Acoustic Side Channel Attack Against DNA Synthesis Machines: Poster Abstract

Sina Faezi[*], Sujit Rokka Chhetri[*], Arnav Vaibhav Malawade[*], John Charles Chaput[*], William Grover[†],
Philip Brisk[†], and Mohammad Abdullah Al Faruque[*]
[*]University of California, Irvine, Email: {sfaezi, schhetri, malawada, john.chaput, alfaruqu}@uci.edu
[†]University of California, Riverside, Email: wgrover@engr.ucr.edu, philip@cs.ucr.edu

***Abstract*—Synthetic DNA molecules play an essential role in genomics research and are a promising, high-capacity data storage medium. Currently, researchers use automated DNA synthesizers to custom-build sequences of oligonucleotides (short DNA strands) using the nucleobases: Adenine (A), Guanine (G), Cytosine (C), and Thymine (T). Research laboratories invest large amounts of capital to engineer unique oligonucleotide sequences. In our work, we demonstrate the vulnerability of commonly used DNA synthesizers to acoustic side-channel attacks, where confidentiality can be breached to steal precious DNA sequences. We introduce a novel methodology to reverse engineer the acoustic noise generated by the DNA synthesizer and extract the type and order of the nucleobases delivered to the output. To the best of our knowledge, this is the first work which highlights the possibility of physical-to-cyber attacks in DNA synthesis technologies.**

***Index Terms*—DNA synthesise,cyber-physical systems, side-channel, confidentiality, statistical modeling**

I. Introduction

In this modern era of synthetic biology, researchers are able to accurately create custom oligonucleotides. Oligonucleotides are short custom-made single or double-stranded Deoxyribonucleic acid (DNA)or Ribonucleic acid (RNA) molecules. Ever since scientists first crafted methods to artificially synthesize these molecules in a laboratory in the 1950s, there has been a boom in the field of medicine, gene-editing, etc., to name a few. This has led to the production of efficient DNA synthesis machines which are fully automated and can synthesize a large number of oligonucleotides. These molecules are eventually used for research and development of new drugs for preventing diseases, creating new biological agents, new species of disease resistant plants, etc. In fact, with the advance of this technology, the global market for synthetic biology is expected to reach $18.9 billion by 2024 [1].

Although the cost of synthetic DNA production is continuously dropping, the total cost of engineering a sequence of bases in DNA that has certain characteristics is still high. Tremendous amounts of laboratory research and repeated expensive experiments are needed to assure the correct functionality of a genetically engineered organism. Investors in this area start gaining financial benefits only after the engineered organism passes all required tests from corresponding official administrations (such as the Food and Drug Administration) and an official notice of intellectual property ownership for that particular sequence is issued in the form of a patent or copyright. However, before the sequence is officially recognized, it is vulnerable to theft by competitors which can result in huge financial losses. Additionally, once DNA becomes a common data storage medium, the assurance of complete privacy during data transfer will become a top priority.

Countries implement regulations to assure the safety of DNA synthesis technologies usage (physical domain) [2], [3] while the cyber security communities address the concerns related to the computation core of these systems (cyber domain) [4]. However, the integration of cyber and physical components may create a surface for a new cross-domain set of possible attacks [5]. Hence, one of the challenges for securing modern DNA synthesizers is being able to recognize of such threats. To this end, in this work, we investigate the possibility of using the acoustic side-channel for stealing the order of nucleobases in a custom-built DNA molecule while it is being synthesized.

II. Attack Model

In our attack model, the attacker is a disgruntled employee or a visitor who can place an audio recorder in close proximity to the target DNA synthesizer. The attacker passes the recorded signal through the attack process presented in Figure 1 to determine the order and type of nucleobases delivered to the output of the machine. The recorded noise is first preprocessed to remove the ambient noise and is segmented to individual base delivery sections (based on the timing information provided in the DNA synthesizer user manual and local maxima in the signal). Then, various frequency and time domain features are extracted from each section of the signal. During the training phase, the extracted features are mapped to the training nucleobases and fed to the classifiers. In this phase, the classifiers learn the relationship between the extracted features and the corresponding delivered nucleobases. In the attack phase, the model predicts the type of nucleobase delivered at each section, and is ultimately able to determine the whole structure of the synthesized DNA sequence. Furthermore, we suggest that the attacker may use a post-prcessing step to eliminate bad predictions from the results and further improve the accuracy of the attack methodology.

This research is supported in part by National Science Foundation awards CMMI-1739503 and CMMI-1763795.

DOI 10.1109/ICCPS48487.2020.00026

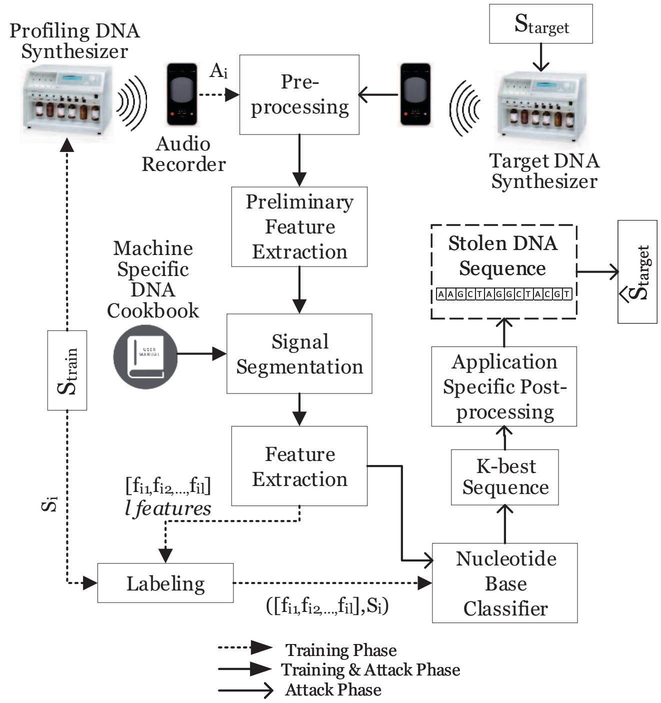


Fig. 1: Side-channel attack methodology on DNA synthesizers.

III. Experimental Setup and Summery of Results

As shown in Figure 2, our experimental setup consists of an AB 3400 DNA synthesizer, and a Zoom H6 audio recorder placed in close proximity to the DNA synthesizer.

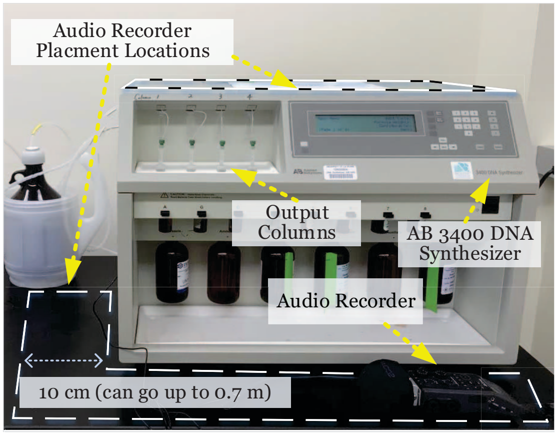


Fig. 2: Experimental setup.

We use various learning algorithms for the classification task. Figure 3 represents the classification accuracy based on number of training samples used for training each type of classifier. As the results in this figure shows, Neural Networks (NN, two 1D convolutional layers + one dense layer) and the voting classifier (voting based on confidence of NN and random forest) can both capture the relationship between the nucleobase types and the extracted features with about 180 training samples and result in %87.51 and %88.07 accuracy on average.

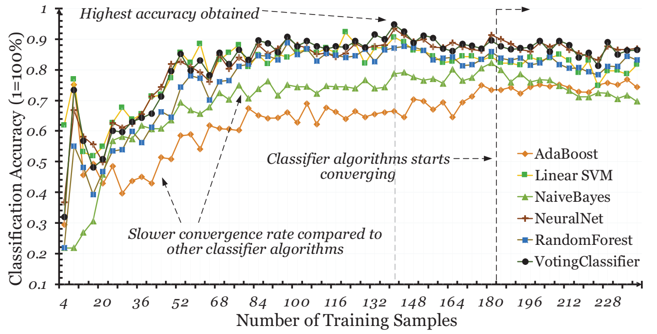
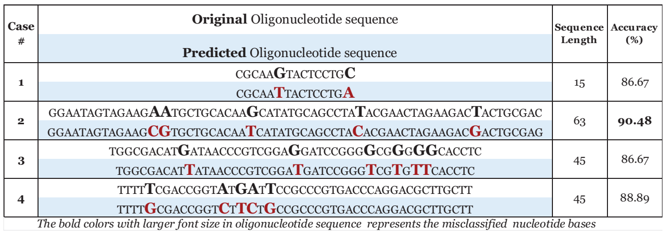


Fig. 3: Learning curve of various classifiers for nucleotide base classification

To validate our proposed attack methodology further, we asked the target DNA synthesizer operator to synthesize four oligonucleotide sequences without giving us any clues about them. As Figure 4 shows, our attack methodology was able to predict the order and type of nucleobase with high accuracy which matches with our estimated accuracy.



Case #	**Original** Oligonucleotide sequence **Predicted** Oligonucleotide sequence	Sequence Length	Accuracy (%)
1	CGCAA**G**TACTCCTG**C** CGCAA**T**TACTCCTG**A**	15	86.67
2	GGAATAGTAGAAG**AA**TGCTGCACAA**G**CATATGCAGCCTA**T**ACGAACTAGAAGAC**T**ACTGCGAC GGAATAGTAGAAG**CG**TGCTGCACAA**T**CATATGCAGCCTA**C**ACGAACTAGAAGAC**G**ACTGCGAG	63	**90.48**
3	TGGCGACAT**G**ATAACCCGTCGGA**G**GATCCGGG**G**CG**G**G**GG**CACCTC TGGCGACAT**T**ATAACCCGTCGGA**T**GATCCGGG**T**CG**T**G**TT**CACCTC	45	86.67
4	TTTT**T**CGACCGGT**A**T**GA**T**T**CCGCCCGTGACCCAGGACGCTTGCTT TTTT**G**CGACCGGT**C**T**TC**T**G**CCGCCCGTGACCCAGGACGCTTGCTT	45	88.89

The bold colors with larger font size in oligonucleotide sequence represents the misclassified nucleotide bases

Fig. 4: Results for reconstructing the test cases.

IV. Poster Presentation

The poster presents more details on the proposed attack methodology and discusses possible post-processing options. It also provides further results to evaluate the robustness of the attack methodology against various factors such as ambient noise and the distance of the microphone from the target DNA synthesizer.

References

[1] J. Bergin. (2020) Synthetic biology: Global markets. [Online]. Available: https://www.bccresearch.com/market-research/biotechnology/synthetic-biology-global-markets.html

[2] C. Sorenson and M. Drummond, “Improving medical device regulation: the united states and europe in perspective,” *The Milbank Quarterly*, vol. 92, no. 1, pp. 114–150, 2014.

[3] *Screening framework guidance for providers of synthetic double-stranded DNA*. US Department of Health and Human Services, 2010.

[4] P. A. Williams and A. J. Woodward, “Cybersecurity vulnerabilities in medical devices: a complex environment and multifaceted problem,” *Medical Devices (Auckland, NZ)*, vol. 8, p. 305, 2015.

[5] A. Faruque, M. Abdullah, S. R. Chhetri, A. Canedo, and J. Wan, “Acoustic side-channel attacks on additive manufacturing systems,” in *Proceedings of the 7th International Conference on Cyber-Physical Systems*. IEEE Press, 2016, p. 19.