UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

Fábio Sousa de Sant'Ana

Avaliação de desempenho de mecanismos de segurança em ambientes PACS (*Picture Archiving and Communication System*) baseados em computação em nuvem.

Ribeirão Preto - SP
2016

Fábio Sousa de Sant'Ana

Avaliação de desempenho de mecanismos de segurança em ambientes PACS (*Picture Archiving and Communication System*) baseados em computação em nuvem.

Dissertação de Mestrado apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do Título de Mestre em Ciências, Programa de Pós-Graduação Mestrado Profissional em Gestão de Organizações de Saúde.

Orientador: Prof. Dr. Paulo Mazzoncini de Azevedo Marques

FICHA CATALOGRAFICA

Sant'Ana, F. S.

Avaliação de desempenho de mecanismos de segurança em ambientes PACS (Picture Archiving and Communication System) baseados em computação em nuvem.

62p.:il.;30cm

Dissertação de Mestrado apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do Título de Mestre em Ciências, Programa de Pós-Graduação Mestrado Profissional em Gestão de Organizações de Saúde.

Orientador: Marques, Paulo Mazzoncini de Azevedo

1. PACS over *cloud computing* 2. Segurança em nuvem 3. SSH over cloud.

FOLHA DE APROVAÇÃO

Nome: Fábio Sousa de Sant'Ana

Titulo: Avaliação de desempenho de mecanismos de segurança em ambientes PACS (Picture Archiving and Communication System) baseados em computação em nuvem.

Dissertação de Mestrado apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do Título de Mestre em Ciências, Programa de Pós-Graduação Mestrado Profissional em Gestão de Organizações de Saúde.

Aprovado em:

Banca Examinadora

Prof. Dr._____

Instituição:_____Assinatura:_____

Prof. Dr._____

Instituição:_____Assinatura:_____

Prof. Dr._____

Instituição:_____Assinatura:_____

Ao meu pai João, minha mãe Lena e minha irmã Natália pelo melhor exemplo de perseverança, por serem meus incentivadores e por serem a personificação do amor incondicional.

À minha amada esposa Veridiana pela paciência e incentivo durante toda esta etapa de minha vida. Você foi a melhor escolha que eu fiz até hoje. Ao meu filho João Lucas, razão do todos os meus esforços.

Agradecimentos

Ao Prof. Dr. **Paulo Mazzoncini de Azevedo**, meu orientador, por ter aceitado me orientar e principalmente por ter confiado em meu trabalho e minha capacidade. Obrigado por tornar esta jornada mais fácil e contribuir para meu crescimento profissional e científico.

À. Dra. **Kátia Mitiko Suzuki** pelo apoio e confiança, conselhos e por abrir as portas da Seção Técnica de Informática - STI/FMRP para o meu aprendizado pessoal, compreendendo meus prazos e obrigações. Obrigado por ser exemplo de profissional, pesquisadora e chefe.

À **toda equipe da STI/FMRP** pela contribuição e boa vontade em me ajudar em tudo que fosse necessário. Obrigado por tornarem nossa convivência tão agradável, descontraída e construtiva ultrapassando as paredes da seção.

Aos colegas Msc. **Saulo Cordeiro** e MSc. **Hugo Rodrigues**, do Centro de Ciência das Imagens e Física Médica do HCFMRPUSP, pela colaboração, dedicação e sugestões neste projeto.

Aos amigos **César e Gabriella Ushiro, Derek e Thaisa Bragueto, Patrick Ribeiro, Paulo e Priscila Torres** pela amizade, momentos descontraídos, palavras de compreensão e incentivo durante toda esta estapa.

Ao meu sogro **Nehemias Suazo,** minha sogra **Maria José Kiill** e minha cunhada **Josuelle Kill Suazo** pelo incentivo em todas as etapas da minha vida.

"Enquanto o homem não souber para que porto quer ir, nenhum vento será o vento certo."

Sêneca

RESUMO

Sant'Ana, F. S. **Avaliação de desempenho de mecanismos de segurança em ambientes PACS (Picture Archiving and Communication System) baseados em computação em nuvem.** 2016. Dissertação de Mestrado – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto.

Introdução: A adoção de um Sistema de Arquivamento e Distribuição de Imagens (PACS, do inglês Picture Archiving and Communication System) é condição fundamental para a estruturação de um ambiente radiológico sem filme. Um PACS é composto basicamente por equipamentos e sistemas informatizados interconectados em rede, direcionados à aquisição, armazenamento (ou arquivamento), recuperação e apresentação de imagens médicas aos especialistas responsáveis por avaliá-las e laudá-las. A computação em nuvem vem ao encontro dos PACS e surge como uma maneira de simplificar o compartilhamento de imagens entre organizações de saúde e promover a virtualização de espaços físicos e para garantir o seu funcionamento ininterrupto.**Objetivo:** Este estudo teve como objetivo implementar um *PACS* simplificado em ambiente *cloud computing* privado, com foco nas funcionalidades de arquivamento e disponibilização de imagens médicas e avaliar questões de segurança e performance. **Metodologia:** As imagens que compuseram o PACS do ambiente *cloud* foram obtidas através do PACS físico atualmente em uso no Centro de Ciência das Imagens e Física Médica do Hospital das Clinicas da Faculdade de Medicina de Ribeirão Preto – CCIFM/HCFMRP. Para os procedimentos da avaliação de segurança foram construídos cenários que possibilitavam a: 1) anominização de dados de identificação dos pacientes através de criptografia computacional em base de dados utilizando o algoritmo de criptografia *Advanced Encryption Standard - AES*, 2) transferência de imagens médicas seguras através de conexão com a Internet utilizando *Virtual Network Private* – VPN sobre o protocolo *Internet Protocol Security* – IPSec (VPN/IPSec) e 3) envio seguro através de tunelamento baseado em *Secure Shell* – SSH. **Resultados:** Foi identificada uma queda de performance no envio de informações para a nuvem quando submetidos aos níveis de segurança propostos, sugerindo a relação entre aumento de segurança e perda de performance, apontando para a necessidade de estudos de desempenho quando da condução de projetos envolvam a adoção em ambientes clínicos de solução PACS baseada em *cloud computing*.

Palavras Chaves: PACS; Nuvem; PACS over *cloud computing*, segurança em nuvem, SSH.

ABSTRACT

Sant'Ana, FS. **Performance evaluation of security mechanisms in PACS environments (Picture Archiving and Communication System) based on cloud computing.** *2016. Master's Dissertation - Ribeirão Preto Medical School , University of São Paulo, Ribeirão Preto.*

Introduction: the adoption of a PCAS (Picture Archiving and Communication System) is fundamental for the structuring of a radiological environment without film. A PACS comprises, essentially, hardware and information systems interconnected in a network, oriented towards acquisition, storage (or archiving), retrieving and presentation of medical images to specialists entrusted with analyzing and assessing them. Cloud computing comes to support of PCAS, simplifying medical imaging sharing between health care organizations and promoting the virtualization of physical infrastructure to assure uninterrupted availability of the PCAS. **Goal**: This study aimed to implement a simplified PCAS in a private cloud computing environment, and subsequently to evaluate its security and performance. **Methodology**: The images that formed the new PCAS were obtained from the exiting PCAS of Centro de Ciência das Imagens e Física Médica of Hospital das Clinicas da Faculdade de Medicina de Ribeirão Preto – CCIFM/HCFMRP. To evaluate its security, scenarios were built within the following framework: 1) patient identification data anonymization through computational database cryptography, using the AES (Advanced Encription Standards) algorithm ; 2) transfer of encrypted medical images on the Internet using VPN (Virtual Private Network) over IPSec (Internet Protocol Security); and 3) safe traffic through Secure Shell (SSH) tunneling. **Results**: There was a performance drop on traffic of information to the cloud under the proposed security levels that suggests a relationship between increase in security and loss of performance, pointing to the need for performance studies when the project involving driving adoption in clinical environments PACS solution based on cloud computing.

KeyWords:*PACS; Cloud; PACS over cloud computing, cloud security, SSH.*

LISTA DE ILUSTRAÇÕES

LISTA DE TABELA

LISTA DE ABREVIATURAS E SIGLAS

CCIFM	Centro de Ciência das Imagens e Física Médica
DICOM	*Digital Imaging and Communications in Medicine*
HCFMRPUSP	Hospital das Clinicas da Faculdade de Medicina de Ribeirão Preto
USP	Universidade de São Paulo
IAAS	*Infrastructure as a Service*
PAAS	*Plataform as a Service*
PACS	*Picture Archiving and Communication System*
RAID	*Reduntant Array of Inexpensive Disk*
SAAS	*Software as a Service*
SFTP	*Secure File Transfer Protocol*
TIC	Tecnologia da Informação e Comunicação
USP	Universidade de São Paulo
VHD	*Dinamic Virtual Hard Disk*
SGBD	*Sistema Gerenciador de Banco de Dados*
VPN	*Virtual Private Network*
VLAN	*Virtual Local Area Network*
ACL	*Access Control List*
CPU	*Central Processing Unit*
CRC	*Cyclic Redundancy Check*

SUMÁRIO

1. Introdução

A adoção de um Sistema de Arquivamento e Distribuição de Imagens *(PACS, do inglês Picture Archiving and Communication System)* ajuda aos hospitais, ou unidades de saúde em geral, na aquisição, gerenciamento, armazenamento, recuperação e visualização de imagens médicas. Um *PACS* é composto basicamente por equipamentos e sistemas informatizados interconectados em rede, direcionados à aquisição, armazenamento (ou arquivamento), recuperação e apresentação de imagens médicas aos especialistas responsáveis por avaliá-las e laudá-las.

As imagens provenientes das diversas fontes de aquisição, tais como: radiologia, tomografia computadorizada, ressonância magnética, ultrassonografia, medicina nuclear etc. são direcionadas ao *PACS* em formato eletrônico, permitindo que especialistas avaliem-nas e realizem seus respectivos diagnósticos. Um dos benefícios de uso de *PACS* é a economia de tempo, uma vez que os especialistas não precisam aguardar as impressões dessas imagens em filmes, podendo se dedicar integralmente à leitura e interpretação dessas imagens e à realização de diagnósticos, sem contar a economia de espaço físico que seria necessário, além da logística associada, para o armazenamento, manutenção e recuperação desses filmes **[1].**

As imagens médicas em *PACS* possuem um padrão de formatação que permite que sejam trocadas entre equipamentos e sistemas informatizados. Um dos padrões mais conhecidos e difundidos atualmente é o *Digital Imaging and Communications in Medicine (DICOM)*. Trata-se de uma norma que facilita a interoperabilidade dos sistemas, garantindo a conformidade em um ambiente de múltiplos fornecedores e que tem o potencial de facilitar a implementação de soluções envolvendo PACS **[2].**

Além dos padrões que garantem a conformidade entre os diversos ambientes envolvidos, outros aspectos devem ser observados no uso de *PACS*. Sob a óptica da Tecnologia da Informação e Comunicação (TIC), planos de controle de qualidade e de manutenção preventiva são fundamentais em *PACS***.**

O desempenho é um aspecto muito importante em *PACS*. Antes de construir uma rede de dados para o gerenciamento de imagens médicas, é imprescindível uma análise rigorosa de requisitos, tais como: largura de banda disponível, volume de dados, padrões de trabalho esperado e necessidades operacionais **[3]**. Tais requisitos podem influenciar diretamente no desempenho esperado.

Um ambiente com alto desempenho possibilita que as imagens médicas transitem em alta velocidade. A forma como as imagens eletrônicas são transportadas entre os vários componentes desse sistema também deve ser definida pela arquitetura empregada na criação do ambiente. A arquitetura mais comum em hospitais é a arquitetura centralizada. Nela, os exames são enviados diretamente para o servidor de arquivamento e, a partir destes, para as estações de trabalho onde serão analisadas (Figura 1). **[4]**

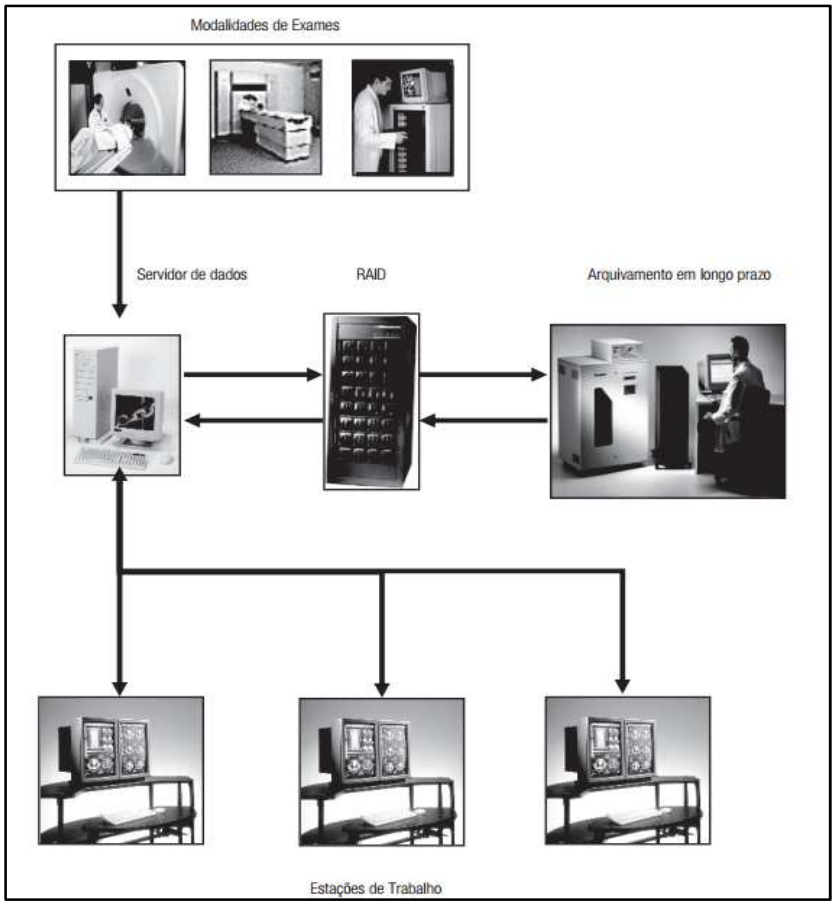

Figura 01: Exemplo de arquitetura centralizada de PACS.

Desde a sua geração, as imagens médicas possuem invariavelmente resoluções altíssimas, e, consequentemente, consomem bastantes recursos de redes de comunicação para trafegarem entre os equipamentos. Os computadores que recebem essas imagens, normalmente, precisam ser dotados de alto poder computacional. Neste caso, exige-se que o processador (componente de *hardware* dedicado aos cálculos e decisões lógicas) seja capaz de processar a imagem no mínimo tempo possível, permitindo ao responsável realizar um diagnóstico mais rápido.

Além disso, considera-se que a alta disponibilidade seja um fator relevante no uso de *PACS*. Algumas instituições sugerem estratégias que incluem esquemas de rede completamente redundantes com equipamentos multiplamente conectados **[3]**. A alta disponibilidade é composta por um ambiente de computação tolerante a falhas de *hardware*, *software* e de energia elétrica, tendo como objetivo principal manter os serviços disponíveis o máximo de tempo possível. Os equipamentos operam em redundância e, quanto maior a redundância, menores serão as probabilidades de interrupções no serviço. O uso ininterrupto de *PACS* confere maior rapidez à realização dos diagnósticos, fluidez nos atendimentos, evitando maiores problemas.

Um *PACS* também precisa ser seguro. Segurança é um componente crucial para qualquer rede de dados que trafegue imagens médicas e pode ser geralmente dividida em 3 camadas de proteção: segurança física, segurança administrativa e segurança eletrônica. **[3]**.

Por serem utilizados em muitos hospitais, os *PACS* vêm tornando-se uma missão crítica e requerendo alta disponibilidade **[4]** de armazenamento. Os principais dispositivos responsáveis pelo armazenamento em *PACS* incluem discos magnéticos (*Hard disks*, *HDs*), conjunto redundante de discos independentes (*Reduntant Array of Inexpensive Disks, RAID),* fitas magnéticas e, também, discos compactos (CDs e DVDs). **[5]**. A redundância de equipamentos de armazenamento garante que, em caso de falhas no sistema de armazenamento, haja sempre um sistema capaz de assumir a operação em curso. Entretanto, sempre existe a necessidade de realização de cópias de segurança (*backups*) rotineiras para garantir que as informações, uma vez indisponíveis, sejam recuperadas.

2. Considerações Teóricas

2.1 Computação em nuvem

Hoje, o termo computação em nuvem descreve uma tecnologia que permite o acesso ubíquo, conveniente e sob demanda a um conjunto compartilhado de recursos computacionais configuráveis, composto por redes, servidores, armazenamento, aplicações e serviços, que podem ser rapidamente provisionados ou liberados com um esforço mínimo de gerenciamento.

A computação em nuvem se refere, essencialmente, à noção de utilizar, em qualquer lugar e independente de plataforma, as mais variadas aplicações ou serviços por meio da internet com a mesma facilidade de tê-las instaladas em infraestrutura de rede locais, sem que seja preciso conhecer a infraestrutura que existe por detrás das aplicações. A evolução constante da tecnologia computacional e das telecomunicações têm feito com que o acesso à internet se torne cada vez mais amplo e rápido. Esse cenário cria a condição perfeita para a popularização da computação em nuvens, pois faz com que o conceito se dissemine no mundo todo.

Segundo **[6]**, existem alguns modelos de implantação para *cloud computing* descritos a seguir

Nuvem pública: É considerada uma nuvem pública quando a infraestrutura disponível para contratação consiste em recursos compartilhados, padronizados e com autoatendimento pela Internet. É considerada um modelo pague-por-uso e são oferecidas por grandes organizações que possuem grande capacidade de armazenamento e processamento.

- **Nuvem privada:** É a infraestrutura que utiliza as características da computação em nuvem, como a virtualização, mas na forma de uma rede privada. Geralmente, os serviços são oferecidos para serem utilizados pela própria organização, não estando publicamente disponível para uso geral.

- **Nuvem híbrida:** É a infraestrutura composta por nuvem privada e por nuvem pública, que continuam a ser entidades únicas, porém conectadas através de tecnologias proprietárias.

De uma maneira geral, as nuvens privadas tendem a ser utilizadas quando há necessidade de níveis mais rigorosos de segurança e privacidade, ou de garantia de disponibilidade da aplicação, enquanto que as nuvens públicas estão diretamente relacionadas à necessidade de redução de custos, uma meta constante dos gestores de organizações, pois sabe-se que a adoção de nuvens públicas exige menor investimento.

A computação em nuvem não representa uma tecnologia em si e, sim, um modelo de TI, que tem como base serviços e não produtos, sendo disponibilizada em diferentes camadas: *Software as a Service* (SAAS), *Plataform as a Service* (PAAS) e *Infrastructure as a Service* (IAAS).

Resumidamente, a *SAAS* utiliza a Internet para oferecer aplicativos que são gerenciados por um terceiro e utiliza-se de uma interface que pode ser acessada pelos clientes para interação com as aplicações. A *PAAS* é um camada da nuvem designada à criação, hospedagem e controle de software, oferecendo um ambiente pronto para rodar aplicações, por exemplo, um provedor de acesso. Já o *IAAS* oferece uma infraestrutura pronta para que se instale servidores seja qual for a plataforma, portanto, toda a infraestrutura de servidores, sistemas de rede, armazenamento, e todo o ambiente necessário para o funcionamento são disponibilizados como serviços.

A computação em nuvem vem ao encontro dos *PACS* e surge como uma maneira de simplificar o compartilhamento de imagens entre organizações de saúde **[6]** e vem demonstrando ser uma poderosa ferramenta para promover a virtualização de espaços físicos **[7]** e para garantir o funcionamento ininterrupto de *PACS*.

As “nuvens”, como também é simplificadamente referenciada a computação em nuvem, proporcionam escalabilidade (ou elasticidade), onde, através das mesmas torna-se possível ampliar a disponibilidade de recursos conforme a necessidade (demanda) **[8]**. Esta funcionalidade apresenta-se como uma solução de grande interesse em *PACS* devido ao seu crescimento, dificilmente mensurável, ao longo do tempo.

Tecnicamente, nuvens dizem respeito a um modelo de provisionamento para processamento virtualizado e capacidade de armazenamento digital**[9]**.

Basicamente, são compostas por vários computadores que trabalham conjuntamente para realizar processamentos de maior complexidade, como se fossem uma única máquina. Este tipo de solução também é conhecido como *cluster*. Neste caso, as nuvens podem oferecer poder computacional muito superior ao obtido em servidores de *hardware* comuns. Outras características, como memória e disco, também podem ser provisionadas com capacidades superiores às convencionais.

As nuvens também garantem operações ininterruptas, sendo que problemas comuns de energia elétrica dificilmente acontecem, pois, assim como acontece com os equipamentos de rede, o fornecimento de energia elétrica aos *datacenters* (nome com que são chamados os centros de computação onde os servidores de *hardware* que compõem as nuvens estão alojados) geralmente é redundante e estável, garantido por equipamentos do tipo *no-breaks* de grande capacidade e, além disso, apoiados por geradores de energia elétrica alternativa.

O estudo sobre a utilização de *PACS* em nuvens ainda é bastante recente, mas nota-se uma grande expectativa por parte da comunidade científica sobre o tema. Além disso, várias empresas têm percebido tais necessidades e têm direcionado estratégias comerciais para atender aos *PACS* através da tecnologia de nuvens **[10]**.

Em processos de migração de ambientes PACS para a nuvem, na maioria das vezes a informação é o único componente preservado. Ativos tecnológicos como software, hardware e rede são essenciais para o funcionamento dos PACS, mas cada um desses componentes pode ser substituído no momento que um ambiente é transferido para a nuvem, exceto a informação. Sabe-se que, por se tratarem de infraestruturas que geralmente funcionam através de conexões de Internet, as informações que trafegam do cliente até o servidor PACS na nuvem e da nuvem para o cliente, podem ser expostas a riscos, agravando-se mais quando de tais informações pertencem a terceiros, como acontece em hospitais. Os possíveis desafios de segurança para computação em nuvem, segundo **[11]** são: roubo de informações sigilosas e garantia da integridade das informações armazenadas. Por outro lado, quando se tenta diminuir os riscos adotando mecanismos de segurança através de ferramentas de tecnologia da informação, esses ambientes podem sofrer

uma considerável queda de performance, podendo muitas vezes inviabilizar o projeto ao qual se destinam.

2.2 VPN

As comunicações entre dispositivos em um ambiente PACS são feitas seguindo as regras do padrão DICOM, um protocolo de aplicação que é executado sobre o protocolo de transporte TCP. Essas tecnologias de comunicação por si só não definem um canal de comunicação seguro. A principal alternativa para solucionar esse problema é a adoção de VPNs. Basicamente, a VPN é uma rede privada construída sobre a Internet, utilizando tecnologias de tunelamento e criptografia para manter seguros os dados trafegados. Todo o tráfego gerado e transmitido é protegido com chaves de acesso, e somente pessoas e computadores de posse dessa chave podem destravar o bloqueio e acessar o conteúdo transmitido **[12]**. Dessa forma, caso o pacote de dados seja interceptado durante o transporte, seu conteúdo estará codificado e inelegível, sendo desencapsulado e decriptografado no destino, retornando ao seu formato original e garantindo confidencialidade, autenticação e integridade da informação.

2.3 Túnel SSH

O protocolo SSH foi criado para ser um substituto do protocolo Telnet e de outros protocolos inseguros de acesso remoto. Para isso passou a utilizar criptografia dos dados, provendo dessa forma confidencialidade e integridade dos dados mesmo que eles estejam trafegando em uma rede insegura como a internet **[13].** O protocolo possui duas versões, conhecidas como SSH1 e SSH2. A versão SSH2 é uma versão melhorada da versão anterior, possui novas funcionalidades que garantem uma segurança e integridade melhor dos dados. A segurança foi fortalecida através da troca de chaves *Diffie–Hellman* e também pela checagem de integridade utilizando algoritmos de MAC (*Message Authentication Code*)**[14]**.

Um túnel SSH é uma conexão criptografada entre duas máquinas (o servidor SSH e o cliente) que tem como objetivo redirecionar o tráfego entre a máquina remota (o servidor SSH) e uma terceira. Dessa forma, as informação

enviadas do cliente para o servidor SSH estarão codificadas e serão enviadas para um terceiro servidor, destino das informações.

2.4 Criptografia AES

Conforme o documento que a normaliza *(disponível em http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf, acessado em 27/06/2015),* a criptografia *Advanced Encryption Standard* - AES foi homologada pelo instituto americano NIST - *National Institute of Standards and Technolog,* em 2001 como um algoritmo de criptografia utilizado para proteger dados eletrônicos através da conversão de dados para uma forma ilegível chamada *ciphertext*, e decriptografando a *ciphertext* para o seu formato original, chamado *plaintext.* A criptografia AES é capaz de usar chaves de 128, 192 ou 256 bits para criptografar e decriptografar dados em blocos de 128 bits. Feita corretamente, a encriptação protege os dados tornando difícil, ou quase impossível em alguns casos, para um invasor obter um arquivo em seu texto plano. A ideia é que qualquer tempo de computação e perícia necessários para quebrar uma encriptação sejam mais custosos que o valor percebido da informação sendo protegida **[15]**.

2.5 O projeto *cloud* USP

O projeto Nuvem USP foi desenvolvido pela Superintendência de Tecnologia de Informação da USP em 2010. De acordo com a Agência USP de Notícias, o projeto teve por objetivo a criação de servidores virtuais para o apoio na gestão da tecnologia da informação na Universidade de São Paulo, facilitando o monitoramento e o gerenciamento dos servidores, o controle dos backups e o dimensionamento da real necessidade de processamento e armazenamento de cada unidade, o que proporcionou a eliminação de gastos com equipamentos ociosos. Desde 2014, o Serviço de Informática de Imagens Médicas do CCIFM, em parceria com a Seção Técnica de Informática da Faculdade de Medicina de Ribeirão Preto, vem realizando testes de simulação de ambientes PACS sobre a infraestrutura oferecida como serviço pela Nuvem USP. O objetivo dos administradores do PACS e da coordenação do CCIFM é avaliar a viabilidade da implantação de todos os servidores de imagens, aplicações e bancos de dados do

PACS no ambiente de nuvem da USP, sem degradação de seus requisitos não funcionais.

2.6 Armazenamento de dados DICOM em PACS

Em modelos tradicionais de armazenamento PACS, as imagens DICOM são armazenadas em arquivos DICOM. Neste caso, a tabela *Imagem* do banco de dados conterá apenas os nomes dos arquivos. Quanto o PACS necessita obter uma imagem de determinado paciente, o registro do estudo, a série e o registro do arquivo de imagem será encontrada no banco de dados, porém os arquivos da imagem propriamente dita serão localizados em discos rígidos (ou storages) e carregados no PACS como um objeto DICOM. A maior vantagem desse modelo é a simplicidade e também o fato de estarem separados que pode ser bastante facilitador em uma futura migração, quando por exemplo o antigo PACS precisa ser substituído por um novo.

O estudo sobre o armazenamento de grandes blocos binários de imagens diretamente em tabelas de banco de dados assim como são armazenados os registros de pacientes e ids de imagens, ainda é muito recente, portanto a maioria dos sistemas ainda armazenam os arquivos DICOM em dispositivos de armazenamento. No entanto, a principal vantagem desse modelo é a possibilidade de desfrutar de todas as ferramentas que um banco de dados pode oferecer. Por exemplo: O cabeçalho DICOM passa a ser criptografado com as ferramentas de criptografia do banco de dados e imediatamente proporcionar uma camada segura ao arquivo. Além disso, todos os bancos de dados atuais oferecem ferramentas de auditorias para identificar quando, como e por quem qualquer objeto DICOM é acessado ou modificado. O principal problema com o armazenamento de dados baseado em banco de dados é que, a informação também passa a depender do encapsulamento de dados do banco particular, sendo necessário sempre recorrer ao banco de dados específico para ter acesso à informação, além da possibilidade de degradação de performance do banco, ou seja, a visualização de uma imagem armazenada diretamente no banco de dados e não em disco rígido como normalmente os sistemas DICOM atuais operam, pode trazer uma instabilidade, mesmo com as principais tecnologias de indexação de dados, e que pode ser

perceptível ao usuário, além do aumento significante no tamanho da base de dados, podendo gerar instabilidade no sistema se houver uma grande quantidade de acessos simultâneos.

3. Segurança em base de dados DICOM

De acordo com **[16],** qualquer metodologia ou abordagem adotada em uma organização deve considerar três aspectos de segurança da Informação:

Ataques de Segurança: são quaisquer ações que possam comprometer a disponibilidade, integridade, sigilo e autenticidade de uma informação pertencente a uma organização.

Mecanismos de Segurança: são mecanismos projetados para se detectar, prevenir ou se recuperar de um ataque de segurança.

Serviços de Segurança: são funções que aumentam o nível de segurança dos sistemas de processamento de dados e das transmissões de informação em uma Organização. Estes serviços podem utilizar um, ou mais, mecanismos de segurança.

Felizmente, a tecnologia de segurança da informação já superou muitos dos desafios relacionados as questões de integridade, privacidade, autenticidade e disponibilidade de informações para documentos eletrônicos, porém, deve-se entender que quando se trata de segurança, não existe uma solução totalmente segura, imune a ataques; o que existe, é uma informação que possui um nível de segurança maior que outra. E quanto maior o nível de segurança, maior será a confiança depositada.

O banco de dados de um PACS pode ser considerado como um dos componentes mais críticos uma vez que nele estão registradas todas as informações referente às imagens dos pacientes. Um banco de dados inseguro pode comprometer a segurança das informações nele inseridas e trazer graves problemas para as organizações de saúde.

3.1 Conquest

Conquest é um serviço DICOM completo que foi desenvolvido no *Netherlands Cancer Institute,* por Marcel van Herk e Lambert Zijp, com base em um código de domínio público (*UCDMC DICOM*) desenvolvido inicialmente no *Medical Center of the University of Califórnia at Davis* por Mark Oskin.

3.2 Segurança DICOM Conquest

O Conquest armazena em banco de dados somente as informações dos cabeçalhos das imagens, são elas: *ID do paciente, Nome do paciente, Data de Nascimento, Sexo, Data do estudo*, além de informações técnicas da imagens e da sua origem ou de outras informações que podem ser incluídas para que seja possível consulta através de variáveis, como podem ser observadas através da Figura 02.

←T→			PatientID	PatientNam	PatientBir	PatientSex	AccessTime	qTimeStamp	qFlags	qSpare
Editar	Inline Edit	Copiar Remover	52/5398	SILVA^FULANO	19750122	F	1435174624	NULL	NULL	NULL

PhotoMetri	Rows	Colums	BitsStored	ImageType	ImageID	ImagePat	SeriesInst	AccessTime
MONOCHROME1	512	512	16	NULL	NULL	52/5398	1.3.6.1.4.1.29565.1.4.2760659664.9584.1435174460.9...	1435174624

StationNam	Institutio	FrameOfRef	SeriesPat	StudyInsta	AccessTime
Demo Station	USP-TESTE	NULL	52/5398	1.3.6.1.4.1.29565.1.4.2760659664.9584.1435174460.9...	1435174624

Figura 02: Informações DICOM em banco de dados

Como pode ser visto, as informações são inseridas no banco de dados em texto-puro *(plain-text)*, sem nenhum tipo de codificação, sendo representadas por caracteres em formato totalmente legível e interpretável. Com isso, é possível fazer uma relação entre o paciente e a sua imagem de exame ou ao seu conjunto de imagens (séries), uma vez que essa informação também é registrada no banco de dados.

Mesmo as informações registradas podendo ser compreendidas visualmente, é importante ressaltar que as informações em banco de dados não estão à mercê de qualquer pessoa que solicite acesso. Normalmente, os SGBDs primam por métodos seguros de autenticação e autorização. Enquanto a autenticação preocupa-se em confirmar quem o usuário é, a autorização determina o que esse usuário pode fazer **[17]**. Na prática, os bancos dados devem ser acessíveis apenas através de senhas complexas, geralmente em posse de administradores de TI, controladas por rigorosas listas de controle de acessos (*acl's*) que definem quem pode acessar, quais tabelas e quais os direitos do usuário para manipulação do banco.

O Conquest é compatível atualmente com quatro diferentes SGBDs, são eles *DbaseIII*, *MySQL*, *Postgres* e *SqLite* e, mesmo em versões diferentes, a estrutura do banco é preservada, conforme Figura 03. Em computação, a estrutura do banco de dados é um modo particular de armazenamento e organização de dados dentro de tabelas que compõem a estrutura. Resumidamente, a tabela *DICOMImages* armazena as informações das imagens, a tabela *DICOMPatientes* armazena as informações de pacientes, a tabela *DICOMSeries* controla as séries recebidas da mesma forma com quem a tabela *DICOMStudies* controla os estudos. Já a tabela *DICOMWorkList* é utilizada para integração dos dados com outros sistemas, e por fim, a tabela UIDMODS armazena as modificações realizadas através de processos conhecidos por anonimização, permanecendo vazia quando não se opta por anonimizar informações de pacientes.

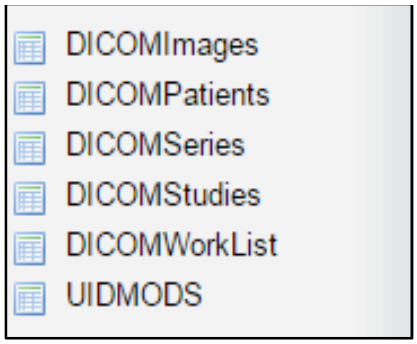

Figura 03: Conjunto de tabelas do Conquest

3.3 Anonimização de dados

No dicionário Aurélio a definição de anonimato é “sem o nome ou assinatura do autor; sem nome ou nomeada; obscuro”. Portanto, podemos dizer que no contexto da informatização dos dados a informação anônima é aquela que não seja possível identicar a quem ela se refere. Diante disso, a anonimização de dados é o processo de retirar as informações que possam levar à identicação dos usuários. Mais abrangentemente, essa anonimização engloba também o conteúdo da informação.

Para **[18],** o processo de anonimização contempla uma lista de atributos confidenciais, sejam eles ligados aos pacientes ou ligados à imagem, e os remove do arquivo DICOM. Como resultado, é produzido um arquivo DICOM anônimo que ainda contém a imagem e os dados não confidenciais, mas desprovidas de qualquer

informação que identifique o paciente. Dessa forma a informação anônima é inserida no banco de dados. O processo de desanonimização da imagem, que consiste na reversão das informações da imagem para o seu estado original compreensível pelo ser humano, também é realizado pela aplicação DICOM e é um procedimento imprescindível quando as informações estão codificadas.

Um arquivo DICOM não-anonimizado pode ser aberto por um leitor de textos (*notepad*, do *Microsoft Windows*, por exemplo) e as informações de seu cabeçalho podem ser acessíveis como apresentadas nas Figuras 04 e 05. A Figura 04 apresenta a visualização de informações não codificadas do cabeçalho da imagem, portanto, as informações do paciente, como por exemplo o seu nome (*Silva4^FULANO*) na 10ª linha, o ID do paciente (*71/225*), data de nascimento (*19910722*) e Sexo (*F*) na 11ª linha, podem ser lidas. Por outro lado, considerando a Figura 05, os dados do paciente como Nome e ID estão complementa ilegíveis (*U2FsdGVkX181yjeUfJTSWQ41bTYLqK67x0c0nW*+2YKQ= e *4063390747.2659023964* respectivamente).

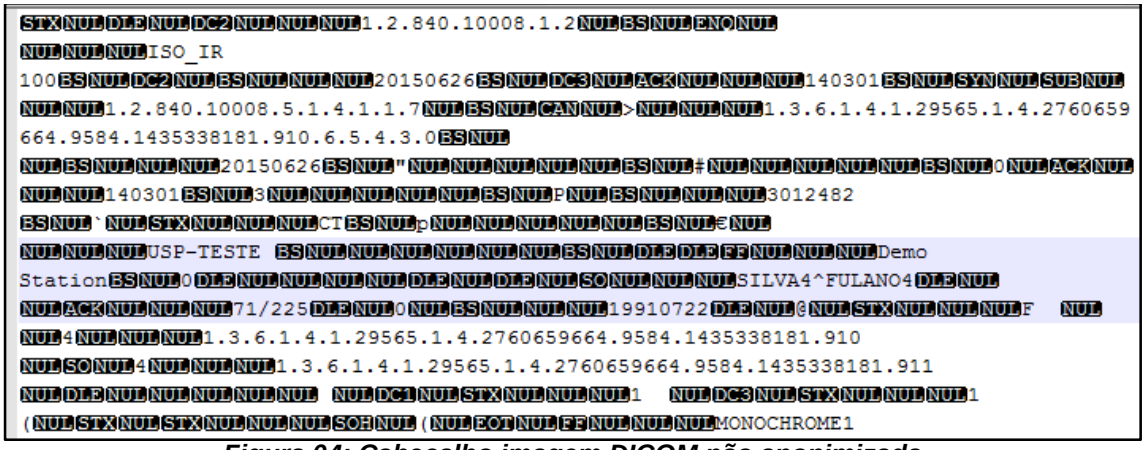

Figura 04: Cabeçalho imagem DICOM não anonimizada

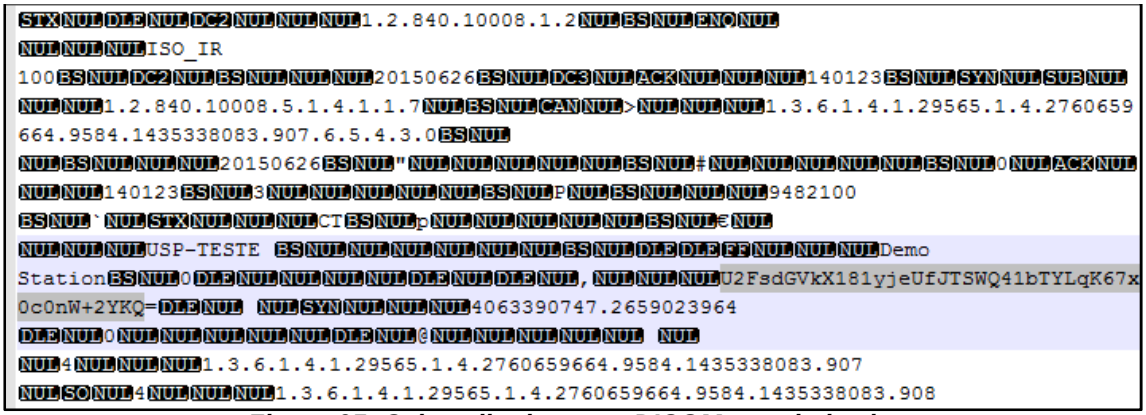

Figura 05: Cabeçalho imagem DICOM anonimizada.

3.4 Anonimização pelo Conquest

Diante da necessidade de se manter a privacidade e segurança dos dados, surgiram as chamadas ferramentas de anonimização de dados, que definem um conjunto de políticas e técnicas para tentar garantir o anonimato das informações ou dos sujeitos a quais elas pertencem, sem, no entanto, afetar a qualidade das informações.

O Conquest não possui um processo nativo de anonimização e desanonimização. Essas tarefas ficam por conta de dois *scripts* externos que são distribuídos juntamente com os arquivos fonte do Conquest e que foram desenvolvidos em uma linguagem de programação de alto-nível, multi-paradigma criada em 1993 no Brasil, conhecida por LUA. Em computação, *script* é o nome genérico para um programa externo escrito em linguagem de programação que, essencialmente, são interpretados em tempo de execução e adicionam funções a outros programas.

Na prática, o Conquest faz referência ao dois *scripts* em seu arquivo de configuração *dicom.ini* e permite que sempre que uma imagem for recebida pelo servidor (*IMPORT*) ou enviada à partir dele (*EXPORT*) os dados sejam anonimizados e desanonimizados, respectivamente. Posteriormente, o Conquest, registra em banco de dados as referências de cada modificação realizada.

Para que se possa entender como este processo acontece, criou-se uma imagem de um paciente fictício chamado JOÃO SOUZA cujo ID é 123456. Ao ser recebida pelo servidor, o Conquest dispara a execução do *script* de anonimização, que abre o cabeçalho DICOM da imagem e insere os dados do paciente, além dos dados da imagem, em uma tabela específica do banco de dados, como apresentados na Figura 06.

PatientID	PatientNam	PatientBir	PatientSex	AccessTime	qTimeStamp	qFlags	qSpare
3098571458.4279037106	PAT3098571458	*NULL*	*NULL*	1435343328	*NULL*	*NULL*	*NULL*

Figura 06: Imagem anonimizada em banco de dados

Como se observa, o campo *PatientID* que anteriormente era *123456* foi convertido para *3098571458.4279037106,* já o campo *Nome do Paciente*, que anteriormente era *JOÃO SOUZA*, agora se apresenta como *PAT3098571458,* impedindo qualquer correlação entre o nome original e o transformado. Quando a mesma imagem é enviada para a estação de laudo para então ser visualizada, o registro é convertido novamente para o seu formato original através do processo de desanonimizacao que também ocorre no servidor, e portanto se torna possível a visualização das informações do paciente e da imagem recebida, como pode ser observado na Figura 07.

Modality	Status	Patient name	Patient ID	Date of birth	Sex	Study date	Study time
CT		SOUZA, JOAO	123456	27/04/1996	M	26/06/2015	152555

Figura 07: Informação visualizada na estação de laudo

Questão: Como o Conquest consegue reverter as informações anonimizados para que seja possível o envio para a estação de laudo?

O Conquest armazena em uma tabela do banco de dados chamada *UIDMODS* todas os registros de imagens que são recebidas pelo servidor DICOM e que passaram pelo processo de anonimização via *script*. No caso do paciente fictício JOÃO SOUZA, o processo de anonimização do Conquest gerou através de uma função *CRC32* um conjunto de 8 caracteres hexadecimaisque passam a identificar o Nome do paciente (*NewUID*), como pode ser observado na Figura 08. Ou seja, o nome do paciente é automaticamente transformado na junção da *string* "PAT" com os novos caracteres; no exemplo o paciente JOÃO SOUZA teve seu nome alterado para *PAT30988571458.*

	MODTime	OldUID	MODType	NewUID
Editar Inline Edit Copiar Remover	1435343340	1.3.6.1.4.1.29565.1.4.2760659664.9584.1435343155.9...	StudyInstanceUID	99999.99999.7.1435343340.1
Editar Inline Edit Copiar Remover	1435343340	1.3.6.1.4.1.29565.1.4.2760659664.9584.1435343155.9...	SOPInstanceUID	99999.99999.7.1435343340.0
Editar Inline Edit Copiar Remover	1435343340	1.3.6.1.4.1.29565.1.4.2760659664.9584.1435343155.9...	SeriesInstanceUID	99999.99999.7.1435343340.2
Editar Inline Edit Copiar Remover	1435343340	123456	lua	3098571458.4279037106
Editar Inline Edit Copiar Remover	1435343340	3098571458.4279037106.bd.19960427	lua	3098571458.4279037106.bd.
Editar Inline Edit Copiar Remover	1435343340	3098571458.4279037106.ps.M	lua	3098571458.4279037106.ps.
Editar Inline Edit Copiar Remover	1435343340	SOUZA^JOAO	lua	PAT3098571458

Figura 08: Informações tabela UIDMOD Conquest

O propósito da função CRC32 neste caso, além da sua utilização para codificar o nome do paciente, é utilizada também para checagem da integridade do arquivo feita através da geração de *checksum*, ou seja, a função CRC32 produz um código correspondente ao dado original, de modo que quando o dado original é transmitido junto a rede este código passa a ser transmitido com ele. O destinatário recebe o dado e, primeiramente, computa o *checksum*, sendo possível averiguar se houve alguma alteração durante a transmissão. Se os *checksums* diferente um do outro, provavelmente existiu alguma manipulação do dado durante o percurso e a imagem que foi enviada certamente não é a mesma que está sendo entregue ao destinarário.

Já o processo de desanonimização do Conquest acontece de forma mais simples, pois conhecendo a informação anonimizada, é possível obter através dos registros da tabela UIDMOD as informações do paciente e, enfim, disponibilizá-las.

Considerando o mecanismo que tornam anônimos os dados de pacientes, pode-se dizer que apenas relacionando duas de suas tabelas (*DICOMPatients* e *UIDMODS)* as informações originais podem ser acessadas. Dada essa natureza do processo que anonimiza as informações no Conquest, pode-se considerar que, trata-se de um processo que pode comprometer a segurança das informações dos pacientes quando se considera a possibilidade de invasão do banco de dados por um terceiro ou até mesmo por alguém dentro da organização de saúde. Em outras palavras, basta o "atacante" obter acesso às tabelas de banco de dados e conhecer a lógica adotada pelo Conquest para conseguir as informações reais.

Visando uma maior segurança das informações contidas em banco de dados dos PACS, um antigo conceito de criptográfica assimétrica pode ser aplicado. A criptografia assimétrica, também chamada de “criptografia de chave pública” consiste basicamente em utilizar duas chaves distintas, uma é publicada (chave pública) e a outra mantida em segredo (chave privada). A informação codificada com a chave privada somente é decifrada com a chave pública e vice-versa.

Este modelo possibilita ao PACS um nível maior de segurança, assegurando que a informação real seja acessível apenas com a utilização de duas

chaves (pública e privada). Na prática, as informações do banco de dados Conquest poderiam ter mais de um responsável pelas chaves, por exemplo. Em hospitais, o setor que administra o servidor DICOM e/ou também o setor responsável pelo banco de dados podem ser detentores de chaves distintas, que combinadas podem permitir o acesso à informação original.

O compartilhamento de chaves em ambientes PACS baseados em nuvem pode ser um recurso bastante favorável, uma vez que as informações estarão codificadas no ambiente virtual e a chave de decriptografia pode ser enviada em cada requisição. Principalmente em nuvens públicas, aonde se faz uso compartilhado de recursos com outras organizações e empresas, esta prática pode proporcionar maior segurança, impedindo também que nem mesmo o pessoal de gerência das máquinas virtuais da nuvem possam ter acesso à informação real. Até mesmo em nuvens privadas, esse mecanismo de segurança também pode ser implementado, uma vez que a nuvem pode ser utilizada por todas as áreas do hospital e não somente à área de radiologia.

4 Justificativa e Objetivos

A adoção de soluções baseadas em nuvem para os serviços de saúde tem apresentado grande expansão. Publicação recente estima que o mercado global de computação em nuvem voltada para sistemas de saúde tenha movimentado US$ 1,82 bilhão em 2011 e que, para o ano 2018, esse valor poderá atingir os US$ 6,8 bilhões **[19].** A computação em nuvem além de facilitar a administração de grandes montantes de dados, possibilita a incorporação de modelos mais simples e baratos de distribuição de software, reduz custos de manutenção, amplia as possibilidades de trabalho colaborativo, permite a execução de upgrades com reduzido tempo de inatividade, entre outras vantagens **[20]**.

A necessidade de um teor confidencial em informações hospitalares também se apresenta como fator relevante para os propósitos desse trabalho. Sabe-se que, em hospitais, os dados adquiridos diariamente estão diretamente relacionados à saúde e bem-estar de pacientes. Este tipo de informação apresenta um teor confidencial que torna necessário gerir e proteger essa informação, de modo a garantir que a atenção dada ao paciente não seja comprometida devido a vulnerabilidade de tecnologia da informação.

Ligado à questão da confidencialidade está o fator da imprevisibilidade, que é, sem dúvida, uma “vantagem” para os invasores, pois nunca se sabe quais são os motivos que os levarão à praticar a ação. Mas, seja qual for o motivo, acaba sempre interferindo e causando dano. Existem muitos invasores com tempo e habilidades suficientes para projetarem os diferentes ataques às informações.

Cientes disso, várias entidades reconhecidas mundialmente já se posicionaram sobre a importância de se proteger a informação médica: como a ENISA (*European Union Agency for Network and Information Security*, Agência Europeia para a Segurança das Redes e da Informação) e o Departamento de Saúde e Serviços Humanos (HHS, *Department of Health & Human Services*) e a HIPAA, *Health Insurance Portability and Accountability Act*).

Diante do exposto, o objetivo geral da pesquisa aqui descrita foi avaliar as principais questões de segurança e performance relacionadas ao envio de imagens médicas entre ambientes PACS físicos e virtuais, e dessa forma, verificar com base em solução técnica a introdução do *cloud computing* nesses ambientes. Para tanto, alguns objetivos específicos foram também definidos:

- Estruturar um ambiente PACS em nuvem;
- Avaliar o desempenho do ambiente utilizando conexão VPN;
- Avaliar o desempenho do ambiente utilizando túnel SSH;
- Avaliar o desempenho do ambiente quanto submetido à criptografia dos dados de pacientes, através de algoritmo de criptografia AES;
- Propor uma análise comparativa dos métodos de segurança abordados.

5 Material e Métodos

5.1 Considerações Éticas

Este projeto de pesquisa não necessitou de aprovação do Comitê de Ética em Pesquisa do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto, de acordo com o despacho da “Carta de Encaminhamento de Documentos para Apreciação pelo CEP do HCFMRPUSP” datada de 11 de maio de 2015 (em anexo).

5.2 O ambiente PACS

A primeira etapa do projeto consistiu na análise do ambiente *PACS* do Centro de Ciências das Imagens e Física Médica do HCFMRPUSP. A análise considerou as configurações e especificidades dos equipamentos de informática que compõem o ambiente, o conjunto de servidores de *hardware* em uso, o ambiente de rede *de dados* disponível, a capacidade, a modalidade dos dispositivos de armazenamento e o fluxo de trabalho adotado. Esta etapa compreendeu também o mapeamento das aplicações instalados nos servidores, assim como nas estações de trabalho destinadas à visualização das imagens médicas. Além disso, através de reuniões com o grupo responsável pela gestão do serviço e seus utilizadores, foram mapeadas as dificuldades de operações encontradas e as insatisfações com o ambiente atual.

Essa etapa de levantamento de dados foi de extrema importância para que fosse possível reproduzir em ambiente de nuvem o ambiente do Centro de modo a ter um serviço com características iguais ou, ao menos, parecidas às do ambiente físico, possibilitando sua compatibilidade e integração.

Na etapa seguinte, um servidor foi definido, basicamente analisando-se o volume de informação armazenada e a quantidade de acessos ao servidor, para se criar o servidor de imagem na nuvem privada da Universidade de São Paulo – C*loud-USP*.

A próxima etapa consistiu na criação de cenários para que fosse possível a investigação proposta pelo presente trabalho, são eles: 1) cenário para utilização de conexão segura VPN, 2) cenário para utilização tecnologia de tunelamento SSH e 3) cenário para utilização de anonimização e criptografia de dados.

5.3 Cenário para conexão segura VPN

Esse cenário do estudo é composto por um servidor DICOMimplementado no *Cloud*-USP e também por uma estação de trabalho instalada em rede local, interligados através de uma conexão dedicada de 600 megabits/segundo.

O servidor DICOM, responsável por realizar o gerenciamento e armazenamento das imagens médicas, foi instalado em um sistema operacional *Linux Ubuntu® 12.04.4.* O software para gerenciamento DICOM implementado foi o Conquest *1.4.17*, e como *Data Base Management System - DBMS,* utilizou-se o *Mysql® Server,* responsável pela indexação e catalogação de arquivos de imagens recebidos pelo servidor.

A escolha pelos softwares que compõem o estudo foi decidida após a observação de suas matrizes de compatibilidades e de suas formas de licenciamento. Considerou-se que, tanto o Conquest, o *Linux Ubuntu®* e o *Mysql®*, eram apropriados para a construção do ambiente por serem compatíveis entre si e com o *hardware* do servidor virtual instalado na nuvem e também por serem *de de licença de uso livre*.

Basicamente, o hardware do servidor DICOM virtual foi estruturado com 4 processadores modelo *Intel® Xeon(R) CPU E7- 2870 @ 2.40GHz*, 8 g*igabytes* de memória RAM e disco de armazenamento de 2 *Terabytes*. A taxa máxima de transferência da interface de rede do servidor era de 1 g*igabit* por segundo.

A instalação do *Conquest* foi realizada conforme o manual obtido junto ao arquivo de instalação e a sua configuração foi realizada através de dois arquivos (*dicom.ini e acrnema.map)* ambos localizados no diretório de instalação escolhido. Basicamente, foram alterados o nome do servidor (variável *MyACRNema*) e as configurações de acesso ao banco de dados (*SQLHost, SQLServer, Username* e *Password*). No arquivo *acrnema.map*, responsável pela "ligação" do servidor com as estações de laudo determinando as permissões e formas de acesso, foi realizada a configuração de conexão entre o servidor e a estação de laudo localizado na rede local.

A estação de trabalho possui processador *Intel® Core I5-2450M 2.50Ghz* e 6 g*igabytes* de memória, rodando sobre Sistema Operacional *MicrosoftWindows Seven®*.

Para o estudo de desempenho utilizou-se o software *Dcmflow* (), que permite simular a geração de imagens em padrão DICOM. Basicamente, o *Dcmflow* permite a geração randômica e personalizável de grandes lotes de imagens para testes de conectividade e desempenho de soluções PACS.

Na estação de lado, configurou-se um discador de acesso à rede VPN disponível no *Cloud-USP*, utilizando-se de credenciais de acesso (*login* e senha) além de uma chave pré-compartilhada para autenticação. Dessa forma, a estrutura foi montada e os envios de imagens partindo da rede local até a nuvem foram iniciados.

As imagens foram enviadas para o servidor virtual em um primeiro momento conectado ao túnel e, em um segundo momento, desconectado da VPN. Dessa forma foi possível comparar as taxas de transferências obtidas em cada envio. Considerou-se a importância de se realizar os envios em horários variados e aleatórios, para que o resultado fosse independente de possíveis instabilidades na infraestrutura de rede.

Os exames da simulação foram gerados segundo um modelo de modalidade de CT (*ComputedTomography)*, sendo utilizada a mesma configuração em todos os envios: um total de 1050 imagens por envio, com tamanho somado de 1126,4 Megabytes (ou 1.1 Gigabytes). Cada imagem possuía resolução de *512 x 512* pixels, do tipo *MONO* e 2 bytes (16 bits) para os níveis de cinza de cada pixel.

O valor da velocidade média de cada envio foi obtido através da equação:

$$(\text{speed}) \quad A = \frac{x1+x2+xn}{y} \qquad (1)$$

A média de cada envio *(A)* foi conseguida somando-se os tamanhos das imagens *(x)* (obtido através do comando *du* em linha de comando do servidor Linux) e dividindo-os pelo tempo total de envio *(y)* registrado no *log* do Conquest.

5.4 Cenário para tunelamento SSH

O cenário do estudo foi composto por dois servidores hospedados em locais diferentes, um servidor hospedado na rede local da Faculdade de Medicina de Ribeirão Preto – FMRP-USP de onde partiram os envios de imagens e por um

servidor criado no ambiente de nuvem da USP. A função principal dos servidores foi a criação de um túnel SSH entre os dois ambientes, viabilizando a transmissão segura das imagens. Além disso, criou-se um outro servidor no *Cloud-USP* com a finalidade de receber as imagens médicas no padrão DICOM enviadas a partir da rede local, conforme a Figura 09.

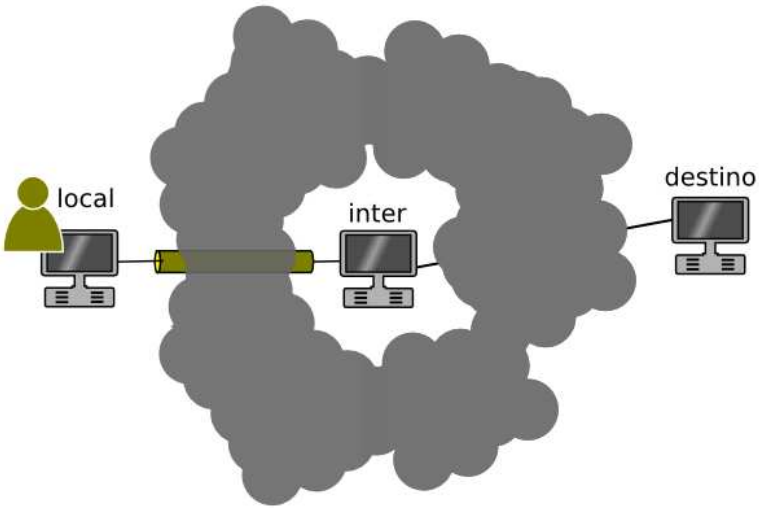

Figura 09: Tunel SSH entre duas redes

O tubo em amarelo da Figura 9, que atravessa parte da nuvem, representa justamente a conexão SSH criptografada entre o servidor da rede local (indicado pelo texto *local*) e o servidor SSH (indicado por *inter*, de intermediário). Por sua vez, o servidor SSH se conecta ao servidor de destino (*destino*) sem qualquer criptografia, e age como intermediário nas comunicações entre a rede local e o servidor PACS da nuvem. Dessa forma, todas as imagens enviadas para o servidor localizado na rede local, serão automaticamente encaminhadas de forma segura para o servidor PACS dentro da nuvem.

Com relação ao servidor DICOM, foi configurado com sistema operacional *Linux Ubuntu® 12.04.4*, serviço Conquest 1.4.17 e o Mysql® Server 5.5 e suas características de hardware eram compostas por 4 processadores modelo Intel® Xeon(R) CPU E7- 2870 @ 2.40GHz, 8 gigabytes de memória RAM e disco com capacidade de armazenamento de 2 Terabytes. O servidor SSH disponível no Cloud-USP possui as mesmas especificações técnicas.

Os envios foram realizados à partir das estações de laudo da Faculdade de Medicina e do HCFMRPUSP.

O túnel SSH entre o servidor local e o servidor SSH foi realizado utilizando o software *Putty*. As configurações realizadas no software são bastante simples, podendo citar:

1. Endereço IP do servidor SSH hospedado na nuvem e número da porta de comunicação do SSH (utilizou-se a porta padrão 22);
2. Endereço IP do servidor DICOM e número da porta de comunicação associada ao software gerenciador de imagens médicas Conquest DICOM (utilizou-se a porta padrão número 5678) como destino das informações;
3. Número da porta de origem do servidor local (utilizou-se a porta 5678) utilizada para receber as conexão provenientes da rede local, para envio das imagens até o servidor DICOM.

Com relação à infraestrutura de rede local, a rede do HCFMRPUSP é bastante específica, diferindo praticamente da maioria dos modelos atuais do campus da Universidade, adotando em sua rede determinadas políticas para controle de acesso à Internet (bloqueios). Tais políticas de bloqueios são implementadas utilizando servidores que atuam como uma espécie de árbitro, checando se determinada informação pode trafegar. Trata-se de um tipo de configuração conhecido como *ACL – Access Control Lists* (Listas de Controle de Acesso) que é constituída por conjuntos de regras para filtrar pacotes, permitindo ou negando que eles sigam em frente. Além da política restritiva do hospital, existe também uma política de acesso totalmente liberal para determinados computadores, ou seja, os acessos provenientes desses equipamentos não são arbitrados e a eles são atribuídos endereços IP válidos na internet (também conhecidos como endereços públicos), o que quer dizer que o equipamento pode ser acessado por qualquer outro computador em redes externas.

O HCFMRPUSP incorpora essas políticas de segurança dentro de uma estrutura de rede virtual conhecidas como VLANs (do inglês, virtual LANs). Tecnicamente, essa segmentação de rede lógica divide uma rede local (física) em mais de uma rede (virtual), criando domínios de comunicação separados, o que quer dizer que, mesmo estando fisicamente instalados na mesma sala, dois computadores podem estar em redes virtuais incomuns e possuir políticas de acesso totalmente diferentes.

A observação dessas políticas de rede estabelecidas pela equipe de tecnologia do hospital se fez necessária para este trabalho uma vez que o cenário de rede poderia trazer alguma limitação nos envios de imagens para a nuvem USP.

Diante do exposto, para que pudesse ter certeza de que os resultados dos testes não sofreriam nenhuma influência advinda das características técnicas dos locais, realizou-se envios de imagens paralelos, tanto da FMRP quanto do HCFMRPUSP e, como resultado, observou-se que as taxas de transferência média de ambos os locais até a nuvem foram similares, concluindo que a infraestrutura da rede de dados, neste caso, não impactaria diretamente nos resultados.

Finalmente, foram realizadas transferências de imagens para o servidor DICOM estando conectado e, posteriormente, desconectado do túnel SSH, para que fosse possível comparar as taxas de transferências obtidas após a incorporação da segurança.

5.5 Cenário para Criptografia e Anonimização dos dados demográficos.

Como já observado, a anonimização e desanonimização do cabeçalho DICOM das imagens no Conquest é realizada através de chamadas aos *scriptsanonymize_script.lua* e *deanonymize_script.lua* no arquivo *dicom.ini*, responsável pela configuração do Conquest.

Em sua instalação padrão, o Conquest não ativa os processos de anonimização, ou seja, esta passa a ser uma configuração opcional que pode ser ativada após a instalação do Conquest, incluindo-se algumas linhas no seu arquivo de configuração *dicom.ini*:

- ***ImportExportDragAndDrop = 1***
- ***ImportConverter0 = lua/anonymize_script.lua***
- ***QueryResultConverter0 = lua/deanonymize_script.lua***
- ***RetrieveResultConverter0 = lua/deanonymize_script.lua***

A primeira opção habilita os *Converters* que são responsáveis por atribuir ações sempre que alguma imagem é recebida pelo servidor (*IMPORT*) ou em casos

de consultas à base de dados (*QUERY*) ou em solicitações de envio de imagem (*RETRIEVE*). Todas as opções, como podem ser vistas, apontam para um determinado *script*. Tecnicamente, assim que a imagem é recebida pelo servidor é executado o *script* de anonimização localizado no diretório *lua*. De forma semelhante, sempre que houver uma solicitação de consulta (por exemplo, um *C-FIND*), ou de envio (*C-MOVE*), o Conquest disparará o *script* que desanonimiza as informações, enviando ao solicitante as informações em seu formato original.

O Conquest permite as execuções de *script* em linguagem de programação LUA mas por padrão não instala um software capaz de traduzir (ou compilar) o algoritmo em uma linguagem que possa ser interpretada pelo computador. Como o trabalho foi realizado utilizando servidor construído sobre a plataforma *Linux*, a instalação do compilador foi possível através do comando "*apt-get install lua"*. *Este* comando está disponível apenas em distribuições *Linux* baseados em *Debian*, como por exemplo o *Ubuntu*. Em versões baseadas em plataforma *RedHat*, o comando será o*"yum install lua"*. Já em ambientes *Microsoft Windows*, não é necessária instalação de compilador.

Posteriormente, realizou-se a programação dos *scripts* LUA:

Basicamente, manteve-se toda a estrutura original dos scripts que já anonimizavam os dados do paciente, porém, mantendo as informações do paciente em segunda tabela (*UIDMODS*), possibilitando, em posse da base de dados, a sua identificação.

A anonimização realizada pelo script LUA original do Conquest é realizada por um algoritmo de criptografia chamado *CRC32* que permanece sendo utilizado para criptografar os dados dos pacientes, com exceção do nome do paciente, que será criptografado utilizando o algoritmo AES. A grande vantagem em se manter a anonimização pela função CRC32 é que ela proporciona um menor custo computacional, ou seja, requer menor processamento de CPU para realização dos cálculos matemáticos, quando comparados ao algoritmo de criptografia AES que será utilizado para criptografia do nome do paciente.

Não é objetivo do trabalho a explicação de toda a estrutura dos *scripts*, contudo, pode se dizer que as principais alterações ocorrem nas linhas 7, 8 e

de 30 a 36 (para o arquivo de anonimização) e nas linhas 3, 4 e de 15 a 20 (para o arquivo de desanonimização), conforme Figura 10 e Figura 11. As informações do procedimento para incorporação da criptografia no nome do paciente são bastante técnicas e encontram-se detalhadas no *APÊNDICE A – Configuração dos scripts*, bem como o seu esquema que encontra-se no *APÊNDICE B – Esquema de anonimização Conquest.*

5.6 Processo de captura de dados nos cenários

A captura pode pacote, também conhecida como *sniffing*, corresponde ao processo de interceptar e registrar o tráfego de dados em redes de computadores. Conforme o fluxo de dados trafega na rede, o *sniffer*, ferramenta utilizada para captura de pacote, captura os dados e eventualmente decodifica-os, possibilitando a análise do conteúdo. Essa prática foi concebida com o objetivo de oferecer uma maneira de diagnosticar problemas no tráfego de dados de entre rede de computadores, mas também pode ser utilizada com propósitos maliciosos por invasores para roubo de informações. Na Figura 10 é possível observar como as informações quando eventualmente interceptadas são exibidas.

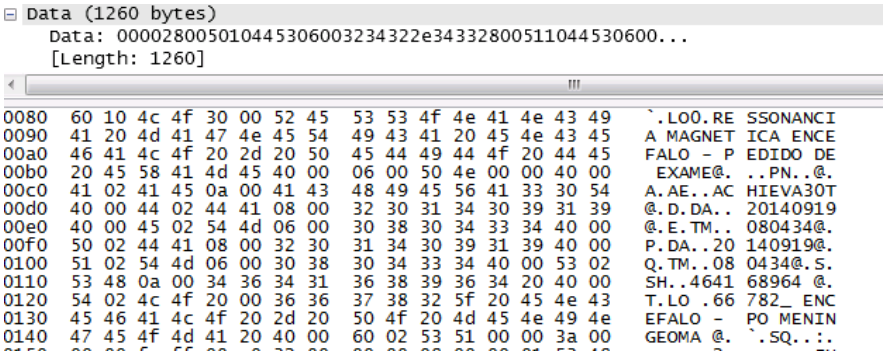

Figura 10: Pacote interceptado

Como pode ser observado, os dados que foram interceptados pelo programa são exibidos em texto puro, possibilitando a sua leitura. Por outro lado, quando submetidos a processos de criptografia de dados, essas informações são codificadas fazendo com que, mesmo interceptadas, não seja possível a leitura, como pode ser visto através da Figura 11.

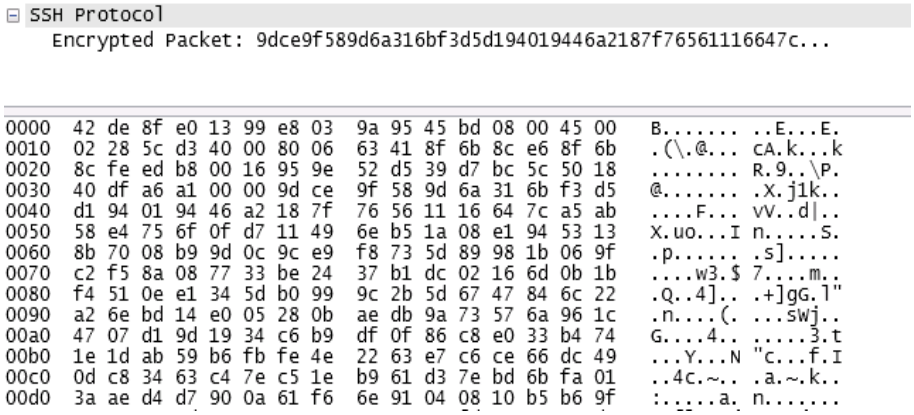

Figura 11: Tráfego criptografado interceptado

Ambas as imagens (Figura 10 e Figura 11) possuem invariavelmente as mesmas informações, porém, a segunda imagem é composta por informações criptografadas em sua origem através de protocolos de criptografia, neste caso o SSH. Caso esse pacote de dados fosse interceptado por um invasor, a sua visualização seria praticamente impossível.

Em todos os cenários propostos por este trabalho, várias destas interceptações foram realizadas com o objetivo de certificar sobre a consistência e eficiência dos métodos de segurança aplicados.

6 Resultados e Discussão

6.1 Resultados obtidos com a Conexão SeguraVPN

Na primeira transferência de imagens, sem nenhum mecanismo de segurança aplicado, foi realizado o envio de 1126,4 megabytes de imagens em um tempo total de 420 segundos, obtendo-se uma taxa de transferência média de 2.66 megabytes/segundo.

A mesma transferência foi reaplicada usando uma conexão segura do tipo VPN e obteve-se 671 segundos de duração para a transmissão das imagens e uma taxa de transferência de 1.68 megabytes/segundo.

Realizou-se outras três transferências de imagens dentro e fora da VPN. Os resultados dos testes realizados, incluindo o teste inicial, são apresentados nas tabelas 1 e 2.

Tabela 1: Transferências fora da VPN

Envio	Tamanho(MB)	Tempo(s)	Velocidade (MB/s)
1	1126,4	420	2,68
2	1126,4	422	2,67
3	1126,4	424	2,66
4	1126,4	426	2,64
		Média:	**2,66**
		Desvio Padrão: 0,017	

Tabela 2: Transferências dentro da VPN

Envio	Tamanho(MB)	Tempo(s)	Velocidade (MB/s)
1	1126,4	671	1,68
2	1126,4	667	1,69
3	1126,4	672	1,68
4	1126,4	672	1,68

	Média:	**1,68**
	Desvio Padrão:	**0,005**

Como se observa nas tabelas, a média das transferências realizadas na VPN foi de 2,66 megabytes/segundo com desvio padrão de 0,017 e, fora da VPN, a média foi de 1,68 megabytes/segundo eu desvio padrão 0,005. A relação de perda entre os dois ambientes é dada pela equação:

$$(\text{loss}) \quad B = 1 - \left(\frac{x}{y}\right)$$

Sendo *B* a queda do desempenho, *x* a média utilizando VPN e *y* a média não utilizando VPN, obtém-se uma queda no desempenho de 36,8%.

A taxa de utilização de placa de rede na estação de laudo variou entre 20% e 26%, o que mostra que não houve limitações do hardware de rede da estação de laudo que pudesse interferir nas transferências.

6.2 Resultados obtidos com tunelamento SSH

O tempo médio observado para envio das imagens dentro do túnel foi de 08 minutos e 27 segundos, obtendo uma taxa de transferência média de 1,82 Megabytes/segundo, como pode ser observado na Tabela 3.

De igual modo, foram reenviadas as imagens diretamente ao Cloud-USP, sem utilização de tunel SSH uma taxa de transferência de 2,46 megabytes/segundo observada na tabela 4.

Tabela 3: Envio conectado ao Túnel SSH

Envio	Tamanho (MB)	Tempo (s)	Velocidade (MB/s)
1	1126,4	720	1,56
2	851	409	2,08
3	532	289	1,84
		Média:	**1,82**

Desvio Padrão: 0,26

Tabela 04: Envio desconectado do Túnel SSH

Envio	Tamanho (MB)	Tempo (s)	Velocidade (MB/s)
1	851	341	2,49
2	254	90	2,82
3	408	239	1,70
		Média:	**2,33**
		Desvio Padrão: 0,57	

Como se observa nas tabelas, a média das transferências realizadas através do túnel SSH foi de 1,68 megabytes/segundo com um desvio padrão de 0,26 e, diretamente para a nuvem sem utilizar o túnel, a média foi de 2,33 megabytes/segundo com desvio padrão de 0,57. A relação de perda entre os dois ambientes é dada pela mesma equação previamente utilizada para teste com a conexão VPN:

$$(\text{loss}) \quad \text{B} = 1 - (\frac{x}{y})$$

Sendo B a queda do desempenho, x a média conectado ao túnel e y a média não utilizando o túnel, obtém-se uma queda no desempenho de 21,8%.

6.3 Resultados obtidos com a Criptografia e Anonimização de dados de pacientes, através de algoritmo de criptografia AES.

A implementação de criptografia em informações inseridas na base de dados de pacientes é bastante relevante sob o ponto de vista de segurança da informações e de sigilo. Entretanto, já se sabe que implementações de segurança podem apresentar interferências no desempenho e até mesmo inviabilizar a sua utilização.

Dessa forma, propôs-se a realização de alguns testes de envio de imagens originadas à partir de estações de trabalho da FMRP e do HCFMRPUSP com destino ao ambiente *cloud computing* da Universidade, submetendo-os a criptografia em banco de dados *Mysql* e comparando o seu desempenho ao encontrado quando não se utiliza a criptografia.

Tabela 5: Relação de perda entre envio criptografado e não criptografado.

	Com Criptografia	Sem Criptografia	Perda Performance
envio1	2,67	3,25	17,85%
envio2	2,67	3,29	18,84%
envio3	2,76	3,34	17,37%
		MÉDIA	**18,02%**

Com relação aos dados apresentados na Tabela 5, conclui-se que o processo de envio de imagens médicas para a nuvem, submetendo-o à criptografia do cabeçalho DICOM utilizando o protocolo do criptografia AES, pode oferecer queda de desempenho em até 18%.

7 Discussão

Em computação, quando se observa um aumento da segurança no tráfego de informações, sejam elas médicas ou não, é praticamente impossível não considerar o seu desempenho. Sabe-se atualmente que segurança e desempenho são itens que precisam andar juntos, pois uma rede com um alto desempenho e não segura é tão ineficiente quanto uma muito segura com desempenho abaixo do esperado.

Dentro de um ambiente PACS, essas duas questões se tornam ainda mais importantes. Em geral, exames de imagens médicas tendem a apresentar grandes volumes de dados (seja devido à sua matriz de aquisição, na mamografia digital, por exemplo; seja devido a um grande número de cortes, em exames de tomografia de tórax de alta resolução, por exemplo), exigindo infraestrutura de comunicação de alto desempenho mas, também, um transporte seguro dessas informações.

A resolução número 1997 de 2012 do Conselho Federal de Medicina – CFM, considera que “as informações constantes do prontuário médico de pacientes possuem amparo constitucional, pois se ligam à ideia de preservação da intimidade”. Dessa forma, considera-se também que as imagens médicas, mesmo que em vias digitais, são partes do prontuário do paciente, pois fornecem informações para prestação do serviço médico e, desta forma, devem estar seguras quanto a possíveis ameaças à sua informação. Nesse contexto, a norma ABNT NBR ISO/IEC 27002, que trata sobre segurança da informação, relaciona o crescente número de ameaças e vulnerabilidades à informação com o incrível aumento da interconectividade nos últimos anos. Sendo assim, a adoção de métodos computacionais seguros, que garantem principalmente o sigilo do paciente, durante as comunicações em ambientes PACS se torna fundamental.

Com base nos resultados obtidos através dos testes, tem-se, quanto às quedas de desempenho:

- Utilizando conexão segura VPN: **36,8%**
- Utilizando tunelamento SSH: **21,8%**

- Utilizando criptografia de dados por protocolo AES: **18,02%**

Em relação à queda de performance apresentada em virtude da utilização de segurança nos procedimentos de envios de imagens, a utilização de criptografia no cabeçalho DICOM mostrou-se satisfatória. A taxa de perda de performance se dá, basicamente, pela aplicação dos processos de criptografia AES na base de dados do Conquest e no cabeçalho DICOM da imagem, uma vez que esse mecanismo não realiza a criptografia da estrutura da imagem em si mas apenas de seu cabeçalho. Com isso, quando enviada de um local para outro, a imagem preserva visualmente as mesmas características, entretanto não é possível realizar a correlação entre a imagem e o paciente, uma vez que os dados pessoais estão totalmente indecifráveis. Entretanto ela não inviabiliza o acesso do atacante até o dado em si, que ficará possível de ser obtido no caso de uma rede insegura, mas que mesmo que obtido, não revela a identidade do paciente.

Considerando os resultados da utilização do tunelamento SSH, observou-se ser um método de simples implementação, o que vale destacar, pois não é necessário grandes conhecimentos técnicos para configuração de um túnel SSH, entretanto, verificou-se um importante oscilação dos resultados, considerados os valores de desvio padrão encontrados, o que pode apontar para um mecanismo inconsistente e inapropriado para transferência de grandes volumes de dados.

Por fim, a VPN, com o pior desempenho, demonstrou ser um mecanismo mais sofisticado, agregando bastante segurança à transferência, mas com limitações de velocidade que devem ser consideradas.

A relação entre performance e segurança é bastante subjetiva, portanto, determinar em quais situações se deve optar por um melhor desempenho e em quais situações se deve optar por uma maior segurança, exige bastante conhecimento do projeto. É necessário que os gestores, em parceria com a equipe de Tecnologia da Informação, avaliem os impactos diretos dessas duas variáveis dentro de um contexto mais amplo. Em outras palavras, a agregação da segurança proposta, mesmo com queda de performance para alguns projetos pode ser intolerável, enquanto que para outros projetos, principalmente para os que utilizam nuvem como um segundo ambiente, de replicação de dados de produção por exemplo, pode ser aceitável.

Ainda assim, quando mínimas perdas de desempenho se tornam insustentáveis, é possível também atribuir várias outras abordagem aos PACS em nuvens, utilizando-os como ambientes apenas para replicação ou de contingência, possibilitando que haja uma réplica dos dados na nuvem para conter eventuais problemas na infraestrutura de produção ou desastres. Ou até mesmo, fazendo dele um repositório virtual de imagens para outras instituições de saúde (ou de ensino, que utilizam-se das informações para pesquisa, como é o caso deste trabalho), de outros municípios por exemplo, enviando ao ambiente da nuvem apenas as informações que são de interesse específico.

Conclusão

A pesquisa aqui apresentada teve como objetivo geral o estudo do impacto da implementação de mecanismos de segurança de dados sobre a performance de ambientes PACS implementados em nuvem. Os resultados obtidos apontam para uma perda de performance quando da adoção dos mecanismos de segurança estudados, evidenciando a importância de estudos detalhados antes da adoção de solução em nuvem em ambientes clínicos reais. Conclui-se que o objetivo proposto foi alcançado e que o trabalho desenvolvido pode contribuir na condução de projetos de implantação de ambientes PACS em nuvem.

Como possíveis trabalhos futuros sugere-se a aplicação de novas avaliações de impacto de desempenho aos cenários de PACS em computação em nuvem, como outros protocolos de criptografia, utilização de Secure File Trasport Protocol (SFPT), análise de desempenho mediante inserção de *firewalls*, diagnóstico do desempenho de PACS baseados em banco de dados, como o *Oracle DICOM Database*, por exemplo, e também testes de desempenho em ambientes PACS com Qualidade de Serviço (QOS).

Resultados parciais relacionados à pesquisa desenvolvida foram publicados em anais de congressos, a saber:

- SANT'ANA, F. S.; SUZUKI, K. M. F. ; CORDEIRO, S. S. ; MARQUES, P. M. A. . Performance impact of the transfer of DICOM images from PACS physical and virtual environments with implementation of VPN security.. In: INFOLAC, 2014, Montevideo. INFOLAC 2014. Conferencia Latinoamericana de Informática Médica, 2014.

- SANT'ANA, F. S.; SUZUKI, K. M. F. ; MARQUES, P. M. A. . PACS in the cloud: impact of tunneling and encryption security protocols implementation over image communication-archival performance.. In: A Telessaúde para a Universalização da Saúde, 2015, Rio de Janeiro.

- CORDEIRO, S. S. ; SANT'ANA, F. S. ; SUZUKI, K. M. F. ; MARQUES, P. M. A. . A Risk Analysis Model for PACS Environments in the Cloud. In: Workshop for Ongoing Projects on Computer-based Medical Systems, 2016. Workshop for Ongoing Projects on Computer-based Medical Systems.

Referências Bibliográficas

[1] YIU, EDLIC NGA-LIK; EDWOOD NGA-WOOD. YIU**. Network Management for Picture Archiving and Communication Systems**. Burnaby B.C.: Simon Fraser University, 2007. Print.

[2]Digital Imaging and Communications in Medicine (DICOM) standard 2008, National Electrical Manufacturers Association. Available at:

[3] CARRINO JOHN, **Digital Imaging Overview, Seminars in Roentgenology**, Volume 38, Issue 3, July 2003, Pages 200-215, ISSN 0037-198X, http://dx.doi.org/10.1016/S0037-198X(03)00062-2.
()

[4]Azevedo-Marques, P.M.; Salomão, S.C. PACS: **Sistemas de Arquivamento e Distribuição de Imagens PACS: Picture Archiving and Communication Systems**,Revista Brasileira de Física Médica. 2009;3(1):131-9.

[5] Brent J. Liu, M.Z. Zhou, J. Documet, **Utilizing data grid architecture for the backup and recovery of clinical image data**, Computerized Medical Imaging and Graphics, Volume 29, Issues 2–3, March–April 2005, Pages 95-102, ISSN 0895-6111,

[6] Veras, M**; Arquitetura de Nuvem: Amazon Web Services (AWS).** 1º Edição Rio de Janeiro, Brasport, 2013, p15.

[7] Cristen, B; **Roadmap to cloud-based PACS**. Applied Radiology. 41.6 (June 2012) p22.

[8] Ferraro-Souza,R et al. **Challenges of Operationalizing PACS on Cloud Over Wireless Networks**, The Ninth International Conference on Wireless and Mobile Communications - ICWMC 2013.

[9] consulta realizada em 05/11/2013 às 23:30hs.

[10] Philbin, J.; Prior, F.; Nagy, P.**Will the next Generation of PACS Be Setting on a Cloud?**, Journal of Digital Imaging, Vol 24, No2, April, 2011, p 180

[11] Wang, Q., Wang, C., jin, L., Ren, K., e Lou, W. (2009). Enabling and Verificability and Data Dinamics for Storage Security in Cloud Computing. Criptology ePrint Archive.

[12] Feilner, M. (2006). **OpenVPN.** Packt Publishing.

[13] RFC 4251: **The Secure Shell (SSH) Protocol Architecture**. Disponível em: http://tools.ietf.org/html/rfc4251 . Acesso em: 11 Julho 2016.

[14] STALLINGS, William. **Data & Computer Communications**, Sixth Edittion, Ed. Prentice Hall, USA, 1999a.

[15] Loshin, P. (2013). **Simple Steps to Data Encryption**. Elsevier Science; Syngress.

[16] STALLINGS, William. **Criptografia e Seguranca de Redes**. 4. ed. São Paulo: Pearson, 2008.

[17] Rhodes-Ousley, M. (2013). **Information Security The Complete Reference**, Second Edition. McGraw-Hill.

[18]Pianykh, O. S. (2008). ***Digital Imaging and Communications in Medicine (DICOM): A Practical Introduction and Survival Guide.*** Leipzig/Germany: Springer Science & Business Media.

[19] TMR. (2014). **Cloud Computing Market In Healthcare Industry (IAAS, SAAS, PAAS, CIS, NCIS, PACS, EMR, RIS) - Global Industry Analysis, Size, Share, Trends And Forecast 2012 - 2018**. (Transparency Market Research) Acesso em 27 de Julho de 2014, disponível em Transparency Market Research: http://www.transparencymarketresearch.com/healthcare-cloud-computing.html

[20] Bolan, C. (2013). **Cloud PACS and mobile apps reinvent radiology workflow**. Applied Radiology, 42(6):24-6.

Anexos

CARTA DE ENCAMINHAMENTO DE DOCUMENTOS PARA APRECIAÇÃO PELO COMITÊ DE ÉTICA EM PESQUISA DO HCFMRP-USP

Ribeirão Preto, 11 de maio de 2015

À Dra Marcia Guimarães Villanova

Coordenadora do Comitê de Ética do HCRP e FMRP-USP

Venho pela presente submeter ao CEP/Comitê de Ética em Pesquisa do HCRP e FMRP-USP, as informações referentes ao projeto de pesquisa do programa de Mestrado Profissional em Gestão de Organizações de Saúde da FMRP-USP, abaixo citado, **para avaliação da necessidade de apreciação ética do projeto** que utiliza-se de imagens médicas (radiológicas) digitais de pacientes do HCFMRP-USP.

Informações do Projeto de Pesquisa:

Aluno: Fábio Sousa de Sant'Ana

Orientador: Prof. Dr. Paulo Mazzoncini de Azevedo Marques

Título do Projeto: *"Análise e Projeto para a implantação de PACS (Picture Archiving and Communication System) no Centro de Ciências das Imagens e Física Médica do Hospital das Clínicas de Ribeirão Preto, através da tecnologia de computação em nuvem privada"*

Objetivo: Dedica-se a analisar e projetar a implementação de PACS em ambientes de computação em nuvem privado, a partir de um estudo científico e técnico de sua viabilidade.

Metodologia Resumida: Realizar a criação de um servidor de imagens DICOM no Cloud-USP (ambiente de nuvem computacional privado da Universidade de São Paulo), configurá-lo e interligá-lo ao ambiente PACS do Centro de Ciências das Imagens e Física Médica do HCFMRP-USP, propondo um cenário para pesquisa e avaliando através de testes onde se manipula o processo de envio e de recepção de imagens digitais de exames médicos entre ambientes PACS, com o objetivo de analisar o seu comportamento.

Não é proposta do projeto a análise, interpretação ou divulgação de quaisquer informações das imagens médicas, bem como dos pacientes a elas atreladas, cabendo ao projeto, apenas, utilizá-las como carga para realização de testes de caráter computacional (velocidade de transmissão, velocidade de recepção, integridade, segurança, etc) aos quais o projeto se destina.

Fábio Sousa de Sant'Ana
Pós graduando em Gestão em Organizações de Saúde

Drª. MARCIA GUIMARÃES VILLANOVA
Coordenadora do Comitê de Ética em Pesquisa
do HCFMRP-USP e da FMRP-USP

COMITÊ DE ÉTICA EM PESQUISA
10:25 h
Recebido
11/05/15
HCRP

Não necessita ser submetido à apreciação do CEP
11/05/2015

APÊNDICE A – Configuração de *scripts*

```
local scriptversion = "1.5; date 20140309"

local MaintainAge = false
local MaintainSex = false
local reversible  = true
local logroot     = "DicomAnonymized_Log"
local passwd = 'DICOM'
local file = '/tmp/crypt'

local function CRC32(val)
  return crc(tostring(val))
end;

-- remove characters that are not allowed in a filename
local pid = string.gsub(Data.PatientID, '[\\/:*?"<>|]', '_')

-- set nill if patient birthdate is empty
if Data.PatientBirthDate==nil or Data.PatientBirthDate=='' then
  Data.PatientBirthDate='00000000'
end

-- the changes in patient ID etc are hardcoded
local pre = CRC32(Data.PatientID)..'.'..CRC32(Data.PatientBirthDate);
--local pne = 'PAT'..CRC32(Data.PatientID) #not used after crypt

if version and command_line and command_line~='' then pre=command_line end
if version and command_line and command_line~='' then pne=command_line end

-- always crypt the name of patient based on a passwd and save it to filesystem.
if Data.PatientName~='' then
    os.execute("echo "..Data.PatientName .." | openssl enc -aes-256-cbc -a -salt -pass 'pass:"..passwd.."' > "..file)
    f = assert (io.open (file , "r"))
    s = f:read ()
    f:close ()
    Data.PatientName = changeuid(s, s) --set both fields to the encryted patient's name.
end

if Data.PatientID~='' then
  if reversible==true then
    Data.PatientID = changeuid(Data.PatientID, pre)
  else
    Data.PatientID = pre;
  end
end
```

Arquivo de anonimização Conquest

```
1  local scriptversion = "1.1; date 20130802"
2  local pre = Data.PatientID
3  local passwd = 'DICOM'
4  local file = '/tmp/decrypt.txt'
5
6  if Data.PatientID~=nill then
7    Data.PatientID = changeuidback(pre)
8  end
9
10 if true then
11   local s= changeuidback(pre..'.bd.'..Data.PatientBirthDate)
12   Data.PatientBirthDate = string.sub(s, string.find(s, '%.', -10)+1);
13 end
14
15 if Data.PatientName~='' then
16     os.execute("echo "..Data.PatientName.." | openssl enc -aes-256-cbc -d -a -pass 'pass:"..passwd.."' > "..file)
17     g = assert (io.open (file , "r"))
18     g = g:read ()
19     Data.PatientName = g;
20 end
21
22 if (Data.PatientSex=='') then
23   local s = changeuidback(pre .. '.ps.' .. Data.PatientSex)
24   Data.PatientSex = string.sub(s, string.find(s, '%.', -3)+1);
25 end
```

Arquivo de desanonimização Conquest

No arquivo que anonimiza as informações incluiu-se a função na linha 31:

os.execute("echo "..Data.PatientName .." | openssl enc -aes-256-cbc -a -salt -pass 'pass:"..passwd.."'>"..file)

que permite enviar ao sistema operacional *Linux* a informação do nome do paciente para seja enviado de volta o dado criptografado, que é calculado através de uma chave informada através da variável *passwd* declarada na linha 7 (*local passwd = 'DICOM'*) do arquivo. Por fim, a varíavel *file*, declarada na linha 8, informa em qual arquivo o resultado será armazenado temporariamente, neste caso em */tmp/crypt*.

Já as informações:

f = assert (io.open (file , "r"))

s = f:read ()

f:close ()

manipulam o arquivo com a string criptografada e armazena o resultado em uma variável *s*, fechando o arquivo posteriormente.

Em posse da string, que agora é representada através da variável *s*, basta acionar a função que realiza a gravação dos dados nas tabelas do Conquest:

Data.PatientName = changeuid(s, s)

A função *changeuid()* faz com que o Conquest insira o nome criptografado na tabela UIDMODS sem manter o nome real do paciente.

O arquivo de desanonimização também funciona de modo similar ao arquivo de anonimização, porém utiliza o parâmetro *–d* na função de decriptografia, informando que se trata de uma decriptografia.

APÊNDICE B – Esquema de anonimização Conquest.