PROPOSTA DE FLUXO DE ATAQUES POR CANAIS LATERAIS BASEADO EM APRENDIZADO PROFUNDO

RODRIGO NUEVO LELLIS[1]; GUILHERME PERIN[2]; RAFAEL IANKOWSKI SOARES[3]

[1]*Universidade Federal de Pelotas – UFPel – rn.lellis@inf.ufpel.edu.br*
[2]*Delft University of Technology – G.Perin@tudelft.nl*
[3]*Universidade Federal de Pelotas – UFPel – rafael.soares@inf.ufpel.edu.br*

1. INTRODUÇÃO

Ataques por canais laterais, ou *Side Channel Attacks* - SCA (KOCHER, P. C., 1996), exploram vazamentos físicos de informações não intencionais de dispositivos criptográficos. Consumo de potência e radiação eletromagnética são exemplos mais comuns de canais laterais explorados, resultando em ataques diferenciais por análise de consumo de potência (*Differential Power Analysis* – DPA) (KOCHER, P.; JAFFE; JUN, 1999) e a radiação eletromagnética (*Differential Electromagnetic Analysis* – DEMA) (AGRAWAL *et al.*, 2002), respectivamente. O objetivo de um ataque por canal lateral é a obtenção de informações confidenciais, como chaves criptográficas.

Contudo, para que os ataques tenham sucesso, são necessários milhares (ou mesmo milhões) de traços de canais laterais. Uma etapa comum após medições é o alinhamento temporal dos traços. Essa fragilidade foi explorada para a criação de diversas proteções, chamadas de contramedidas (BOEY *et al.*, 2010). Entretanto, essas contramedidas podem ser neutralizadas através de etapas de pré-processamento no fluxo dos ataques (LE *et al.*, 2007).

Esses ataques baseiam-se em modelos de consumo estáticos como *Hamming Distance* – HD ou *Hamming Weight* – HW e realizam comparações com os traços por meio da diferença das médias. Analogamente, o CPA (*Correlation Power Analysis*) (BRIER; CLAVIER; OLIVIER, 2004) realiza os mesmos passos de DPA/DEMA, com a diferença de que a comparação entre os traços e os modelos de consumo é realizada através do coeficiente de correlação de *Pearson*. Os ataques mencionados fazem parte de uma categoria chamada *non-profiled*.

Por outro lado, os *Template Attacks* (TA) (CHARI; RAO; ROHATGI, 2002) necessitam de um dispositivo idêntico ao dispositivo atacado para criar um modelo ou perfil, que será usado nos ataques. Por isso, esse ataque é classificado como *profiled attack*. Aqui, os traços medidos são comparados ao modelo gerado através de parâmetros estatísticos como a média e a covariância. Contudo, todos estes ataques são determinísticos, ou seja, utiliza-se as mesmas variáveis em todas as etapas. Embora o TA construa um modelo mais ajustado aos traços medidos, este modelo baseia-se em valores estatísticos que dependem de técnicas refinadas de seleção de amostras com maior vazamento ou pontos de interesse. Caso esta etapa seja limitada, o ataque por templates torna-se ineficiente.

Algoritmos de *Machine Learning* – ML e *Deep Learning* – DL, têm sido aplicados às mais diferentes áreas dentro da computação, incluindo SCA. Assim, ML e DL foram recentemente empregadas em *profiled attacks* em diversos trabalhos encontrados na literatura (HETTWER; GEHRER; GÜNEYSU, 2020). Atualmente existem também trabalhos que atacam dispositivos dotados de contramedidas de desalinhamento temporal e proteção por mascaramento, como (PROUFF *et al.*, 2018). Vale destacar, que somente as redes neurais

convolucionais (*Convolutional Neural Networks* – CNNs) são mais eficazes na realização de ataques contra dispositivos dotados de contramedidas temporais. Apesar da possibilidade de atacar dispositivos protegidos, todos os trabalhos que envolvem ML e DL aplicados à SCA possuem um ponto em comum: o esforço computacional excessivo para a realização do ataque. Dos trabalhos encontrados na literatura, alguns relatam a duração de semanas para esse processo, o que dificulta muito a realização do ataque reais na maioria dos casos.

Métodos para reduzir o tamanho das redes neurais, como poda (*pruning*) são encontrados na literatura nas mais diversas áreas (KNIGHT; LEE, 2021). Embora apresente benefícios que podem ser diretamente relacionados à eficiência dos ataques, a aplicação desta técnica no contexto de SCA foi muito pouco explorada. Em (PERIN; WU; PICEK, 2021), os autores mencionam que existem muitas possibilidades de implementar a técnica de poda em redes neurais. Dessa forma, essa técnica pode ser bastante explorada para sua aplicação aos SCAs.

Destaca-se ainda, que técnicas de pré-processamento como em (LE *et al.*, 2007) diminuem a quantidade de amostras dos traços, o que reduz o esforço computacional da rede neural que realiza o ataque de potência.

Com base no exposto, o presente trabalho apresenta como proposta o desenvolvimento de um fluxo de ataques por canais laterais baseado em redes neurais, com esforço computacional reduzido.

2. METODOLOGIA

Com o intuito de desenvolver o fluxo de ataques proposto, primeiramente serão realizados experimentos aplicando-se diferentes técnicas de processamento de sinais aos traços do consumo, antes de inseri-los como entrada de uma rede neural (CNN) consolidada, como a apresentada em (PROUFF *et al.*, 2018). Com isto, analisar-se-á o impacto de cada uma das técnicas empregadas no tempo de treinamento da rede, escolhendo-se a que obtiver o melhor resultado. Essa etapa de Pré-Processamento é mostrada no fluxograma da Figura 1 (esquerda).

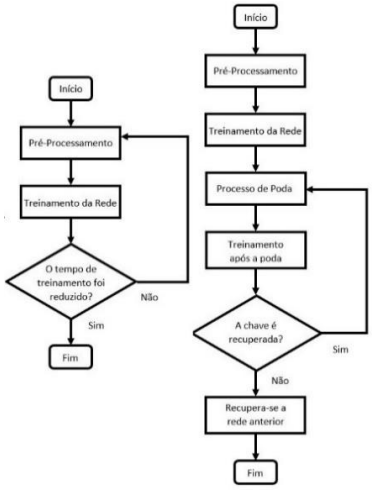

Figura 1 - Fluxograma de experimentos para a etapa de Pré-Processamento (esquerda) e poda (direita). Fonte: Própria.

Em seguida, aplica-se o procedimento iterativo de poda à CNN e realiza-se o treinamento. A partir disso, analisa-se o resultado obtido, a fim de verificar se a rede resultante foi capaz de revelar a chave secreta. Se o resultado for positivo, aplica-se novamente o procedimento de poda, removendo-se mais elementos da rede. Esse processo continua até que a chave criptográfica não seja mais revelada.

Então resgata-se a rede anterior, ou seja, a menor rede que foi capaz de obter sucesso no ataque. O fluxograma dessa etapa é mostrado na Figura 1 (direita).

3. RESULTADOS E DISCUSSÃO

Até o presente momento, foram realizados experimentos com o intuito de fazer uma prova de conceito sobre os métodos de ML e DL aplicados aos SCAs. Para isto, foram utilizados os traços do *dataset* ASCAD (PROUFF *et al.*, 2018). Foram realizados experimentos com as redes MLP_{best} e CNN_{best} apresentadas no trabalho de Prouff et. al.

O primeiro experimento consistiu em verificar a capacidade da rede MLP_{best} de realizar um ataque. Utilizou-se um total de 50000 traços para a fase de treinamento com rótulos (ou classes) calculadas a partir do consumo hipotético para a chave correta. Para isto, foram utilizados traços oriundos de uma execução do algoritmo AES (*Advanced Encryprion Standard*) em um dispositivo desprotegido. O resultado é mostrado na Figura 2 (esquerda). O termo *rank* indica a posição da chave dentro de um vetor contendo todas as hipóteses de chave. Se o *rank* for 0, significa que a chave correta é recuperada com sucesso.

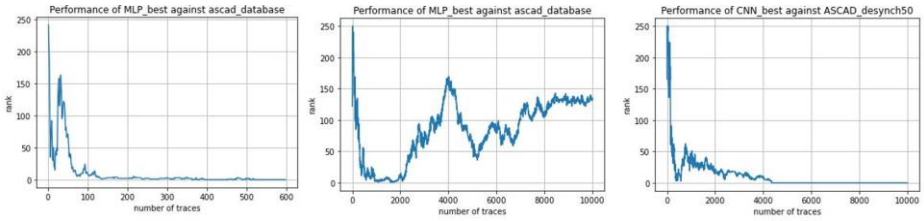

Figura 2 - *Ranking* da chave secreta para traços alinhados MLP_{best} (esquerda) e traços desalinhados de 50 amostras MLP_{best} (centro) e CNN_{best} (direita): Própria.

Como podemos observar a partir da Figura 2 (esquerda), com um pouco mais de 100 traços o ataque foi bem-sucedido (*ranking* 0 da chave secreta).

A seguir, a rede MLP_{best} foi testada sob um conjunto de traços desalinhados de até 50 amostras. O resultado é mostrado na Figura 2 (centro), a qual mostra que o ataque não foi bem-sucedido. O *ranking* da chave ficou entre 100 e 150.

Esse experimento confirmou a necessidade de uma rede mais complexa, como uma CNN, para realizar-se um ataque em um dispositivo dotado de contramedida temporal. Isso foi feito no experimento seguinte, ou seja, aplicou-se a rede CNN_{best} aos traços desalinhados. O resultado pode ser visto na Figura 2 (direita), a qual confirma a capacidade da rede CNN de obter sucesso no ataque, mesmo sobre dispositivos dotados de contramedida temporal. Entretanto, deve-se levar em conta que o tempo de treinamento da MLP_{best} foi de 4 minutos e 51 segundos, enquanto para a CNN_{best} o treinamento levou notáveis 29 horas e 56 minutos em um computador pessoal sem GPUs. Isto nos faz concluir que a eficiência das CNNs está atrelada a um custo computacional elevadíssimo, o que motiva a busca por um fluxo de ataques que possua um esforço computacional menos expressivo.

4. CONCLUSÕES

SCAs constituem uma ameaça real aos dispositivos criptográficos. Ataques clássicos como DPA, DEMA, CPA e TA baseiam-se em modelos de consumo estáticos e ferramentas matemáticas que resultam em processos determinísticos, o que impossibilita ajustes mais precisos em relação aos traços medidos.

O avanço dos métodos de inteligência artificial trouxe um novo horizonte para os SCAs, criando modelos mais compatíveis com esta aplicação resultando em uma maior eficiência. No entanto, o esforço computacional é excessivo ao se realizar ataques em dispositivos protegidos com contramedidas temporais sobre essa abordagem. Assim, a busca por um fluxo de ataques que apresente um esforço computacional reduzido, mantendo-se a eficiência dos ataques é bastante relevante para esta área. Reduzir esforços computacionais em ataques é importante na avaliação de implementações criptográficas, acelerando o processo de melhorias de segurança com contramedidas eficazes.

5. REFERÊNCIAS BIBLIOGRÁFICAS

AGRAWAL, Dakshi *et al.* The EM Side–Channel(s):Attacks and Assessment Methodologies. **Cryptographic Hardware and Embedded Systems - CHES 2002**, [*s. l.*], v. 6917, p. 29–45, 2002.

BOEY, Kean Hong *et al.* Random clock against differential power analysis. **IEEE Asia-Pacific Conference on Circuits and Systems, Proceedings, APCCAS**, [*s. l.*], p. 756–759, 2010.

BRIER, Eric; CLAVIER, Christophe; OLIVIER, Francis. Correlation power analysis with a leakage model. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, [*s. l.*], v. 3156, p. 16–29, 2004.

CHARI, Suren; RAO, Josyula R.; ROHATGI, Pankaj. Template Attacks. **4th International Workshop on Cryptographic Hardware and Embedded Systems (CHES'02)**, [*s. l.*], p. 13–28, 2002.

HETTWER, Benjamin; GEHRER, Stefan; GÜNEYSU, Tim. Applications of machine learning techniques in side-channel attacks: a survey. **Journal of Cryptographic Engineering**, [*s. l.*], v. 10, n. 2, p. 135–162, 2020.

KNIGHT, Autumn; LEE, Byeong Kil. Performance Analysis of Network Pruning for Deep Learning based Age-Gender Estimation. [*s. l.*], p. 1684–1687, 2021.

KOCHER, Paul C. Timing Attacks on Implementations of Diffie-Hellman. **CRYPTO - Annual International Cryptology Conference**, [*s. l.*], p. 104–113, 1996.

KOCHER, Paul; JAFFE, Joshua; JUN, Benjamin. Differential Power Analysis. **Wiener M. (eds) Advances in Cryptology — CRYPTO' 99. CRYPTO 1999. Lecture Notes in Computer Science, vol 1666. Springer, Berlin, Heidelberg**, [*s. l.*], p. 336–338, 1999.

LE, Thanh Ha *et al.* Efficient solution for misalignment of signal in side channel analysis. **ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings**, [*s. l.*], v. 2, p. 257–260, 2007.

PERIN, Guilherme; WU, Lichao; PICEK, Stjepan. Gambling for Success : The Lottery Ticket Hypothesis in Deep Learning-based SCA. [*s. l.*], p. 1–29, 2021.

PROUFF, Emmanuel *et al.* Study of Deep Learning Techniques for Side-Channel Analysis and Introduction to ASCAD Database. **CoRR**, [*s. l.*], p. 1–45, 2018.

VAN DER VALK, Daan *et al.* Learning from a big brother - Mimicking neural networks in profiled side-channel analysis. **Proceedings - Design Automation Conference**, [*s. l.*], v. 2020-July, 2020.