# Defacebot: Uma ferramenta de detecção e notificação de ataques de desfiguração utilizando mecanismos gerenciados por bot de aplicativo de mensagens instantâneas Uma abordagem como ferramenta de apoio ao CSIRT

**Patrícia do Amaral Gurgel M. Gonçalves[1], Felipe R. Coutinho[1], Guilherme D. Jaime[2]**

[1]Diretoria de Segurança da Informação e Governança - SegTIC
Universidade Federal do Rio de Janeiro (UFRJ)
Caixa Postal: 68571 – 21941-972 – Rio de Janeiro – RJ – Brazil

[2]Instituto de Engenharia Nuclear
Caixa Postal 68550 – 21941-972 – Rio de Janeiro – RJ – Brazil

patriciaamaral@tic.ufrj.br, felipecoutinho@tic.ufrj.br, jaime@ien.ufrj.br

***Abstract.*** *The high number of attacks websites has taken, in recent years, government organizations and private institutions to consider defacement attacks as one of the main threats to cybersecurity. These events can generate serious losses, including unavailability of the operation and even damage to the institutional reputation. While many defacement detection techniques have been proposed in recent years, the availability of tools is still scarce. This paper presents the Defacebot tool. Its goal is to detect defacements on monitored websites and immediately notify those responsible through messages sent to their smartphones. To the best of our knowledge, this is the first proposal of a totally free software based tool, integrated with a free notification solution.*

***Resumo.*** *O alto número de ofensivas a websites tem levado, nos últimos anos, organizações governamentais e instituições privadas a considerar os ataques de desfiguração como uma das principais ameaças à segurança da informação. Estes eventos podem gerar sérios prejuízos, incluindo indisponibilidade da operação e até mesmo dano à reputação institucional. Embora muitas técnicas de detecção de desfiguração vêm sendo propostas nos últimos anos, a disponibilização de ferramentas ainda é escassa. Este trabalho apresenta a ferramenta Defacebot. Seu objetivo é detectar desfigurações em websites monitorados, e imediatamente notificar os responsáveis através de mensagens enviadas para seus smartphones. Para o melhor do nosso conhecimento, esta é a primeira proposta de ferramenta totalmente baseada em software livre, integrada a uma solução gratuita de notificação.*

## 1. Introdução

Devido ao crescimento e popularização das tecnologias de informação e comunicação, a utilização da rede mundial de computadores e seus serviços têm sido cada vez maior. A cada dia, mais dispositivos estão conectados à grande rede. Aliado a este crescimento, o número de ataques que exploram este meio também aumentou.

Considerando o crescimento da demanda e do número de ataques, a necessidade e a importância da segurança da informação vêm crescendo na mesma proporção. No Brasil, o CERT.br (2018) - Centro de Estudos, Resposta e Tratamento de Incidentes de Segurança no Brasil, mantido pelo NIC.br do Comitê Gestor da Internet no Brasil, que atende a qualquer rede brasileira conectada à Internet – demonstra, em suas estatísticas, exponencial aumento de incidentes reportados. Em 2018, foram 676.514, 2017 foram 833.755 e em 2016 647.112 notificações.

Dos incidentes de segurança da informação, podemos destacar o ataque de desfiguração de páginas web (deface ou defacement). Tal técnica de ataque consiste na realização de modificações de conteúdo de páginas web. Em geral, o conteúdo original é substituído por conteúdos embaraçosos, imagens perturbadoras, manifestações ofensivas e assinaturas do invasor [Bartoli et al. 2009]. Em alguns casos os ataques são realizados apenas com o intuito de causar indisponibilidade do serviço. De qualquer forma, o resultado é o mesmo: prejuízos financeiros e impacto negativo na reputação da instituição. Ainda segundo o Cert.BR, este tipo de incidente de segurança cresceu 168% em 2018.

A lentidão no processo de identificação e resposta a esse tipo de ataque é um fator desafiador, uma vez que a permanência da desfiguração pode culminar em grandes prejuízos ou até mesmo em processos judiciais.

Conforme apresentado por [Silva Júnior 2017] existem padrões de comportamento de indivíduo/grupos que praticam desfiguração em sites que atuam no Brasil. Segundo o autor é possível identificar os principais indivíduos ou grupos atuantes, os perfis de redes sociais e até mesmo padrões de desfiguração. E justamente devido a identificação desses padrões, é possível o desenvolvimento de mecanismos de detecção automática de desfiguração de páginas.

Nos últimos anos, surgiram na literatura diversas propostas com técnicas e ferramentas para detecção automática de desfiguração de páginas web, cada uma com seu grau de eficiência.

Este trabalho apresenta o Defacebot, uma ferramenta desenvolvida pelo time de respostas a incidentes de segurança computacional (CSIRT - Computer Security Incident Response Team) da universidade, e usada para detectar desfigurações em websites monitorados. A ferramenta é capaz de emitir alertas ao CSIRT através de mensagens enviadas para seus smartphones, usando soluções totalmente baseadas em software livre. Para o envio de alertas foi desenvolvido um módulo que usa a API do Telegram bot.

O fato de ser totalmente baseada em software livre e incluir um serviço gratuito de emissão de alertas são atrativos, pois estão em conformidade com as limitações orçamentárias por quais passam as instituições públicas na atualidade.

Para descrever melhor a ferramenta e suas funcionalidades, organizamos o trabalho da seguinte forma: na seção 2, apresentamos os principais trabalhos relacionados disponíveis na literatura. A seção 3 traz uma breve descrição da ferramenta Defacebot, incluindo sua arquitetura, processo de varredura, técnicas de análise e funcionalidades. Comentários sobre a demonstração planejada e a documentação da ferramenta estão disponíveis na seção 4. A seção 5 traz as considerações finais e, por fim, a seção 6 discute novas funcionalidades a serem implementadas no futuro.

## 2. Trabalhos relacionados

Com o aumento da literatura sobre o tema, é possível encontrar diversas técnicas desenvolvidas para detecção de desfiguração, porém há prevalência das técnicas de comparação por *hash* (função de resumo). De acordo com [Oriyano 2014], em qualquer alteração, por menor que seja, o *hash* será impactado, pois resultará em um *hash* diferente do original.

Já a análise comparativa de páginas HTML consiste em comparar o conteúdo HTML de dois arquivos em momentos diferentes, o arquivo previamente coletado e o arquivo do tempo de execução da ferramenta. Devido ao conteúdo dinâmico dos sites, é definido um limite de aceitação de mudanças. Esta técnica foi apresentada no trabalho [Gurjwar 2013], onde o objetivo era detectar mudanças nos componentes de um website. A solução proposta em tal trabalho armazena em um repositório os componentes da página web e verifica a integridade com as versões originais.

Há ainda a abordagem por similaridade em imagens, de [Jesus and Brandão 2017], onde a detecção de desfiguração de páginas é baseada na análise de imagens, identificando a simetria entre o estado inicial da página web e o estado final.

Além das técnicas, existem diversos trabalhos publicados sobre ferramentas de monitoramento de ataques de desfiguração. O Web Defacement and Intrusion Monitorin (WDIMT) [Masango et al. 2017], por exemplo, detecta a desfiguração e oferece a regeneração do conteúdo original do site após uma série de testes. Uma das técnicas de identificação de ataques é através do *hash*. O Defacebot, ferramenta proposta neste trabalho, não oferece a opção de regeneração do site original como oferecido pelo WDIMT, porém o defacebot oferece uma interface simplificada, tanto para alertas quanto para as demais funcionalidades, utilizando aplicativo de mensagens no smartphone.

Em [Hoang and Nguyen 2019] é proposto um modelo híbrido de detecção de desfiguração baseado na combinação de detecção baseada em aprendizagem de máquina e a detecção baseada em assinatura. Este modelo se diferencia da ferramenta proposta, uma vez que o Defacebot não utiliza aprendizagem de máquina. As análises do HTML, CSS e *hash* são realizadas baseadas em parâmetros predefinidos.

Há ainda uma abordagem de detecção baseada em programação genética (PG), apresentado em [Medvet 2007], onde a PG cria um algoritmo baseado em leituras de página remota, a ser monitorada, e em um conjunto de exemplos de ataques. A partir disso, efetua o monitoramento em intervalos regulares emitindo notificações quando necessário. Esta abordagem se diferencia da proposta neste trabalho pois não é implementado algoritmo genético em sua análise.

As técnicas aplicadas nesta ferramenta foram escolhidas devido à facilidade de implementação e por permitir que os padrões de desfiguração sejam identificados, conforme defendido por [Silva Júnior 2017]. O diferencial da ferramenta proposta é o desenvolvimento totalmente baseado em software livre e a integração com um aplicativo gratuito de mensagens, o que permite a utilização via smartphone.

## 3. Ferramenta

O Defacebot é um protótipo de ferramenta desenvolvida, implementada e utilizada pelo CSIRT da universidade para monitorar sites previamente cadastrados e hospedados nos

servidores da universidade, a fim de identificar ataques de desfiguração e emitir alertas à equipe de segurança. Além do monitoramento e alerta, o protótipo também provê funcionalidades para suporte ao time de segurança, detalhadas na seção Funcionalidades.

O processo de monitoramento é realizado de duas maneiras: análise e comparação dos sites da instituição (site x registro de estado do site no momento do cadastramento ou atualização) e varredura por publicação, conforme detalhado abaixo:

- Varredura de sites: São analisadas alterações nos arquivos CSS e HTML (*hash* linhas de código) e identificação de palavras-chaves frequentemente utilizadas por atacantes ao efetuar a desfiguração.
- Varredura de publicação: Varredura em sites onde são publicados domínios que foram atacados. Quando encontrada a publicação relacionada ao site monitorado, uma notificação é enviada para os analistas.

Caso um site apresente características que tenha sofrido um ataque de desfiguração, uma mensagem de notificação é enviada ao bot, como por exemplo: "O site: xxx.universidade.br pode ter sofrido uma desfiguração."

A utilização da ferramenta exige cadastro para que o analista possa receber notificações e utilizar todas as funcionalidades. Assim como, para que um site seja monitorado, é necessário o cadastramento prévio, o qual pode ser realizado através do próprio bot.

### 3.1. Arquitetura

O protótipo foi desenvolvido utilizando linguagem PHP e utiliza um sistema gerenciador de banco de dados MySQL. A integração com a plataforma de mensagens instantâneas (Telegram) é provida por uma API disponibilizada por seus próprios desenvolvedores. A escolha de tais tecnologias foi devido ao fato de serem software livre e com vasta documentação técnica.

### 3.2. Execução da varredura

A análise do site pode ocorrer de duas formas: automatizada, através de agendamento para execução da verificação, ou através da utilização de funcionalidades disponíveis na interface do aplicativo de mensagens. Assim, todos os sites cadastrados são analisados, e seus conteúdos atuais comparados com o conteúdo gravado previamente.

#### 3.2.1. Técnicas da análise

O processo de análise possui várias técnicas para avaliar se um site foi desfigurado. O objetivo da análise é identificar indícios de ocorrência de ataque de desfiguração. Os métodos utilizados são: *Hash*, alteração de código HTML, busca por palavra, alteração de CSS e busca em sites externos, conforme detalhado na Tabela 1.

| Método | Descrição |
|---|---|
| **Hash** | Comparação de *hash* atual com o coletado no cadastro |
| **Alteração** | Verifica se houve alteração maior que 30% do conteúdo do site, comparando site atual e amostra coletada no cadastro |
| **String** | Busca por strings do dicionário do sistema que caracterizam possíveis ataques |
| **CSS** | Alterações em mais de 80% do arquivo CSS do estado atual e do coletado no cadastro |
| **Externa** | Verificações em sites como Zone-H e DefacerID onde é comum os atacantes publicarem os sites desfigurados. |

**Tabela 1. Métodos de varredura utilizado pelo protótipo.**

Para cada análise realizada pela aplicação, os métodos descritos na tabela 1 são executados. Os processos de análise estão exemplificados na Figura **??**.

### 3.3. Funcionalidades

Através da interface do Telegram Bot API, foram disponibilizadas as seguintes funcionalidades, conforme a Tabela 2.

| Comando | Descrição |
|---|---|
| /init senha | Inicia sessão, adicionando novo usuário ao banco de dados |
| /add url | Adiciona URL para ser monitorado |
| /log | Retorna os últimos 20 incidentes |
| /check | Verifica os sites cadastrados |
| /update | Atualiza as páginas no sistema |
| /leave | Termina a sessão, removendo o usuário do sistema |
| /blacklist | Adicionar uma nova palavra na blacklist |
| /remove | Remove uma URL do sistema |
| /status | Retorna o status do serviço |

**Tabela 2. Funcionalidades da ferramenta.**

## 4. Descrição da demonstração planejada

A demonstração da ferramenta será disponibilizada aos participantes utilizando a ferramenta com acesso online no local do evento.

- Ferramenta disponível em: https://defacebot.tic.universidade.br
- Documentação da ferramenta disponível em: https://defacebot.tic.universidade.br/wikis/home

## 5. Conclusão

Os ataques de desfiguração vêm se popularizando pelo fato de serem fáceis de implementar e também pela potencialização da exposição dos nomes dos atacantes, uma vez que,

geralmente, ataques a sites de instituições públicas, têm grande visibilidade no mundo virtual. Tais ataques, por sua vez, consistem em uma ameaça à imagem dessas instituições, afinal, se os ataques se tornarem muito prevalentes, a imagem da instituição pode ser gravemente degradada a ponto de se tornar impossível voltar a se comunicar utilizando web sites para tal finalidade. Assim, é fundamental o reconhecimento da prevalência destes ataques, e o domínio das formas de identificar e mitigá-los.

Neste trabalho apresentamos um protótipo que auxilia a detecção, monitoramento e notificação deste tipo de ataque. Em seguida, indicamos as principais técnicas utilizadas para esta finalidade.

Apresentamos também, o modelo de fluxo do processo de análise que a ferramenta utiliza como base para executar as varreduras. Finalmente, indicamos também formas utilizar a ferramenta e suas funcionalidades.

Identificamos as detecções de falso positivo como um ponto controverso. Grandes atualizações nos sites monitorados podem resultar em identificações como um incidente de desfiguração.

Acreditamos que este trabalho abre portas para inúmeros desenvolvimentos futuros. A utilização do aprendizado de máquinas aliadas às técnicas dispostas neste trabalho, sugerem um considerável avanço e que pode auxiliar muito, em especial as instituições governamentais, a mitigar esse tipo de ataque.

## 6. Trabalhos futuros

Novas funcionalidades estão sendo testadas. A versão atual aponta necessidade de redução de alertas de falso positivo, a qual pode ser reduzida com a implementação de técnicas, como por exemplo, a abordagem de detecção de desfiguração similaridade por imagens, apresentada no trabalho de [Jesus and Brandão 2017], a implementação de técnica que permita a ferramenta automatizar a recuperação um site desfigurado, e, por fim, a utilização de técnicas de aprendizado de máquina no aperfeiçoamento das análises.

## Referências


Bartoli, A., Davanzo, G., and Medvet, E. (2009). The reaction time to web site defacements. *IEEE Internet Computing*, 13:52–58.

Gurjwar, Rajiv Kumar, S. D. R. T. D. S. (2013). An approach to reveal website defacement. *Computers*, 11(6):73.

Hoang, X. D. and Nguyen, N. T. (2019). Detecting website defacements based on machine learning techniques and attack signatures. *Computers*, 8(2):35.

Jesus, H. M. d. and Brandão, J. E. M. d. S. (2017). Detecção de desfiguração de sites por similaridade em imagens. *XVII SBSeg*, page 8.

Masango, M., Mouton, F., Antony, P., and Mangoale, B. (2017). Web defacement and intrusion monitoring tool: Wdimt. In *2017 International Conference on Cyberworlds (CW)*, pages 72–79. IEEE.

Medvet, Eric ; Fillon, C. . A. B. (2007). Detection of web defacements by means of genetic programming. *IEEE Internet Computing*.

Oriyano, J. (2014). Ceh v8: Certified ethical hacker version 8 study guide. In *CEH v8: Certified Ethical Hacker Version 8 Study Guide*. Wiley & Sons, USA.

Silva Júnior, N. V. d. (2017). Estudo e análise dos grupos hackers que realizam desfiguração de páginas web no brasil. B.S. thesis, Universidade Tecnológica Federal do Paraná.