UNIVERSIDADE FEDERAL DE SANTA CATARINA

João Vitor Maia Neves Cordeiro

Desenvolvimento de uma técnica de esteganografia explorando arquivos binários de código compilado.

Florianópolis,
2022/2

João Vitor Maia Neves Cordeiro

Desenvolvimento de uma técnica de esteganografia explorando arquivos binários de código compilado.

Proposta submetida ao Programa de Graduação em Ciência da Computação para obtenção do Grau de Bacharel.
Orientador: Prof. Jean Everson Martina

Florianópolis,
2022/2

FOLHA DE APROVAÇÃO DE PROPOSTA DO TCC

Acadêmico(s)	João Vitor Maia Neves Cordeiro
Título do Trabalho	Desenvolvimento de uma técnica de esteganografia explorando arquivos binários de código compilado
Curso	Ciências da Computação/INE/UFSC
Área de Concentração	**SEGURANÇA COMPUTACIONAL**

Instruções para preenchimento pelo *ORIENTADOR DO TRABALHO*:

- Para cada critério avaliado, assinale um X na coluna SIM apenas se considerado aprovado. Caso contrário, indique as alterações necessárias na coluna de Observação.

Critérios	**Aprovado**				**Observação**
	Sim	**Parcial**	**Não**	**Não se aplica**	
O trabalho é adequado para um TCC em CCO (relevância / abrangência)?	X				
O título é adequado?	X				
O Tema de pesquisa está claramente descrito?	X				
O problema/hipóteses de pesquisa do trabalho está claramente identificado?	X				
A relevância da pesquisa é justificada?	X				
Os objetivos descrevem completa e claramente o que se pretende alcançar neste trabalho?	X				
É definido o método a ser adotado no trabalho? O método condiz com os objetivos e é adequado para um TCC?	X				
Foi definido um cronograma coerente com o método definido (indicando todas as atividades) e com as datas das entregas (p.ex. Projeto I, II, Defesa)?	X				
Foram identificados custos relativos à execução deste trabalho (se houver)? Haverá financiamento para estes custos?				X	
Foram identificados todos os envolvidos neste trabalho?		X			
As formas de comunicação foram definidas?	X				
Riscos potenciais que podem causar desvios do plano foram identificados?	X				
Caso o TCC envolva a produção de um software ou outro tipo de produto e seja desenvolvido também como uma atividade realizada numa empresa ou laboratório, consta na proposta uma declaração (Anexo 3) de ciência e concordância com a entrega do código fonte e/ou documentação produzidos?				X	

Avaliação	☒ Aprovado	☐ Não Aprovado	
Professor Responsável:	Jean Everson Martina	18/07/2022	
Orientador Externo:			

RESUMO

Em um momento onde a tecnologia evolui tão rapidamente, novas vulnerabilidades em aplicações web são encontradas todos os meses enquanto outras são remediadas, isso faz com que o estado da arte da segurança computacional seja atualizado constantemente. Neste contexto, este trabalho se propõe a desenvolver um ambiente para aprendizado e prática de técnicas de segurança ofensiva em aplicações web com vulnerabilidades de alto risco nos últimos anos.
Palavras-chave: segurança computacional. vulnerabilidades. aplicações web. ambiente para aprendizado.

Sumário

1 Introdução

Esteganografia é a nomenclatura utilizada para um conjunto de técnicas, inicialmente desenvolvidas para espionagem, que atuam escondendo uma informação dentro de uma outra mídia, de modo que a mídia criada após a manipulação seja visualmente indistinguível da original para um observador externo.

A crescente disseminação de computadores a partir do século 20 recuperou diversas técnicas analógicas de espionagem que predatam a área da computação, adicionando o poder de processamento digital para aprimorar a efetividade e as aplicações de tais técnicas. Notóriamente temos a criptografia como a principal expoente dessa transformação, estando presente na maior parte das aplicações modernas. Apesar de ser o maior caso a criptografia não foi a única área a passar por isso, as técnicas de esteganografia também foram afetadas pela introdução do fator digital [**JohnsonJajodia**].

Representações digitais de mídias diversas permitem que a informação seja processada e escondida de formas que uma mídia física por sua natureza não conseguiria suportar, fornecendo uma gama maior de técnicas e aplicações esteganográficas, além de permitir a combinação das técnicas mais modernas de criptografia para esconder uma mensagem cifrada.

As novas técnicas esteganográficas desenvolvidas no meio digital circulam por diversas áreas da computação, nos últimos anos destaca-se o uso no ramo de marcas d'água digitais, onde são inseridas no arquivo informações para ajudar na verificação de sua integridade, autenticidade e dados autorais.

Nessas aplicações, os dados embutidos não são necessariamente secretos portanto a invisibilidade nesses casos é apenas uma propriedade opcional do algoritmo e outros fatores surgem como prioridade. Uma dessas novas características que ganham destaque é a capacidade de resistir a ataques de distorção do arquivo, mantendo os dados embutidos seguros de um atacante que queira danificar a integridade do arquivo fonte.

Apesar de o comum ser utilizar tais técnicas em imagens, áudios e outros formatos de mídia de uso geral, os mesmos conceitos podem ser aplicados para qualquer formato de arquivo que contenha informação, ajustando ou criando novos algoritmos. Uma aplicação pouco explorada atualmente é a esteganografia em arquivos de código binário, que poderia ser utilizada tanto para transmissão de mensagens secretas quanto para marcas d'água digitais em software.

Os softwares que consumimos diariamente muitas vezes são compilados em arquivos binários que codificam as linhas de código fonte em instruções binárias compreendidas pelo processador alvo, esses arquivos compilados também possuem extensões e formatos, sendo os principais formatos utilizados o **ELF (Executable Linkable File)** e **EXE**, sendo que o último possuí diversas versões.

Tendo em vista a atual falta de algoritmos específicos para a utilização de arquivos binários de código fonte como estego-objetos, esse trabalho se propõe a desenvolver um algoritmo que atenda aos requisitos necessários para esse objetivo.

O algoritmo proposto se baseia na modificação do arquivo após a compilação, aproveitando de brechas na arquitetura alvo para codificar informação dentro das próprias instruções do programa, sem alterar a semântica inicial do código em nenhuma forma. Por aplicar as transforações após o processo de compilação, não é necessário a preocupação com a linguagem de programação do código fonte, tampouco com qual o compilador utilizado, melhorando a portabilidade do algoritmo.

Para o escopo inicial dessa implementação, será utilizada como alvo a arquitetura RISC-V por possuir uma especificação aberta e um conjunto de instruções reduzido, entretanto os conceitos utilizados podem ser adaptados para algumas outras arquiteturas seguindo a especificação. Como formato de arquivo, o **ELF** foi escolhido por possuir um conjunto maior e mais acessível de ferramentas para uso e compilação na plataforma RISC-V.

2 Objetivos

2.1 Objetivos gerais

São tidos como objetivos gerais do trabalho o desenvolvimento e validação de uma técnica de esteganografia que utilize arquivos binários de código compilado, especificamente no formato `ELF` como meio de transporte da mensagem, sem alterar a semântica do programa fonte. Além disso, deseja-se que a técnica desenvolvida possa ser expandida para diversas arquiteturas de processadores comercialmente utilizadas.

2.2 Objetivos específicos

Os objetivos específicos que podem ser descritos são os seguintes:

- Documentar possíveis pontos de exploração de esteganografia, sejam eles dependentes ou não da arquitetura alvo da compilação;
- Comparar os pontos documentados a fim de encontrar os mais eficientes;
- Desenvolver uma técnica capaz de explorar os pontos encontrados;
- Implementar a técnica desenvolvida como uma biblioteca de código aberto para uso geral;
- Utilizar a implementação para validar a técnica, buscando métricas de eficiência e escalabilidade do método.

3 Método de Pesquisa

4 Cronograma

Etapas	2022					2023				
	ago	set	out	nov	dez	jan	fev	mar	abr	mai
Desenvolvimento da solução										
Relatório projeto I										
Rascunho projeto II										
Defesa										
Ajustes e envio Final										

5 Recursos Humanos

Nome	Função
João Vitor Maia Neves Cordeiro	Autor
Jean Everson Martina	Orientador
a definir	Membro da banca I
a definir	Membro da banca II

6 Comunicação

O que precisa ser comunicado	Por quem	Para quem	Melhor forma de comunicação	Quando ou com que frequencia
Entrega da proposta de TCC	Autor	Coordenador	Sistema de TCC	única vez
Entrega do relatório I	Autor	Coordenador	Sistema de TCC	única vez
Entrega do relatório II	Autor	Coordenador	Sistema de TCC	única vez
Reuniões com o orientador	Autor	Orientador	Pessoalmente/videochamada	quando necessário

7 Riscos

Risco	Probabilidade	Impacto	Prioridade	Estratégia de resposta	Ações de Prevenção
Perda de dados	baixa	alto	alta	Recuperação da versão mais atual do código	manter o GitHub do projeto atualizado
Alteração do cronograma	baixa	médio	média	redefinição do croonograma	não se aplica