Segurança da Informação

OpenSSL e Dicionário

Arthur do Prado Labaki - 11821BCC017

19-05, 2023

GBC083

Exercício

Vários arquivos foram cifrados usando o SSL. Sabe-se que o individuo que fez isso não é muito cuidadoso e é um super fâ do livro 1984 (limite a sua busca por palavras ao Wikipedia em portugues sobre o livro)

O processo de ciframento usado foi esse:

openssl enc -aes-256-cbc -pbkdf2 -salt -in file.txt -out file0.enc -pass pass:teste

Voce deve usar um ataque de dicionario para recuperar as senhas usadas no ciframento dos arquivos 1 a 15.

O arquivo 0 foi cifrado com a senha teste...para voce testar. Voce deve decifrar até 10 arquivos para nota integral.

A entrega deve ser a lista das senhas encontradas e uma descrição breve de como voce as obteve. (manualmente? script...incluir? alguma ferramenta...qual?)

O objetivo deste exercicio é um primeiro contato com o OpenSSL e mostrar que senhas ruins - sejam elas de ciframento ou de login - podem ser obtidas de forma "facil"por um adversário.

Informações importantes

Esse trabalho foi separado em duas partes, uma antes do dia 11 e outra depois do dia 11, quando foi dado mais alguns arquivos e tempo para resolver o exercício.

Todos os códigos e dicionários criados estão no meu repositório disponível no link abaixo.
Link do Repositório

Vale mostrar que, ao executar a descriptografia do arquivo, pode existir 4 tipos de saídas:

1. bad decrypt 140737348081472:error:06065064:digital envelope ... (senha invalida)
2. 'Facecrime', 'file6.enc', b'E¡q[\x16\xe7\xbdJ\x1az\xde3\xfeL\xc8' (senha invalida)
3. 'M1n1$try', 'file4.enc', b" (senha com caractere invalido)
4. 'teste', 'file0.enc', b'teste \n' (senha correta - senha, arquivo, conteúdo)

Primeira metade do trabalho

Para gerar o dicionário inicial, foi utilizado 3 métodos:

- Copiar todo o site da Wikipédia e usar uma função para separa-lo por palavras (muitas palavras repetidas);
- Escolher manualmente palavras relacionadas ao livro do site;
- Usar a ferramenta Cewl para obter as palavras do site.

Apos obter esses dicionários, foi criado um código base para testar a descriptografia dos arquivos. Esse script se resume em tentar a descriptografia de um arquivo usando uma lista de senhas e salvar sua descriptografia em um arquivo (salva somente se ocorreu a descriptografia).

Também foi criado um arquivo contendo funções para criar variações do dicionário. Essas funções são:

- Todas as letras em minúsculo;
- Todas as letras em maiúsculo;
- Todas as palavras capitalize (só a primeira letra em maiúsculo);
- Palavras invertidas;
- Palavras com letras maiúsculas e minusculas alternadas;
- Palavras com letras minusculas e maiúsculas alternadas;
- Transformar palavras pra Leet (A vira 4, E vira 3, ...).

Como essas tentativas não obtiveram bons resultados (somente 3 senhas), foi utilizado uma nova estrategia. Essa estrategia consiste em criar outro dicionário manualmente nomes ou frases que possam ter algum sentido, como nome de personagens, frases usadas, locais, entre outros. Com a adição desse novo dicionário, foi necessário criar mais algumas variações, complementado as antigas, que são:

- Remover espaço das frases;
- Substituir espaço por underline;
- Substituir espaço por underline e letra maiúscula;
- Transformar frase em CamelCase (múltiplas palavras começando com letra maiúscula);
- Remover acentos e ç;
- Remover caracteres especiais (como ",", "' ", ":", -")
- Concatenar 2 palavras;
- Gerar 1 letra em maiúsculo da palavra (como "Teste", "tEste", "teSte", ...);
- Outras versões de Leet;

Vale lembrar que cada uma função pode ser usada junto com outra, gerando outras variações de dicionários. Com tudo isso, foi gerado mais de 10 dicionários diferentes, contendo no maior deles 61 mil palavras. Como o ambiente usado foi uma máquina virtual Kali Linux, cada execução do dicionário demora algumas horas, sendo que esse maior demorou 9 horas para ser concluído.

Segunda metade do trabalho

Apos adiar a data em 1 semana e adicionar mais arquivos para serem descriptografados, foi necessário realizar novamente as buscas, porem por conta de um problema com meu hd externo (ele tinha parado de funcionar), foi necessário refazer as funções. Ao invés de refazer as funções já criadas, foi feito o indicado pelo professor:

- Duas palavras simples concatenadas;
- Maiúsculas (primeira letra de das palavras, ultima letra, qualquer letra...);
- Substituição gráfica (a vira @ e similares);
- Substituição fonética (to (em inglês) vira 2 (two)... e similares).

Com essas mudanças, foram criados outros diversos dicionários, todos derivados de um que contem algumas palavras do livro em minusculo. Porem em uma das funções, foi criado um dicionário que contem todas as variações possíveis de maiúsculo em uma palavra (Casa, CAsa, CaSa...) e com isso, resultou em um dicionário inviável de caracteres (14.151.665 palavras). Então a variação descrita maiúsculas, sera realizada somente com as 1 letra maiúscula por palavra (Casa, cAsa, caSa, casA).

Após todas essas exaustivas buscas nos 24 arquivos, foram obtidos 10 senhas:

1. file11 - guerraepaz
2. file12 - novilingua
3. file14 - duplipensar
4. file15 - crimideia
5. file16 - Winston
6. file18 - grande1rmao
7. file19 - minip@z
8. file20 - miniamo
9. file21 - Bempensante
10. file23 - miniveR

Observação: Meu HD voltou a funcionar, então vou colocar o código completo do projeto (ambas as partes) no repositório.