UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA BIOMÉDICA

Trabalho de Conclusão de Curso

DICOM ANONYMOUS: SISTEMA WEB PARA ANONIMIZAÇÃO DE IMAGENS MÉDICAS

Paulo Victor de Azevedo Guerra

Natal/RN
2016

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

CENTRO DE TECNOLOGIA

DEPARTAMENTO DE ENGENHARIA BIOMÉDICA

DICOM ANONYMOUS: SISTEMA WEB PARA ANONIMIZAÇÃO DE IMAGENS MÉDICAS

Trabalho de conclusão de curso apresentado ao Departamento de Engenharia Biomédica da Universidade Federal do Rio Grande do Norte para obtenção do título de Graduado em Engenharia Biomédica.

Graduando: Paulo Victor de Azevedo Guerra

Orientador: Profª. Dra. Karilany Dantas Coutinho

Natal/RN

2016

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

CENTRO DE TECNOLOGIA

DEPARTAMENTO DE ENGENHARIA BIOMÉDICA

DICOM ANONYMOUS: SISTEMA WEB PARA ANONIMIZAÇÃO DE IMAGENS MÉDICAS

Banca Examinadora do Trabalho de Conclusão de Curso:

Profª. Dra. Karilany Dantas Coutinho _____

DEB/UFRN - Orientadora

Eng. Marcel da Câmara Ribeiro Dantas _____

LAIS/UFRN – Co-Orientador e Avaliador Externo

Prof. Dr. Ricardo A. M. Valentim _____

DEB/UFRN – Avaliador Interno

Natal/RN

2016

AGRADECIMENTOS

Aos meus pais Custódio e Marilu, por todo apoio e participação fundamental durante minha formação acadêmica e pessoal.

Ao meu irmão Gustavo, com quem sei que sempre poderei contar.

A minha namorada Ana Karenina, por apoiar e me dar forças em todas as decisões.

Aos meus orientadores Karilany Coutinho e Marcel Ribeiro Dantas por confiarem no meu trabalho.

A todos os meus colegas de estudo e trabalho, por compartilharem seus conhecimentos e experiências. Além de promover um ambiente de aprendizagem excepcional, onde parte significativa desse projeto foi desenvolvido. Sem vocês, este trabalho não seria possível.

Muito obrigado a todos.

SUMÁRIO

AGRADECIMENTOS 4

SUMÁRIO 5

LISTA DE FIGURAS 6

LISTA DE ABREVIAÇÕES, SIGLAS E SÍMBOLOS 7

RESUMO 8

ABSTRACT 9

1. INTRODUÇÃO 10

2. OBJETIVOS 11

3. FUNDAMENTAÇÃO TEÓRICA 12

4. METODOLOGIA 19

5. RESULTADOS E DISCUSSÕES 23

6. CONCLUSÕES 26

7. REFERÊNCIAS 27

LISTA DE FIGURAS

Figura 1 – Wilhelm Conrad Röntgen e radiografia da mão 12
Figura 2 – Procedimento de Fluoroscopia em 1909 13
Figura 3 – DICOM Dataset 14
Figura 4 – DICOM Element PatientName 15
Figura 5 – Fluxograma Criptografia 15
Figura 6 – Processo de Criptografia Simétrica 16
Figura 7 – Processo de Criptografia Assimétrica 17
Figura 8 – Trecho DICOM Dataset via PyDICOM 19
Figura 9 – Fluxograma da Metodologia Aplicada 19
Figura 10 – Interface do Sistema 23
Figura 11 – Visualização do Dataset durante procedimento 24
Figura 12 – Imagem Anonimizada 25

LISTA DE ABREVIAÇÕES, SIGLAS E SÍMBOLOS

CR – Computed Radiography

CT – Computed Tomography

DICOM – Digital Imaging and Communications in Medicine

DA – DICOM Attributes

DE – DICOM Elements

DO – DICOM Objects

DP – DICOM Protocol

DS – DICOM Services

DX – Digital Radiography

ECG – Electrocardiography

HUOL – Hospital Universitário Onofre Lopes

IV – Initialization Vector

MR – Magnetic Resonance

NEMA – National Electrical Manufacturers Association

PACS – Picture Archiving and Communications System

PN – Person Name

RSA – Rivest, Shamir, Adleman

US - Ultrasound

VF – Value Field

VL – Value Length

VR – Value Representation

XA – X-Ray Angiography

XC – External-camera Photography

GUERRA, Paulo Victor de Azevedo. **DICOM Anonymous: Sistema WEB para Anonimização de Imagens Médicas.** Trabalho de Conclusão de Curso, Graduação em Engenharia Biomédica, Universidade Federal do Rio Grande do Norte, 29p., 2016.

RESUMO

As imagens médicas do padrão DICOM são difundidas mundialmente e amplamente utilizadas em todas as modalidades de exames médicos. Uma característica fundamental dessas imagens é sua capacidade de armazenar dados do paciente, médico, instituição e vários outros parâmetros característicos a determinado exame. Por esse motivo, também, a imagem DICOM se torna um arquivo bastante sensível em relação a questão ética médica. Ao mesmo tempo, possui uma riqueza de detalhes que poderiam muito bem proporcionar grandes melhorias na educação, caso fosse permitido compartilha-las. O sistema proposto neste trabalho tem como objetivo criptografar todas as informações necessárias para assegurar a não identificação de qualquer indivíduo ou instituição relacionados ao exame, tornando possível, assim, a disponibilização e compartilhamento de imagens médicas sem risco de comprometimento ético. Este sistema seguiu a documentação disponibilizada pelos desenvolvedores do padrão DICOM, e conseguiu com êxito, utilizando apenas soluções próprias e de código aberto, anonimizar todos os dados relevantes em relação a identidade dos envolvidos no exame.

Palavras-chave: DICOM. Criptografia. Imagens Médicas. Anonimização. Anonimizar. Privacidade.

GUERRA, Paulo Victor de Azevedo. **DICOM Anonymous: WEB System for Anonymizing Medical Images.** Conclusion Work Project, Biomedical Engineering Bachelor Degree, Federal University of Rio Grande do Norte, 29p., 2016.

ABSTRACT

Medical images of the DICOM standard are widely distributed and largely used in all medical examination modalities. A key feature of these images is their ability to store data of patient, physician, institution, and various other characteristic parameters for a given examination. For this reason, also, the DICOM image becomes a very sensitive file regarding the medical ethical issue. At the same time, it has a wealth of detail that could provide great improvements in education if it were possible to share them. The system proposed in this work aims to encrypt all information necessary to ensure the non-identification of an individual or institution related to the examination, thus making it possible to provide and share medical images without risk of ethical commitment. This system followed the documentation provided by the developers of the DICOM standard, and was successful in anonymizing all relevant data regarding the identity of those involved in the exam using only open source solutions.

Keywords: DICOM. Cryptography. Medical Images. Anonymity. Anonymize. Privacy.

1. INTRODUÇÃO

O padrão DICOM (*Digital Imaging and Communications in Medicine*) surgiu a partir da necessidade de possibilitar e facilitar a comunicação entre dispositivos médicos, além de reduzir os custos e complexidade em integrar hardware e software de diferentes fabricantes. Antes do padrão DICOM, cada fabricante de equipamentos de imagens médicas possuía seu próprio formato de arquivo e protocolos de comunicação (FIRMINO, 2012).

Uma característica fundamental do padrão DICOM é a capacidade de armazenar informações, seja do equipamento, exame, paciente ou hospital. Apesar de todas as possibilidades positivas obtidas através da sua aplicação, podemos destacar uma questão negativa bastante relevante, a ausência de segurança ao acesso dessas informações. Principalmente, por tratarmos de imagens médicas que possuem informações a nível pessoal dos pacientes. De acordo com a documentação do padrão DICOM, disponibilizado pela NEMA (*National Electrical Manufacturers Association*), não há mecanismos de segurança nativos, mas são disponibilizados meios para seu desenvolvimento (W. DEAN BIDGOOD, JR., STEVEN C. HORII, 1997).

Assim, este trabalho tem o objetivo de desenvolver um software que possa ser integrado a um sistema de arquivamento de imagens médicas, e tenha a capacidade de proteger qualquer informação contida nos dados de arquivos DICOM. Além de obter maior controle sobre os dados do exame, as imagens protegidas poderão ser disponibilizadas para estudos e pesquisas, sem preocupação com violação de privacidade das informações pessoais do paciente.

2. OBJETIVOS

2.1. GERAL

O objetivo deste trabalho é desenvolver uma aplicação web com a funcionalidade de anonimizar imagens médicas no padrão DICOM, seguindo a documentação disponibilizada pela NEMA.

2.2. ESPECÍFICO

Como objetivos específicos tem-se:

- Obter os dados presentes na imagem DICOM e encriptar os campos referentes a identidade do paciente, profissional e instituição ligados ao exame.
- Escrever os campos encriptados novamente na imagem. Salvar a chave para decriptografia associada a aquele exame na própria imagem. E por último retornar a imagem, agora protegida, para o usuário.
- Além disso, esse sistema deve ser capaz de facilmente migrar para qualquer outro projeto compatível, permitindo assim compor e aprimorar outros sistemas desenvolvidos com a mesma tecnologia.

3. FUNDAMENTAÇÃO TEÓRICA

3.1. IMAGENS MÉDICAS

3.1.1. HISTÓRICO

Desde a descoberta dos raios X em Novembro de 1895 pelo cientista alemão Wilhelm Conrad Röntgen (Figura 1), as imagens médicas têm avançado significativamente. Não demorou para físicos começaram a explorar o potencial médico de uma radiografia. Devido a esta descoberta, diversos traumas ósseos puderam ser visualizados de maneiras que previamente seriam impossíveis. Fazendo com que a radiografia se tornasse inclusive parte importante da medicina durante a primeira guerra mundial. (SCATLIFF, 2014).

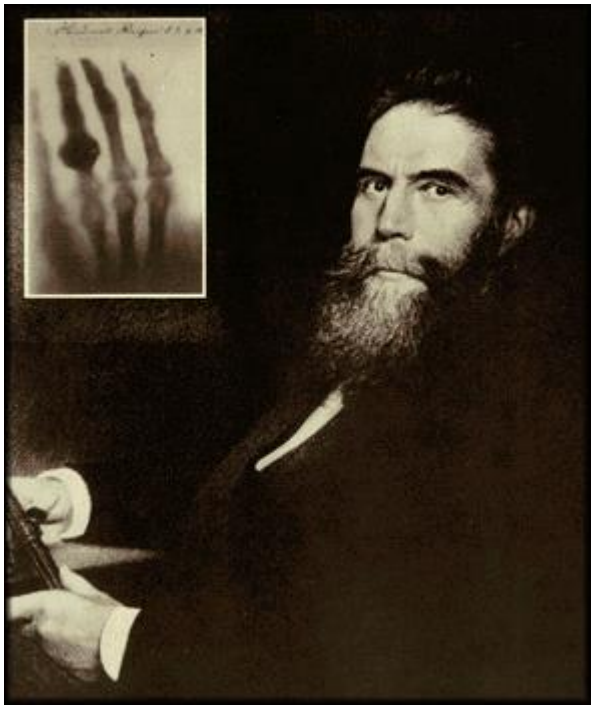

Figura 1 - Wilhelm Conrad Röntgen e a radiografia da mão de sua esposa Anna Bertha Ludwig.
Fonte: http://web.calstatela.edu/faculty/kaniol/f2000_lect_nuclphys/lect1/roentgen.htm

Com o avanço da tecnologia relacionada aos raios X, a movimentação do paciente passou a ser visualizada, não apenas a imagem estática, e a fluoroscopia se tornou possível. O processo de fluoroscopia consiste da obtenção de imagens em tempo real das estruturas internas de um paciente. A partir de 1920, radiologistas começaram a dar contraste para os pacientes tomarem, normalmente bário, por ser opaco aos feixes do

rádio. Foi dessa forma que cânceres como de esôfago, estômago e intestino começaram a ser diagnosticados, assim como, úlceras, diverticulites e apendicites. Naquela época os riscos em relação a exposição aos raios X não eram conhecidos, e os exames eram realizados sem nenhuma proteção (Figura 2). Atualmente, esses diagnósticos são realizados por CT (*Computed Tomography*), que apesar de ainda emitir radiação ionizante, possuem menor risco de causar câncer. (BRADLEY, 2008).

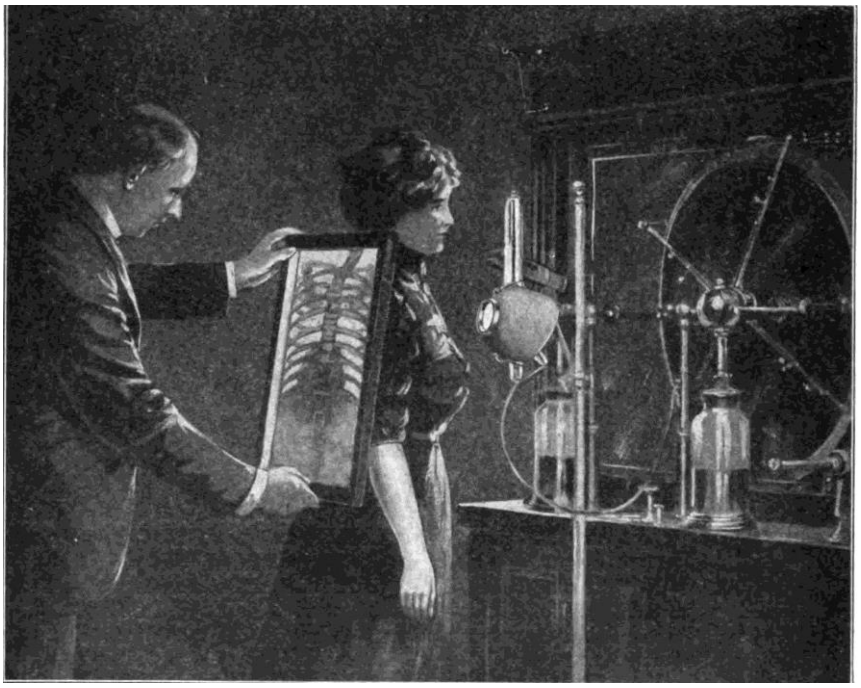

Figura 2 - Procedimento de Fluoroscopia em 1909

Fonte: https://upload.wikimedia.org/wikipedia/commons/0/01/Fluoroscopy_procedure_1909.jpg

Computadores entraram no mundo das imagens médicas no início dos anos 70, com o advento da CT e MR. O grande diferencial da CT, em relação ao raio-x, é permitir múltiplas imagens (fatias), antes da CT havia apenas filmes planos. Na CT, um tubo de raios X rotacionam em volta do paciente enquanto detectores percebem os raios não absorvidos, refletidos ou refratados assim que passam pelo corpo. (BRADLEY, 2008).

3.1.2. DICOM

DICOM é um padrão universal de imagens médicas, fornecendo todas as informações necessárias para o diagnóstico e processamento de dados dessas imagens.

Entretanto, é importante não confundir DICOM com uma imagem ou, simplesmente, formato de arquivo. O padrão consiste em um protocolo a nível de aplicação projetado para cobrir todos os aspectos da medicina contemporânea, envolvendo transferência de dados, armazenamento e visualização das imagens médicas. (PIANYKH, 2011).

O padrão DICOM pode ser dividido em três partes, para um melhor entendimento do seu funcionamento. *DICOM Protocol* (DP), *DICOM Services* (DS), e *DICOM Objects* (DO). No entanto, para este trabalho, precisamos nos aprofundar apenas em DO.

Um DO é compreendido em *DICOM Elements* (DE) ou *DICOM Attributes* (DA), como mostra a figura 3. Cada DE possui uma *DICOM Tag* (DT), uma representação para o tipo de dado, denominada *Value Representation* (VR), um campo para especificar o tamanho máximo permitido do elemento, *Value Length* (VL), e por último, o campo de valor do elemento *Value Field* (VF). (DICOM Standard PS3.3).

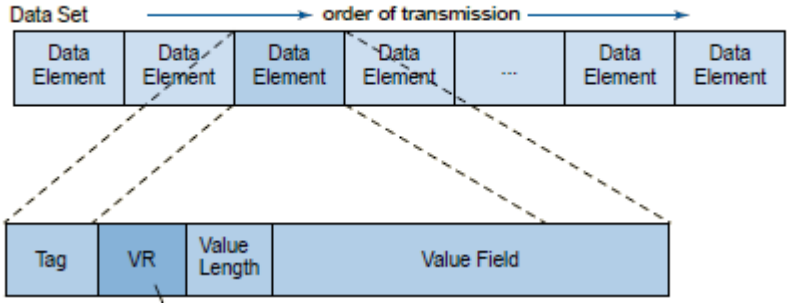

Figura 3 – DICOM Dataset e estrutura

Fonte: Dicom Standard PS3.5

Para descrever cada um desses campos, vejamos como exemplo o DE, *PatientName*, figura 4. A DT é composta por um par de dois grupos de número inteiros. O primeiro número representa o grupo a qual o elemento pertence, ou seja, todas as tags com início (0010), por exemplo, estão relacionadas a uma informação de identificação do paciente. Enquanto o segundo número representa o elemento em si, dentro daquele grupo.

O VR define o tipo de dado de determinado elemento. No exemplo da figura 4, temos PN (*Person Name*). O VR é um campo opcional na representação do elemento, sendo relevante apenas para o protocolo de transferência e na definição do VL. O tamanho do VL é definido exclusivamente pelo VR.

(0010,0010)	PN	64 (bytes)	Nome Paciente Exemplo

Figura 4 – DE PatientName (DT, VR, VL, VF) respectivamente

Fonte: Autoria Própria

3.2. CRIPTOGRAFIA E DECRIPTOGRAFIA

Segurança é um dos maiores desafios quando falamos sobre internet e sistemas web. Essas tecnologias estão crescendo rapidamente, e isso implica num aumento na quantidade de dados trafegando pela internet e na probabilidade desse dado ser interceptado. Visto isso, a criptografia é a melhor solução para que os dados transitem sem que haja preocupação com a segurança, caso tais dados sejam sigilosos (THAKUR, J., KUMAR, N., 2011).

A criptografia em si, é o processo de conversão de texto simples (Figura 5), legível, em um texto criptografado, ilegível, normalmente com o objetivo de se proteger dados (SINGH, S.P., 2011).

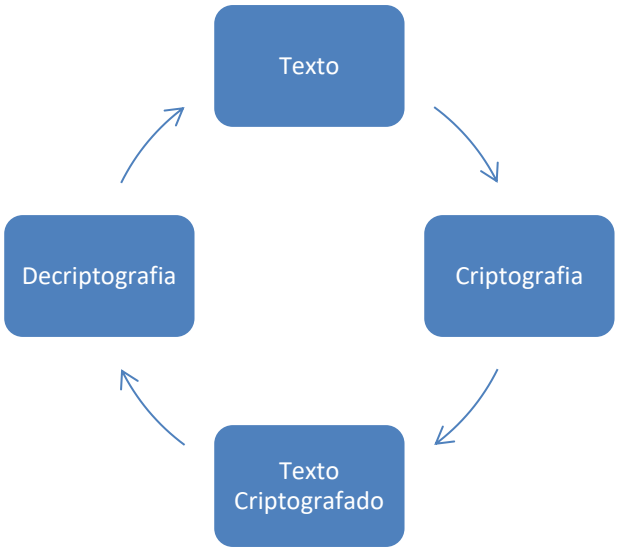

Figura 5 – Fluxograma exemplificando a criptografia

Fonte: Autoria própria

Existem dois tipos principais de criptografias, simétrica ou assimétrica. Também conhecidos por Criptografia de Chave Secreta ou Pública, respectivamente. A escolha desses tipos depende do objetivo de quem for utilizar, e não há nada que impeça uma combinação dessas técnicas.

A criptografia simétrica utiliza uma chave secreta e única (Figura 6). Nesse processo, tanto quem recebe quanto quem escreve a mensagem compartilham a mesma chave secreta. Uma característica marcante desse tipo de criptografia é que o texto original e o texto criptografado possuem o mesmo tamanho (THAKUR, J., KUMAR, N., 2011).

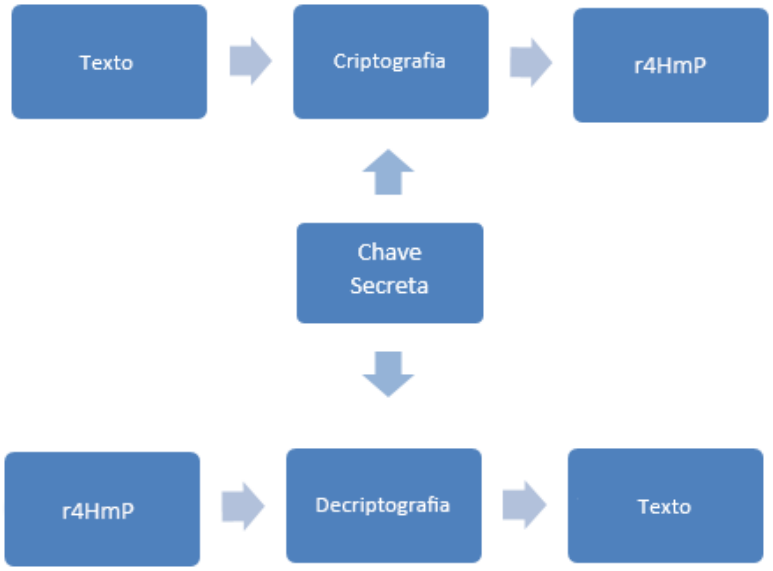

Figura 6 – Processo de Criptografia Simétrica

Fonte: Autoria Própria

Na criptografia assimétrica, duas chaves são utilizadas (Figura 7). Uma dessas chaves deve ser conhecida por quem deseja criptografar ou assinar o texto, é a chave pública, enquanto a outra é privada e serve para descriptografar o texto (THAKUR, J., KUMAR, N., 2011). Uma característica bastante interessante da criptografia assimétrica, é a possiblidade de criar uma assinatura única, a partir da sua chave privada, que pode ser verificada a partir do uso da chave pública, identificando assim, a origem de determinado bloco criptografado.

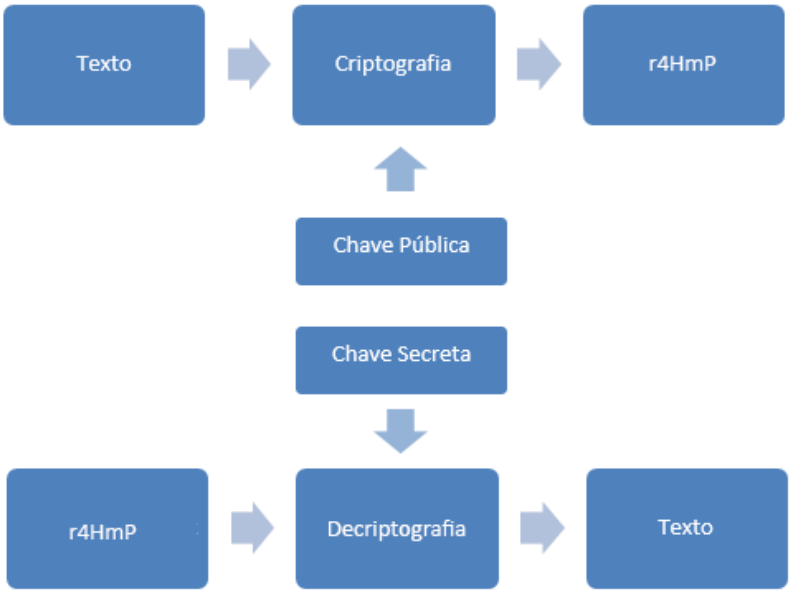

Figura 7 – Processo de Criptografia Assimétrica – Autoria Própria

O algoritmo escolhido para este trabalho foi o RSA, um algoritmo de criptografia assimétrica, ou seja, possui uma chave pública e outra, privada. Ambas as chaves podem ser utilizadas para criptografar e decriptografar uma mensagem. Esta característica é um dos motivos do RSA ter se tornado o algoritmo de criptografia assimétrica mais amplamente utilizado, pois fornece confidencialidade, integridade e autenticidade (ROUSE, M., 2016).

3.3. DESENVOLVIMENTO DE SISTEMAS WEB

Para desenvolver sistemas web é necessário conhecimento em diversas linguagens, como, por exemplo:

- HTML;
- CSS;
- JavaScript;
- PHP;
- Python;
- Ruby.

Essas linguagens podem ser divididas em dois conceitos a partir de suas funções no sistema. As linguagens de *front-end* (HTML, CSS, JavaScript), aquelas responsáveis

pelas funções do sistema que irão interagir diretamente com o usuário através do navegador. E as de *back-end* (PHP, Python ou Ruby), que realizam suas funções junto ao servidor onde o sistema será executado.

Para facilitar e acelerar o processo de desenvolvimento surgiram os frameworks web. Frameworks, em relação a desenvolvimento web, podem ser definidos como um conjunto de códigos, específicos de terminada linguagem, e bibliotecas prontas para serem utilizadas. Esses recursos proveem toda uma série de funcionalidades pré-definidas, permitindo um rápido desenvolvimento inicial do projeto. Frameworks existem tanto para a área de front quanto de *back-end*.

Neste projeto, as linguagens utilizadas foram, respectivamente, HTML, CSS, JavaScript e Python. E os frameworks, Django (Python) e Bootstrap (CSS).

3.3.1. DJANGO E BIBLIOTECAS

Ao utilizar Django é possível criar sistemas web verdadeiramente densos e dinâmicos em um curto espaço de tempo. Django foi desenvolvido para prover atalhos para as tarefas comuns e básicas presentes em todos os sistemas, e ao mesmo tempo permitir que o desenvolvedor crie suas próprias soluções (HOLOVATY, A., KAPLAN-MOSS, J., 2009).

Um dos motivos fundamentais da escolha desse framework foi uma característica herdada do Python conhecida por *batteries included*. Desta forma, com Django é possível obter uma coleção de recursos que auxiliam o desenvolvimento do sistema, e ainda fragmentar problemas complexos em tarefas mais simples (DANTAS, M.C.R., 2016). Para a problemática em questão, anonimizar imagens médicas do tipo DICOM, existem duas bibliotecas desenvolvidas em Python que foram essenciais para a realização do projeto. PyDICOM e PyCrypto, ambas tecnologias livres e de código aberto.

PyDICOM permite a manipulação de imagens DICOM através de funções Python. Com essa biblioteca é possível realizar um CRUD (*Create, Retrieve* ou *Read, Update, Delete*) completo com todo e qualquer dado presente nas imagens médicas. (PyDICOM, 2016). A figura 8 é um exemplo de como os dados estão organizados na imagem DICOM e são visualizados a partir de comandos disponibilizados no PyDICOM.

```
(0012, 0062) Patient Identity Removed            CS: 'Yes'
(0018, 0022) Scan Options                        CS: 'SURVIEW'
(0018, 0050) Slice Thickness                     DS: '0.625'
(0018, 0060) KVP                                 DS: '120'
(0018, 0090) Data Collection Diameter            DS: '500'
(0018, 1000) Device Serial Number                LO: ''
(0018, 1020) Software Version(s)                 LO: '2.6.2'
(0018, 1030) Protocol Name                       LO: ''
(0018, 1100) Reconstruction Diameter             DS: '500'
```

Figura 8 – Trecho de visualização de dataset DICOM via PyDICOM

Fonte: Autoria Própria

PyCrypto é uma biblioteca bastante completa que possui uma coleção de algoritmos de criptografia (AES, DES, Blowfish, RSA, ...) e funções hash (SHA256, MD5, ...). Ao utilizar o módulo RSA do PyCrypto, é possível criar ambas as chaves, pública e privada, realizar a verificação da assinatura, importar e exportar as chaves. (PyCrypto, 2016).

4. METODOLOGIA

O fluxograma apresentado na figura 9, representa de modo simplificado os passos realizados pelo sistema até atingir o objetivo final, isto é, retornar a imagem com os dados criptografados.

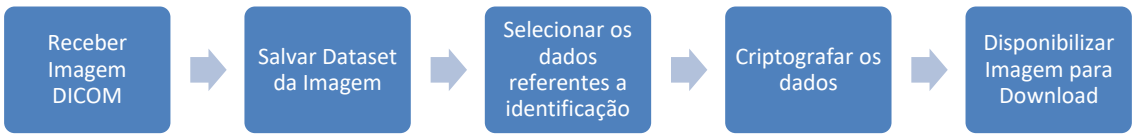

Figura 9 – Fluxograma da Metodologia aplicada

Fonte: Autoria Própria

4.1 CARACTERÍSTICAS GERAIS DO SISTEMA

O sistema foi desenvolvido utilizando principalmente a linguagem de programação Python com o framework Django. As principais bibliotecas Python utilizadas foram PyDICOM e PyCrypto. O front-end foi desenvolvido usando as linguagens HTML, CSS e JavaScript, incluído a biblioteca jQuery e o framework Bootstrap. O Git foi utilizado como sistema de versionamento de código, sendo o GitLab a interface web para

interagir com o repositório. Todas as funcionalidades foram testadas em servidor local de uma máquina com sistema operacional Ubuntu 16.04 x64, processador Intel i5 3337U 1.80GHz, 6GB RAM.

4.2. OBTENDO IMAGEM E SALVANDO DATASET

Seguindo as normas propostas pela NEMA, na documentação do DICOM. (DICOM Supplement 55, NEMA). Existem determinados conjuntos de DE essenciais (Tabela 1) para que possamos considerar a imagem médica anônima. Ainda assim, é importante ressaltar que, na própria documentação, fica clara a possibilidade de identificação do paciente, médico ou instituição, mesmo diante de tais medidas. Apesar de bastante improvável. (DICOM Standard PS3.15, apêndice E, NEMA).

Ao receber a imagem médica enviada pelo usuário do sistema, o dataset completo de tal imagem pode ser obtido através de funções disponibilizadas pela biblioteca PyDICOM. A partir deste ponto, os dados presentes na tabela 1 são selecionados para a etapa de anonimização.

Tabela 1 – DE necessários para anonimizar a imagem

Nome do Elemento	**Tag**
Instance Creator UID	(0008,0014)
SOP Instance UID	(0008,0018)
Accession Number	(0008,0050)
Institution Name	(0008,0080)
Institution Address	(0008,0081)
Referring Physician's Name	(0008,0090)
Referring Physician's Address	(0008,0092)
Referring Physician's Telephone Numbers	(0008,0094)
Station Name	(0008,1010)
Study Description	(0008,1030)
Series Description	(0008,103E)
Institutional Department Name	(0008,1040)
Physician(s) of Record	(0008,1040)

Performing Physicians' Name	(0008,1050)
Name of Physician(s) Reading Study	(0008,1060)
Operators' Name	(0008,1070)
Admitting Diagnoses Description	(0008,1080)
Referenced SOP Instance UID	(0008,1155)
Derivation Description	(0008,2111)
Patient's Name	(0010,0010)
Patient ID	(0010,0020)
Patient's Birth Date	(0010,0030)
Patient's Birth Time	(0010,0032)
Patient's Sex	(0010,0040)
Other Patient Ids	(0010,1000)
Other Patient Names	(0010,1001)
Patient's Age	(0010,1010)
Patient's Size	(0010,1020)
Patient's Weight	(0010,1030)
Medical Record Locator	(0010,1090)
Ethnic Group	(0010,2160)
Occupation	(0010,2180)
Additional Patient's History	(0010,21B0)
Patient Comments	(0010,4000)
Device Serial Number	(0018,1000)
Protocol Name	(0018,1030)
Study Instance UID	(0020,000D)
Series Instance UID	(0020,000E)
Study ID	(0020,0010)
Frame of Reference UID	(0020,0052)
Synchronization Frame of Reference UID	(0020,0200)
Image Comments	(0020,4000)
Request Attributes Sequence	(0040,0275)
UID	(0040,A124)
Content Sequence	(0040,A730)

Storage Media File-set UID	(0088,0140)
Referenced Frame of Reference UID	(3006,0024)
Related Frame of Reference UID	(3006,00C2)

4.3. SELEÇÃO DE DADOS E CRIPTOGRAFIA

Sabendo exatamente quais dados são necessários para obtermos o objetivo de proteger a identidade de todos os envolvidos, podemos partir para o processo de obtenção da imagem.

Foi desenvolvido um formulário, conforme visualizado na figura 10, onde o usuário seleciona a imagem desejada. De posse da imagem, todo o processo de anonimização se inicia.

Utilizando função '*dicom.read_file()*' disponibilizada pela biblioteca PyDICOM o dataset completo da imagem é obtida. Nesse ponto, é importante informar que foi notada uma grande falta de homogeneidade nos datasets de diferentes imagens DICOM. Assim, a solução adotada foi desenvolver uma função com tratamento de exceção e, dessa forma, criar os campos necessários que faltavam.

Após obter todos os dados, os mesmos são criptografados com a chave pública gerada pela função RSA presente no PyCrypto. Todos os DE da imagem são apagados e dois DE são criados, conforme indicado na documentação. (0012,0062) *PatientIdentityRemoved* recebe o valor '*Yes'*, e os dados criptografados foram convertidos em formato string e salvos em (0400,0520) *EncryptedContent*. Utilizando a função '*dicom.write_file()*' do PyDICOM, as modificações são salvas em num novo arquivo e disponibilizado para download.

O processo de anonimização foi realizado com sucesso em oito diferentes modalidades de exames, todas retiradas do sistema OpenPACS do HUOL (Hospital Universitário Onofre Lopes):

- CT
- DX
- ECG
- US
- XA
- MR

- XC
- CR

5. RESULTADOS E DISCUSSÕES

O sistema DICOM Anonymous foi desenvolvido visando ser uma aplicação Django capaz de ser integrada em qualquer projeto compatível com este framework. Entretanto, também é possível executá-lo independentemente.

5.1. FRONT-END

Uma única interface *front-end* (Figura 10) foi desenvolvida para realização dos testes necessários para garantir todas as funcionalidades propostas. Essa mesma interface pode ser utilizada em produção. Todo o processo de anonimização ocorre numa única tela e pode ser completado com apenas três ações.

Figura 10 – Interface do sistema

Fonte: Autoria Própria

Para anonimizar a imagem basta entrar com um novo título, selecionar a imagem desejada e confirmar clicando no botão ‘Anonimizar’, conforme ilustrado na figura 9.

Ao realizar esse procedimento, uma cópia da imagem é salva no banco de dados, enquanto o usuário recebe por download a imagem anônima desejada.

5.2. BACK-END

Na figura 11 alguns DE foram alterados para preservar a identidade do paciente e exemplificar como o conjunto de dados pode ser visualizado enquanto todo o processo ocorre.

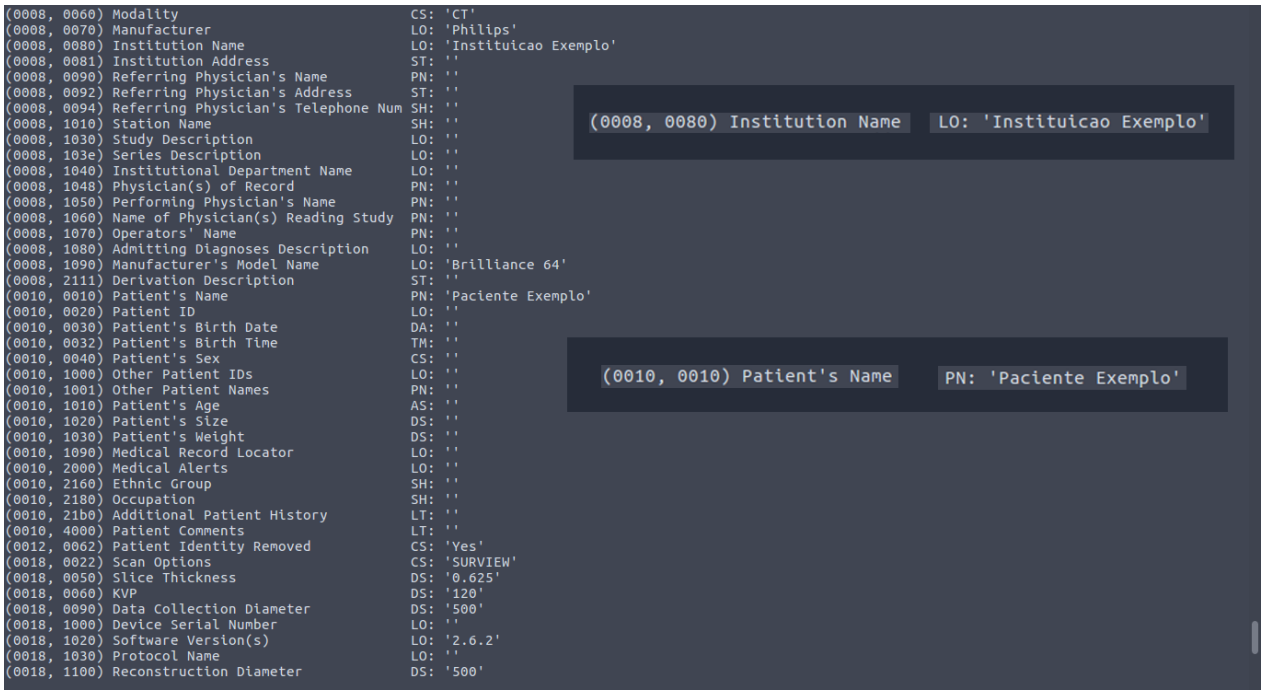

Figura 11 – Visualização do dataset durante o procedimento. Destaque para os nomes, Paciente Exemplo e Instituição Exemplo - Fonte: Autoria Própria

Na figura 12 podemos notar que todas as informações que poderiam servir para associar a imagem médica a algum paciente, médico ou instituição foram removidas com sucesso. Assim como, o DE *PatientIdentityRemoved* foi criado e preenchido com o valor ‘*Yes’*.

```
(0008, 0060) Modality                          CS: 'CT'
(0008, 0070) Manufacturer                      LO: 'Philips'
(0008, 0080) Institution Name                  LO: ''
(0008, 0081) Institution Address               ST: ''
(0008, 0090) Referring Physician's Name        PN: ''
(0008, 0092) Referring Physician's Address     ST: ''
(0008, 0094) Referring Physician's Telephone Num SH: ''
(0008, 1010) Station Name                      SH: ''
(0008, 1030) Study Description                 LO: ''
(0008, 103e) Series Description                LO: ''
(0008, 1040) Institutional Department Name     LO: ''
(0008, 1048) Physician(s) of Record            PN: ''
(0008, 1050) Performing Physician's Name       PN: ''
(0008, 1060) Name of Physician(s) Reading Study PN: ''
(0008, 1070) Operators' Name                   PN: ''
(0008, 1080) Admitting Diagnoses Description   LO: ''
(0008, 1090) Manufacturer's Model Name         LO: 'Brilliance 64'
(0008, 2111) Derivation Description            ST: ''
(0010, 0010) Patient's Name                    PN: ''
(0010, 0020) Patient ID                        LO: ''
(0010, 0030) Patient's Birth Date              DA: ''
(0010, 0032) Patient's Birth Time              TM: ''
(0010, 0040) Patient's Sex                     CS: ''
(0010, 1000) Other Patient IDs                 LO: ''
(0010, 1001) Other Patient Names               PN: ''
(0010, 1010) Patient's Age                     AS: ''
(0010, 1020) Patient's Size                    DS: ''
(0010, 1030) Patient's Weight                  DS: ''
(0010, 1090) Medical Record Locator            LO: ''
(0010, 2000) Medical Alerts                    LO: ''
(0010, 2160) Ethnic Group                      SH: ''
(0010, 2180) Occupation                        SH: ''
(0010, 21b0) Additional Patient History        LT: ''
(0010, 4000) Patient Comments                  LT: ''
(0012, 0062) Patient Identity Removed          CS: 'Yes'
(0018, 0022) Scan Options                      CS: 'SURVIEW'
(0018, 0050) Slice Thickness                   DS: '0.625'
(0018, 0060) KVP                               DS: '120'
(0018, 0090) Data Collection Diameter          DS: '500'
(0018, 1000) Device Serial Number              LO: ''
(0018, 1020) Software Version(s)               LO: '2.6.2'
(0018, 1030) Protocol Name                     LO: ''
(0018, 1100) Reconstruction Diameter           DS: '500'
```

Figura 12 – Imagem completamente anônima

Fonte: Autoria Própria

5.3. CRIPTOGRAFIA SIMÉTRICA E ASSIMÉTRICA

Conforme citado anteriormente na metodologia, o método de criptografia selecionado foi o RSA, um método criptográfico de chave-pública, ou, assimétrico. As principais razões desta escolha foram as possibilidades de criar uma assinatura única para imagem criptografada, assim como, pelas características do método assimétrico, as chaves para criptografar e decriptografar serem diferentes.

Entretanto, durante os testes realizados em todas as imagens DICOM, ficou claro uma limitação da criptografia RSA. Quanto maior o número de dados que se deseja criptografar, maior será o tempo gasto para a realização do processo, e o tamanho do texto a ser criptografado não pode ultrapassar o tamanho da chave. Portanto, caso fosse necessário anonimizar o dataset completo, a criptografia RSA se torna inviável.

Uma solução proposta, seria a utilização de criptografia simétrica ou uma combinação dos dois tipos. Pois, a criptografia simétrica não possui limitação quanto a quantidade de caracteres que se deseja criptografar.

6. CONCLUSÕES

Imagens DICOM possuem informações sensíveis inerentes a ética médica, e por este motivo, devem ser armazenadas com bastante cuidado e segurança. Por outro lado, essas mesmas imagens contêm dados muito valiosos que poderiam ser compartilhados e/ou analisados em larga escala, sem ferir a privacidade de nenhum indivíduo.

Assim, o processo de anonimização da imagem DICOM, seguindo a própria documentação oficial do padrão, foi realizada com sucesso. Independente da modalidade do exame.

Entretanto, quando se trata de segurança nunca devemos ficar satisfeitos. Algumas novas camadas de proteção podem e devem ser adicionadas ao DICOM Anonymous futuramente. São elas, uma combinação de criptografias simétrica e assimétrica associada com esteganografia.

A esteganografia é a arte de esconder a informação, diferentemente da criptografia, onde a informação está visível, apesar de ilegível. (JUDGE, 2001). Assim, seria uma nova camada de proteção bastante interessante, pois, além do esforço necessário para quebrar a criptografia ainda seria necessário encontra-la dentro da imagem.

Por fim, o DICOM Anonymous é um sistema web desenvolvido em Django/Python que atingiu todos os principais objetivos propostos quanto a anonimização de imagens médicas, independentemente da modalidade do exame.

7. REFERÊNCIAS

BRADLEY, WILLIAM G., **History of Medical Imaging.** Proceedings of the American Philosofical Society, v.152, nº3, 2008.

DANTAS, M.C.R., **Sistema de Telemonitoramento para Paciente com Esclerose Lateral Amiotrófica**, 2016.

FIRMINO FILHO, J.M, VALENTIM, R.A.M, **OpenPACS – Sistema OpenSource para Comunicação e Arquivamento de Imagens Médicas: Relato de Experiência em um Hospital Universitário.** RECIIS, Rio de Janeiro, v. 7, nº 2, 2013.

JUDGE, JAMES C., **Steganography: Past, Present, Future.** SANS Institute Reading Room, GSEC Version 1.2f, 2001.

KAPLAN-MOSS, J., HOLOVATY, A., **The Definitive Guide to Django: Web Development Done Right.** 1ª edição, 2007.

MAHAJAN, P., SACHDEVA, A., **A Study of Encryption Algorithms AES, DES and RSA for Security.** Global Journal of Computer Science and Technology Network, Web & Security, v.13, nº15, 2013.

MARQUES, P.M.A, SALOMÃO, S.C, **PACS: Sistema de Arquivamento e Distribuição de Imagens.** Revista Brasileira de Física Médica, v.3, 2009.

MOLIN, RICHARD A., **RSA and Public-Key Cryptography.** 1ª edição, 2002.

NEMA – National Electrical Manufacture Association. **Digital Imaging Communications in Medicine (DICOM).** Disponível em <http://dicom.nema.org>. Acesso em 15 de setembro 2016.

NEMA – National Electrical Manufacture Association. **DICOM PS3.3 – Information Object Definitions.** 2016.

NEMA – National Electrical Manufacture Association. **DICOM PS3.5 – Data Structures and Encoding.** 2016.

NEMA – National Electrical Manufacture Association. **DICOM PS3.15 – Security and System Management Profiles.** 2016.

NEMA – National Electrical Manufacture Association. **DICOM Supplement 55: Attribute Level Confidentiality (including De-identification).** 2002.

PIANYKH, OLEG S., **Digital Imaging and Communications in Medicine (DICOM): A Practical Introduction and Survival Guide.** 2ª edição, 2011.

PyCrypto, 2016. **PyCrypyo Cryptography Toolkit.** Disponível em <http://pythonhosted.org/pycrypto/>. Acesso em 10 de outubro de 2016.

PyDICOM, 2016. **PyDICOM User Guide**. Disponível em <http://pydicom.readthedocs.io/en/latest/pydicom_user_guide.html>. Acesso em 07 de outubro 2016.

ROUSE, M., **RSA Algorithm (Rivest-Shamir-Adleman).** Disponível em <http://searchsecurity.techtarget.com/definition/RSA>. Acesso em 16 de novembro 2016.

SCATLIFF, JAMES H., MORRIS, PETER J., **From Röntgen to Magnetic Resonance Imaging: The History of Medical Imaging.** NCMedicalJournal, v.75, nº2, 2014.

SINGH, S.P., MAINI, R., **Comparison of Data Encryption Algorithms.** International Journal of Computer Science and Communication, v.2, nº1, 2011.

THAKUR, J., KUMAR, N., **DES, AES and Blowfish: Symmetric Key Cryptography Algorithms Simulation Based Performance Analysis.** International Journal of Emerging Technology and Advanced Engineering, v.1, nº2, 2011.

W. DEAN BIDGOOD, JR., STEVEN C. HORII, **Understanding and Using DICOM, the Data Interchange Standard for Biomedical Imaging.** Journal of the American Medical Informatics Association, v. 4, nº 3, 1997.