta e cinco pessoas, sendo dez docentes, dez discentes, dez técnicos e cinco terceiros, totalizando trinta e cinco pessoas com vínculo ativo com a UFPR. O critério de seleção foi intencional, de forma que foi enviado o questionário para pessoas da rede de contatos do pesquisador. O questionário foi enviado individualmente para cada pessoa, via correio eletrônico, cujo conteúdo informava a motivação do questionário, informações da pesquisa, pesquisador e *link* para acesso.

Na segunda etapa, o questionário foi disponibilizado a toda comunidade da instituição pesquisada que, no ano de 2009, somava cerca de 36.742, entre servidores docentes, funcionários técnico-administrativos e alunos, segundo dados

da Pró-reitoria de Planejamento, Orçamento e Finanças (PROPLAN, 2009). Os recursos de divulgação do questionário se deram, por meio de notícia no portal da instituição e envio de mensagens de correio eletrônico. O processo de divulgação foi realizado com apoio da Assessoria de Comunicação Social (ACS) da instituição pesquisada.

3.1.2 Levantamento dos sistemas de informação e serviços de TI

Este levantamento tem como objetivo verificar se há conformidade entre os principais sistemas de informação da instituição pesquisada e controles da norma NBR ISO/IEC 27002, principalmente, os controles relacionados ao controle de acesso. O estudo não possui a pretensão de realizar um processo completo de auditoria, visto que se trata de um estudo simplificado com intuito de identificar algumas lacunas no processo de segurança, mais especificamente os relacionados ao controle de acesso e, assim, sugerir melhorias no processo.

3.1.2.1 Método utilizado para o levantamento

O levantamento de dados é um estudo exploratório, baseado na técnica de observação. Segundo Laville e Dionne (1999), a observação como coleta de dados proporciona uma visão detalhada e aprofundada do ambiente pesquisado.

O levantamento se deu por meio de consultas SQL (*Structured Query Language* ou Linguagem de Consulta Estruturada) executadas diretamente nas bases de dados, para os sistemas que utilizam banco de dados relacional, observação não estruturada sob a visão de usuário e conhecimento prévio a respeito dos sistemas pesquisados.

3.1.2.2 Amostra para o levantamento

O critério de seleção dos sistemas e serviços para avaliação se deu por dois motivos: a) os sistemas de informação e serviços fazem parte do rol de serviços suportados pelo Centro de Computação Eletrônica (CCE); centro, ao qual o autor desta dissertação pertence, sendo que, assim, possui conhecimento e afinidade com os mesmos; b) os sistemas e serviços que representam se destacam como os principais recursos de tecnologia da informação disponibilizados para a instituição,

abrangendo as áreas de ensino, pesquisa e extensão.

3.1.2.3 Sistemas e serviços analisados

Para melhor entendimento do cenário foi necessário identificar os fluxos de comunicação (figura 15) que os sistemas executam no processo de autenticação de usuários.

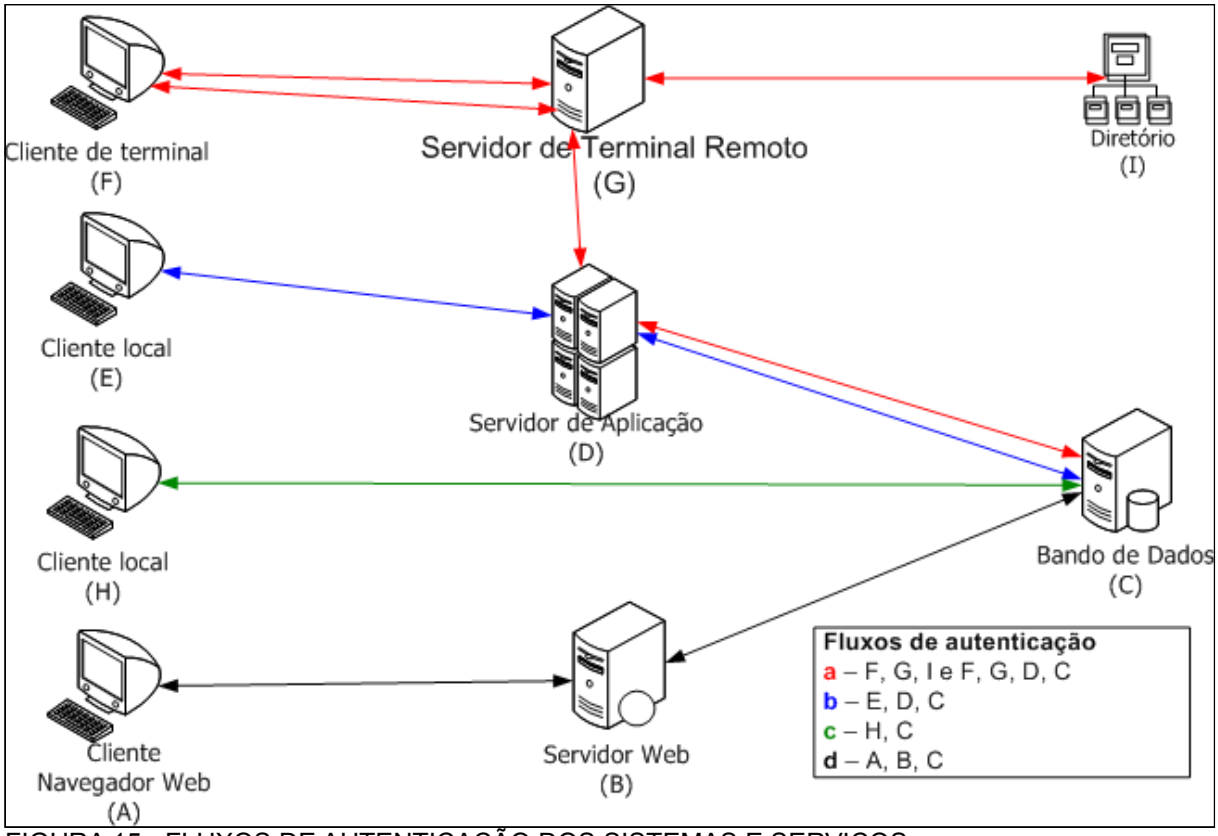

FIGURA 15 - FLUXOS DE AUTENTICAÇÃO DOS SISTEMAS E SERVIÇOS
FONTE: O autor (2010)

Os fluxos de autenticação apresentados (figura 15) representam as formas possíveis de comunicação dos sistemas de informações e serviços disponibilizados pela instituição pesquisada aos seus usuários. O ambiente tratado na figura 15 pode ser, tanto a intranet quanto a internet, ou seja, podem se dispostos fisicamente em locais distintos. Os fluxos de informação são representados pelas setas bidirecionais, que simbolizam uma requisição ou uma resposta. De modo geral os fluxos representados são quatro:

1. fluxo de autenticação “a” F ↔ G ↔ I e F ↔ G ↔ D ↔ C, intermediado por servidor de terminal de acesso remoto e servidor de aplicação;
2. fluxo de autenticação “b” E ↔ D ↔ C, intermediado por servidor de aplicação;

3. fluxo de autenticação “c” H ↔ C, sem intermediação, comunicação direta com a base de dados;
4. fluxo de autenticação “d” A ↔ B ↔ C, intermediado por servidor de aplicação, especificamente por servidor para Internet.

Os elementos contidos na figura 15 são:

1. o elemento F representa um cliente de terminal remoto, esse tipo de cliente normalmente é um software que possui a função específica de acessar um recurso disponível em um servidor de terminal remoto (G);
2. os elementos E e H são aplicações cliente, esses clientes normalmente são *softwares* instalados localmente nos computadores dos usuários e configurados para acessar tal qual D ou C;
3. o elemento A representa um cliente para Internet (*browser*), é um aplicativo instalado no computador do cliente, normalmente utilizado para acessar conteúdos oriundos da intranet ou internet;
4. o elemento G representa um servidor de terminal remoto, possui objetivo de disponibilizar algum recurso para um cliente de terminal (F). O uso desse tipo de tecnologia é justificado, quando há necessidade de disponibilização de algum recurso que não é possível instalar no computador do cliente, seja por questões de desempenho ou por incompatibilidade com o sistema operacional do cliente;
5. o elemento D representa um servidor de aplicação, cujo objetivo é processar e responder às requisições a ele efetuadas;
6. o elemento B representa um servidor de aplicação semelhante ao D, entretanto fornece um serviço mais especializado, o de processar conteúdos para Internet;
7. o elemento I representa um serviço de diretórios, que pode ser entendido como um banco de dados hierárquico utilizado para armazenamento de dados de contas de usuários;
8. o elemento C representa um banco de dados relacional é utilizado tanto para armazenamento de dados de contas de usuários quanto dados de aplicativos;

O fluxo de autenticação pode ocorrer de quatro formas como pode ser observado, na figura 15:

1. fluxo “a” comunicação intermediada por servidor de terminal remoto e servidor de aplicação. Este fluxo ocorre, quando o usuário deseja acessar determinada funcionalidade restrita de um sistema de informação por meio de um cliente de terminal de acesso remoto (G), o terminal solicita as credencias de acesso (usuário e senha) ao usuário, estas credenciais são verificadas no servidor de diretórios (I), o servidor de terminal libera acesso à aplicação específica, que novamente solicita outras credenciais do usuário, estas credenciais são verificadas no banco de dados (C) então a funcionalidade pode ser acessada. Um exemplo desse fluxo pode ser observado quando um professor executa a rotina de lançamento de notas no sistema SIE a partir da sua casa;
2. fluxo “b” comunicação intermediada por um servidor de aplicação. Este fluxo ocorre quando o usuário deseja acessar determinada funcionalidade restrita de um sistema de informação por meio de um cliente local (E), o cliente solicita as credenciais do usuário que as envia ao servidor de aplicação (D), que por sua vez verifica no banco de dados. Um exemplo desse fluxo pode ser observado quando um funcionário deseja criar um processo administrativo no sistema SIE;
3. fluxo “c” comunicação direta com o banco de dados. Esse fluxo ocorre quando se deseja acessar uma funcionalidade protegida de uma aplicação local (H), a aplicação solicita as credenciais do usuário e verifica diretamente no banco de dados. Um exemplo pode ser observado quando um funcionário deseja emitir um relatório de bens patrimoniais, por local, no sistema SAP;
4. fluxo “d” comunicação intermediada por servidor de aplicação para Internet. Este fluxo ocorre de maneira semelhante ao fluxo “b”, porém o cliente utilizado pelo usuário é um navegador de Internet. Um exemplo desse fluxo pode ser observado quando um aluno acessa o portal do aluno para obter uma declaração de matrículas, no sistema SIE.

Os sistemas de informação selecionados para a pesquisa são detalhados conforme segue.

Sistema acadêmico (Sistema de Informação para o Ensino – SIE)

É o principal sistema da instituição e também o maior, em número de usuários, módulos e funcionalidades. O sistema acadêmico provê funcionalidade para armazenamento e gerenciamento dos processos de gestão acadêmica, além dos módulos de controle de processos administrativos e controle de almoxarifado. Portanto é um sistema utilizado em todos os setores e departamentos da instituição. Este sistema possui um banco de dados relacional IBM DB2[7] , que armazena dados de alunos, docentes, funcionários, fornecedores etc. Os principais usuários do sistema são: alunos, docentes e funcionários técnico-administrativos. Os alunos utilizam o sistema por meio de um subsistema disponibilizado na internet chamado portal do aluno, onde é possível efetuar a solicitação de matrícula, entre outras funcionalidades disponíveis. Os docentes e funcionários técnico-administrativos utilizam para trâmite de documentos, gerenciamento do sistema acadêmico, rotinas de controle de estoque, solicitações de materiais etc., os docentes utilizam para obter relatórios, lançamentos de notas etc.

O sistema possui, aproximadamente, sessenta e quatro mil contas de usuário, que o acessam de três formas: a) por meio do portal do aluno (acesso efetuado por alunos); b) por meio de aplicação local (docentes e corpo técnico-administrativos); e c) por meio de aplicativo cliente de terminal remoto. Neste caso o cliente de terminal funciona como mediador para prover acesso remoto a um aplicativo instalado localmente no servidor de terminal remoto.

É importante observar que o sistema acadêmico possui a função de registrar e manter o registro acadêmico de todos os alunos da instituição.

O fluxo de autenticação pode ocorrer de duas formas, conforme pode ser observado no quadro 1, sexta coluna e na figura 15, fluxo de autenticação “a”.

Sistema de controle de recursos humanos (Sistema Integrado de Gestão de Pessoal – SIGEPE)

O sistema tem objetivo de gerenciar o cadastro de funcionários da instituição, que são: docentes, técnico-administrativos, médicos residentes e estagiários. Basicamente o sistema provê o gerenciamento de todos os trâmites administrativos acerca dos servidores e colaboradores da instituição. Para

---

[7] DB2 é um banco de dados relacional criado e mantido pela empresa IBM®

armazenamento dos dados o sistema utiliza um banco de dados PostgreSQL, um banco de dados relacional de código aberto.

Possui aproximadamente cem contas de usuários que acessam exclusivamente por meio da Intranet institucional ou Internet, caso estejam fora da rede da instituição.

Neste sistema é efetuado o registro e manutenção de dados cadastrais de funcionários e, portanto, configura-se como um dos dois sistemas responsáveis pela manutenção de vínculos dos funcionários com a instituição. Desta forma, o sistema de recursos humanos trata os vínculos de servidores e colaboradores e o sistema acadêmico trata os vínculos de alunos.

O fluxo de autenticação ocorre seguindo o fluxo “a” (figura 15).

Serviço de correio eletrônico e serviço de Proxy[8]

O serviço de correio eletrônico ou *e-mail*, disponibilizado pela instituição, caracteriza-se como um serviço de troca de mensagens eletrônicas, tanto no ambiente interno quanto externo. Este serviço possui um banco de dados MySQL[9], onde são armazenados os dados de contas de usuários. O banco de dados, assim como nos outros sistemas, funciona de forma isolada e independente.

O serviço possui em torno de dez mil contas de correio que são acessadas tanto pela intranet quanto pela internet. A autenticação pode ocorrer de duas formas: acesso por meio de um navegador de internet (*browser*) ou um cliente de correio eletrônico tal como Microsoft® Outlook ou similares. Seu fluxo de autenticação é semelhante ao do SIGEPE.

O fluxo de autenticação ocorre de duas formas, “c” ou ”d” (figura 15).

O serviço de Proxy funciona com uma espécie de “túnel”, ou seja, os usuários utilizam este serviço para acessar sites da internet como se estivessem na intranet da instituição. Alguns portais de periódicos, como I*nstitute of Electrical and Electronics Engineers* (IEEE – http://www.ieee.org), *Association for Computing Machinery* (ACM – http://portal.acm.org), entre outros. Disponibiliza acesso aos periódicos somente a partir do ambiente institucional. Neste contexto o serviço de

---

[8] Proxy é um servidor que atende as requisições de usuários e as repassa, funcionando com espécie de “ponte”.

[9] MySQL – é um SGBD relacional de propriedade da MySQL®. O MySQL se configura como um dos bancos mais rápidos do mercado de código aberto. É utilizado por grandes empresas como: Yahoo!, Alcatel-Lucent, Google, Nokia etc (MYSQL, 2010).

Proxy atua como um “túnel” para que o pesquisador (aluno, docente ou técnico-administrativo) não tenha que se deslocar até a instituição para obter acesso aos portais de periódicos. Este serviço acessa o mesmo banco de dados do serviço de correio eletrônico, utilizando-o como base para o processo de autenticação.

O fluxo de autenticação ocorre seguindo o fluxo “d” (figura 15).

Serviço de VoIP (Voice over Internet Protocol ou Voz sobre Internet)

O serviço de VoiP é disponibilizado a todos os funcionários da instituição (docentes e técnico-administrativo) e possibilita a realização de chamadas telefônicas pela internet entre outras instituições. Este serviço é resultado de um projeto chamado fone@rnp, elaborado pela Rede Nacional de Ensino e Pesquisa (RNP), cujo objetivo é promover a interconexão entre as instituições clientes. Por meio deste serviço, potencialmente, todos os usuários das instituições clientes que compõem esta rede conseguem se comunicar por voz (via telefone comum, telefone IP ou software) pela Internet“ (RNP-VoIP, 2010).

O serviço possui cerca de cento e sessenta contas de usuários. Para armazenamento dos dados, como contas de usuários, número de telefone IP etc., é utilizado um serviço de diretório baseado em LDAP.

O fluxo de autenticação ocorre seguindo o fluxo “d” (figura 15).

Serviço de terminal de acesso remoto (*Windows Terminal Server* – WTS)

Este serviço funciona como uma espécie de ponte, visto que os usuários, por meio de um cliente, acessam o servidor de terminal para obter acesso a outro sistema. O serviço de terminal remoto foi criado com objetivo de prover acesso ao sistema acadêmico por meio da Internet, principalmente, pelos usuários docentes que necessitam efetuar tarefas no sistema em locais externos à instituição. Este serviço também possibilita acesso ao sistema acadêmico, por meio de sistema operacional não compatível com o sistema acadêmico, sendo assim um computador que possui GNU/Linux[10] pode acessar o sistema acadêmico passando pelo servidor de terminal.

O sistema possui cerca de cinco mil contas de usuários e utiliza Microsoft® Active Directory[11], para armazenamento de dados.

O fluxo de autenticação ocorre seguindo o fluxo “b” (figura 15).

---

[10] GNU/Linux – Sistema operacional de código livre.

[11] *Active Directory* – implementação do protocolo LDAP e de propriedade da Microsoft®.

Sistema de administração de patrimônio (SAP)

Sistema utilizado para gerenciamento de bens patrimoniais da instituição. Para armazenamento dos dados o sistema utiliza um banco de dados relacional Firebird[12].

O sistema possui cerca de 530 usuários. O acesso é feito por meio de aplicativo local ou pela intranet.

O fluxo de autenticação ocorre de duas formas, “c” ou ”d” (figura 15).

Sistema de registro de pesquisa (Thales)

Sistema utilizado pelos pesquisadores docentes ou técnicos para registro e controle de pesquisas. O sistema possui importância fundamental na instituição, uma vez que todas as pesquisas vinculadas à Pró-Reitoria de Pesquisa e Pós-Graduação (PRPPG) são registradas neste sistema. O armazenamento de dados é feito em um banco de dados relacional Oracle[13], que funciona de maneira independente dos outros bancos de dados da instituição.

O sistema possui cerca de duas mil e trezentas contas de usuário que o acessam via internet ou aplicativo instalado localmente.

O fluxo de autenticação ocorre de duas formas, “c” ou ”d” (figura 15).

Plataforma de educação a distância (Moodle)

A instituição pesquisada possui dois portais principais com o mesmo objetivo de prover mecanismos para auxiliar à educação presencial, realizar cursos a distância ou semipresenciais. Os sistemas utilizados em ambas as plataformas é o Moodle, que é um pacote de software desenvolvido para internet e distribuído gratuitamente por meio da internet. O objetivo principal do projeto Moodle é construir um arcabouço para auxiliar a educação.

A plataforma MoodleUFPR é administrada pelo Centro de Computação Eletrônica (CCE) e o segundo, MoodleCIPEAD é mantido pela Coordenadoria de Integração de Políticas de Educação a Distância (CIPEAD). Ambos os sistemas utilizam banco de dados relacional PostgreSQL, para armazenamento de usuários, cursos, disciplinas, docentes, etc. Apesar de utilizarem a mesma plataforma, ambos utilizam bancos separados, funcionando isoladamente entre si e os demais sistemas da instituição.

---

[12] Firebird – banco de dados relacional de código livre.

[13] Oracle – banco de dados relacional comercial de propriedade da Oracle®.

O MoodleUFPR possui cerca de três mil e quinhentas contas de usuários, já o MoodleCIPEAD não foi possível verificar o quantitativo de contas de usuários que o acessam por meio da intranet ou internet.

O fluxo de autenticação ocorre seguindo o fluxo “d” (figura 15).

Sistema de bibliotecas (SOPHIA)

Utilizado pelo Sistema de Bibliotecas (SIBI) da instituição pesquisada para gerenciamento do acervo. O sistema possui duas formas principais de acesso: acesso via cliente instalado localmente e por meio da internet. O sistema utiliza um banco de dados relacional Microsoft® SQL Server, para armazenamento dos dados.

O fluxo de autenticação ocorre seguindo o fluxo “d” (figura 15).

3.2 INTEGRAÇÃO DE BANCO DE DADOS

O processo de integração de dados de usuários se deu seguindo as metodologias vistas no referencial teórico, e foi realizado em seis etapas: a) considerações iniciais; b) pré-integração; c) identificação de correspondências; d) identificação de conflitos; e) integração; e f) união e reestruturação. A arquitetura de integração selecionada foi a abordagem material, baseada em *data warehouse*.

O critério de seleção das bases se deu pela representatividade. Trata-se da base de dados de recursos humanos, base de dados de alunos e base de dados de contas de correio eletrônico.

O processo de integração se restringe a integrar apenas dados pessoais, considerados como pertencentes à identidade digital de cada pessoa.

3.3 GESTÃO DE IDENTIDADES

As recomendações quanto à gestão de identidade se dividem em: a) provisionamento de usuário, cujo principal objetivo é a criação de sistema de gestão de identificação única; b) manutenção do ciclo de vida da identidade; e c) recomendações quanto aos sistemas de informações e serviços, para que estes estejam inseridos no contexto de gestão de identidade e, assim, contribuir no processo de segurança da informação.

4 RESULTADOS ALCANÇADOS E CONSIDERAÇÕES

Pesquisas exploratórias subsidiaram a introdução deste trabalho e pesquisas subsequentes.

4.1 ANÁLISE DOS RESULTADOS OBTIDOS (QUESTIONÁRIO)

O questionário, em sua versão final, foi amplamente divulgado na instituição pesquisada e ficou disponível na internet, por quarenta e cinco dias, entre os meses de junho e julho do ano de 2010. Foi obtido um total de 266 respostas, sendo 199 de alunos, 19 de docentes, 66 de técnico-administrativos e 1 de funcionário terceirizado. O percentual de respostas foi aproximadamente 0,72%, que estatisticamente representa uma amostra pouco significativa, entretanto, representa uma amostra compatível com o prazo da pesquisa. De qualquer forma, o percentual de respostas indicam uma tendência, principalmente pela qualidade das respostas e por envolver um número considerável de alunos, docentes e técnicos.

A análise considera dois focos principais: a) a adequação dos serviços oferecidos pela instituição em relação às recomendações da norma NBR ISO/IEC 27002 de 2005; e b) o comportamento dos usuários quanto ao uso de senhas, de forma a verificar, se seus hábitos e atitudes estão adequados às recomendações da norma NBR ISO/IEC 27002 de 2005 e RFC 2196.

A questão 2 (figura 16) demonstra a quantidade de serviços e sistemas para os quais o usuário possui acesso. Aproximadamente, 45% dos respondentes utilizam três ou mais senhas para acessar os serviços oferecidos pela instituição. O que significa três ou mais formas diferentes de controle de acesso e, consequentemente, três ou mais cadastros duplicados, visto que os sistemas de informação e serviços possuem mecanismos isolados de controle de autenticação.

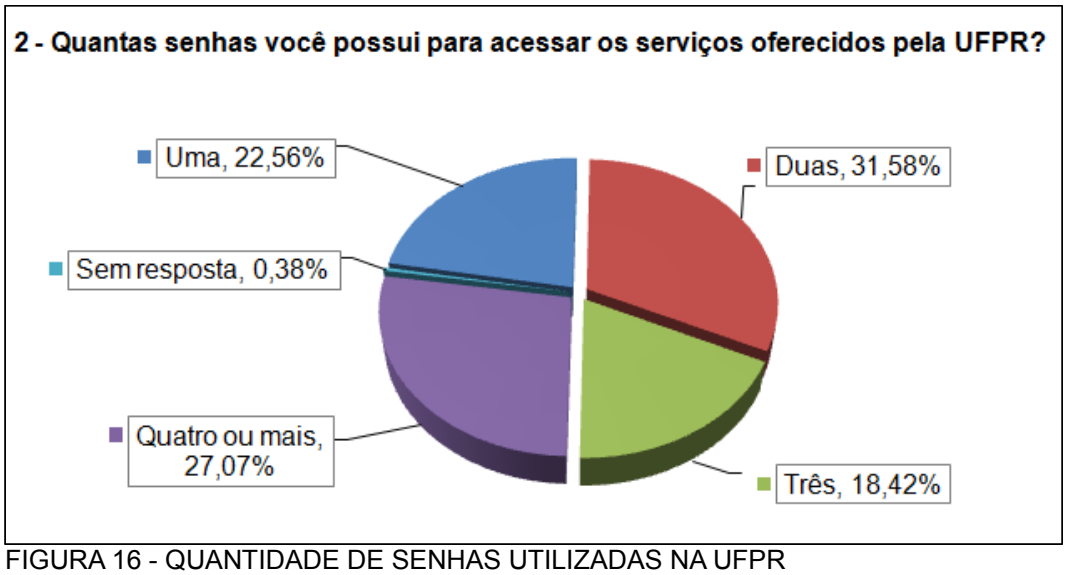

FIGURA 16 - QUANTIDADE DE SENHAS UTILIZADAS NA UFPR
FONTE: O autor (2010)

A questão 3 (figura 17) indaga ao usuário sobre a utilização de senhas distintas ou iguais para acesso aos serviços disponibilizados pela instituição. Os resultados apurados mostram que quase 25% dos respondentes utilizam a mesma senha para acesso a todos os serviços; mais de 49% utilizam senhas distintas e quase 25% utilizam senhas iguais para alguns serviços. Embora a quantidade mais expressiva seja a dos usuários que responderam que utilizam senhas distintas (49%), se somarmos os 25% que utilizam a mesma senha para todos os serviços com os 25% que utilizam senhas iguais para alguns serviços, o total é quase de 50%, o que representa inadequação tanto com a norma NBR ISO/IEC 27002 quanto com a RFC 2196, que, em resumo, recomendam o uso de senhas distintas para serviços distintos.

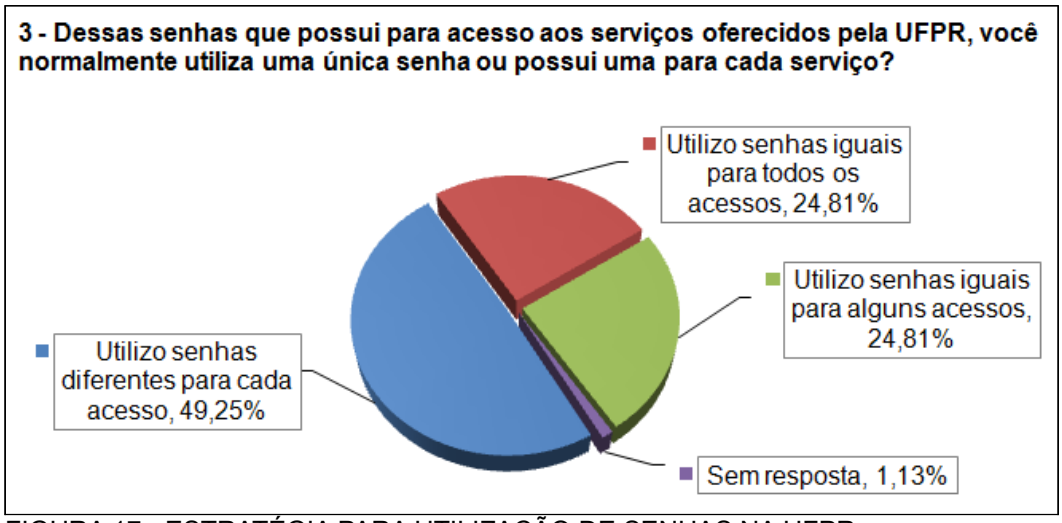

FIGURA 17 - ESTRATÉGIA PARA UTILIZAÇÃO DE SENHAS NA UFPR
FONTE: O autor (2010)

A questão 4 trata, especificamente, do processo de registro de usuários e concessão de acesso. Foi observado que 41% (figura 18) responderam que não houve procedimento formal, para a concessão de acesso aos serviços e 45% responderam que houve tais procedimentos, em alguns casos. Esse resultado apresenta não conformidade com a norma, que prevê o estabelecimento de um processo formal, que inclui assinatura de um termo de responsabilidade por parte do usuário e orientações detalhadas a respeito do serviço disponibilizado por parte de quem o oferece ao usuário, no caso, a instituição pesquisada. Normalmente, esse processo de registro e concessão de acesso deve ser tratado como gestão de identidade, e portanto, apoiado por mecanismos de segurança da informação, como normativa de registro de usuários e política que regulamenta o uso de recursos da instituição.

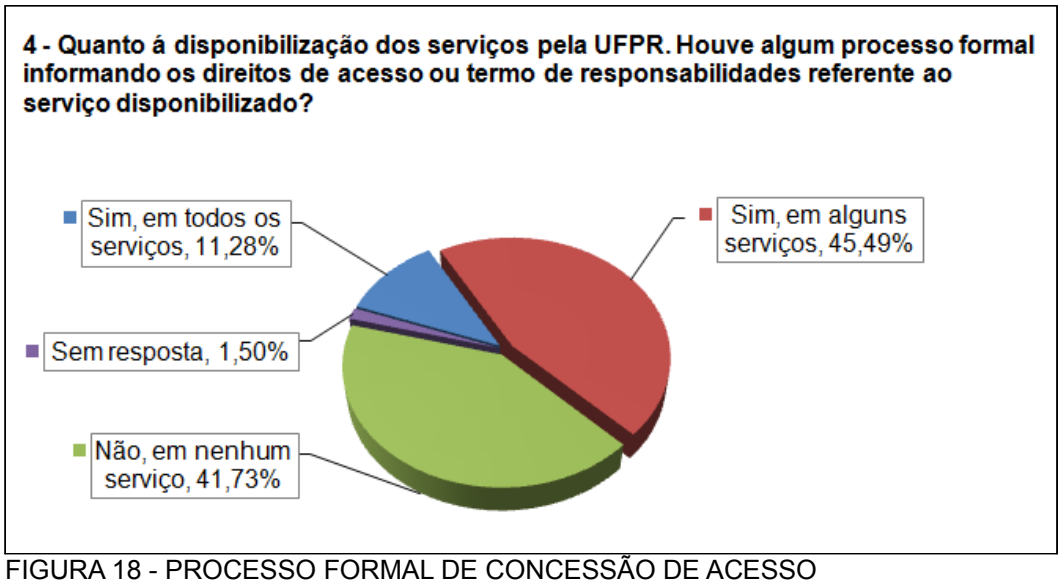

FIGURA 18 - PROCESSO FORMAL DE CONCESSÃO DE ACESSO
FONTE: O autor (2010)

A questão 5 (figura 19) trata do compartilhamento de senhas entre usuários (figura 19). Foi observado que a maioria (67%) nunca compartilha suas senhas pessoais com outras pessoas, e 25%, raramente, as compartilham. Os percentuais observados representam conformidade com a norma estudada, que recomenda o compartilhamento de senha, em casos que não envolvam dados sigilosos, ou se a descoberta desta não resultar em grandes prejuízos à organização.

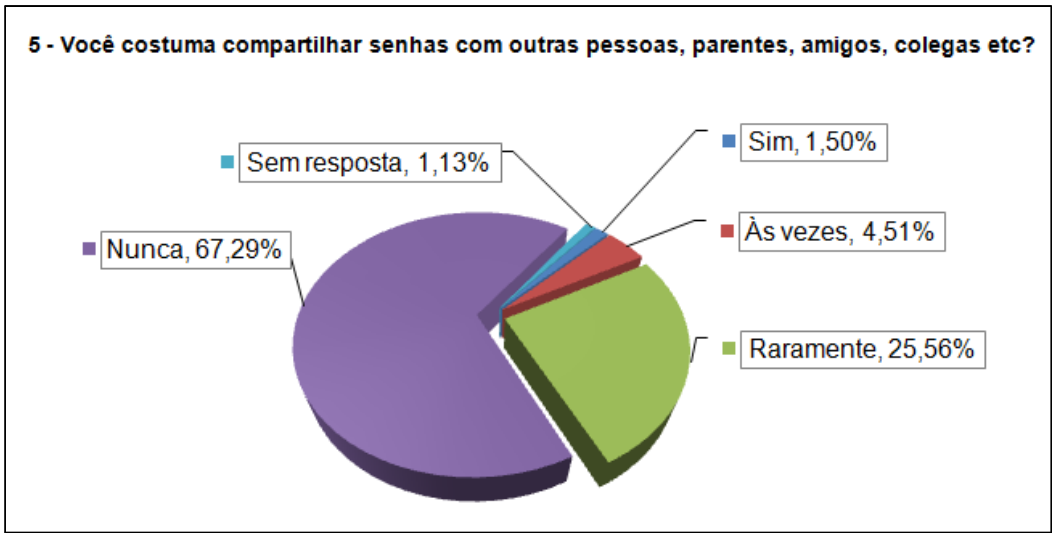

FIGURA 19 - COMPARTILHAMENTO DE SENHAS
FONTE: O autor (2010)

A questão 6 (figura 20) trata da estratégia usada pelo usuário para armazenamento de senhas. Cerca de 71% indicaram que utilizam a memorização para guardar suas senhas, 11% que anotam e armazenam em local protegido, e cerca de 10% também utilizam meios seguros de armazenamento. Apenas uma

minoria de quase 8% não toma os devidos cuidados, armazenando as senhas em caixas de correio eletrônico. Em geral, o comportamento do usuário, relacionado ao armazenamento de senhas está em concordância com a norma NBR ISO/IEC 27002 e quanto com as recomendações RFC 2196, que, em linhas gerais, recomendam que os dados sensíveis, tais como senhas, devem ser mantidos de maneira protegida.

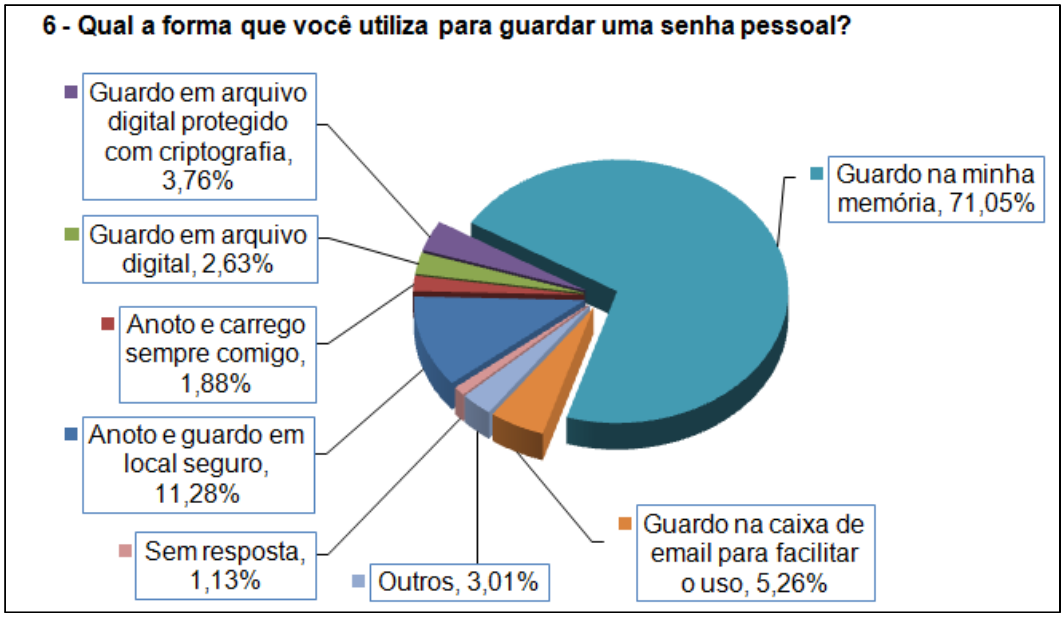

FIGURA 20 - FORMAS PARA GUARDAR UMA SENHA PESSOAL
FONTE: O autor (2010)

As questões 7 e 8 estão relacionadas à alteração de senhas. A questão 7 trata do comportamento do usuário em relação ao uso de senhas temporárias, e a questão 8 está relacionada à frequência de trocas de senhas. Na questão 7 (figura 21) a maioria dos respondentes (64%) afirma que altera a senha no primeiro acesso, já 25% afirmam que às vezes alteram e 9% nunca alteram.

FIGURA 21 - ALTERAÇÃO DE SENHAS TEMPORÁRIAS
FONTE: O autor (2010)

Na questão 8 (figura 22), cerca de 48% não possuem o hábito de alterar a senha, 12% alteram a cada dois anos, 31% alteram em intervalos que variam de três meses a um ano, e 7% não responderam.

Os resultados das duas questões (7 e 8) são bastante preocupantes e não remetem somente aos hábitos dos usuários, mas ao planejamento dos sistemas de informação e serviços, visto que estes devem possuir mecanismos de controle e gerenciamento do uso de senhas, garantindo que o usuário altere sua senha temporária no primeiro acesso, assim como solicite ao usuário a troca de senhas em intervalos regulares.

A norma NBR ISO/IEC 27002 de 2005 e RFC 2196 orientam para a troca de senhas, que deve ocorrer, após o procedimento de concessão ou recuperação de senhas. Quanto à regularidade de mudança de senhas, a norma recomenda expressamente que as senhas comuns devem ser trocadas a cada seis meses, já as senhas de missão mais críticas, senhas que protegem informações mais importantes devem ser trocadas a cada três meses.

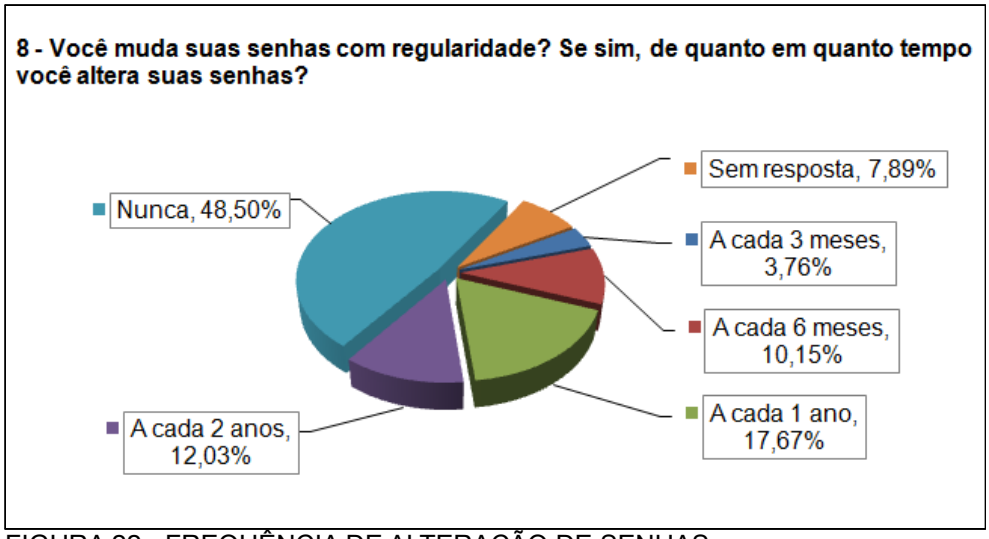

FIGURA 22 - FREQUÊNCIA DE ALTERAÇÃO DE SENHAS
FONTE: O autor (2010)

A questão 9 (figura 23) indaga sobre o processo de gestão de senhas dos sistemas e serviços oferecidos pela instituição pesquisada. Cerca de 70% dos respondentes afirmam que não há solicitação de alteração de senhas com regularidade e 26% responderam que alguns sistemas solicitam esse tipo de alteração com regularidade. Esta questão serve para confirmar o que foi constatado nas questões anteriores (questão 8 e 9), cuja percepção do usuário denota a falta de

mecanismos adequados de gestão de senhas. A constatação revela não conformidade com a norma NBR ISO/IEC 27002, que recomenda o gerenciamento de senhas pelos sistemas e serviços, tais como trocas regulares de senhas, de acordo com o grau de sigilo dos dados mantidos pelo sistema, manter histórico de senhas para que o usuário não reutilize senhas, evitar que usuário cadastre senhas fracas (senhas fáceis de serem adivinhadas) etc.

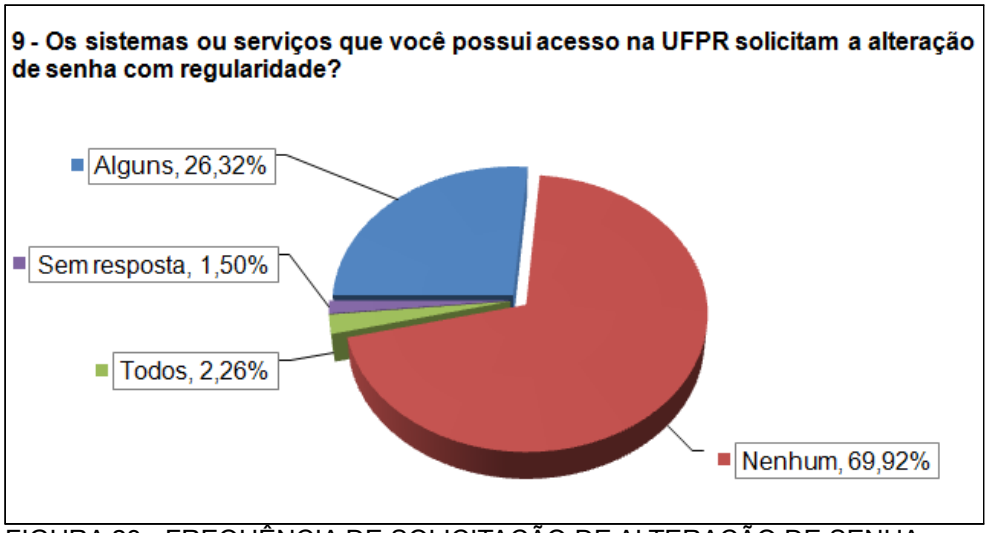

FIGURA 23 - FREQUÊNCIA DE SOLICITAÇÃO DE ALTERAÇÃO DE SENHA PELOS SISTEMAS OU SERVIÇOS DA UFPR
FONTE: O autor (2010)

Na questão 10 (figura 24) o usuário foi levado a selecionar a estratégia de composição de uma senha segura. Em geral, as respostas foram bastante satisfatórias, quando relacionadas às boas práticas encontradas na norma NBR ISO/IEC 27002 e RFC 2196. Os respondentes, em sua maioria, indicaram que uma senha para ser segura deve conter letras, números, símbolos, possuir seis ou mais caracteres. Os usuários também consideram importante que as senhas sejam fáceis de memorizar e digitar.

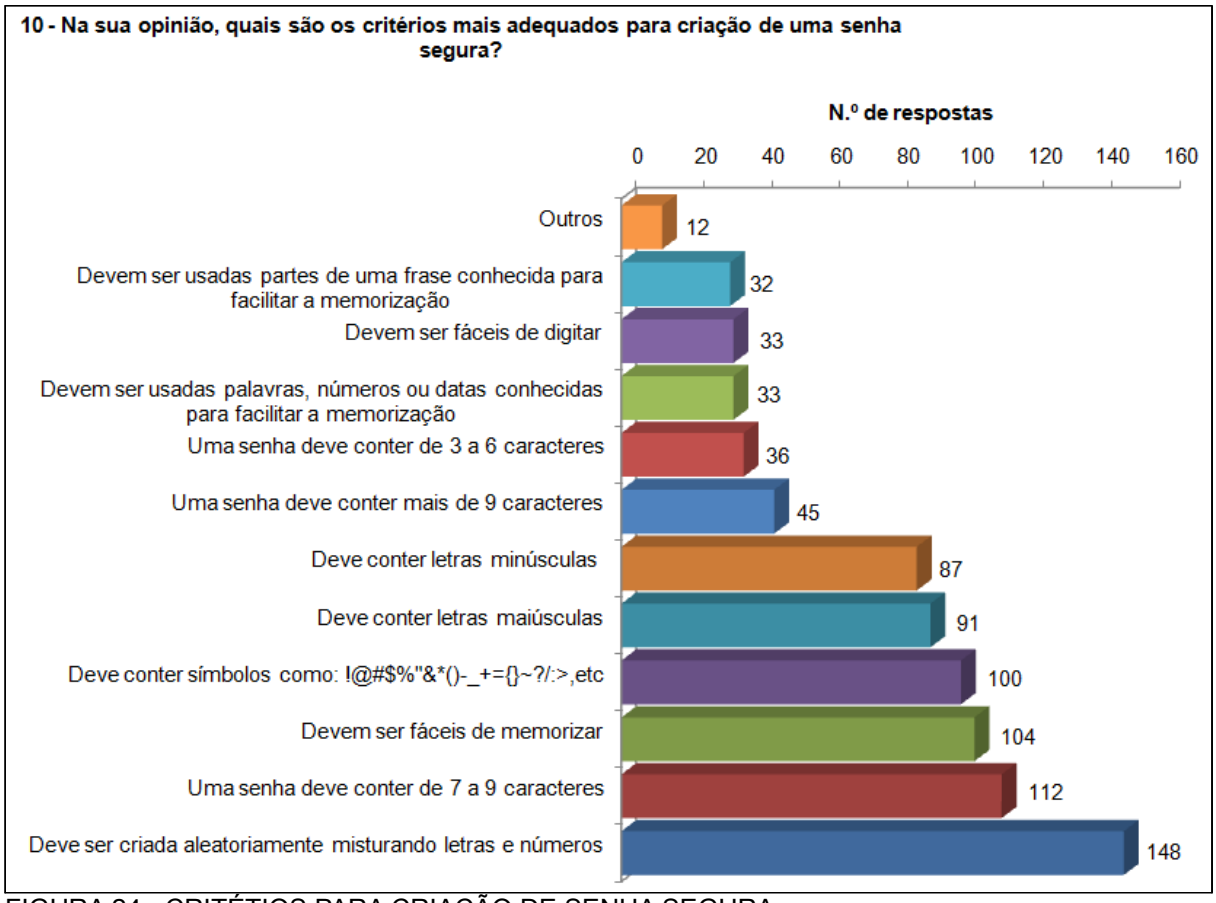

FIGURA 24 - CRITÉTIOS PARA CRIAÇÃO DE SENHA SEGURA
FONTE: O autor (2010)

Alguns mecanismos disponibilizados pelas ferramentas de navegação na internet (*browsers*) podem ser verdadeiras armadilhas, quando utilizados sem os devidos cuidados. Na questão 11 (figura 25) os usuários foram indagados sobre um desses mecanismos, cuja proposta é a memorização de senhas pelos *softwares* navegadores para que, em outro momento da navegação, o usuário não tenha que redigitar a senha. Esse mecanismo pode ser bastante perigoso, quando utilizado em locais públicos, visto que outras pessoas podem utilizar a senha armazenada para obter acesso. Para esta questão, 65% dos usuários responderam que não utilizam esse mecanismo de memorização de senhas, já 27% responderam que o utilizam. O resultado de 65% representa conformidade com a norma, entretanto há uma parcela (27%) que preocupa e deve ser alvo de conscientização, com objetivo de dar esclarecimentos sobre os perigos que essa prática representa.

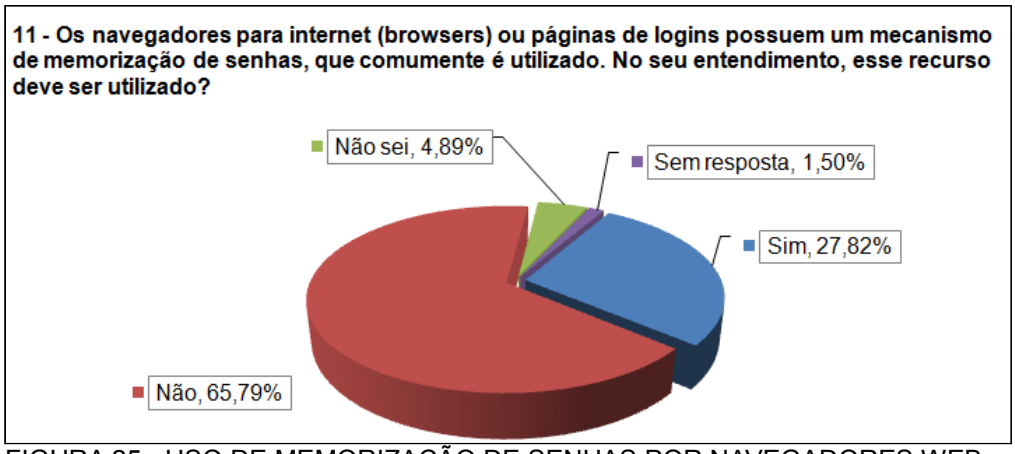

FIGURA 25 - USO DE MEMORIZAÇÃO DE SENHAS POR NAVEGADORES WEB
FONTE: O autor (2010)

A questão 12 (figura 26) é semelhante à questão 3, contudo não está restrita ao ambiente interno da instituição. O objetivo é identificar qual a estratégia utilizada pelos usuários diante da gama de *sites* que requer autenticação e, consequentemente, uso de diversas senhas. Para esta questão a resposta mais significativa foi de , cerca de 47% que afirmaram que utilizam senhas iguais para finalidades semelhantes, 5% utilizam estratégias similares aos 47%, há uma porção significativa (25%) que utiliza senhas diferentes a cada *site* e uma parcela menor, cerca de 10% que não utilizam estratégias para manter as senhas.

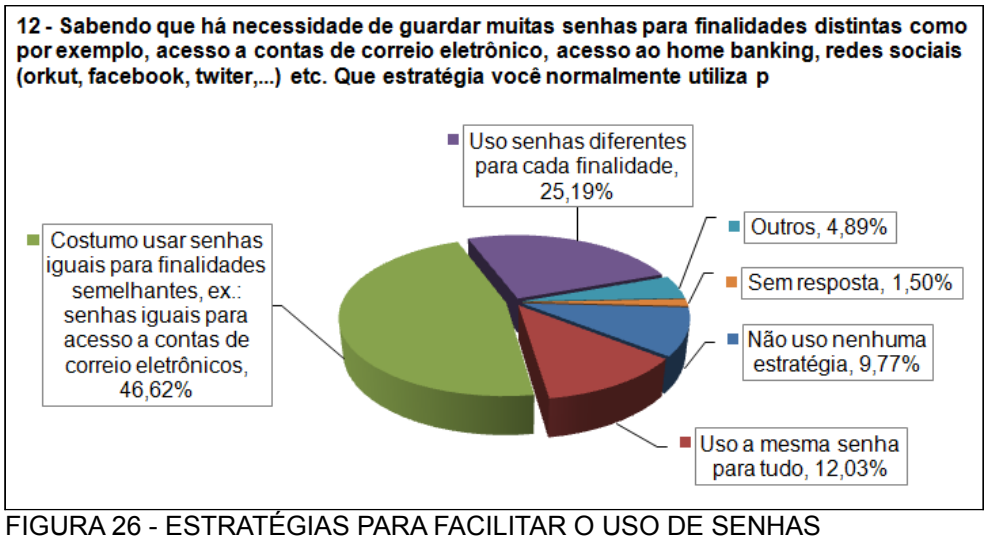

FIGURA 26 - ESTRATÉGIAS PARA FACILITAR O USO DE SENHAS
FONTE: O autor (2010)

4.1.1 Discussão (questionário)

Percebe-se, através das questões analisadas, que a maioria, principalmente as relacionadas ao comportamento do usuário, apresenta conformidade com as normas pesquisadas, como pode ser observado nas seguintes questões: a) questão 5 (figura 19) onde cerca de 67% dos usuários nunca compartilham senhas; b) questão 6 (figura 20) mais de 80% utilizam meios seguros para armazenamento de senhas; c) questão 7 (figura 21) mais de 60% costumam alterar as senhas temporárias no primeiro acesso; d) questão 10 (figura 24) a maioria dos usuários respondeu que uma senha para ser forte deve atender os requisitos de boas práticas propostos, tanto pela norma quanto pelo guia pesquisado; e) questão 11 (figura 25) onde cerca de 66% responderam que não utilizam mecanismos de memorização de senhas, configuração oferecida pelos *browsers*; f) questão 12 (figura 26) onde quase 90% utilizam algum tipo de estratégia para armazenar as diversas senhas que possuem. Por outro lado, ainda permanecem alguns aspectos que devem ser alvo de conscientização e orientação, como pode ser notado na questão 8 (figura 22), onde mais da metade dos usuários não possuem o hábito de trocar de senhas ou a frequência de troca de senha é muito maior que a recomendada. Os casos de gerenciamento de senhas poderiam ser minimizados, se os sistemas de informação e serviços tivessem mecanismos de gerenciamento de senhas.

Mesmo com bons índices persiste, ainda, uma boa percentagem de usuários que não demonstrou cuidados com a senha pessoal, o que sugere medidas de esclarecimentos, como a elaboração de um manual de orientações, quanto aos cuidados com a senha pessoal, como, por exemplo:

1. como criar senhas fortes e fáceis de lembrar;
2. estratégias para guardar as senhas em segurança;
3. como proceder ao utilizar senhas em ambientes públicos como laboratórios e outros locais públicos.

As questões relacionadas aos sistemas de informação e serviços, questões (4, 6, 7, 8 e 9) apresentam não conformidade, tanto com a norma quanto com a RFC, sob a visão dos respondentes. Dessa forma se fazem necessários processos de restruturação, visto que a maioria não leva em consideração as normas

estabelecidas pela ABNT ISO/IEC 27002 de 2005.

Os resultados do questionário sugerem investigação mais aprofundada, com o objetivo de identificar principalmente os processos de concessão de acesso, assim como os controles de acessos baseados em senha. Dessa forma será possível sugerir medidas de segurança da informação que poderão ser incorporadas às aplicações. Essas medidas visam padronizar o processo de gerenciamento de senhas e, assim, proporcionar maior segurança aos usuários, além de proporcionar maior conformidade com os padrões e normas.

4.2 ANÁLISE DOS RESULTADOS DO LEVANTAMENTO DE SISTEMAS E SERVIÇOS

O principal resultado do instrumento de pesquisa está sintetizado em uma matriz, representada no quadro 2, onde as linhas enumeradas de 1 a 9 representam os controles e as colunas representam os sistemas de informação e serviços disponibilizados pela instituição a seus usuários.

A matriz do quadro 2 representa uma ferramenta para visualização de conformidades e não conformidades dos sistemas e serviços relacionados aos controles da norma NBR ISO/IEC 27002 de 2005. Além da matriz também é possível observar o percentual aproximado de usuários que possui potencial de acesso aos sistemas de informação e serviços, mesmo não mantendo vínculo formal de ativo com a instituição pesquisada, representando possíveis acessos não autorizados.

Matriz de conformidades e não conformidades												
	Controles da norma NBR ISO/IEC 27002:2005 aplicáveis ao controle de acesso	**SIE**	**SIGEPE**	**Webmail**	**Proxy**	**VOIP**	**WTS**	**SAP**	**Thales**	**MoodleUFPR**	**MoodleCIPEAD**	**SOPHIA**
1	10.6.2 Segurança dos serviços de rede	**N**	**C**	**C**	**N**	**N**	**C**	**N**	**N**	**N**	**N**	**N/A**
2	10.10.1 Registros de auditoria	**P**	**C**	**C**	**C**	**N**	**C**	**N**	**N**	**C**	**C**	**N/A**
3	11.1.1 Política de controle de acesso	**P**	**P**	**C**	**N**	**N**	**P**	**P**	**N**	**P**	**P**	**N/A**
4	11.2.1 Registro de usuário	**N**	**P**	**P**	**P**	**N**	**N**	**N**	**N**	**N**	**P**	**N/A**
5	11.2.2 Gerenciamento de privilégio	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**N**	**C**	**C**	**N/A**
6	11.2.3 Gerenciamento de senha do usuário	**P**	**N**	**N**	**N**	**N**	**C**	**N**	**N**	**P**	**P**	**N/A**
7	11.2.4 Análise crítica dos direitos do usuário	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**C**	**N/A**
8	11.3.1 Uso de senhas	**P**	**N**	**N**	**N**	**N**	**C**	**N**	**N**	**P**	**P**	**N/A**
9	11.6.1 Restrição de acesso a informação	**N**	**C**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**C**	**N/A**
	Total de usuários com potencial de acesso	64022	84	9827	9827	167	N/A	535	2311	3491	N/A	N/A
	Total de usuários com vínculo inativo ou sem vínculo	28749	0	1787	1787	56	N/A	144	541	N/A	N/A	N/A
	Percentual aproximado de usuários com vínculo inativo ou sem vínculo	45%	0%	18%	18%	34%	N/A	27%	23%	N/A	N/A	N/A

Legenda
N – NÃO CONFORMIDADE
P – CONFORMIDADE PARCIAL
C – CONFORMIDADE
N/A – NÃO SE APLICA

QUADRO 2 - LEVANTAMENTO DOS PRINCIPAIS SISTEMAS DE INFORMAÇÃO E SERVIÇOS DA INSTITUIÇÃO, LISTANDO OS CONTROLES DA NORMA NBR ISO/IEC 27002:2005 RELACIONADOS AO CONTROLE DE ACESSO
FONTE: O autor (2010)

Para melhor entendimento do quadro 2, serão detalhados os sistemas e serviços representados no quadro. Para facilitar a descrição, os controles foram representados, sequencialmente de 1 a 9, primeira coluna do quadro 2, dessa forma a linha 1 da matriz representa o controle 10.6.2 Segurança dos serviços de rede.

SIE – Sistema de Informação para o Ensino

1. não conformidade, visto que não utiliza mecanismos de criptografia no processo de autenticação;
2. conformidade parcial, visto que apresenta registo de auditoria parcial, ou seja, armazena apenas o último usuário que executou a ação, desta forma não é possível obter trilhas de auditoria;
3. conformidade parcial, visto que não apresenta documento formal de política de controle de acesso, no entanto possui bem definidos os papéis dos usuários;

4. não conformidade, no qual, apesar do registro de usuário ser vinculado ao registro acadêmico, para alunos, foi observado percentual bastante significativo de usuários com potencial de acesso, mesmo não possuindo vínculo formal de ativo com a instituição, como pode ser visto no quadro 2;
5. conformidade, visto que as funcionalidades do sistema são agrupadas em perfis bem definidos;
6. conformidade parcial, na qual o sistema provê alguns mecanismos de gerenciamento de senhas, contudo não há processo formalizado em documento;
7. não conformidade, pelo mesmo motivo do observado no controle 4, que evidencia a ausência de análise crítica nos direitos dos usuários;
8. conformidade parcial, visto que o sistema garante troca de senhas com certa regularidade, contudo não possui todos os mecanismos para garantir senhas fortes ou senhas usadas anteriormente pelo usuário;
9. não conformidade, pelo mesmo motivo do controle de número 7, ou seja, pelo grande percentual de usuários com acesso quando não o deveria possuir;

Sistema Integrado de Gestão de Pessoal (SIGEPE)

1. conformidade, visto que utiliza mecanismos de criptografia na comunicação do processo de autenticação;
2. conformidade, visto que apresenta registros de acessos, assim como trilhas de auditoria;
3. conformidade parcial, de modo semelhante ao SIE;
4. conformidade parcial, visto que não apresenta controle formalizado de registro, contudo o controle é efetivo, de forma que não foi detectada a presença de usuários com acesso e sem vínculo formal de ativo (quadro 2), como foi observado nos outros sistemas de informação e serviços;
5. conformidade, de modo semelhante ao SIE;
6. não conformidade, visto que não apresenta mecanismos de gerenciamento de senhas como recomenda a norma;
7. não conformidade, apesar de não haver usuários com acesso indevido, não foi observado nenhum procedimento formal de análise periódica dos direitos dos usuários;

8. não conformidade, visto que não foi observado nenhum mecanismo de troca de senha, propiciando que o usuário fique sempre com a senha temporária;
9. não conformidade, uma vez que o objetivo desse controle é restringir o acesso, que não foi observado, devido ao alto índice de usuários sem vínculo formal de ativo, que possuem acesso;

Serviço de correio eletrônico (*Webmail*)

1. conformidade, de modo semelhante ao SIGEPE;
2. conformidade, visto que apresenta registros das atividades dos usuários conforme recomenda a norma;
3. conformidade, visto que apresenta normativa formalizada para regulamentar o uso do serviço;
4. conformidade parcial, visto que apresenta processo formal de registro de usuários, no entanto apresenta índice bastante elevado de usuário sem vínculo formal de ativo, cerca de 18%;
5. conformidade, de modo semelhante ao SIE;
6. não conformidade, de modo semelhante ao SIGEPE;
7. não conformidade, pelo mesmo motivo do observado, no controle 4, que evidencia a ausência de análise crítica nos direitos dos usuários;
8. não conformidade, visto que não foi observado nenhum mecanismo de auxílio para que o usuário mantenha uma senha forte;
9. não conformidade, de modo semelhante ao SIE;

Serviço de *Proxy*

O serviço segue de maneira semelhante ao serviço de correio eletrônico, considerando-se que utiliza a mesma base de dados para autenticação. A diferença está nos controles 1 ao 3.

1. não conformidade, visto que não utiliza criptografia no processo de autenticação;
2. conformidade, visto que apresenta registros das atividades dos usuários, de forma bastante detalhada, registrando não só o acesso, mas todas as atividades de navegação do usuário. Isso ocorre porque o serviço funciona como uma “ponte” de acesso aos *sites* da internet;
3. não conformidade, pois não apresenta normativa formalizada para

regulamentar o uso do serviço;

Serviço de VoIP (*Voice over Internet Protocol* ou Voz sobre Internet)

1. não conformidade, pois não utiliza mecanismos de criptografia no processo de autenticação;
2. não conformidade, visto que não fornece nenhum tipo de registro de atividades ou trilhas de auditoria;
3. não conformidade, visto não que apresenta normativa formalizada para regulamentar o uso do serviço pela instituição;
4. não conformidade, no qual o processo de registro é feito de forma quase que manual, a solicitação de nova conta é feita por mensagem de correio eletrônico, cria-se uma senha temporária, esta é encaminhada ao usuário via correio eletrônico. Esta senha fica em definitivo, uma vez que o sistema não permite a troca de senhas. A evidência de não conformidade pode ser observada no alto índice de usuários sem vínculo formal de ativo com a instituição e que possui acesso ao serviço, cerca de 34%, como pode ser visto no quadro 2;
5. conformidade, de modo semelhante ao SIE;
6. não conformidade, de modo semelhante ao SIGEPE;
7. não conformidade, pelo mesmo motivo do observado no controle 4, que evidencia a ausência de análise crítica nos direitos dos usuários;
8. não conformidade, de modo semelhante ao SIGEPE;
9. não conformidade, de modo semelhante ao SIE;

***Windows Terminal Server* (WTS)**

Não foi possível obter uma listagem das contas de usuário do serviço, devido a limitações técnicas, consequentemente, não foi possível obter o percentual de usuário sem vínculo formal de ativo.

1. conformidade, de modo semelhante ao SIGEPE;
2. conformidade, pois os registros de atividades dos usuários são armazenados;
3. conformidade parcial, de modo semelhante ao SIE;
4. não conformidade, visto que o registro ocorre de forma manual com a peculiaridade da conta não possuir vínculo com a pessoa à qual o acesso é concedido, dessa forma não é possível verificar se uma determinada conta

pertence a uma pessoa;

5. conformidade, de modo semelhante ao SIE;
6. conformidade, visto que atende os requisitos propostos pela norma NBR ISO/IEC 27002 de 2005;
7. não conformidade, pelo mesmo motivo do observado no controle 4, que inviabiliza a ausência de análise crítica nos direitos dos usuários;
8. conformidade, visto que o serviço oferece mecanismo que assegura a manutenção de senha forte pelo usuário;
9. não conformidade, pois não é possível detectar de maneira segura a que pessoa determinada conta pertence;

Sistema de Administração Patrimonial (SAP)

1. não conformidade, de modo semelhante ao SIE;
2. não conformidade, visto que não fornece nenhum tipo de registro de atividades ou trilhas de auditoria;
3. conformidade parcial, de modo semelhante ao SIE;
4. não conformidade, no qual o registro de contas de usuário ocorre de forma manual, não está vinculado a nenhum processo cadastral de pessoa, dessa forma cerca de 27% das contas de usuário do sistema não possuem vínculo formal de ativo na instituição;
5. conformidade, de modo semelhante ao SIE;
6. não conformidade, de modo semelhante ao SIGEPE;
7. não conformidade, pelo mesmo motivo do observado no controle 4, que evidencia a ausência de análise crítica nos direitos dos usuários;
8. não conformidade, de modo semelhante ao SIGEPE;
9. não conformidade, de modo semelhante ao SIE;

Sistema de Registro de Pesquisa (Thales)

1. não conformidade, de modo semelhante ao SIE;
2. não conformidade, de modo semelhante ao SAP;
3. conformidade parcial, de modo semelhante ao SIE;
4. não conformidade, de modo semelhante ao SAP, diferindo apenas no percentual de contas de usuário sem vínculo formal de ativo na instituição, que atinge cerca de 23% das contas;

5. não conformidade, visto que as funcionalidades do sistema não estão agrupadas em perfis;
6. não conformidade, de modo semelhante ao SIGEPE;
7. não conformidade, pelo mesmo motivo do observado no controle 4, que evidencia a ausência de análise crítica nos direitos dos usuários;
8. não conformidade, de modo semelhante ao SIGEPE;
9. não conformidade, de modo semelhante ao SIE;

Plataforma de Educação a Distância (MoodleUFPR)

Para o MoodleUFPR, diferentemente do WTP, foi possível obter a lista de contas de usuários, mas como não há critérios para criação de contas de usuários (o próprio usuário que a cria) não foi possível verificar o vínculo dessas contas com as pessoas da instituição.

1. não conformidade, de modo semelhante ao SIE;
2. conformidade, visto que apresenta registro das atividades do usuário;
3. conformidade parcial, de modo semelhante ao SIE;
4. não conformidade, de modo semelhante ao WTS;
5. conformidade, de modo semelhante ao SIE;
6. conformidade parcial, de modo semelhante ao SIE;
7. não conformidade, pelo mesmo motivo do observado no controle 4, que inviabiliza a ausência de análise crítica nos direitos dos usuários;
8. conformidade parcial, visto que o sistema apresenta mecanismos que possibilita que o usuário mantenha uma senha forte, contudo não atende todas as recomendações previstas na norma, tais como histórico de senhas, vencimento de senhas em tempos regulares, entre outras;
9. não conformidade, de modo semelhante ao WTS;

Plataforma de Educação a Distância (MoodleCIPEAD)

Não foi possível obter o quantitativo de contas de usuário por falta de padrões para criação de contas de usuário. Os usuários deste sistema estão divididos em dois grandes grupos: a) usuários de cursos regulares; e b) usuários de cursos eventuais que não constam no sistema acadêmico. Essa segmentação de grupos não possibilitou a verificação de contas de usuário com os sistemas institucionais principais (acadêmico e recursos humanos).

1. não conformidade, de modo semelhante ao SIE;
2. conformidade, visto que apresenta registro das atividades do usuário;
3. conformidade parcial, de modo semelhante ao SIE;
4. conformidade parcial, no qual o registro da conta do usuário é feito por meio de processo formal adotando como chave de vínculo um documento oficial da pessoa vinculada à conta.
5. conformidade, de modo semelhante ao SIE;
6. conformidade parcial, de modo semelhante ao SIE;
7. conformidade, visto que há análise periódica dos direitos dos usuários com objetivo de evitar acessos não autorizados. Um dos fatores que contribuem para essa prática é que os usuários, principalmente alunos e docentes (tutores), estão vinculados diretamente à atividade acadêmica. Dessa forma ao encerrar uma turma, por exemplo, todos perdem o acesso;
8. conformidade parcial, de modo semelhante ao MoodelUFPR;
9. conformidade, de modo semelhante ao WTS;

4.2.1 Comentários

O maior problema, detectado nos sistemas de informação e serviços, está relacionado ao controle 4 da matriz (11.2.1 Registro de usuários), posto que nenhum sistema apresentou conformidade com este controle. O alto índice de usuários com acesso, sem vínculo formal de ativo com a instituição indica que o processo de retirada de acesso não ocorre de maneira adequada. Esse comportamento pode ser observado, em quase todos os recursos, pois apresentam discordância com todos os controles relacionados ao gerenciamento de acesso do usuário, contido na seção 11.2 da norma NBR ISO/IEC 27002:2005, ou controles de 3 a 9 da matriz de conformidades e não conformidades (quadro 2).

Outra questão, bastante preocupante, está relacionada aos fluxos de autenticação. Somente três dos dez recursos pesquisados apresentam uso de criptografia nos processos de comunicação.

A falta de gestão de usuários e, consequentemente, os problemas de controle de acesso, a falta de mecanismos, que garantam segurança no tráfego de dados

sensíveis, aliados a maus hábitos de usuários, principalmente, os relacionados à falta de trocas regulares de senhas, contribuem para insegurança do ambiente de tecnologia da informação e comunicação. Esta insegurança, por sua vez pode ser convertida em incidentes de segurança.

4.3 PROTÓTIPO INTEGRAÇÃO DE BASE DE DADOS

O protótipo de integração de base de dados foi elaborado com fundamento no referencial teórico, seguindo a metodologia estabelecida na seção 2.1 desta pesquisa.

4.3.1 Etapa 1 – considerações iniciais

Esta fase é marcada pela seleção das bases a serem integradas, assim como a definição do modelo comum de dados. Para integração de bases heterogêneas foram selecionados três SGBDs, que representam os esquemas:

1. esquema acadêmico – armazena dados de alunos e demais dados administrativos do sistema;
2. esquema recursos humanos – armazena servidores técnicos e docentes;
3. esquema contas de correio eletrônico – armazena dados de contas de usuários, tais como identificador e senha.

A visão integrada foi composta por oito entidades, ou tabelas de banco de dados relacional, para representar a visão integrada dos dados (figura 27). Esse modelo foi elaborado com base no modelo de dados proposto pela ferramenta de importação e exportação de dados (EID), versão 1.2.1.

A figura 27 representa o diagrama de entidade e relacionamentos do modelo comum de dados. As entidades do diagrama serão detalhadas, nos quadros 3 a 8. A representação utilizada é uma forma simplificada de dicionário de dados.

FIGURA 27 - VISÃO INTEGRADA – MODELO COMUM DE DADOS
FONTE: O autor (2010)

A principal entidade do modelo é a identificação (quadro 3), considerando-se que todas dependem dela para existir. Seu objetivo é armazenar dados pessoais, portanto pode também ser considerada entidade pessoa por armazenar dados relativos à pessoa.

Identificação (pessoa)			
Campo	**Tipo de dado**	**Tamanho**	**Descrição**
ID	INTEIRO	-	Identificador único para cada pessoa
CPF	CARACTERE	11	Número de Cadastro Nacional de Pessoa Física (campo chave)
NOMECOMPLETO	CARACTERE	120	Nome completo
NUMEROIDENTIDADE	CARACTERE	20	Número do Registro Geral da pessoa
ORGAOEMISSORIDENTIDADE	CARACTERE	4	Órgão que emissor do Registro Geral da pessoa
UFIDENTIDADE	CARACTERE	2	Unidade da Federação que emitiu o Registro Geral da pessoa
PASSAPORTE	CARACTERE	20	Número do passaporte para estrangeiros
ESTADOCIVIL	CARACTERE	20	Estado Civil
SEXO	CARACTERE	1	Sexo
NACIONALIDADE	CARACTERE	120	Nacionalidade
DATANASCIMENTO	DATA	-	Data de nascimento
CIDADENASCIMENTO	CARACTERE	120	Cidade de nascimento
ESTADONASCIMENTO	CARACTERE	2	Unidade da Federação de nascimento
PAISNASCIMENTO	CARACTERE	120	País de nascimento

QUADRO 3 - DICIONÁRIO DE DADOS - ENTIDADE IDENTIFICAÇÃO
FONTE: O autor (2010)

A entidade conta (quadro 4) está disposta separadamente da identificação, apesar do tipo de relação ser de um para um (uma conta possui uma identificação ou uma identificação possui uma conta). Nos casos de relacionamento um para um, normalmente, usa-se representar os atributos na mesma entidade, exceto em casos

que haja necessidade de segregação. A opção pelo modelo foi separar, visto que a arquitetura de implementação da integração adotada requer sincronizações mais frequentes em algumas entidades que em outras. Esta entidade possui a função de armazenamento de dados relativos a nome de usuário e senha.

Conta			
Campo	**Tipo de dado**	**Tamanho**	**Descrição**
ID	INTEIRO	-	Identificador único para cada pessoa
USUARIO	CARACTERE	30	Identificação do usuário
SENHA	CARACTERE	35	Senha do usuário

QUADRO 4 - DICIONÁRIO DE DADOS - ENTIDADE CONTA
FONTE: O autor (2010)

A entidade aluno (quadro 5), professor (quadro 6) e técnico (quadro 7) representam os tipos de pessoas vinculadas à instituição: aluno, professor e técnico, respectivamente. Cada pessoa representada possui seus respectivos dados pertinentes a cada papel desempenhado na instituição. O relacionamento desses tipos de pessoas com identificação (figura 27) é do tipo é um, isto é, um aluno é uma pessoa, da mesma forma, professor é uma pessoa, assim com técnico também é uma pessoa.

Aluno			
Campo	**Tipo de dado**	**Tamanho**	**Descrição**
ID	INTEIRO	-	Identificador único para cada pessoa
MATRICULA	CARACTERE	20	Matrícula do aluno
NOMECURSO	CARACTERE	120	Nome do curso
DATAVINCULO	DATA	-	Data de início do vínculo com a instituição
DATAAFASTAMENTO	DATA	-	Data de fim do vínculo com a instituição
NIVELCURSO	CARACTERE	20	Nível de curso (graduação, pós-graduação e ensino médio)

QUADRO 5 - DICIONÁRIO DE DADOS - ENTIDADE ALUNO
FONTE: O autor (2010)

Professor			
Campo	**Tipo de dado**	**Tamanho**	**Descrição**
ID	INTEIRO	-	Identificador único para cada pessoa
MATRICULA	INTEIRO	-	Matrícula do funcionário
DATAADMISSAO	DATA	-	Data de admissão
DATAAFASTAMENTO	DATA	-	Data de afastamento
ESTADOVINCULO	CARACTERE	20	Edtado do vínculo do funcionário (ativo, afastado, aposentado)
CLASSE	CARACTERE	20	Classe do professor (auxiliar, assistente, adjunto, associado, titular)
NIVEL	CARACTERE	1	Nível da classe do professor (1, 2, 3 e 4)
TITULACAO	CARACTERE	120	Maior titulação do professor

QUADRO 6 - DICIONÁRIO DE DADOS - ENTIDADE PROFESSOR
FONTE: O autor (2010)

Técnico			
Campo	**Tipo de dado**	**Tamanho**	**Descrição**
ID	INTEIRO	-	Identificador único para cada pessoa
MATRICULA	INTEIRO	-	Matrícula do técnico
DATAADMISSAO	DATA	-	Data de admissão
DATAAFASTAMENTO	DATA	-	Data de afastamento
ESTADOVINCULO	CARACTERE	20	Edtado do vínculo do funcionário (ativo, afastado, aposentado)
FUNCAOPRINCIPAL	CARACTERE	20	Classe do professor (auxiliar, assistente, adjunto, associado, titular)
NIVELCAPACITACAO	CARACTERE	1	NÍvel de capacitação (superior, especialização, mestrado etc)
CLASSE	CARACTERE	1	Classe do técnico (A, B, C, D e E)
PADRAO	CARACTERE	3	Padrão de vencimento (100 a 116)

QUADRO 7 - DICIONÁRIO DE DADOS - ENTIDADE TÉCNICO
FONTE: O autor (2010)

As entidades correio eletrônico (quadro 8), telefone (quadro 9) e endereço (quadro 10) representam dados complementares de uma pessoa. A entidade correio eletrônico armazena dados relativos às contas de correio eletrônico de uma identificação (pessoa), da mesma forma a entidade telefone armazena dados relativos aos contatos de telefone de uma identificação (pessoa), assim como a entidade endereço armazena dados relativos aos endereços de uma identificação (pessoa). O relacionamento dessas entidades com identificação (figura 27) é do tipo um para muitos, ou melhor, uma pessoa possui várias contas de correio eletrônico, números de telefone ou endereços.

Correio eletrônico			
Campo	**Tipo de dado**	**Tamanho**	**Descrição**
ID_EMAIL	INTEIRO	-	Identificador único para telefone
ID	INTEIRO	-	Identificador único para cada pessoa
EMAIL	CARACTERE	120	Endereço de correio eletrônico
TIPO	CARACTERE	20	Tipo do email (principal, alternativo)

QUADRO 8 - DICIONÁRIO DE DADOS - ENTIDADE CORREIO ELETRÔNICO
FONTE: O autor (2010)

Telefone			
Campo	**Tipo de dado**	**Tamanho**	**Descrição**
ID_TELEFONE	INTEIRO	-	Identificador único para telefone
ID	INTEIRO	-	Identificador único para cada pessoa
TELEFONE	CARACTERE	20	Número de telefone
TIPO	CARACTERE	20	Tipo do telefone (residencial, comercia, celular ou recados)

QUADRO 9 - DICIONÁRIO DE DADOS - ENTIDADE TELEFONE
FONTE: O autor (2010)

Endereço			
Campo	**Tipo de dado**	**Tamanho**	**Descrição**
ID_ENDERECO	INTEIRO	-	Identificador único para endereco
ID	INTEIRO	-	Identificador único para cada pessoa
LOGRADOURO	CARACTERE	120	Logradouro, tipo e nome
NUMERO	CARACTERE	10	Número do logradouro
COMPLEMENTO	CARACTERE	30	Complemento do logradouro
BAIRRO	CARACTERE	120	Bairro
CIDADE	CARACTERE	120	Cidade
ESTADO	CARACTERE	2	Unidade da Federação
PAIS	CARACTERE	120	Sigla do país com duas letras
CEP	CARACTERE	10	CEP
TIPO	CARACTERE	20	Tipo do endereço (residencial, comercial etc)

QUADRO 10 - DICIONÁRIO DE DADOS - ENTIDADE ENDEREÇO
FONTE: O autor (2010)

O vínculo funcionário terceirizado não foi contemplado no modelo. Embora seja de grande importância construir um modelo que contenha todos os tipos de pessoas que possuem algum tipo de relacionamento com a instituição, não foi possível obter dados formalizados e consistentes em bancos de dados desse tipo de vínculo.

4.3.2 Etapa 2 – Pré-Integração

Nesta fase, define-se o escopo da integração, quais esquemas serão integrados e qual é a estratégia de integração.

Os esquemas fonte selecionados para integração são provenientes dos bancos de dados do sistema acadêmico, do sistema de recursos humanos e do banco de dados de contas de correio eletrônico. Neste caso trata-se de integração parcial, visto que o foco é integração de identidades, cujo modelo comum de dados é composto de oito entidades ou tabelas (identificação, conta, aluno, professor, técnico, correio eletrônico, telefone e endereço – figura 27).

A estratégia de processamento de integração foi do tipo n-ária, em um único passo (*one-shot*), de modo que o processo de integração foi realizado em uma única etapa.

4.3.3 Etapa 3 – identificação de correspondências

Buscou-se, nesta fase, o estabelecimento de correspondências entre o modelo comum e as bases fonte e também possibilitou a identificação de conflitos e resolução de parte destes.

É importante destacar que o objetivo principal dessa integração é promover uma base integrada de identidades, que será utilizada somente para leitura. Isso possibilitou que a correspondência entre os esquemas fosse feita, por meio de visões de banco de dados construídas nas bases fontes (apêndice B). Dessa forma cada fonte de dados possui um conjunto de visões que representa o modelo comum de dados. A exceção está na entidade conta, posto que, somente, o esquema contas de correio eletrônico possui a entidade conta representada no modelo comum. Assim, a correspondência, foi direta, de modo que a visão conta do esquema contas de correio eletrônico equivale à entidade conta do modelo comum. No caso dos

outros dois esquemas fonte (acadêmico e recursos humanos), as visões foram criadas para corresponder exatamente às entidades do modelo comum de dados. Sendo assim a visão identificação do esquema acadêmico equivale à entidade identificação do modelo comum, assim como, a visão identificação do esquema recursos humanos equivale à entidade do modelo comum e assim sucessivamente. Portanto o mapeamento de integração foi feito, com base em visões de banco de dados.

4.3.4 Etapa 4 – identificação de conflitos

Durante a etapa de identificação de correspondências, foi possível identificar e resolver grande parte dos conflitos, principalmente, os estruturais – relativos a sinônimos, tipo de dados, unicidade, precisão, valor padrão, incompatibilidade de união, isomorfismo, falta de atributo etc. Os conflitos relativos a dados, tais como duplicação de identidade, foram tratados, na etapa de união e reestruturação. A descrição detalhada dos conflitos encontrados e resoluções estão detalhadas no apêndice B.

4.3.5 Etapa 5 – integração

Nesta etapa é realizada a integração propriamente dita, que pode ser realizada de forma manual por um profissional de banco de dados, ou por meio de ferramentas que automatizam o processo.

Foi selecionada, para esta fase, a arquitetura de integração baseada na abordagem de visões materializadas, também chamada de *data warehouse.* A ferramenta utilizada para sincronização das bases foi o EID (*Export / Import Directory Tool*), versão 1.2.1, desenvolvida pela Universidade Federal de Minas Gerais (UFMG). EID é uma ferramenta para auxiliar a importação e exportação de dados entre fontes de dados distintas. Suas principais características são: a) importação de dados de bases legadas; b) detecção de registros duplicados, atuando como agente conciliador de dados; c) exposição dos dados importados via serviço web; d) conciliação automática e manual de registros duplicados; e) manutenção manual dos registros importados (UFMG, 2009).

O EID atua com agente integrador, conectado nas bases selecionadas e

extraindo os dados para a base integrada (figura 27), tratada pelo EID como base intermediária ou metabase, na qual ocorre processamentos de conciliação de dados, que consiste em duplicação dos dados oriundos das bases fonte.

Esta etapa será mais detalhada no protótipo de SSO, seção 4.5 deste capítulo.

4.3.6 Etapa 6 – união e reestruturação

Ao final do processo de integração, é possível comparar os dados fonte com os dados no modelo conceitual, de modo que o quantitativo de identidade do modelo integrado pode ser comparado como quantitativo de identidade das bases fonte.

4.4 GESTÃO DE IDENTIDADE

As melhorias sugeridas, para o processo de gestão de identidade, estão divididas em dois grupos: a) recomendações quanto ao registro de usuários; b) adoção de um sistema de gerenciamento de identidade.

Para padronizar o registro de usuários a proposta é estabelecer um processo de provisionamento de usuário, indicado aqui como ***kit* de provisionamento de usuários**.

O *kit* de provisionamento será composto de um conjunto de documentos personalizados que apresentará quais recursos estarão disponíveis e terá objetivo de auxiliar os novos usuários a usarem de maneira adequada os recursos de TI oferecidos pela instituição. Os documentos sugeridos para este *kit* são: a) descritivo dos recursos de TI; b) recomendações quanto ao uso de senha; c) documento com dados da identificação única de acesso.

O descritivo dos recursos de TI deverá conter informações relativas aos serviços e sistemas de informação oferecidos, como e onde obter acesso a esses recursos e que nível de acesso será disponibilizado. Cada recurso descrito deve conter a importância deste para o usuário, a quem recorrer, caso haja dúvidas sobre o recurso, os níveis de acessos disponibilizados e a política de uso, caso haja.

O documento de recomendações, quanto ao uso de senha pessoal, deve ser bastante consistente, de modo a evidenciar a importância desse dado confidencial. O documento deverá conter exemplos de como proteger a senha pessoal, exemplos

de como elaborar uma senha forte, e deve conter, ainda, de maneira destacada, as consequências da perda ou vazamento dessa informação, assim como quais procedimentos adotar, caso ocorra tal incidente.

O *kit* não está limitado à entrega de documentos, mas a um processo de provisionamento de usuário. Além dos documentos físicos entregues é o momento em que ocorre a criação da identificação única de acesso institucional, com propósito de disponibilizar acesso aos recursos de TI, disponibilizados pela instituição. Às pessoas que, por ventura, já possuem esta credencial, ocorre a renovação da mesma. A criação do identificador único implicará na criação de conta no sistema de gestão de identidade institucional e, consequentemente, emissão em formato impresso do identificador e senha temporária para primeiro acesso.

O *kit* deverá ser entregue para os alunos, no momento da confirmação do registro acadêmico; para os servidores no momento da posse e para os funcionários terceirizados no primeiro dia de trabalho. Na entrega do kit é necessária a apresentação de documentos que comprovem a identidade do usuário, assim como a assinatura do termo de responsabilidade, comprovando a retirada do *kit* e compromisso de zelo pelas regras e normativas institucionais.

Como apoio ao *kit* de provisionamento, é necessário um sistema de informação para gerenciamento de identidade, que dê suporte a este processo, de modo que seja possível a implementação e cumprimento dos controles de segurança estabelecidos pela norma NBR ISO/IEC 27002:2005, tais como rotinas periódicas de checagem de vínculos, a fim de desativar as contas de usuários com vínculos inativos, gerenciamento acerca das senhas de usuário, entre outros controles descritos na norma.

Uma das primeiras atividades necessárias, para implementação de controle de acesso centralizado, é a consolidação de procedimentos de gestão de identidade, pois, um simples vazamento de uma senha, por exemplo, pode comprometer vários recursos de TI, enquanto que, no modelo descentralizado, o impacto pode ser menor. Portanto este processo de gestão de identidade deverá contemplar, tanto as questões relacionadas à conscientização das pessoas, quanto do sistema de informação responsável por gerenciar a identidade digital propriamente dita.

4.5 PROTÓTIPO DE SSO DO MODELO PROPOSTO

O protótipo consiste na implementação de um projeto de SSO e disponibilização em caráter de teste para a comunidade acadêmica. Para este protótipo é necessária a implementação do modelo conceitual de integração de base de dados proposto na seção 4.3 deste capítulo, uso de ferramentas para manutenção das visões materializadas de identidade e ferramenta segura de SSO. Outro fator, a ser levado em consideração, é a adoção de procedimentos formais de segurança acerca do uso de senha pessoal e provisionamento de usuários, para regulamentar o ciclo de vida de identidade do usuário.

Para implementação do protótipo, assumiu-se que a base de identidade de usuários e processos de gerenciamento são efetuados, na base de dados de Contas de Correio Eletrônico.

O modelo de autenticação centralizado, que pode ser observado na figura 28, consiste na integração de três bases de dados (SGBD Acadêmico, SGBD Recursos Humanos e SGBD Contas de Correio Eletrônico) em uma base centralizada de identidades (Metabase), por meio da ferramenta de integração EID, que faz a sincronização e conciliação dos dados dos SGBDs para a Metabase e disponibiliza os dados conciliados por meio de *web service.* Posteriormente, a ferramenta EID2LDAP (*Export Import Data Tool* – módulo de exportação para bases LDAP), versão 1.1.1, acessa os dados contidos na Metabase por meio do *web service* do EID e sincroniza os dados para o Diretório, que se configura como o repositório de identidades, podendo ser acessado por provedores de identidades (IDPs), ou mesmo por outros recursos de TI, tais como sistemas, equipamentos de rede ou sistemas operacionais.

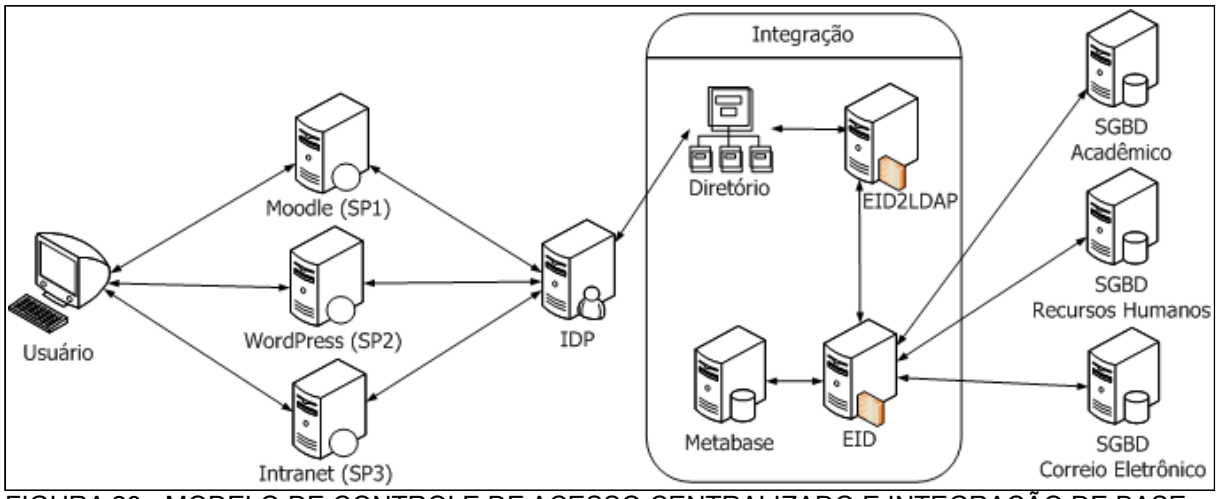

FIGURA 28 - MODELO DE CONTROLE DE ACESSO CENTRALIZADO E INTEGRAÇÃO DE BASE DE DADOS
FONTE: O autor (2010)

A figura 28 mostra o processo de autenticação de maneira simplificada, de modo que o usuário acessa um dos provedores de serviço (Moodle (SP1), WordPress (SP2) ou Intranet (SP3)). O SP por sua vez redireciona o usuário para autenticação no provedor de identidade (IDP). O IDP confere as credenciais do usuário e redireciona para o serviço requisitado. Uma vez autenticado, o usuário poderá acessar qualquer provedor de serviços que esteja disponível, sem a necessidade de realizar novamente o processo de autenticação.

O EID atua como agregador de identidades. Acessando as bases relacionais, recupera os dados, realiza processamentos de conciliação de dados duplicados por meios de algoritmos, por fim deposita os dados resultantes no esquema de banco dados definido, anteriormente, neste caso, no modelo definido na seção 4.3 (figura 27). O EID possui ainda a função de disponibilização dos dados por meio de *web servisse.* Dessa forma os dados disponibilizados pelo EID podem ser acessados por outros sistemas, como é o caso do EID2LDAP, que possui o papel de sincronizador, de modo que acessa os dados disponibilizados pelo EID e os sincroniza para uma base de dados LDAP (Diretório) preparada para receber os dados relativos à identidade.

O modelo de dados utilizado no diretório LDAP foi elaborado pelo projeto Comunidade Acadêmica Federada (CAFe), criado com o objetivo de padronizar classes e atributos utilizados pelas Instituições de Ensino Superior (IES) e centros de pesquisas.

O projeto CAFe, sob a coordenação da Rede Nacional de Ensino e Pesquisa

(RNP) possui o objetivo de promover a integração entre as instituições de ensino e centros de pesquisas brasileiros, por meio de uma rede federada de confiança. Essa integração possibilita que haja compartilhamento de recursos e serviços entre as instituições parceiras. Além da padronização da estrutura de diretórios, o projeto possui infraestrutura para treinamento e operação dos recursos, para implementação do ambiente federado na instituição interessada. O projeto recomenda o uso dos softwares EID, MySQL, EID2LDAP, OpenLDAP[14] e Shibboleth2 para implementação do ambiente federado.

Este protótipo limita-se à preparação do ambiente de autenticação centralizada, que é um dos requisitos para implementação do ambiente de federação CAFe. O ambiente federado envolve questões de políticas de divulgação de dados para outras instituições e entidades parceiras. Requer, portanto, autorização da alta administração da instituição, e dessa forma fica fora do escopo deste trabalho acadêmico.

4.5.1 Recursos utilizados

O protótipo foi elaborado, tendo como ambiente uma universidade pública federal brasileira, que disponibiliza diversos tipos de serviços e sistemas de informação (quadro 1) a seus usuários.

Os recursos disponibilizados neste protótipo para testes serão três sistemas, distribuídos em três provedores de serviços (SPs) distintos, conforme o modelo da figura 28:

1.Moodle – plataforma de educação a distância, utilizado por vários órgãos públicos brasileiros, tais como Ministério da Educação (MEC), Universidade Federal do Paraná (UFPR), Universidade Federal da Bahia (UFBA), entre outros. O Moodle é uma ferramenta de código aberto desenvolvido, que utiliza a linguagem PHP (Personal Home Page), também de código aberto;

2.WordPress – sistema utilizado para disponibilizar e padronizar páginas *web* , em uma única plataforma. Dentre as principais características da ferramenta estão a facilidade de criação de páginas e a disponibilização de conteúdo de forma gerenciável. A ferramenta é de código aberto e desenvolvida em PHP, podendo ser

[14] OpenLDAP – é um *software* de código aberto que implementa o protocolo *Lightweight Directory Access Protocol* (LDAP).

adaptada sem pagamento de licenças ou direitos autorais;

3.Intranet – sistema disponibilizado pela instituição pesquisada para seus usuários, que possibilita a criação de contas de correio eletrônico, atualização de dados pessoais, visualização de dados pessoais, acadêmicos, patrimoniais etc. O sistema, desenvolvido utilizando a tecnologia Java, é de uso restrito à instituição pesquisada.

Os *softwares* utilizados foram EID, MySQL, EID2LDAP, OpenLDAP e Shibboleth, todos *softwares* de código aberto, portanto sem custos de licenças ou direitos autorais. O processo de instalação destes *softwares* consta no apêndice C e apêndice D. A descrição detalhada segue:

1. EID – aplicativo *web* , utilizado para importação dos dados relativos a identidades dos bancos de dados relacionais e armazenamento no *data warehouse*. No EID também é possível fazer a gestão manual de identidade, de forma que é possível adicionar, ou remover uma identidade do EID;
2. MySQL – utilizado pelo EID como *data warehouse ,*para armazenamento dos dados oriundos das fontes de dados selecionadas (sistema recursos humanos, sistema acadêmico e contas de correio eletrônico). O MySQL atua como base intermediária de agregação e manutenção de identidades;
3. EID2LDAP – aplicativo *web,* que atua como sincronizador, de forma que os dados disponibilizados pelo EID são sincronizados com a base LDAP gerenciada pelo OpenLDAP;
4. OpenLDAP – serviço que gerencia a base LDAP de identidades. Por meio desse serviço qualquer sistema operacional, equipamento de rede, ou sistema de informação poderá fazer uso da base LDAP para seus processos de autenticação. Duas características são mais expressivas em uma base LDAP: a) o alto desempenho em operações de busca; b) a quantidade de recursos de TI capaz conectar com esse serviço;
5. Shibboleth2 – ferramenta de *Single Sign-On* utilizada para prover controle de acesso centralizado. Shibboleth2 se divide em dois serviços:
   a) IDP – provedor de identidades, responsável pelo processo de autenticação e busca de atributos na base de identidades;
   b) SP – provedor de serviços, que atua como agente, fazendo a

comunicação entre a aplicação e o IDP;

Uma das principais características do Shibboleth2 está relacionada à segurança de comunicação, visto que todos os processos de comunicação entre o IDP e SP são efetuados, usando criptografia, de forma que, somente, os SPs registrados no IDP poderão acessá-lo. Outro fator bastante interessante, ainda na questão de segurança, está relacionado ao nível de proteção do aplicativo. Neste caso, quem faz a proteção da aplicação é o próprio provedor de serviço, atuando com um filtro, possibilitando que o usuário tenha acesso à determinada aplicação, somente se tiver passado pelo processo de autenticação, pelo menos uma vez. Portanto retira a responsabilidade da instituição desenvolver mecanismo de autenticação, uma vez que este mecanismo está disponível no SP e IDP.

Para instalação da infraestrutura, a instituição pesquisada disponibilizou seis servidores com sistema operacional GNU/Linux. Os servidores foram configurados para prover acesso somente no ambiente interno da instituição, por medida de segurança. A disposição dos serviços e domínios pode ser observada no quadro 11.

Nº	Serviço(s)	Domínio	Aplicação *web*
1	Apache HTTPD e Apache Tomcat6	idp.cce.ufpr.br	Shibboleth2 IDP
2	Shibboleth2 - SP, Apache HTTPD	moodle.cce.ufpr.br	Moodle
3	Shibboleth2 - SP, Apache HTTPD	wordpress.cce.ufpr.br	WordPress
4	Shibboleth2 - SP, Apache HTTPD e Apache Tomcat6	intranet.cce.ufpr.br	Intranet
5	MySQL, Apache HTTPD e Apache Tomcat6	eid.cce.ufpr.br	EID-1.2.1 e EID2LDAP-1.1.1
6	OpenLDAP	ldap.cce.ufpr.br	------

QUADRO 11 - SERVIDORES E SERVIÇOS UTILIZADOS NO PROTÓTIPO DE AUTENTICAÇÃO CENTRALIZADA
FONTE: O autor (2010)

4.5.2 Execução da integração

A execução se deu conforme o modelo descrito (figura 28), o EID conectando diretamente às fontes de dados e sincronizando com a metabase.

A ferramenta selecionada para sincronização das bases não atendeu como esperado, comprovando o que foi observado no referencial teórico, a respeito da abordagem de visões materializadas, cuja tarefa mais difícil desse tipo de implementação está na sincronização das fontes com o *data warehouse*, aqui representado pela metabase.

Foram efetuadas tentativas e, devido ao volume de dados para conciliação, cerca de 54 mil registros de identificação para a base de dados acadêmicos, 14 mil

da base de recursos humanos e cerca de 10 mil da base de contas de correio eletrônico, o processo de integração não foi finalizado.

Durante os testes, foi agendado o processamento de dados de identificação da base acadêmica, cerca de 54 mil registros de alunos. Esse processo foi executado, demorou oito dias e não foi concluído. Então, o processo foi parado, feita a limpeza da metabase e nova tentativa, nesta com cerca de 400 registros. Para esta amostra, a conciliação ocorreu rapidamente em menos de 5 minutos, portanto a causa foi a quantidade de registros para conciliação, o que inviabilizou o uso da ferramenta como integrador de dados. Como a ferramenta está em constante desenvolvimento, é de se esperar que nas próximas versões esta questão de performance seja resolvida.

A estratégia encontrada foi a criação de programas de replicação, de modo que a base de dados acadêmicos e contas de correio eletrônico foram replicados para a base de recursos humanos. A decisão de replicar os dados para a base de recursos humanos se deu porque esta foi modelada para suportar todas as pessoas da instituição, ou seja, alunos, docentes, técnicos administrativos e pessoal terceirizado. Desta forma o modelo de controle de acesso e integração (figura 28) sofreu algumas adaptações para que a integração fosse viabilizada.

O modelo adaptado (figura 29) sofreu alteração de modo que o EID passa a funcionar como sincronizador de apenas uma base de dados para a metabase.

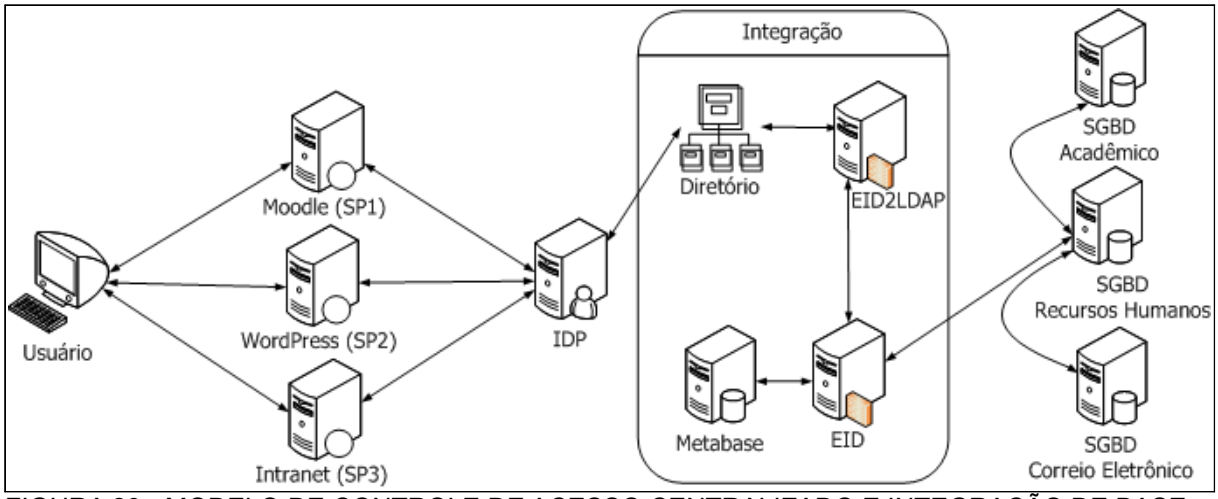

FIGURA 29 - MODELO DE CONTROLE DE ACESSO CENTRALIZADO E INTEGRAÇÃO DE BASE DE DADOS ADAPTADO PARA REPLICAÇÃO
FONTE: O autor (2010)

Na figura 29, as duas setas que partem de SGBD acadêmico e correio

eletrônico representam o fluxo dos processos de integração desenvolvidos, utilizando a linguagem de programação Java. Esses processos realizam operações de união e conciliação, ao final se tem uma base de dados pessoais, completa e coesa. Os processos são executados duas vezes por semana e duram cerca de 50 minutos.

Outra medida também necessária foi a adaptação do EID para funcionar sem a necessidade de aplicar algoritmos de conciliação, portanto foi retirada a capacidade de conciliação, uma vez que os dados já haviam passado por esse processo.

Com todas as adaptações realizadas, foi possível a criação dos processos individuais no EID para população da metabase e, posteriormente, o diretório de identidades por meio do EID2LDAP. Ao final dos processamentos, tem-se um diretório LDAP com dados atualizados de contas de usuários.

4.5.3 Testes de acesso

Uma vez construído o diretório LDAP, implementado o provedor de identidades e de serviços, foi possível realizar os testes da arquitetura de SSO. Os testes apresentam, de forma simplificada, as ações de um usuário para acessar os serviços disponibilizados, conforme figura 29.

Se o usuário deseja acessar o ambiente de apoio ao ensino (Moodle – SP1), abre um navegador para internet, digita o endereço do serviço (http://moodle.cce.ufpr.br/wordpress), em seguida é mostrada a página principal do sistema com os cursos disponíveis. O usuário seleciona o curso desejado e, imediatamente, aparece a mensagem, solicitando o aceite do certificado de segurança autoassinado do moodle.cce.ufpr.br. Confirmado o aceite do certificado, SP1 direciona o usuário para autenticação no IDP, que requisita também o aceite do certificado de segurança autoassinado do idp.cce.ufpr.br. Confirmado o aceite, o provedor de identidades exibe a tela de autenticação. O usuário informa as credenciais de acesso (nome de usuário e senha), IDP faz o processo de validação e redireciona o usuário para SP1 com os atributos como nome, cpf, endereço eletrônico, matrícula etc.

Uma vez realizado o processo de autenticação, se o usuário deseja acessar

outro serviço, sem sair do navegador, o usuário seleciona o endereço http://wordpress.cce.ufpr.br/wordpress, que representa a aplicação Wordpress (SP2), serviço provedor de páginas pessoais. O usuário seleciona a opção "*Login*" para acessar a área administrativa da página pessoal, imediatamente, aparece a mensagem de segurança, solicitando o aceite do certificado de segurança autoassinado do wordpress.cce.ufpr.br. Confirmado o aceite do certificado, SP2 verifica que já existe uma sessão autenticada com IDP e direciona o usuário para SP2, área administrativa com os atributos da identidade do usuário.

O mesmo procedimento ocorre ao acessar SP3, sem a necessidade de realizar novo processo de autenticação. Uma vez autenticado, o usuário poderá navegar todas as aplicações disponíveis na infraestrutura do Shibboleth, todos os provedores de serviços vinculados ao IDP que o usuário autenticou.

Caso o usuário queira encerrar a navegação dos serviços, basta selecionar a opção "Sair" ou "Logout" de qualquer um dos serviços, então usuário é direcionado a uma página com a mensagem, indicando que a finalização da sessão somente será efetivada em todos os serviços, se fechar o navegador.

4.5.4 Problemas encontrados

Durante o estudo e implementação do protótipo de SSO, observou-se que a infraestrutura do Shibboleh não provê, na versão utilizada (versão 2.2.0), *Single Log-Out* (SLO). Esta questão é frequentemente objeto de discussões na área de suporte ferramenta (http://shibboleth.internet2.edu/lists.html), entretanto ainda não há uma solução oficial pelos desenvolvedores da ferramenta.

Para contornar o problema de SLO, foi testada uma implementação do Shibboleth (provedor de identidade) pelo instituto de pesquisa Húngaro, NIIF (*National Information Infrastructure Development*), que agrega instituições acadêmicas, centros de pesquisas e comunidade. A versão desenvolvida pelo NIIF não é suportada pela Internet2, projeto responsável pelo Shibboleth, mas se configura como uma alternativa de SLO (NIIF-AAI, 2010).

Os testes com a versão disponibilizada pelo NIIF foram realizados. No entanto não atendeu de forma plena, pois durante os testes foram observados problemas de instabilidade, apesar da solução funcionar na maioria dos casos.

A solução definitiva, recomendada pela Internet2 para a questão do SLO, é o fechamento do navegador, após o uso de sessões autenticadas pelo Shibboleth. A solução proposta pela Internet2 pode ser incorporada aos hábitos dos usuários, por meio de conscientização e, não inviabiliza, portanto, o uso da infraestrutura de SSO.

5 CONSIDERAÇÕES FINAIS E CONCLUSÃO

5.1 CONTRIBUIÇÕES

A presente pesquisa traz como contribuições o protótipo de SSO e banco de dados integrados, configurando-se como soluções possíveis de serem adotadas para melhoria dos processos de autenticação, contribuindo, assim, para o aumento da segurança da informação. Além do protótipo de SSO e base de dados integrada, traz a contribuição de mais duas pesquisas exploratórias, uma que traça o perfil dos usuários quanto ao uso de senhas, e outra que mostra um panorama dos principais serviços oferecidos pela instituição.

Um dos maiores ganhos obtidos com a pesquisa é a consolidação de uma base de identidades armazenadas em LDAP, base de acesso rápido e flexível, visto que qualquer aplicação, até mesmo aplicações legadas, tem a possibilidade de conectar com esta base e utilizá-la em seus processos de autenticação. Sistemas operacionais, equipamentos de rede e serviços de rede, como *proxyes*, serviço de correio eletrônico, equipamentos de VoIP, e equipamentos de redes sem fio são exemplos de recursos que podem se beneficiar dessa base de identidades sem a necessidade de estarem inseridos na arquitetura de SSO.

5.2 DIFICULDADES

As principais dificuldades estão relacionas à implementação da arquitetura de SSO, principalmente, na implementação do Shibboleth. Para implementação dessa arquitetura, além de pesquisas realizadas no próprio portal da ferramenta, também foi necessária a realização de cursos externos e auxílio de outras instituições de ensino que possuem experiência com essa tecnologia.

Além disso, outra dificuldade está relacionada ao processo de integração de banco de dados, uma vez que a ferramenta selecionada (EID) não atendeu às necessidades da pesquisa, sendo necessário o desenvolvimento de módulo de *software*, para viabilização da base integrada.

5.3 TRABALHOS FUTUROS

O presente trabalho não se dá por concluído, e de modo algum tem a pretensão de esgotar temas tão amplos, como os abordados nesta pesquisa. Abre perspectivas para novas pesquisas em diversas áreas. No entanto, consideradas pelo autor desta pesquisa, como importantes para a continuação e aprimoramento do trabalho, são enumeradas três frentes: 1) pesquisa aprofundada de autenticação multifator[15]; 2) estudos sobre autenticação federada e suas implicações e segurança da informação, especificamente, a questão de privacidade dos dados contidos na identidade; 3) governança, mais especificamente, definições de políticas de segurança da informação.

A autenticação multifator possibilita maior segurança nos processos de autenticação. A exemplo dessa prática, grandes instituições financeiras, como bancos, por exemplo, utilizam tais mecanismos para transações financeiras. A forma mais conhecida é o uso de *tokens*, entretanto existem outros, como chaves criptográficas, dados biométricos, *smart cards* etc.

O uso de autenticação centralizada abre possibilidades de autenticação federada, de modo que um grupo de organizações podem selar relações de confiança e compartilhamento dos seus recursos por meio da autenticação federada. Exemplos dessas arquiteturas podem ser observados em grandes organizações e instituições de ensino, tais como cafe.rnp.br, incommonfederation.org, google.com, microsoft.com entre outras.

Esse modelo de autenticação federada traz novas questões acerca da segurança, estas podem ser de ordem técnica e social, uma vez que envolvem o uso de recursos compartilhados e distribuídos e compartilhamento de dados pessoais, que, por sua vez, remetem à questão de privacidade.

Outra área, também importante para continuidade deste estudo, é a de segurança da informação, principalmente o tópico relacionado à governança, que envolve a definição de diretrizes para políticas de segurança da informação.

---

[15] Multifator – mecanismos de autenticação alternativos ao tradicional uso de nome de usuário (*login*) e senha. *Tokens*, biometria, *smart card* e chaves criptográficas são exemplos de autenticação multifator.

5.4 CONCLUSÃO

O questionário possibilitou a verificação de que, na maioria dos casos, os usuários praticam a segurança da informação, no que se refere ao uso de senhas, uma vez que a maioria das respostas foi afirmativa quanto à proteção desse tipo de informação. O que sugere aceitação de controles de segurança por parte dos usuários. Sendo assim, a implementação de políticas de segurança teria o aval dos usuários, aumentando, dessa forma, as possibilidades de sucesso em um projeto desta natureza.

A visão apresentada pelo levantamento dos principais recursos, disponibilizados pela instituição, proporcionou a identificação de causas de possíveis incidentes de segurança, além de apontar em detalhes a questão de duplicidade de dados e esforços para mantê-los. Foi possível, ainda, observar de maneira sintética, quais pontos devem sofrer intervenções, para aumentar a segurança, e assim diminuir os riscos de incidentes, principalmente, os relacionados ao controle de acesso.

Outro ponto, bastante relevante da pesquisa, foi o processo de integração de base de dados, muito embora a implementação não tenha seguido à risca o referencial teórico, como previa a proposta inicial. Mesmo assim, foi possível o estabelecimento de uma base de dados coesa, que contempla todos os dados de identidade em um único ponto, do qual é replicado para uma base de identidades armazenada em diretório LDAP, abrindo novas possibilidades de acesso e, consequentemente, maior integração entre os recursos, principalmente, no que se refere a dados pessoais. Com a base LDAP consolidada, é possível, por exemplo, a autenticação de aplicações antigas, equipamentos e serviços de rede e sistemas operacionais, entre outros.

A implementação do protótipo de SSO permitiu o contato com tecnologias pouco difundidas no Brasil, fato constatado pela falta de conteúdos técnicos e científicos relacionados ao tema. Entretanto essa implementação só foi possível, devido ao auxílio recebido de outras instituições de ensino e pesquisa, que apoiaram diretamente nas questões técnicas. Além do conhecimento das tecnologias, são enumerados os principais benefícios da implementação da arquitetura, por meio do

uso da ferramenta Shibboleth:

1. a segurança no processo de autenticação, pois todos os processos de comunicação entre o provedor de serviço, provedor de identidade e usuários são efetuados, usando criptografia assimétrica, apresentando assim conformidade com a norma de segurança NBR ISO/IEC 27002;
2. disponibiliza, no provedor de serviço, a configuração de contexto protegido, de modo que é retirada a responsabilidade do desenvolvedor de aplicações de criar mecanismos de segurança para prover autenticação e, em alguns casos, autorização. Essa configuração permite a criação de uma aplicação *web* e, simplesmente, adicioná-la na área protegida, disponível no provedor de serviços, automaticamente esta será protegida pela autenticação disponibilizada pelo provedor de identidade; e,
3. usuário e senha únicos para todos os serviços disponibilizados no contexto do SSO, reduzindo, drasticamente, a quantidade de senhas que o usuário necessita manter.

Foi observada, ainda, a necessidade de gestão de identidade, requerida para que todos os benefícios da arquitetura de SSO possam ser aproveitados. Inspirado neste contexto, foi elaborado um conjunto de recomendações, *kit* de provisionamento que, em linhas gerais, prevê um processo de gestão de identidade. De maneira geral, essa arquitetura deve ser fortemente apoiada por processos de gestão de segurança e gestão de identidade para que sejam efetivos.

A solução de SSO não é abrangente, ao ponto de agregar todos os sistemas de informação e serviços disponibilizados pela instituição, no entanto abre possibilidades reais de uso, de modo que alguns recursos, atualmente em funcionamento, como Moodle, SIGEPE, VoIP, Correio Eletrônico e Intranet, já podem utilizar a proposta apresentada, sem grandes ajustes. Da mesma forma, os novos recursos podem ser projetados para utilizar a solução apresentada.

Por fim, constatam-se ganhos reais, tanto na questão de segurança quanto na gestão de recursos, se for comparado o nível de segurança e o potencial de compartilhamento de recursos proporcionados pela solução apresentada, com a arquitetura atual dos sistemas de informação e serviços, atualmente, disponibilizados pela instituição pesquisada.

REFERÊNCIAS

AHN, G.; LAM, J. Managing privacy preferences for federated identity management. In: **Proceedings of the ACM Workshop on Digital identity Management**. Fairfax, VA, USA. 2005. p. 28-36.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR ISO/IEC 27002:2005** Código de prática para a gestão da segurança da informação. Rio de Janeiro, 2005.

BATINI, C; LENZERINI, M; NAVATHE, S. B. A Coperative Analysis of Methodologies for Database Schema Integration. **ACM Computing Surveys**. v. 18, n. 4, p. 323-364, december 1986.

BEAL, A. **Segurança da Informação**: princípios e melhores práticas para proteção dos ativos de informação nas organizações. São Paulo: Atlas S.A. 2008.

BENANTAR, M. Access Control Systems: security, identity management and trutst models. New York: Springer Science + Business Media, 2006.

BOID, M.; MCBRIEN, P.; TONG, N. The automed schema integration repository. In: **19th British National Conference on Databases (BNCOD19), Lecture Notes in Computer Science (LNCS)**. Sheffield, UK, July 2002. p.42-45.

BRASIL. Decreto n. 3.505 de 13 de julho de 2000. Institui a Política de Segurança da Informação nos órgãos e entidades da Administração Pública Federal. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 14 jun. 2000. Disponível em: <http://www.planalto.gov.br/ccivil_03/decreto/D3505.htm>. Acessado em: 12/08/2010.

BRASIL. Decreto n. 4.553 de 27 de dezembro de 2002. Dispõe sobre a salvaguarda de dados, informações, documentos e materiais sigilosos de interesse da segurança da sociedade e do Estado, no âmbito da Administração Pública Federal, e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 30 dez. 2002. Disponível em: <http://www.planalto.gov.br/Ccivil/decreto/2002/D4553.htm Acessado em: 12/08/2010.

BRASIL. Decreto n. 7.174 de 12 de maio de 2010. Regulamenta a contratação de bens e serviços de informática e automação pela administração pública federal, direta ou indireta, pelas fundações instituídas ou mantidas pelo Poder Público e pelas demais organizações sob o controle direto ou indireto da União. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 13 mai. 2010. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2010/Decreto/D7174.htm>. Acessado em: 12/08/2010.

BRASIL. Lei n. 9.983 de 14 de julho de 2000. Altera o Decreto-Lei nº 2.848, de 7 de dezembro de 1940 - Código Penal e dá outras providências. **Diário Oficial [da] República Federativa do Brasil**, Brasília, DF, 17 jul. 2000. Disponível em: <http://www.planalto.gov.br/ccivil_03/Leis/L9983.htm>. Acessado em: 12/08/2010.

BRASIL. Tribunal de Contas da União. **Boas Práticas em Segurança da Informação**. 3.ed. Brasília: TCU, Secretaria de Fiscalização de Tecnologia da Informação, 2008. 70p. Disponível em: <http://portal2.tcu.gov.br/portal/page/portal/TCU/comunidades/biblioteca_tcu/biblioteca_digital/Boas_praticas_em_seguranca_da_informacao_3a_edicao.pdf>. Acessado em 01/10/2009.

BRASIL. Gabinete de Segurança Institucional da Presidência da República. **Instrução Normativa Nº 1, de 13 de junho de 2008**. Disponível em: <http://www.jf.jus.br/cjf/tecnologia-da-informacao/gestao-documental/seguranca-da-informacao/gsi-pr/IN_GSI-PR_01_13-06-2008_Disciplina_Gestao_de_SIC_na_APF/at_download/file>. Acessado em 01/10/2010.

CERT.BR. Cartilha de segurança para internet, versão 3.1, 2006. Disponível em: <http://cartilha.cert.br/download/cartilha-01-conceitos.pdf> Acessado em:07/12/2010.

CHAWATHE, S. *et al.* The TSIMMIS Project: Integration of Heterogeneous Information Sources. In: **10th Meeting of the Information Processing Society of Japan (IPSJ)**. Tokyo, Japan, October 1994. p 7-18.

DAMIANI, E.; VIMERCATI, S.; SAMARATI, P. Managing Multiple and Dependable Identities. **IEEE**. v. 7, n. 6, p. 29-37, dezembro 2003.

DARONCO, E. XML Integrator: interoperabilidade entre base de dados heterogêneas, baseado no mapeamento de esquemas conceituais. 95f. Dissertação (Mestrado em Informática) - Universidade Federal do Rio Grande do Sul, Instituto de Informática. Porto Alegre, 2003. Disponível em: <http://hdl.handle.net/10183/3425>. Acessado em 05/02/2010.

DUPONT, Y.; PARENT, C.; SPACCAPIETRA, S. Model independent assertions for integration of heterogeneous schemas. **The VLDB Journal** – The International Journal on Very Large Data Bases. v. 1, n. 1, july 1992.

ESR-RNP, Escola Superior de Redes, Rede Nacional de Ensino e Pesquisa. Gestão de Segurança da Informação: NBR 27001 e 27002. Rio de Janeiro: RNP. 2010.

FERREIRA, F.; ARAÚJO, M. **Política de Segurança da Informação**: guia prático para elaboração e implementação. Rio de Janeiro: Ciência Moderna. 2006.

FONTES, E. **Praticando a Segurança da informação**. Rio de Janeiro: Brasport. 2008.

GIL, A. **Estudo de caso : fundamentação científica , subsídios para coleta a análise de dados, como redigir o relatório**. São Paulo: Atlas. 2009.

GIOVINAZZO, R. FISCHMANN, A. Delphi Eletrônico - Uma Experiência de Utilização da Metodologia de Pesquisa e seu Potencial de Abrangência Regional. In: **XIV Congresso Latino-americano de Estrategia**. Buenos Aires, Argentina. Maio de 2001.

HASSELBRING, W. Information system integration. **Communications of the ACM**. v. 43, n. 6 , p. 32-38, June 2000.

HEIMBIGNER, D.; MCLEOD, D. A federated architecture for information management. **ACM Transactions on Information Systems**. v. 3, n. 3, p. 253-278, july 1985.

IBM. InfoShare Warehouse. Disponível em: <http://www-01.ibm.com/software/data /infosphere/warehouse/mining.html>. Acessado em: 08/05/2010.

INTERNET2. About Shibboleth. Disponível em: <http://shibbolet.internet2.edu/about. html>. Acessado em: 20/10/2009.

JOSSO. Java *Open Single Sign-On* Project Home. Disponível em: <http://www.josso.org>. Acesso em: 20/10/2009.

KIM, W.; SEO, J. Classifying schematic and data heterogeneity in multidatabase systems. **IEEE**. v. 24, n. 12, p. 12-18, Dec 1991.

KROTH, E. Integração de bases de dados. In: Notas de Aula da Disciplina de Tópicos Avançados em Banco de Dados, Curso de Especialização em Desenvolvimento de Software. Universidade de Passo Fundo, Passo Fundo, Rio Grande do Sul, 2006. Disponível em: <http://www.upf.br/erbd/integracaoDB.pdf>. Acessado em: 05/03/2010.

LAVILLE, C.; DIONNE, J. **A construção do saber**: manual de metodologia da pesquisa em ciências humanas. Belo Horizonte: UFMG, 1999.

LIME SURVEY. Lime Survery Documentation Wiki. Disponível em: <http://docs.limesurvey.org/tiki-index.php> Acessado em: 07/08/2010

MAIA, L; PAGLIUSI, P. Criptografia e certificação digital. Disponível em: <http://www.training.com.br/lpmaia/pub_seg_cripto.htm>. Acessado em: 07/12/2010.

MARCIANO, J.; LIMA-MARQUES, M. O enfoque social da segurança da informação. **Ci. Inf**. 2006, vol.35, n.3, p. 89-98.

MYSQL. Disponível em: <http://www.mysql.com/why-mysql/>. Acessado em: 15/08/2010.

NIIF-AAI. Disponível em: <https://wiki.aai.niif.hu/index.php/Single_Logout_in_ Shibboleth_IdP>. Acessado em: 10/10/2010.

OPENID. OpenID Project Home. Disponível em: <http://www.openid.net>. Acesso em: 20/10/2009.

ORACLE. Oracle Data Integration Suite. Disponível em: <http://www.oracle.com/us/products/middleware/data-integration/data-integration-suite/index.html>. Acessado em: 08/05/2010.

PARENT, C.; SPACCAPIETRA, S. Issues and approaches of database integration. **Communications of the ACM**, v. 41, n. 5es, p. 166-178, may 1998.

PROPLAN – Pró-Reitoria de Planejamento, Orçamento e Finanças. UFPR em Números 2009. Disponível em: <http://www.proplan.ufpr.br/home/CPI/arquivos/UFPR-Numeros2009.pdf>. Acessado em: 06/05/2010.

QUASS, D; WIDOM, J. On-Line Warehouse View Maintenance. In: **Proceedings of the ACM SIGMOD International Conference on Management of Data**. Tucson, Arizona. May 1997. p. 393-404.

RFC2196 - *Site Security Handbook*. Disponível em: <http://www.faqs.org/rfcs/rfc2196.html> acessado em 20/12/2009.

RNP – Rede Nacional de Ensino e Pesquisa. Gestão de Identidades nas Redes Acadêmicas. **RNP notícias**. n. 38, julho 2009.

RNP-VoIP. Disponível em: <http://www.rnp.br/voip>. Acessado em: 28/08/2010.

RODACKI, A. Aplicação de estrategias de integração de bancos de dados: um estudo de caso. 130p. Dissertação (Mestrado em Informática) - Universidade Federal do Paraná, Setor de Ciências Exatas. Curitiba, 2000. Disponível em <http://hdl.handle.net/1884/24749>. Acessado em: 04/02/2010.

SANTOS, A. **Gerenciamento de identidades**. Rio de Janeiro: Brasport. 2007.

SANTOS, L.; AMARAL, L. Estudos Delphi com Q-Sort sobre a web: a sua utilização em sistemas de informação. In: **CONFERÊNCIA DA ASSOCIAÇÃO PORTUGUESA DE SISTEMAS DE INFORMAÇÃO**, 5, Lisboa, 2004 - "CAPSI 2004 : actas da 5ª conferência" [CD ROM]. Lisboa : APSI, 2004. Disponível em: <http://hdl.handle.net/1822/2280>. Acessado em: 12/06/2009.

SCOPIM, K. J-Schema Integrator – Uma Ferramenta para Integração de Esquemas de Bancos de Dados Heterogêneos. 78p. Dissertação (Mestrado em Informática) - Universidade Federal do Paraná, Setor de Ciências Exatas. Curitiba, 2003.

SHETH, A.; LARSON, J. Federated Database Systems for managing heterogeneous, distributed and autonomous Databases. **ACM Computing Surveys**. v. 22, n. 3, september 1990.

SHETH, A.; KASHYAP, V. So far (schematically) yet so near (semantically). In: **Proceedings of the IFIP WG 2.6 Database Semantics Conference on Interoperable Database Systems (DS-5)**. North-Holland, 1993. p. 283-312.

SOARES, H.; MEDEIROS, C. Integração de Sistemas Legados a Banco de Dados Heterogêneos. In: Simpósio Brasileiro de Banco de Dados (SBBD). 14., 1999, **Anais** … [S.1.:s.n] 1999.

SORDI, J. de. **Administração da Informação**: fundamentos e práticas para a nova gestão do conhecimento. São Paulo: Saraiva, 2008.

TATBUL, N; *et al.* Data Integration Services. **Technical report**, Brown University, Computer Science Department, 2001. Disponível em: <http://pubzone.org/servlet/Attachment?attachmentId=289&versionId=1463676>. Acessado em: 02/05/2010.

Universidade Federal de Minas Gerais (UFMG) EID – Manual do Usuário. Belo Horizonte, Abril 2009.

ZAMBONI, A. Uma proposta de sistema de software para auxílio na geração de transformações de formalismo BAV para integração de bancos de dados. 82p. Dissertação (Mestrado em Informática) - Universidade Federal do Paraná, Setor de Ciências Exatas. Curitiba, 2004. Disponível em <http://hdl.handle.net/1884/1884>. Acessado em: 05/02/2010.

ZHUGE, Y. *et al.* View Maintenance in a Warehousing Environment. In: **Proceedings of the ACM SIGMOD International Conference on Management of Data**. San Jose, California, June 1995. p 316-327.

APÊNDICE A – QUESTIONÁRIO SOBRE SEGURANÇA DA INFORMAÇÃO

O questionário faz parte de um projeto de pesquisa realizado pelo acadêmico, Antonio Rodrigues Barros, mestrando do Programa de Pós-Graduação em Ciência, Gestão e Tecnologia da Informação (PPCGI), sob orientação do Professor Dr. José Simão de Paula Pinto, docente do Departamento de Ciência e Gestão da Informação (DECIGI) da UFPR.

A pesquisa tem como proposta a implantação de um projeto de *Single Sign-On*, uma única senha para acesso aos sistemas e serviços disponibilizados pela UFPR. Críticas e sugestões serão bem vindas, podem ser encaminhadas para o endereço eletrônico: antonio@ufpr.br.

O presente instrumento de pesquisa tem por objetivo identificar o comportamento dos usuários da Universidade Federal do Paraná (UFPR), alunos, docentes, técnicos e terceiros, em relação ao tema Segurança da Informação (SI), mais precisamente ao uso de senhas utilizadas para acesso aos serviços disponibilizados pela instituição. Dentre os serviços que a UFPR oferece acesso a seus usuários estão: correio eletrônico, Sistema de Bibliotecas (SophiA), Sistema de Informação para o Ensino (SIE), Portal do Aluno, Sistema de registro de pesquisas (Thales), Sistema de Educação a Distância (Moodle), Sistema de Iniciação Científica, Sistema Integrado de Gestão de Pessoas (SIGEPE), Sistema de telefonia fixa, Sistema de Voz Sobre IP (VOIP), entre outros. Este questionário possibilitará a elaboração de sugestões de melhorias no processo de SI, nos serviços oferecidos pela UFPR, se necessário, a elaboração de documento de orientação aos usuários, se detectada a necessidade deste e por fim sugerir implantação de Políticas de Segurança da Informação, para os processos de concessão de acesso. O questionário é composto de 12 questões.

Uso de senhas na UFPR

1 Qual é o seu vínculo com a UFPR?

Favor escolher apenas uma das opções a seguir:

- ○ Aluno
- ☐ Docente
- ○ Técnico administrativo
- ☐ Terceirizado

2 Quantas senhas você possui para acessar os serviços oferecidos pela UFPR?

Serviços oferecidos pela UFPR: correio eletrônico, VOIP - Voz sobre internet (para funcionários), telefonia (para funcionários), biblioteca, portal do aluno, sistema acadêmico - SIE (para funcionários), sistema de educação à distância (moodle), Thales - registro de pesquisas, Iniciação científica, Sigepe - sistema de recursos humanos etc.

Favor escolher apenas uma das opções a seguir:

- ○ Uma
- ☐ Duas
- ○ Três
- ☐ Quatro ou mais

3 Dessas senhas que possui para acesso aos serviços oferecidos pela UFPR, você normalmente utiliza uma única senha ou possui uma para cada serviço?

Favor escolher apenas uma das opções a seguir:

- ○ Utilizo senhas diferentes para cada acesso
- ☐ Utilizo senhas iguais para todos os acessos
- ○ Utilizo senhas iguais para alguns acessos

4 Quanto á disponibilização dos serviços pela UFPR. Houve algum processo formal informando os direitos de acesso ou termo de responsabilidades referente ao serviço disponibilizado?

Este processo formal refere-se ao momento em que você recebeu acesso ao portal do aluno ou acesso ao sistema de protocolo, para servidores, ou a qualquer outro sistema ou serviço.

Favor escolher apenas uma das opções a seguir:

- ☐ Sim, em todos os serviços

- ○ Sim, em alguns serviços
- ☐ Não, em nenhum serviço

5 Você costuma compartilhar senhas com outras pessoas, parentes, amigos, colegas etc?

Favor escolher apenas uma das opções a seguir:

- ○ Sim
- ☐ Às vezes
- ○ Raramente
- ☐ Nunca

6 Qual a forma que você utiliza para guardar uma senha pessoal?

Favor escolher apenas uma das opções a seguir:

- ○ Anoto e guardo em local seguro
- ☐ Anoto e carrego sempre comigo
- ○ Guardo em arquivo digital
- ☐ Guardo em arquivo digital protegido com criptografia
- ○ Guardo na minha memória
- ☐ Guardo na caixa de *e-mail* para facilitar o uso
- ○ Outros: _____

7 Quando você recebe uma senha temporária, você a altera no primeiro acesso?

Favor escolher apenas uma das opções a seguir:

- ☐ Sim
- ○ Não
- ☐ Às vezes

8 Você muda suas senhas com regularidade? Se sim, de quanto em quanto tempo você altera suas senhas?

Favor escolher apenas uma das opções a seguir:

- ○ A cada 3 meses
- ☐ A cada 6 meses
- ○ A cada 1 ano

☐ A cada 2 anos

○ Nunca

9 Os sistemas ou serviços que você possui acesso na UFPR solicitam a alteração de senha com regularidade?

Favor escolher apenas uma das opções a seguir:

☐ Alguns

○ Nenhum

☐ Todos

10 Na sua opinião, quais são os critérios mais adequados para criação de uma senha segura?

Por favor, selecione os itens que contribuem para criação de uma senha segura:

Uma senha deve conter de 3 a 6 caracteres

☐ Uma senha deve conter de 7 a 9 caracteres

Uma senha deve conter mais de 9 caracteres

☐ Deve conter letras maiúsculas

Deve conter letras minúsculas

☐ Deve conter símbolos como: !@#$%"&*()-_+={}~?/:>,etc

Deve ser criada aleatoriamente misturando letras e números

☐ Devem ser usadas palavras, números ou datas conhecidas para facilitar a memorização

Devem ser usadas partes de uma frase conhecida para facilitar a memorização

☐ Devem ser fáceis de digitar

Devem ser fáceis de memorizar

Outros: ☐

11 Os navegadores para internet (*browsers*) ou páginas de logins possuem um mecanismo de memorização de senhas, que comumente é utilizado. No seu entendimento, esse recurso deve ser utilizado?

Favor escolher apenas uma das opções a seguir:

☐ Sim

○ Não

☐ Não sei

12 Sabendo que há necessidade de guardar muitas senhas para finalidades distintas como por exemplo, acesso a contas de correio eletrônico, acesso ao *home banking*, redes sociais (orkut, facebook, twiter,...) etc. Que estratégia você normalmente utiliza para facilitar o uso de suas senhas?

Favor escolher apenas uma das opções a seguir:

- ○ Não uso nenhuma estratégia
- ☐ Uso a mesma senha para tudo
- ○ Costumo usar senhas iguais para finalidades semelhantes, ex.: senhas iguais para acesso a contas de correio eletrônico
- ☐ Uso senhas diferentes para cada finalidade
- ○ Outros: _____

APÊNDICE B – Integração e resolução de conflitos de banco de dados

As bases de dados fonte utilizadas para integração são: a) base de dados do sistema acadêmico; b) base de dados de recursos humanos; e c) base de dados de contas de correio eletrônico. Estas serão identificadas respectivamente como A, B e C, para facilitar o entendimento.

Criação de visões de banco de dados

Optou-se por deixar as fontes mais homogêneas para facilitar o processo de mapeamento, posteriormente, feito na ferramenta EID. A forma encontrada foi por meio de visões de banco de dados. Dessa forma, as visões foram criadas e, posteriormente, mapeadas para o modelo conceitual (figura 27).

Na fonte de dados **A,** foram criadas quatro visões (quadro 12), viewidentificacao, viewaluno, viewemail e viewendereco, que representam os dados de alunos.

```
viewidentificacao (
 cpf character varying(15),
 "cidadeNascimento" character varying(50),
 "dataNascimento" date,
 "estadoCivil" character varying(255),
 "estadoNascimento" character varying(2),
 nacionalidade character varying(120),
 "nomeCompleto" character varying(150),
 "nomeMae" character varying(150),
 "nomePai" character varying(150),
 "nomeSolteiro" character varying(150),
 "numeroIdentidade" character varying(20),
 "orgaoEmissorIdentidade" character varying(20),
 "paisNascimento" character varying(40),
 passaporte character varying(20),
 sexo character varying(120),
 "ufIdentidade" character varying(2) )
viewaluno (
 cpf character varying(15),
 “codigoCapes” character varying(120),
 “codigoCapes” character varying(120),
 "dataVinculacao" date,
  "dataAfastamento" date,
  "estadoVinculo" character varying(50),
  matricula character varying(120),
  "nivelCurso" text,
  “nomeCurso” character varying(120))
viewemail (
 cpf character varying(15),
 servidor character varying(255),
 email varing(255),
 redirecionamento varing(255),
 tipo character varing(255) )
viewendereco (
  cpf character varying(15),
  bairro character varying(120),
  cep character varying(8),
  cidade character varying(50),
  complemento character varying(120),
  estado character varying(2),
  logradouro character varying(120),
  numero bigint,
  pais character varying(40),
  preferencial bigint,
  tipo character varying(50) )
```

QUADRO 11 - VISÕES CRIADAS NO ESQUEMA A PARA FACILITAR A INTEGRAÇÃO DE ESQUEMAS
FONTE: O autor (2010)

Na fonte de dados **B** foram criadas cinco visões (quadro 12), viewidentificacao, viewdocente, viewtecnico, viewemail e viewendereco, que representam os dados de servidores (docentes e técnicos administrativos).

```
viewidentificacao (
 cpf character varying(15),
 "cidadeNascimento" character varying(50),
 "dataNascimento" date,
 "estadoCivil" character varying(255),
 "estadoNascimento" character varying(2),
 nacionalidade character varying(120),
 "nomeCompleto" character varying(150),
 "nomeMae" character varying(150),
 "nomePai" character varying(150),
 "nomeSolteiro" character varying(150),
 "numeroIdentidade" character varying(20),
 "orgaoEmissorIdentidade" character varying(20),
 "paisNascimento" character varying(40),
 passaporte character varying(20),
 sexo character varying(120),
 "ufIdentidade" character varying(2) )
viewdocente (  cpf character varying(15),
 classe character varying(120),
 "dataAdmissao" date,
 "dataDemissao" date,
 "dataAfastamento" date,
 "estadoVinculo" character varying,
 lotacao character varying(120),
 nivel character varying,
 siape bigint,
 titulacao character varying(120) )
viewtecnico (
  cpf character varying(15),
  classe character varying(120),
  "dataAdmissao" date,
  "dataAfastamento" date,
  "estadoVinculo" character varying(50),
  "funcaoPrincipal" text,
  lotacao character varying(120),
  "nivelCapacitacao" text,
  padrao character varying(30),
  siape bigint,
  titulacao character varying(120) )
viewemail (
 cpf character varying(15),
 servidor character varying(255),
 email varing(255),
 redirecionamento varing(255),
 tipo character varing(255) )
viewendereco (
  cpf character varying(15),
  bairro character varying(120),
  cep character varying(8),
  cidade character varying(50),
  complemento character varying(120),
  estado character varying(2),
  logradouro character varying(120),
  numero bigint,
  pais character varying(40),
  preferencial bigint,
  tipo character varying(50) )
```

QUADRO 12 - VISÕES CRIADAS NO ESQUEMA B PARA FACILITAR A INTEGRAÇÃO DE ESQUEMAS
FONTE: O autor (2010)

Na fonte de dados **C** foi criada uma visão (quadro 13), viewconta, que representa os dados de contas, tanto de alunos quanto servidores.

```
viewidentificacao (
 cpf character varying(15),
 usuario character varying(50),
 senha character varying(50) )
```

QUADRO 13 - VISÃO CRIADA NO ESQUEMA C PARA FACILITAR A INTEGRAÇÃO DE ESQUEMAS
FONTE: O autor (2010)

No processo de criação de visões, adotou-se o cpf (cadastro de pessoa física) como chave natural de ligação entre as entidades.

Identificação de conflitos

A identificação e a resolução dos conflitos estruturais se deu na criação das

visões. Sendo assim, as visões foram criadas diretamente, nas fontes de dados, deixando-os o mais uniforme possível, para facilitar as próximas etapas que compreendem a integração propriamente dita.

A identificação e a resolução de conflitos seguem:

1. conflitos de sinônimos, objetos com nomes distintos, contudo representam a mesma entidade do mundo real. Este caso foi encontrado entre os esquemas A e B, posto que a entidade pessoa era nomeada de duas maneiras (A.PESSOAS e B.PESSOA). Também foi observado o mesmo caso na entidade A.EMAILS e B.EMAIL. Esses conflitos foram resolvidos, atribuindo-se os apelidos de viewidentificacao e viewemail, respectivamente;
2. conflitos de tipos de dados, objetos equivalentes com tipos de dados diferentes. Foi o caso de conflito observado, no atributo matrícula, visto que A.CURSO_ALUNO.MATR_ALUNO do tipo *character varyng* (texto de tamanho variável) corresponde a B.CARREIRA.MATRICULA do tipo *bigint* (inteiro longo). Esse conflito foi resolvido, por meio de conversão de tipos, alterando-se B.CARREIRA.MATRICULA para *character varyng*, no momento de criação da visão;
3. conflito de unidade, quando os atributos armazenam unidades de medidas diferentes. Um caso encontrado que pode ser enquadrado nessa categoria de conflitos é o atributo A.PESSOAS.CPF e B.PESSOA.CPF. No esquema A, o atributo apresentava máscara (NNN.NNN.NNN-NN), já, no esquema B, o atributo apresentava apenas números. Esse conflito foi resolvido por meio de função de substituição de conteúdo aplicada, no esquema A, no momento da criação da visão;
4. conflito de valor padrão, quando são definidos valores padrão para diferentes esquemas. Esse tipo de conflito foi identificado no atributo A.PESSOAS.NOME_PAI e B.PESSOA.NOME_PAI, de forma que no esquema A apresentava nulo para pessoas sem nome de pai informado, já no esquema B apresentava o texto “NÃO INFORMADO” para a mesma situação. A resolução desse conflito foi resolvido atribuindo *null* para onde havia “NÃO INFORMADO” ao criar a visão;
5. conflito de incompatibilidade união, quando duas entidades semanticamente

parecidas são representadas com número diferente de atributos ou atributos não relacionados. Esse tipo de conflito foi identificado no atributo CPF dos esquemas A e B, visto que, no esquema B, a entidade PESSOA possuía o atributo CPF, enquanto em A, o atributo estava presente, em outra entidade chamada DOCUMENTO. O conflito foi resolvido por meio de junção das entidades A.PESSOA com A.DOCUMENTO, na criação da visão;

6. conflito de isomorfismo, quando entidades, semanticamente parecidas, possuem número de atributos diferentes para representar conceitos similares. Esse tipo de conflito foi identificado, na entidade endereço entre os esquemas A e B. No esquema A.ENDERECO.RUA representava informações do tipo de logradouro, logradouro e número, enquanto, no esquema B, os dados eram apresentados de modo normalizado, B.ENDERECO.ID_TIPOLOGRADOURO, que representa chave estrangeira da entidade B.TIPOLOGRADOURO, B.ENDERECO.LOGRADOURO e B.ENDERECO.NUMERO. Esse conflito foi resolvido, por meio de junção e concatenação de atributos, no esquema B, efetuados na criação da visão;
7. conflito de falta de atributo, quando entidades, semanticamente semelhantes, possuem número de atributos diferentes. Esse tipo de conflito foi observado entre as entidades EMAIL dos esquemas A e B, posto que A não apresentava o atributo TIPOEMAIL, que representa, se o *e-mail* é de uso pessoal ou comercial. Esse conflito foi resolvido, atribuindo-se ou adicionando o pseudo atributo TIPOEMAIL, no esquema A, na criação da visão.