4.1 REVISÃO SISTEMÁTICA DA LITERATURA E METODOLOGIA UTILIZADA

A revisão sistemática é o método científico adotado para resumir a literatura sobre um dado assunto, em que protocolos específicos são usados para determinar criteriosamente quais estudos serão incluídos na revisão (CHUNG; BURNS; KIM, 2006). Um de seus objetivos é nortear o desenvolvimento de projetos, indicando novos rumos para futuras investigações e identificando os métodos de pesquisa utilizados em determinada área. Com o emprego da revisão sistemática, é possível obter respostas a uma dada pergunta, utilizando-se métodos sistemáticos (e que podem ser reproduzidos) para identificar, selecionar e avaliar criticamente pesquisas relevantes. Para tanto, realiza-se a revisão da literatura de maneira planejada, estruturada e controlada. O emprego de tais critérios permite delimitar a área de busca e evita a subjetividade na análise dos fatos observados (BRERETON et al., 2007).

A revisão sistemática aqui realizada foi efetuada seguindo quatro etapas:

- Definição do objetivo da revisão;
- Identificação da literatura;
- Seleção dos estudos a serem incluídos;
- Análise e discussão.

Para Sampaio (2007), essas etapas são importantes pois auxiliam os pesquisadores a adequarem a pergunta norteadora da revisão em função da informação disponível sobre o tema de interesse (SAMPAIO; MANCINI, 2007).

Para percorrer as quatro etapas citadas, nesta tese adotou-se o método descrito por Chung (2006), que propõe os seguintes passos: Etapa 1 - Definir objetivo: formular uma pergunta para definir os objetivos da revisão; Etapa 2 - Identificar literatura: selecionar uma estratégia de busca, realizar a busca nas bases de dados disponíveis e recuperar artigos; Etapa 3 - Selecionar estudos: identificar os artigos pelos títulos e resumos, selecionar os estudos primários de acordo com critérios de inclusão e exclusão e extrair os dados; e Etapa 4 - Analisar e discutir.

4.1.1 Etapa 1: objetivo da revisão

O objetivo desta pesquisa foi responder a seguinte pergunta: Quais trabalhos representam o estado da arte em sistemas autonômicos de resposta à intrusão?

Ou seja, a pergunta conduz a uma investigação sobre quais modelos são utilizados para dar respostas a ataques em sistemas de IDS autonômicos. Nesse processo, investigou-se se alguma técnica utiliza *Big Data* e qual o estado da arte em IDS com *Big Data*.

4.1.2 Etapa 2: identificação da literatura

A identificação da literatura engloba as etapas de definir a estratégia de busca (incluindo os critérios para isto); selecionar as fontes de consulta e recuperar os artigos identificados durante o processo. A estratégia adotada foi pesquisar trabalhos publicados em língua portuguesa e inglesa, disponíveis *on-line*, e recuperados nos seguintes bancos de dados:

- IEEEXplore;
- ACM Digital library;
- Google Scholar;
- Compendex Essagem I;
- Wiley InterSciene;
- Elsevier Science Direct;
- AIS eLibrary;
- SpringerLink;

A Tabela 1 apresenta a compilação dos resultados com os artigos recuperados conforme as palavras-chave utilizadas nas buscas, e publicados no período de 2010 até 2017. As palavras-chave foram categorizadas por relevância: começa-se com a Categoria 1, que abrange os artigos mais genéricos em relação ao tema desta tese, como artigos que tratam apenas da aplicação do conceito autonômico em computação; ou apenas abrangem sistemas de detecção de intrusão ou assuntos relacionados com *Big Data*, até chegar à Categoria 7, que inclui artigos

mais específicos, com foco em Sistemas de Detecção Autonômico para Ambientes de Nuvem abrangendo *Big Data*.

O primeiro estágio utilizou os critérios apresentados na Categoria 1. Foram encontrados 85.767 artigos relacionados com *Autonomic Computing*; 89.847 relacionados com *Intrusion Detection System* e 51.486 relacionados com *Big Data*. É um volume expressivo, mas que passa a reduzir-se sensivelmente à medida em que a pesquisa é refinada, com a interpolação de outras palavras-chave. Os diferentes termos, que abrangem os principais conceitos aplicados nesta tese, foram combinados com o uso do operador booleano 'And'. Desse modo, a partir da Categoria 3, que utiliza os termos **"Sistemas de Detecção de Intrusão" E "Big Data"**, o número de artigos cai para 10, chegando a menos de 5 artigos nas Categorias 4 e 6. Na Categoria 7 não foi possível localizar qualquer artigo. A partir destes títulos, foram selecionados os estudos que fariam parte desta revisão.

4.1.3 Seleção dos estudos a serem incluídos

Os trabalhos foram identificados por seus títulos e resumos (*abstracts*) e, a partir disso, os seguintes parâmetros de inclusão foram aplicados para permitir a extração de dados:

- Uso da língua portuguesa ou inglesa;
- Publicação feita entre o período de 2010 e 2017;
- Disponibilidade para *download*;
- Apresentação, no título ou resumo, de uma das palavras-chave;
- Inserção no contexto do tema desta tese.

Para a exclusão de um estudo, foram determinados os seguintes parâmetros:

- Impossibilidade de acesso ao artigo completo;
- Artigo fora do contexto da tese, mesmo apresentando as palavras-chave anteriormente mencionadas, pois nem sempre elas estão relacionadas com a área de pesquisa desta tese.

O título e o resumo de cada estudo foram analisados, conforme os critérios já descritos. Além da análise qualitativa, também foi efetuada

uma meta-análise, com uso dos dados quantitativos: as informações extraídas dos estudos selecionados foram confrontadas com as questões de pesquisa desta tese para que fosse possível detectar similaridades e diferenças em relação às propostas aqui efetuadas. No total, após a aplicação dos fatores de inclusão e exclusão, restaram 58 artigos que se circunscreviam no contexto de segurança computacional e abordavam o uso de sistemas autonômicos em IDS com uso de *Big Data*, os quais foram submetidos a uma pré-análise.

Tabela 1 – Termos utilizados na pesquisa

Palavra chave	**Total**	**Categoria**
Autonomic Computing	85.670	Categoria 1
Intrusion Detection System	89.847	Categoria 1
Big Data	51.486	Categoria 1
Security, Cloud	5.327	Categoria 2
Autonomic, Cloud	474	Categoria 2
Intrusion Detection System, Cloud	228	Categoria 2
Intrusion Detection System, Big Data, Autonomic Computing	31	Categoria 2
Intrusion Detection System, Big Data	10	Categoria 3
Intrusion Detection System, Big Data, Cloud	4	Categoria 4
Intrusion Detection System, Autonomic Computing	11	Categoria 5
Intrusion Detection System, Autonomic Computing, Cloud	2	Categoria 6
Intrusion Detection System, Autonomic Computing, Cloud, Big Data	0	Categoria 7

Fonte: elaborada pelo autor.

4.2 A PESQUISA

Esta pesquisa aborda a evolução dos sistemas de detecção de intrusão como sistemas de segurança em nuvem. Abordou-se as técnicas básicas para IDSs na nuvem, os IDSs baseados como serviços da nuvem e os IDSs baseados em *Big Data*; e chegou-se à proposta de um sistema de resposta autonômico com *Big Data*, que funcione numa nuvem. Sabe-se que, no contexto da evolução da Internet, a oferta de serviços resultou no paradigma de computação nuvem. Assim, o uso do modelo de computação orientada a serviços em nuvem tem experimentado vários problemas, tais como segurança, privacidade e confidencialidade da informação. Um progresso significativo ocorreu em termos da melhora dos sistemas de segurança em um ambiente de nuvem. Mesmo assim, a segurança quanto à privacidade e confidencialidade tem sido explorada, consistindo em um problema diante das vulnerabilidades desses sistemas.

As vulnerabilidades são associadas a riscos e, na maioria desses ambientes, elas levam a ameaças e ataques, podendo ser exploradas tanto por usuários não autorizados quanto por autorizados, de modo que vários tipos de atividades maliciosas precisam ser prevenidos para se garantir a segurança. Os riscos existem em vários aspectos dos sistemas computacionais de rede. Riscos correspondem a probabilidades de ocorrências de ameaças e a análise e gestão de riscos tem sido sempre um tema vital. Em um ambiente de computação em nuvem, na medida em que o número de aplicações aumenta, a exposição da informação também aumenta. Por conseguinte, a proteção da informação sensível é muito importante e diferentes organizações podem sofrer um impacto de maneiras diferentes, dependendo da sensibilidade da informação.

4.3 TRABALHOS COM TÉCNICAS BÁSICAS PARA NUVEM E IDS

Kholidy e Baiardi (2012), em *CIDD: A Cloud Intrusion Detection Dataset for Cloud Computing and Masquerade Attacks*, focalizam em ataques mascarados que representam uma séria ameaça para o sistema de nuvem, devido à enorme quantidade de recursos destes sistemas. Este artigo apresenta um *Intrusion Detection Dataset Cloud* (CIDD), que é o primeiro sistema de nuvem que consiste de dados de auditoria de conhecimento e comportamento, com base em dados recolhidos de ambos os usuários, UNIX e Windows.