PROPOSTA PARA DETECÇÃO DE ANOMALIAS DE DADOS NA NUVEM BASEADO NO COMPORTAMENTO DE ENXAME DE GAFANHOTOS

Tiago Martins Ribeiro*, V. Leonardo Paucar*, Matias Romário Pinheiro dos Santos†, Raimundo Pereira da Cunha Neto‡

*Universidade Federal do Maranhão
Departamento de Pós Graduação em Engenharia de Eletricidade
São Luís, Maranhão, Brasil

†Universidade Federal do Ceará
Departamento de Computação
Fortaleza, Ceará, Brasil

‡Estácio Ceut
Nupaccthe
Teresina, Piauí, Brasil

Emails: tiagomartinz@hotmail.com, Lpaucar@ieee.org, matiasromariop2@gmail.com, netocunhathe@gmail.com

Abstract— This study aims to present a methodology for identification of security anomalies of a network in the cloud considering aspects of traffic. The proposed methodology is based on the application of intelligent behavior of desert locust swarm. The cloud is a computational model quite popular and it stands out as an economical solution for businesses. In the detection process of security anomalies in a cloud computing environment, various techniques such as the use of implemented systems for this, intrusion detection systems or identification data traffic patterns are available. Identifying aspects of vulnerability based on the locust swarm behavior brings understanding and analysis of one of the most important paradigms of cloud computing, the security. For the development of this work it was necessary to understand the basic theory that guides such as information security, cloud computing, particle swarm intelligence and the desert locust behavior.

Keywords— Network security, cloud computing, particle swarm intelligence, artificial intelligence, desert locust.

Resumo— O presente trabalho tem por objetivo apresentar uma metodologia para identificação de anomalias de segurança de uma rede na nuvem considerando os seus aspectos de tráfego. A metodologia proposta está baseada na aplicação da inteligência dos enxames de gafanhotos do deserto. A nuvem é um modelo computacional bastante popular e que se destaca como uma solução econômica para as empresas. No processo de detecção de anomalias de segurança de um ambiente de computação em nuvem existem várias técnicas, tal como o uso de sistemas implementados para isso, sistemas de detecção de intrusão ou a identificação de padrões de tráfego de dados. A identificação de aspectos de vulnerabilidade baseado no comportamento do enxame de gafanhotos, traz consigo o entendimento e a análise de um dos paradigmas mais importantes da computação em nuvem, a segurança. Para o desenvolvimento deste trabalho se fez necessário entender a teoria base que a orienta, como segurança da informação, computação em nuvem, inteligência de enxames de partículas e o comportamento do gafanhoto do deserto.

Palavras-chave— Segurança de redes, computação em nuvem, inteligência de enxames de partículas, inteligência artificial, gafanhoto do deserto.

1 Introdução

Um dos aspectos mais importantes a ser considerado nos sistemas computacionais e de informação é a segurança, na nuvem não é diferente. Virtualizar a infraestrutura e/ou migrar sistemas que possuem dados vitais para empresas e pessoas são tarefas de risco, uma vez que essas informações caindo em mãos erradas podem provocar falência à organização ou pessoas, principalmente em nuvem pública, que possui um risco de segurança maior, McCabe and Aggarwal (2012).

A migração de serviços para a nuvem tende a promover cortes de gastos consideráveis, mas pode elevar riscos de segurança, quando o mesmo é migrado para uma nuvem pública, mesmo que esse esteja sendo migrado para um modelo de nuvem comercial, ou do tipo *Open-Source*, como o Openstack (2015), Eucalyptus (2015), etc. As nuvens do tipo Open-Source são cada vez mais utilizadas como *frameworks* para a criação de nuvens privadas, Donevski et al. (2014).

Segurança em redes compreende a área responsável pela consistência e segurança dos dados que trafegam em uma determinada rede local ou não, mantendo o seu conteúdo protegido de quaisquer alterações sem a devida autorização. O uso de técnicas de inteligência artificial na empregabilidade da segurança das redes, pode auxiliar na obtenção de uma segurança mais eficaz e com menos brechas ou falhas em sua implementação. De fato, há necessidade de criar modelos e estratégias de seguranças que possam assegurar a confiabilidade e disponibilidade de dados, principalmente

nas nuvens, devido ao seu conteúdo existente que é suscetível de sofrer um número grande de ataques.

A computação em nuvem é um paradigma cada vez mais presente no dia a dia das pessoas e nas empresas, Mazzer (2014), a sua popularização é fruto do desenvolvimento de aplicações e plataformas que possibilitam a utilização de recursos e aplicações sem as mesmas estarem instaladas no computador local, no celular, ou em qualquer outro dispositivo

No presente artigo será detalhada uma base conceitual sobre computação em nuvem na seção 2, em seguida, na seção 3 serão apresentadas as técnicas de detecção de anomalias existentes atualmente, na seção 4 serão discutidos alguns trabalhos relacionados à detecção de anomalias na rede, já na seção 5 será mostrado o estudo do comportamento do gafanhoto, na seção 6 será apresentado o modelo proposto, na seção 7 será relatada uma possível aplicação da arquitetura proposta na nuvem de computadores, por fim, na seção 8 a conclusão do trabalho.

2 Computação em nuvem

Segundo o NIST (*National Institute of Standards and Technology*) a computação em nuvem é um modelo que permite acesso ubíquo, com acesso através de uma rede com recursos configuráveis, como servidores, que fornecem um gerenciamento mínimo ou interação com o provedor que oferece o serviço, Mell and Grance (2011). Ela já é realidade nas empresas, uma vez que, principalmente devido a sua capacidade de cortar gastos operacionais, permite as empresas de tecnologia da informação (TI) focarem apenas em questões estratégicas em vez de se preocuparem em manter grandes data centers em sua estrutura.

A nuvem possui diferentes definições, ainda não há um consenso sobre a sua definição, Velte et al. (2011). Pode-se dizer que ela está principalmente relacionada a dispor armazenamento e processamento. A infraestrutura e recursos computacionais da nuvem ficam transparentes aos usuários, tendo o usuário apenas a interface gráfica para seu uso da Silva (2010).

A computação em nuvem permite ter acesso à aplicação sem a mesma estar instalada nos dispositivos, em qualquer lugar que seja, Laureano (2005).

2.1 Tipos de nuvens

Dependendo da necessidade, é possível contratar um tipo de nuvem que se encaixe na necessidade da empresa ou do cliente, existem diversos modelos, dentre elas temos a nuvem pública, na qual o usuário possui acesso amplo ao conteúdo e recursos, mais dependendo da situação ela pode não ser a nuvem mais indicada, essa diversidade de tipo se dá mais pela necessidade de controlar o nível de acesso dos clientes às informações disponibilizadas.

Nuvem Pública: A infraestrutura fica disponível para o público em geral, podendo a mesma ser acessada por qualquer usuário, sendo ela oferecida por organizações públicas ou grandes grupos que disponibilizam armazenamento e processamento, (Sousa et al., 2010; Veras, 2013).

Nuvem Privada: Na implantação privada, a infraestrutura de nuvem é utilizada exclusivamente para uma organização, esse modelo é utilizando quando se necessita de mais segurança, pois nela são implementadas políticas de acesso para adicionar segurança, (Sousa et al., 2010; Veras, 2013).

Nuvem Híbrida: No modelo de implantação híbrido, existe uma composição de duas ou mais nuvens, que podem ser privadas, pública ou comunitária (Sousa et al., 2010; Veras, 2013).

2.2 Segurança na nuvem

Quanto aos aspectos de segurança da nuvem deve-se destacar problemas como o acesso compartilhado, baseado no modelo multi-inquilino (*multi-tenancy*), *exploits* virtuais, indisponibilidade dos serviços, gargalos na transferência dos dados dentre outros, (Velte et al., 2011; Sousa et al., 2010; Armbrust et al., 2010).

3 Detecção de anomalias em redes

Os métodos de detecção de anomalias normalmente recebem um conjunto de informações e separam as saídas selecionadas em duas, a saída normal, que segue o padrão normal estabelecido na rede, e a anormal, que difere do padrão da rede, considerando sempre o ambiente em que os dados estão inseridos e a sua natureza, de Barros Paranhos da Costa (2014).

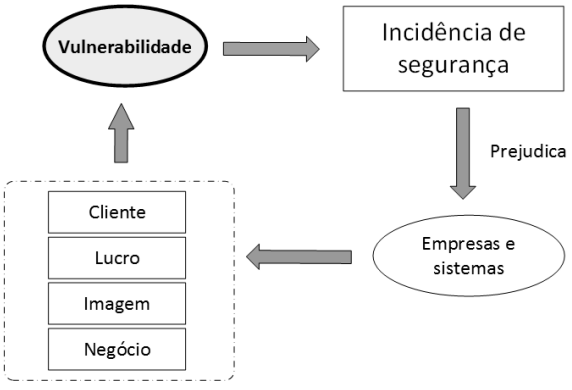

Figura 1: Representação de ocorrências de vulnerabilidade em empresas e sistemas.

Existem muitas técnicas voltadas para a detecção de anomalias, como é o caso da téc-

nica baseada em métodos estatísticos, da Silva (2003), técnicas baseadas em agrupamentos, Nascimento (2012), métodos baseados em teorias, Ando (2007), dentre outras mais.

A detecção de anomalias se torna essencial para a análise de quesitos de segurança, e levando isso para a computação em nuvem o torna ainda mais importante, pois ela possibilita assegurar a segurança e confiabilidade dos dados que trafegam em sua rede.

As anomalias podem ser identificadas a partir de mudanças acentuadas no fluxo da rede, comparando o mesmo com o fluxo normal que a mesma teria, Erramilli et al. (1996).

Um dos desafios quanto à análise do tráfego de rede na nuvem, se dá pela grande quantidade massiva de dados que trafegam na rede. Para isso faz-se necessário implementar regras para a realização da coleta e análise, já que verificar todos os pacotes seria um processo que exigiria muito gasto de recursos computacionais.

Quanto à computação em nuvem é importante levar em consideração as suas características de escalabilidade e elasticidade, assim sendo, a mesma pode ter seu fluxo de dados aumentado ou diminuído de forma instantânea, e isso deve ser levado em consideração no modelo proposto.

A detecção de anomalias pode ser realizada de diversas formas. Além das já apresentadas é possível identificá-la através de informações como tempo gasto pelo processador, comportamento do *host*, consumo excessivo de memória e CPU, dentre outras formas, de Sá Silva et al. (2004).

Na Figura 1, são apresentadas ocorrências de vulnerabilidades, podendo elas estar associadas a ocorrências de anomalias, e todas as implicações de segurança que poderiam ocorrer para as empresas devido a essa fragilidade.

4 Trabalhos Relacionados

Quanto à detecção de anomalias na rede tem-se o trabalho de Zarpelao et al. (2009), que trata de um sistema de detecção de anomalias para redes de computadores baseada em análises em três aspectos, um baseado no tráfego, outro no protocolo SNMP e o restante baseado em heurística.

Kim and Reddy (2008) propõem um sistema que faz a análise baseada na coleta dos pacotes e análise no cabeçalho do mesmo, fazendo uma detecção em tempo real.

Júnior and de Oliveira Dantas (2012) propõem a implementação de um analisador de tráfego de dados realizando a detecção através do uso de entropia.

5 Comportamento dos gafanhotos

O uso do comportamento em colônias de insetos é muito comum no campo de ciência da computação. Algoritmos desenvolvidos baseados nesse tipo de comportamento são utilizados para resolver problemas existentes no campo da computação e afins. São exemplos desses algoritmos: colônia de formigas, Dorigo (1992), colônia de abelhas, Karaboga and Gorkemli (2011), dentre outros, assim como o algoritmo de enxame por partículas, Kennedy and Eberhart (2001).

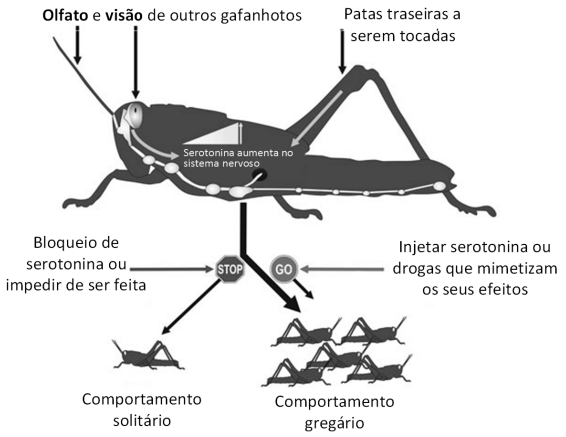

Figura 2: Mudança da fase solitária para a fase de agregação (Yong (2009)).

O gafanhoto do deserto através do seu comportamento na formação em exames de ataques, possui uma relevância para aplicação no campo computacional. Pois através do mesmo pode-se trabalhar com soluções em diversas áreas e propor ideias de melhorias. O foco do presente artigo está em apresentar um modelo baseado no comportamento inteligente do gafanhoto do deserto visando sua aplicação para solucionar problemas de segurança de redes na nuvem. Para representar esse comportamento gregário foi considerado o gafanhoto da espécie *Schistocerca Gregária*, conhecido por gafanhoto do deserto, Anstey et al. (2009), citado por Yong (2009).

Inicialmente o gafanhoto do deserto está em sua fase solitária e possui uma cor esverdeada, mas ao encontrar ou visualizar uma boa quantidade de alimentos, ele ativa em seu sistema neural uma monoamina neurotransmissora denominada serotonina. Essa substância se eleva fazendo que o inseto mude seu aspecto físico, anteriormente de cor esverdeada, para uma cor amarelada levando-o à fase de agregação (Figura 2) no qual sai em busca de outros gafanhotos que estão em sua região em fase solitária.

Ao encontrar esses insetos, o gafanhoto mutante, os induz a passar da fase solitária para a fase de agregação através do olfato, visão ou bater das patas alertando ao parceiro que existe alimento em abundância na região e o induz a aumentar a taxa de serotonina no seu cérebro, seguindo assim em busca de outros gafanhotos. Quando um número de gafanhotos em fase de agregação é grande, eles

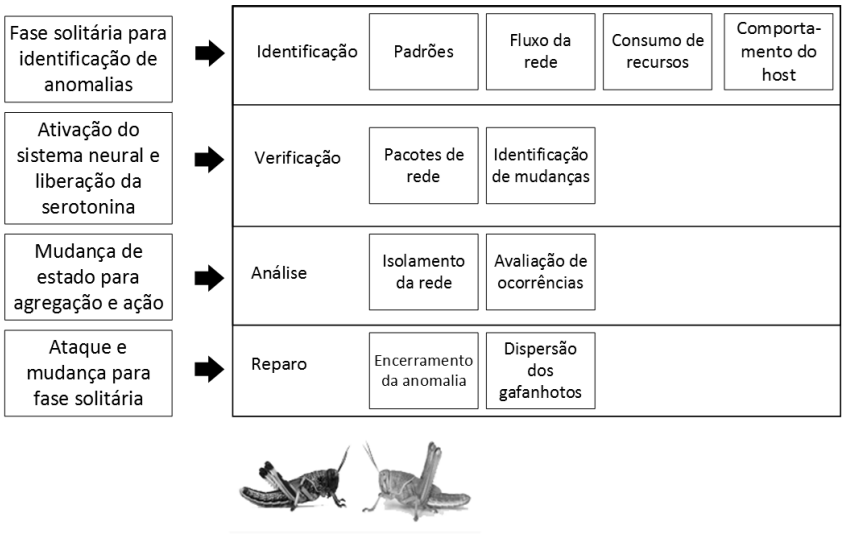

Figura 3: Componentes da arquitetura do modelo do gafanhoto.

partem para o ataque ao alimento. Uma vez satisfeitos ou quando o alimento acaba alguns gafanhotos morrem por ataque de predadores naturais ou por outras causas, servindo assim como adubo ao terreno. Os gafanhotos que sobrevivem à fase gregária simplesmente voltam à sua fase solitária.

6 Modelo proposto

A inteligência de enxame dos gafanhotos possui características peculiares que serão usadas para o processo do modelo proposto, baseado nessas características que se propõe uma metodologia de detecção de anomalias da rede na nuvem baseado na inteligência do enxame do gafanhoto do deserto para garantir a segurança na rede.

Devido à possibilidade de busca individual dos gafanhotos, e sua capacidade de agregação em forma de enxame, quando avistado um potencial alvo (o alimento); sua habilidade se apresenta como uma forma de busca de fluxo normal de dados na rede que pode trazer resultados significativos. Com a implementação dessa característica é possível verificar toda a estrutura da rede na nuvem de forma silenciosa, ou seja, de forma a não realizar um consumo grande de recursos, como CPU, memória, dentre outros mais.

Na Figura 3 detalha-se a arquitetura do modelo proposto, sendo dividido o processo em etapas e que se assemelha à formação do enxame dos gafanhotos aplicado ao volume de dados que trafega na rede da nuvem. O modelo proposto é dividido em quatro etapas, como já mencionadas, essas etapas são os processos da arquitetura, e cada uma possui sua responsabilidade. A seguir a descrição de cada etapa.

Identificação: O processo de identificação é baseado na vida do gafanhoto do deserto, pode ser caracterizado pela busca solitária por alimento se comparado ao inseto. Nele os objetos de identificação vasculham a rede atrás de informações sobre os padrões, o fluxo de rede, o consumo de recursos e o comportamento do *host*, a fim de encontrar possíveis variações no seu processo comum. Após identificar esse desvio de padrões segue o processo de verificação.

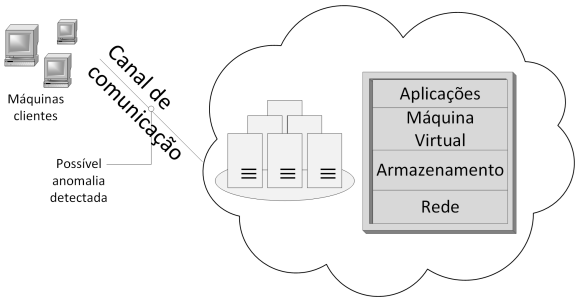

Figura 4: Detecção de anomalia pelo modelo do gafanhoto.

Verificação: Neste processo são analisados os pacotes dos dados a fim de identificar mudanças, principalmente no teor do pacote ou se o mesmo está corrompido depois de identificado algo suspeito. Na verificação usa-se a técnica de *sniffing* para verificar o conteúdo dos dados. Caso identificada a anomalia, passa-se para o processo de análise. A análise é realizada levando em conta o comportamento do gafanhoto e corresponde à etapa de estímulo do aumento da serotonina em outros insetos na região, onde após o estímulo os gafanhotos se agregam e o enxame deles parte para o ataque.

Análise: Processo que ocorre quando se é verificada a anomalia da rede na nuvem, nela, baseado no que se é determinado, o processo pode correr da seguinte forma, ou promove o isolamento da rede até a realização do extermínio da anomalia (caso de tentativa de intrusão), ou realiza apenas a avaliação da ocorrência, a fim de determinar o

problema e a possível solução (dados corrompidos, por exemplo). Então se passa para o processo final. Em comparação ao gafanhoto, esse processo corresponde à fase de agregação, onde um número considerado de gafanhotos já está pronto para o ataque ao alimento.

Reparo: Essa função do objeto está associada ao uso de uma base de conhecimentos para a solução do ocorrido, como seria o caso do bloqueio do tráfego de dados caso seja detectada uma possível intrusão, dentre outras mais. Ao concluir o reparo, ocorre a dispersão dos processos, que, se comparado aos gafanhotos, seria como se eles tivessem terminado seu alimento e retornado ao seu estado solitário.

Os objetos de identificação (gafanhotos solitários) estão dispersos na rede na forma aleatória, a fim de identificar variações da rede. Na Figura 4 representa-se uma possível detecção da anomalia no canal de comunicação da nuvem por parte do modelo do gafanhoto, no qual ele identifica mudança quanto ao padrão da rede e logo começa a analisar.

O canal de comunicação com o cliente deve estar sempre protegido, no caso, ele deve ter objetos de identificação acompanhando em todo momento os processos na rede. A proposta define a existência de pelo menos um objeto de identificação por rede conectada à nuvem, para realizar isso, protegendo os dados na rede, principalmente nos componentes da aplicação, da máquina virtual, do armazenamento e da rede em si.

7 Representação em uma aplicação de nuvem

A fim de apresentar melhor o modelo proposto, será descrita a seguir a forma como o processo deve ocorre em uma nuvem.

Na Figura 5 são apresentados os componentes de análise baseado nos objetos de identificação presentes no canal de comunicação da nuvem. Também é apresentado o gerente de monitoramento dos dados, responsável por verificar e analisar os dados de comunicação nos nós dos componentes da nuvem (recursos estruturais da nuvem), além dos dados recebido no banco de dados da aplicação.

O canal de comunicação também possui os dados trafegados das aplicações dos clientes com a nuvem, como já mencionado, estando presente nela um administrador. Esse administrador contém informações dos dados, que realizam ações paralelas com os nós de recursos (possuem conexão direta). Os dados dos nós de recursos são utilizados pelo gerente de monitoramento dos dados e ele faz todo o processo já explicado com relação aos objetos de identificação (os gafanhotos), ou seja, realiza as ações apresentadas na arquitetura.

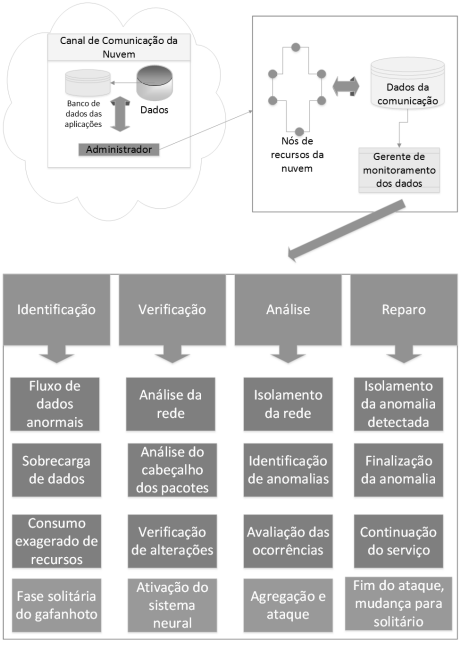

Figura 5: Etapas da ação do modelo proposto na nuvem.

8 Conclusão

O presente estudo se fez em torno do desenvolvimento de uma proposta de um modelo de arquitetura para realizar a análise de dados em tráfego, a fim de encontrar e resolver problemas de anomalias baseado no enxame de gafanhotos. A proposta foi motivada da necessidade de identificar aspectos de segurança da nuvem, pois, como visto, segurança é sempre importante a ser considerada.

A relevância do modelo que aqui foi projetado está ligada à sua contribuição ao analisar os dados e assim possibilitar a identificação de possível carência de segurança na rede da nuvem através de um modelo de inteligência computacional. Segundo o modelo proposto é possível visualizar sua importância no requisito de segurança na computação em nuvem.

Referências

Ando, S. (2007). Clustering needles in a haystack: An information theoretic analysis of minority and outlier detection, *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, IEEE, Yokohama, pp. 13–22.

Anstey, M. L., Rogers, S. M., Ott, S. R., Burrows, M. and Simpson, S. J. (2009). Serotonin mediates behavioral gregarization underlying swarm formation in desert locusts, *Science* **323**: 627–630.

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson,

D., Rabkin, A., Stoica, I. and Zaharia, M. (2010). A view of cloud computing, *Communications of the ACM* **53**(04): 50–58.

da Silva, A. C. B. (2003). Sistema de detecção de intrusão baseado em métodos estatísticos para análise de comportamento, http://goo.gl/DiODq1. Access date: 13 October 2015.

da Silva, F. R. H. (2010). Um estudo sobre os benefícios e os riscos de segurança na utilização de cloud computing, https://goo.gl/h3N3Cx. Access date: 09 July 2015.

de Barros Paranhos da Costa, G. (2014). Detecção de anomalias utilizando métodos paramétricos e múltiplos classificadores, http://goo.gl/22K9vR. Access date: 12 October 2015.

de Sá Silva, L., Montes, A. and da Silva, J. D. S. (2004). Uma solução híbrida para detecção de anomalias em redes, *IV Workshop de Computação Aplicada (WORCAP)*, IV WORCAP, São José dos Campos.

Donevski, A., Ristov, S. and Gusev, M. (2014). Security assessment of eucalyptus' web management interface, *in* IEEE (ed.), *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention*, pp. 1372–1375.

Dorigo, M. (1992). *Optimization, Learning and natural algorithms*, PhD thesis, Politecnico di Milano.

Erramilli, A., Narayan, O. and Willinger, W. (1996). Experimental queueing analysis with long-range dependent packet traffic, *IEEE/ACM Transactions on Networking* **4**(2): 209–223.

Eucalyptus, H. H. (2015). Hewlett packard enterprise development lp, https://www.eucalyptus.com/. Access date: 08 July 2015.

Júnior, A. F. L. and de Oliveira Dantas, A. C. A. (2012). Implementação de um detector de anomalias de tráfego de rede baseado na entropia de métricas para sistemas de computação em nuvem, *Congresso Norte Nordeste de Pesquisa e Inovação (CONNEPI)*, VII CONNEPI, Palmas.

Karaboga, D. and Gorkemli, B. (2011). A combinatorial artificial bee colony algorithm for traveling salesman problem, *Innovations in Intelligent Systems and Applications (INISTA), 2011 International Symposium*, IEEE, Kayseri, Turkey.

Kennedy, J. and Eberhart, R. (2001). *Swarm Intelligence*, Morgan Kaufmann Publishers Inc, San Francisco, CA, USA.

Kim, S. S. and Reddy, A. L. N. (2008). Statistical techniques for detecting traffic anomalies through packet header data, *IEEE/ACM Transactions on Networking* **16**(3): 562–575.

Laureano, M. A. P. (2005). Gestão de segurança da informação, http://goo.gl/Vbuqmu. Access date: 11 July 2015.

Mazzer, L. (2014). A segurança em computação nas nuvens, http://goo.gl/DaIjgh. Access date: 08 July 2015.

McCabe, L. and Aggarwal, S. (2012). Clareando as nuvens para empresas de midmarket, http://goo.gl/2ovUK8. Access date: 07 July 2015.

Mell, P. and Grance, T. (2011). The NIST definition of cloud computing, *Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology Gaithersburg* .

Nascimento, E. G. S. (2012). Um algoritmo baseado em técnicas de agrupamento para detecção de anomalias em séries temporais utilizando a distância de mahalanobis, http://goo.gl/otxdPm. Access date: 27 October 2015.

Openstack (2015). Open source software for creating private and public clouds, https://www.openstack.org/. Access date: 08 July 2015.

Sousa, F. R. C., Moreira, L. O. and Machado, J. C. (2010). Computação em nuvem: Conceitos, tecnologias, aplicações e desafios, *Escola Regional de Computação dos Estados do Ceará, Maranhão e Piauí (ERCEMAPI)*, ERCEMAPI, Parnaíba.

Velte, A. T., Elsenpeter, R. C. and Velte, T. J. (2011). *Cloud Computing - Computação em Nuvem: uma abordagem prática*, Alta Books.

Veras, M. (2013). Arquitetura corporativa de nuvem: amazon web services (aws).

Yong, E. (2009). The swarm-maker molecule: how serotonin transforms solitary locusts into social ones, http://goo.gl/TGoP1i. Access date: 02 March 2016.

Zarpelao, B. B., Mendes, L. S., Abrao, T., Sampaio, L. D. H., Lima, M. F. and Junior, M. L. P. (2009). Detecção de anomalias em redes de computadores, *XXVII Simpósio Brasileiro de Telecomunicações (SBrT 2009)*, SBrT2009, Blumenau, Santa Catarina.