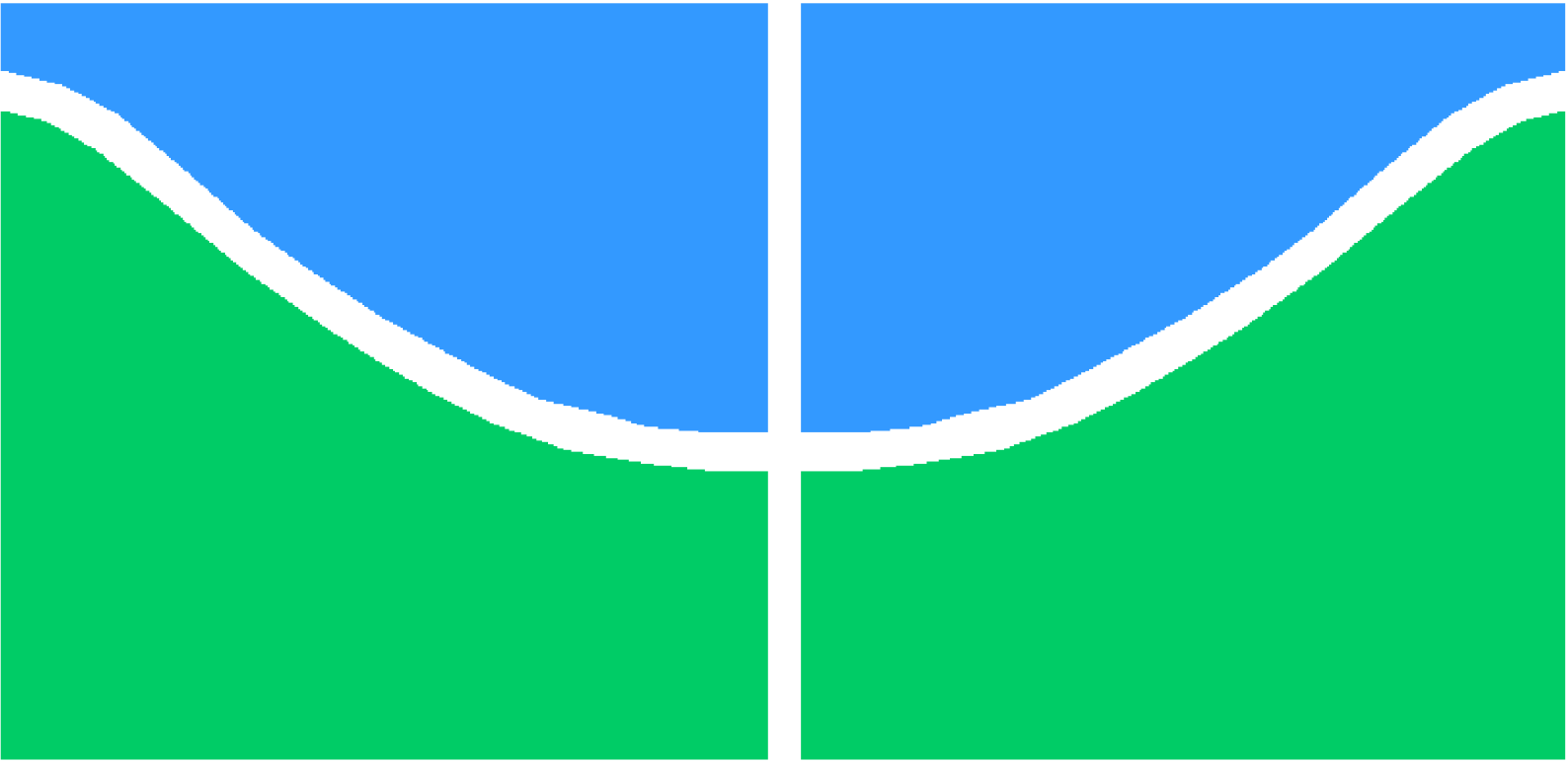

Universidade de Brasília - UnB
Faculdade UnB Gama - FGA
Engenharia de Software

Criação de Serviço para Extração de Conhecimento de Comportamento de Malware

Autor: Fillipe Oliveira Feitosa
Orientador: Dr. Fabricio Ataides Braz

Brasília, DF
2015

Fillipe Oliveira Feitosa

Criação de Serviço para Extração de Conhecimento de Comportamento de Malware

Monografia submetida ao curso de graduação em (Engenharia de Software) da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em (Engenharia de Software).

Universidade de Brasília - UnB

Faculdade UnB Gama - FGA

Orientador: Dr. Fabricio Ataides Braz

Coorientador: (quando houver, Titulação Acadêmica e Nome do Orientador)

Brasília, DF

2015

Fillipe Oliveira Feitosa
Criação de Serviço para Extração de Conhecimento de Comportamento de Malware/ Fillipe Oliveira Feitosa. – Brasília, DF, 2015-
51 p. : il. (algumas color.) ; 30 cm.

Orientador: Dr. Fabricio Ataides Braz

Trabalho de Conclusão de Curso – Universidade de Brasília - UnB
Faculdade UnB Gama - FGA , 2015.

1. Malware. 2. Machine-learning. I. Dr. Fabricio Ataides Braz. II. Universidade de Brasília. III. Faculdade UnB Gama. IV. Criação de Serviço para Extração de Conhecimento de Comportamento de Malware

CDU 02:141:005.6

Errata

Elemento opcional da **??**, 4.2.1.2). **Caso não deseje uma errata, deixar todo este arquivo em branco**. Exemplo:

FERRIGNO, C. R. A. **Tratamento de neoplasias ósseas apendiculares com reimplantação de enxerto ósseo autólogo autoclavado associado ao plasma rico em plaquetas**: estudo crítico na cirurgia de preservação de membro em cães. 2011. 128 f. Tese (Livre-Docência) - Faculdade de Medicina Veterinária e Zootecnia, Universidade de São Paulo, São Paulo, 2011.

Folha	**Linha**	**Onde se lê**	**Leia-se**
1	10	auto-conclavo	autoconclavo

Fillipe Oliveira Feitosa

Criação de Serviço para Extração de Conhecimento de Comportamento de Malware

Monografia submetida ao curso de graduação em (Engenharia de Software) da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em (Engenharia de Software).

Trabalho aprovado. Brasília, DF, 27 de novembro de 2015:

Dr. Fabricio Ataides Braz
Orientador

Titulação e Nome do Professor
Convidado 01
Convidado 1

Titulação e Nome do Professor
Convidado 02
Convidado 2

Brasília, DF
2015

A dedicatória é opcional. Caso não deseje uma, deixar todo este arquivo em branco.

Este trabalho é dedicado às crianças adultas que,
quando pequenas, sonharam em se tornar cientistas.

Agradecimentos

A inclusão desta seção de agradecimentos é opcional, portanto, sua inclusão fica a critério do(s) autor(es), que caso deseje(em) fazê-lo deverá(ão) utilizar este espaço, seguindo a formatação de *espaço simples e fonte padrão do texto (arial ou times, tamanho 12 sem negritos, aspas ou itálico.*

Caso não deseje utilizar os agradecimentos, deixar toda este arquivo em branco.

A epígrafe é opcional. Caso não deseje uma, deixe todo este arquivo em branco.

"Não vos amoldeis às estruturas deste mundo,
mas transformai-vos pela renovação da mente,
a fim de distinguir qual é a vontade de Deus:
o que é bom, o que Lhe é agradável, o que é perfeito.
(Bíblia Sagrada, Romanos 12, 2)

Resumo

O resumo deve ressaltar o objetivo, o método, os resultados e as conclusões do documento. A ordem e a extensão destes itens dependem do tipo de resumo (informativo ou indicativo) e do tratamento que cada item recebe no documento original. O resumo deve ser precedido da referência do documento, com exceção do resumo inserido no próprio documento. (. . .) As palavras-chave devem figurar logo abaixo do resumo, antecedidas da expressão Palavras-chave:, separadas entre si por ponto e finalizadas também por ponto. O texto pode conter no mínimo 150 e no máximo 500 palavras, é aconselhável que sejam utilizadas 200 palavras. E não se separa o texto do resumo em parágrafos.

Palavras-chaves: latex. abntex. editoração de texto.

Abstract

This is the english abstract.

Key-words: latex. abntex. text editoration.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

Fig.	Area of the i^{th} component
456	Isto é um número
123	Isto é outro número
lauro cesar	este é o meu nome

Lista de símbolos

Γ	Letra grega Gama
Λ	Lambda
ζ	Letra grega minúscula zeta
$\in$	Pertence

Sumário

Parte I

Introdução

1 Introdução

Este trabalho e o projeto prático a ele relacionado foram submetidos ao curso de graduação em Engenharia de Software, como requisito parcial para obtenção do título de Bacharel em Engenharia de Software. O documento foi organizado de forma a contextualizar o leitor sobre a problemática dos *malwares* e sua relação com a segurança de informação, as definições, motivação, contexto e justificativa, os objetivos do trabalho e a sua execução. Ao término, serão apresentadas as etapas necessárias para a conclusão do projeto de implantação da solução proposta.

1.1 Contexto

A sociedade moderna se apoia de forma maciça nas tecnologias de informação para nutrir o seu crescimento. Cada dia novas técnicas, ferramentas e métodos surgem e revolucionam a forma como as pessoas e as organizações se relacionam, trocam informações e armazenam dados. Por diversas razões indivíduos ou mesmo grupos mal intencionados criam software maliciosos, ou *malware*, para explorar falhas de vulnerabilidade em sistemas e interferir no processo de segurança de informação das mais variadas formas. Se tornaram uma ameaça crítica para as grandes organizações e usuários domésticos (EETEN; BAUER, 2008). Dos computadores pessoais aos modernos *smartphones*, dados pessoais do usuários podem ser roubados de qualquer lugar do mundo e até mesmo seu terminal pode ser usado como um instrumento de processamento para atividades criminosas.

Compreender o funcionamento de *malware* e técnicas de combate compreendem uma área de estudo ativa mas extremamente desafiadora devido ao contexto tecnológico inovador da atualidade e da própria complexidade computacional deste tipo de aplicação. São diversos tipos de famílias, com inúmeras formas de ataque e técnicas de invasão. O processo tradicional de engenharia reversa do código malicioso envolve a "desmontagem"e descompilação do código, mas esta análise é extremamente complexa e pode não gerar os resultados esperados num tempo razoável.

> "Há dois pré-requisitos em tal situação: uma é que os investigadores devem ter conhecimentos técnicos profundos de linguagem de máquina, a outra é que os processos de análise devem ser eficientes o suficiente para lidar com a renovação constante e variação do *malware*."(QIAO et al., 2014)

Felizmente, existem outras formas de analisar código malicioso além da engenharia reversa do código. Uma alternativa é o estudo do comportamento do *malware*, que

procura entender o funcionamento de uma amostra de *malware* agrupando-os por comportamento. Entende-se por análise do comportamento tudo aquilo que o *malware* realiza, executa ou opera no terminal hospedeiro. Esta análise é realizada num ambiente controlado, conhecido como ambiente "caixa de areia"ou *sandbox* de forma que todo tipo de operação realizada no sistema operacional possa ser capturada e registrada em relatório (QIAO et al., 2014). Tal ambiente deve ser configurado de tal forma que o relatório final que todas as interações do *malware* com o sistema operacional são interceptadas. Dessa forma, o código do malware em si é completamente ignorado para o relatório de comportamento (RIECK et al., 2011). Como resultado, o relatório retorna uma série de registros, que variam de simples modificações nos recursos do sistema e até ao tráfego na rede.

1.2 Problema

1.3 Objetivo Geral

O objetivo geral deste trabalho de conclusão de curso é prover um serviço para extração de conhecimento a partir do comportamento de *malware*, de forma que um usuário possa submeter binários de *malware* e tenha como retorno uma análise completa.

A proposta é que todo o processo fique transparente para o usuário, e que a solução seja capaz de apresentar um retorno com conhecimento agregado.

1.4 Objetivos Específicos

Para alcançar o objetivo geral, este trabalho foi dividido em etapas. Estes objetivos são:

- Definir e consolidar o processo de extração de conhecimento
- Construir um laboratório para extração de comportamento de *malware*
- Engenharia de características
- Aplicar processo de extração de conhecimento na massa de dados de características
- Desenvolver a arquitetura do serviço
- Laboratório com o protótipo do serviço

1.5 Justificativa

1.6 Método

A execução deste trabalho foi dividida em 6 etapas, que tratam da compreensão da problemática relacionada à extração de conhecimento do comportamento do *malware*, até a proposição de uma solução que seja relevante para potenciais usuários.

1.6.1 Etapa 1: Definir e consolidar o processo de extração de conhecimento

1.6.2 Etapa 2: Construir um laboratório para extração de comportamento de *malware*

1.6.3 Etapa 3: Engenharia de características

1.6.4 Etapa 4: Aplicar processo de extração de conhecimento na massa de dados de características

1.6.5 Etapa 5: Desenvolver a arquitetura do serviço

1.6.6 Etapa 6: Laboratório com o protótipo do serviço

1.7 Estrutura do Trabalho

Parte II

Referencial Teórico

2 Referencial Teórico

Estas instruções apresentam um conjunto mínimo de exigências necessárias a uniformidade de apresentação do relatório de Trabalho de Conclusão de Curso da FGA. Estilo, concisão e clareza ficam inteiramente sob a responsabilidade do(s) aluno(s) autor(es) do relatório.

As disciplinas de Trabalho de Conclusão de Curso (TCC) 01 e Trabalho de Conclusão de Curso (TCC) 02 se desenvolvem de acordo com Regulamento próprio aprovado pelo Colegiado da FGA. Os alunos matriculados nessas disciplinas devem estar plenamente cientes de tal Regulamento.

2.1 Composição e estrutura do trabalho

A formatação do trabalho como um todo considera três elementos principais: (1) pré-textuais, (2) textuais e (3) pós-textuais. Cada um destes, pode se subdividir em outros elementos formando a estrutura global do trabalho, conforme abaixo (as entradas itálico são *opcionais*; em itálico e negrito são ***essenciais***):

Pré-textuais

- Capa
- Folha de rosto
- *Dedicatória*
- *Agradecimentos*
- *Epígrafe*
- Resumo
- Abstract
- Lista de figuras
- Lista de tabelas
- Lista de símbolos e
- Sumário

Textuais

- ***Introdução***

- ***Desenvolvimento***
- ***Conclusões***

Pós-Textuais

- Referências bibliográficas
- *Bibliografia*
- Anexos
- Contracapa

Os aspectos específicos da formatação de cada uma dessas três partes principais do relatório são tratados nos capítulos e seções seguintes.

No modelo LaTeX, os arquivos correspondentes a estas estruturas que devem ser editados manualmente estão na pasta **editáveis**. Os arquivos da pasta **fixos** tratam os elementos que não necessitam de edição direta, e devem ser deixados como estão na grande maioria dos casos.

2.2 Considerações sobre formatação básica do relatório

A seguir são apresentadas as orientações básicas sobre a formatação do documento. O modelo LaTeX já configura todas estas opções corretamente, de modo que para os usuários deste modelo o texto a seguir é meramente informativo.

2.2.1 Tipo de papel, fonte e margens

Papel - Na confecção do relatório deverá ser empregado papel branco no formato padrão A4 (21 cm x 29,7cm), com 75 a 90 g/m2.

Fonte – Deve-se utilizar as fontes Arial ou Times New Roman no tamanho 12 pra corpo do texto, com variações para tamanho 10 permitidas para a wpaginação, legendas e notas de rodapé. Em citações diretas de mais de três linhas utilizar a fonte tamanho 10, sem itálicos, negritos ou aspas. Os tipos itálicos são usados para nomes científicos e expressões estrangeiras, exceto expressões latinas.

Margens - As margens delimitando a região na qual todo o texto deverá estar contido serão as seguintes:

- Esquerda: 03 cm;
- Direita : 02 cm;

- Superior: 03 cm;
- Inferior: 02 cm.

2.2.2 Numeração de Páginas

A contagem sequencial para a numeração de páginas começa a partir da primeira folha do trabalho que é a Folha de Rosto, contudo a numeração em si só deve ser iniciada a partir da primeira folha dos elementos textuais. Assim, as páginas dos elementos pré-textuais contam, mas não são numeradas e os números de página aparecem a partir da primeira folha dos elementos textuais que é a Introdução.

Os números devem estar em algarismos arábicos (fonte Times ou Arial 10) no canto superior direito da folha, a 02 cm da borda superior, sem traços, pontos ou parênteses.

A paginação de Apêndices e Anexos deve ser contínua, dando seguimento ao texto principal.

2.2.3 Espaços e alinhamento

Para a monografia de TCC 01 e 02 o espaço entrelinhas do corpo do texto deve ser de 1,5 cm, exceto RESUMO, CITAÇÔES de mais de três linhas, NOTAS de rodapé, LEGENDAS e REFERÊNCIAS que devem possuir espaçamento simples. Ainda, ao se iniciar a primeira linha de cada novo parágrafo se deve tabular a distância de 1,25 cm da margem esquerda.

Quanto aos títulos das seções primárias da monografia, estes devem começar na parte superior da folha e separados do texto que o sucede, por um espaço de 1,5 cm entrelinhas, assim como os títulos das seções secundárias, terciárias.

A formatação de alinhamento deve ser justificado, de modo que o texto fique alinhado uniformemente ao longo das margens esquerda e direita, exceto para CITAÇÕES de mais de três linhas que devem ser alinhadas a 04 cm da margem esquerda e REFERÊNCIAS que são alinhadas somente à margem esquerda do texto diferenciando cada referência.

2.2.4 Quebra de Capítulos e Aproveitamento de Páginas

Cada seção ou capítulo deverá começar numa nova pagina (recomenda-se que para texto muito longos o autor divida seu documento em mais de um arquivo eletrônico).

Caso a última pagina de um capitulo tenha apenas um número reduzido de linhas (digamos 2 ou 3), verificar a possibilidade de modificar o texto (sem prejuízo do conteúdo e obedecendo as normas aqui colocadas) para evitar a ocorrência de uma página pouco aproveitada.

Ainda com respeito ao preenchimento das páginas, este deve ser otimizado, evitando-se espaços vazios desnecessários.

Caso as dimensões de uma figura ou tabela impeçam que a mesma seja posicionada ao final de uma página, o deslocamento para a página seguinte não deve acarretar um vazio na pagina anterior. Para evitar tal ocorrência, deve-se re-posicionar os blocos de texto para o preenchimento de vazios.

Tabelas e figuras devem, sempre que possível, utilizar o espaço disponível da página evitando-se a "quebra" da figura ou tabela.

2.3 Cópias

Nas versões do relatório para revisão da Banca Examinadora em TCC1 e TCC2, o aluno deve apresentar na Secretaria da FGA, uma cópia para cada membro da Banca Examinadora.

Após a aprovação em TCC2, o aluno deverá obrigatoriamente apresentar a versão final de seu trabalho à Secretaria da FGA na seguinte forma:

01 cópia encadernada para arquivo na FGA;

01 cópia não encadernada (folhas avulsas) para arquivo na FGA;

01 cópia em CD de todos os arquivos empregados no trabalho;

A cópia em CD deve conter, além do texto, todos os arquivos dos quais se originaram os gráficos (excel, etc.) e figuras (jpg, bmp, gif, etc.) contidos no trabalho. Caso o trabalho tenha gerado códigos fontes e arquivos para aplicações especificas (programas em Fortran, C, Matlab, etc.) estes deverão também ser gravados em CD.

O autor deverá certificar a não ocorrência de "vírus" no CD entregue a secretaria.

Referências

EETEN, M. J. G. v.; BAUER, J. M. OECD Science, Technology and Industry Working Paper, *Economics of Malware: Security Decisions, Incentives and Externalities*. 2008. Disponível em: <https://ideas.repec.org/p/oec/stiaaa/2008-1-en.html>. Citado na página 29.

QIAO, Y. et al. CBM: Free, automatic malware analysis framework using API call sequences. In: SUN, F.; LI, T.; LI, H. (Ed.). *Knowledge Engineering and Management*. Springer Berlin Heidelberg, 2014, (Advances in Intelligent Systems and Computing, 214). p. 225–236. ISBN 978-3-642-37831-7 978-3-642-37832-4. DOI: 10.1007/978-3-642-37832-4_21. Disponível em: <http://link.springer.com/chapter/10.1007/978-3-642-37832-4_21>. Citado 2 vezes nas páginas 29 e 30.

RIECK, K. et al. Automatic analysis of malware behavior using machine learning. *Journal of Computer Security*, v. 19, n. 3, 2011. Citado na página 30.

Apêndices

APÊNDICE A – Primeiro Apêndice

Texto do primeiro apêndice.

APÊNDICE B – Segundo Apêndice

Texto do segundo apêndice.

Anexos

ANEXO A – Primeiro Anexo

Texto do primeiro anexo.

ANEXO B – Segundo Anexo

Texto do segundo anexo.