

Oligo-Snoop: A Non-Invasive Side Channel Attack Against DNA Synthesis Machines

The Network and Distributed System Security Symposium (NDSS' 2019)

Sina Faezi¹, Sujit Rokka Chhetri¹, Arnav Vaibhav Malawade¹, John Chaput¹, William Grover², Philip Brisk², Mohammad Al Faruque¹

¹ University of California, Irvine

² University of California, Riverside



Deoxyribonucleic Acid (DNA)

➤ DNA molecules

- ❑ Carry biological information
- ❑ Chains of **A**, **G**, **C**, and **T** bases.



- ## ➤ DNA synthesizers are commercially available to produce custom DNA sequences.



Current Security State

➤ Bioterrorism

- Authorities regulation
- Monitoring
- Training



superstarfloraluk.com



Related Work: Other Attacks

- Software vulnerabilities analysis [1]

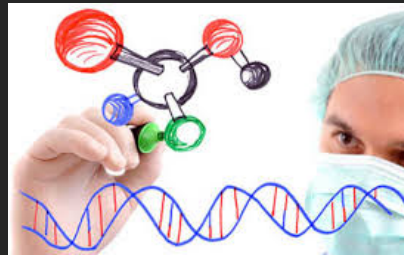


Motivation: A Different Perspective

- Synthetic DNA market \$38.7 billion by 2020 [2].



Drug
Discovery



Medical
Treatment



Crop
Optimization

AGGCAGGTTTTCTAGCTGGAACTCCGA

**Synthetic DNA Sequences are
Intellectual Properties.**

IP (\$\$\$)



Our Contribution

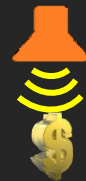
- Novel acoustic side-channel attack

 - ❑ Each base type prediction

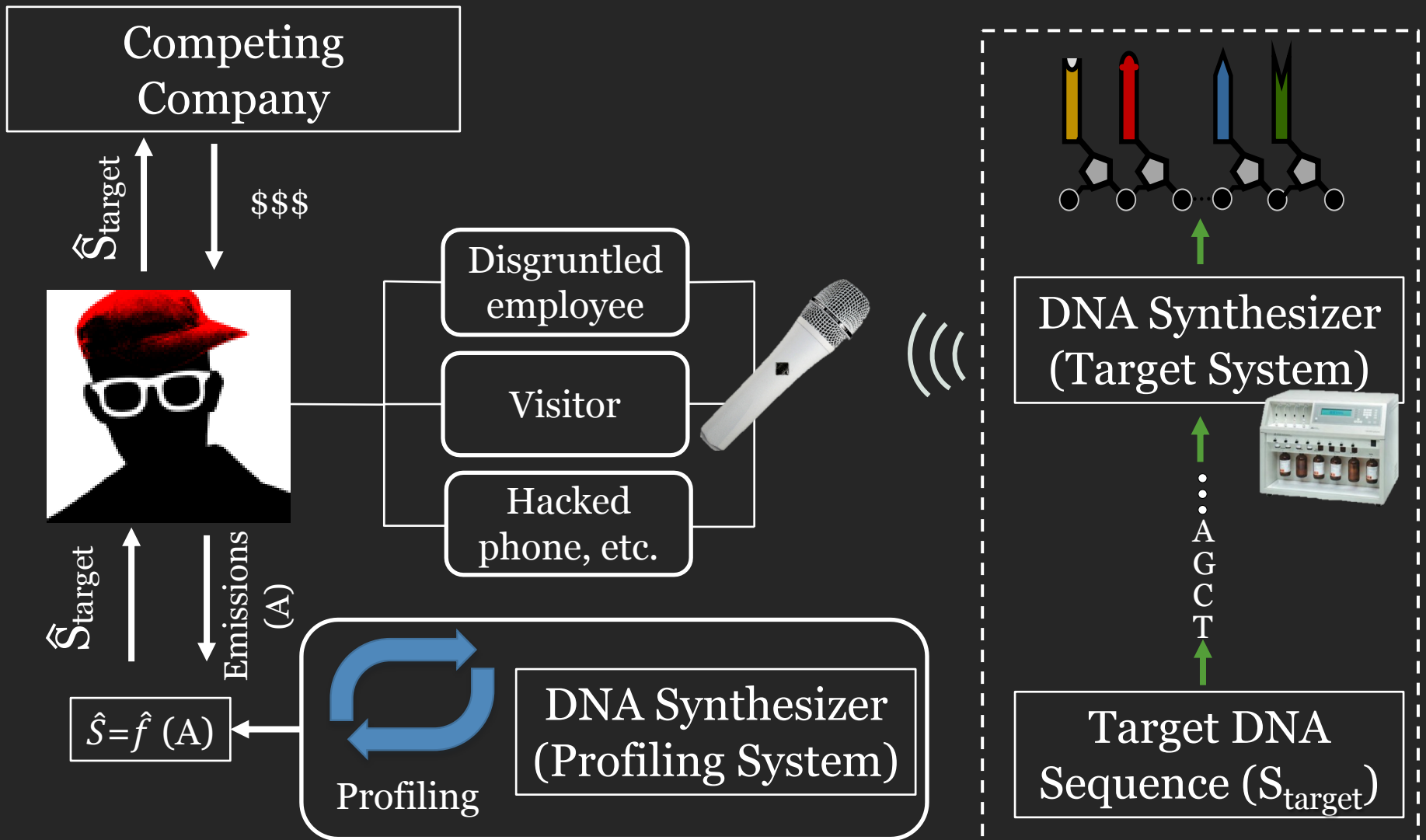
- K-best sequence prediction

- Zero cost post-processing
imperfect predictions

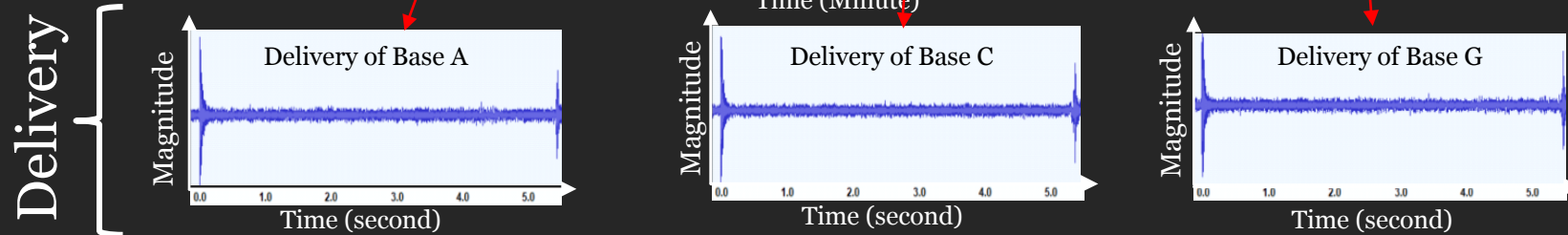
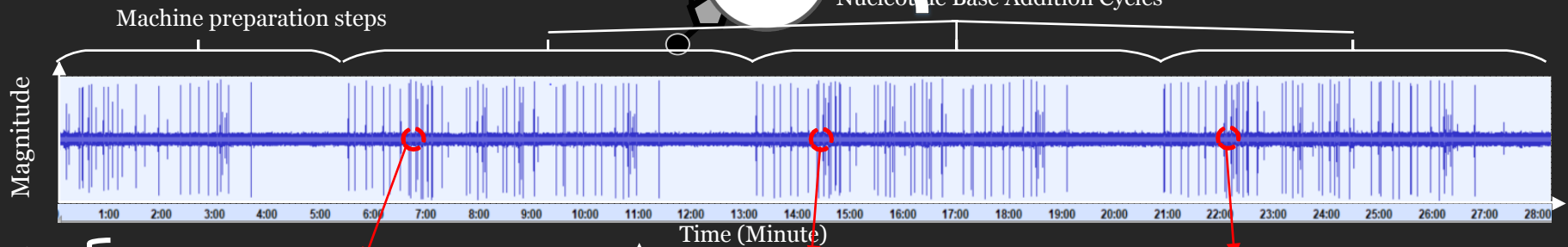
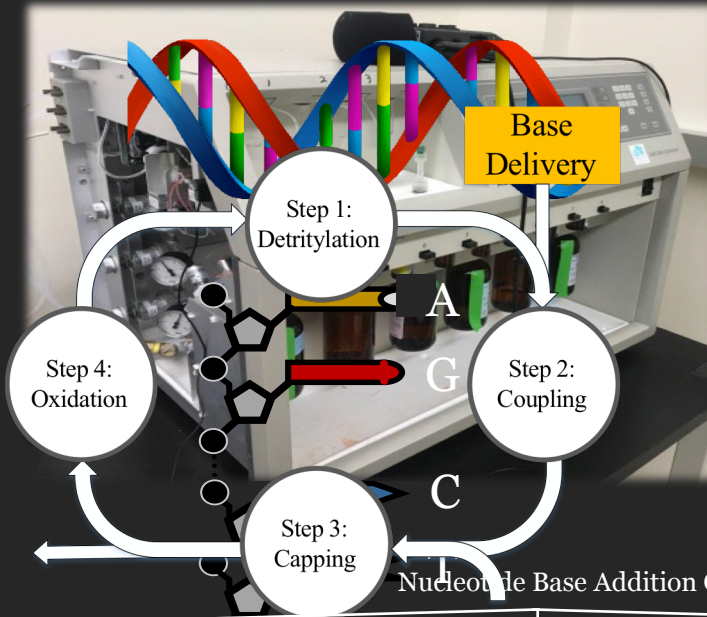
- Countermeasures



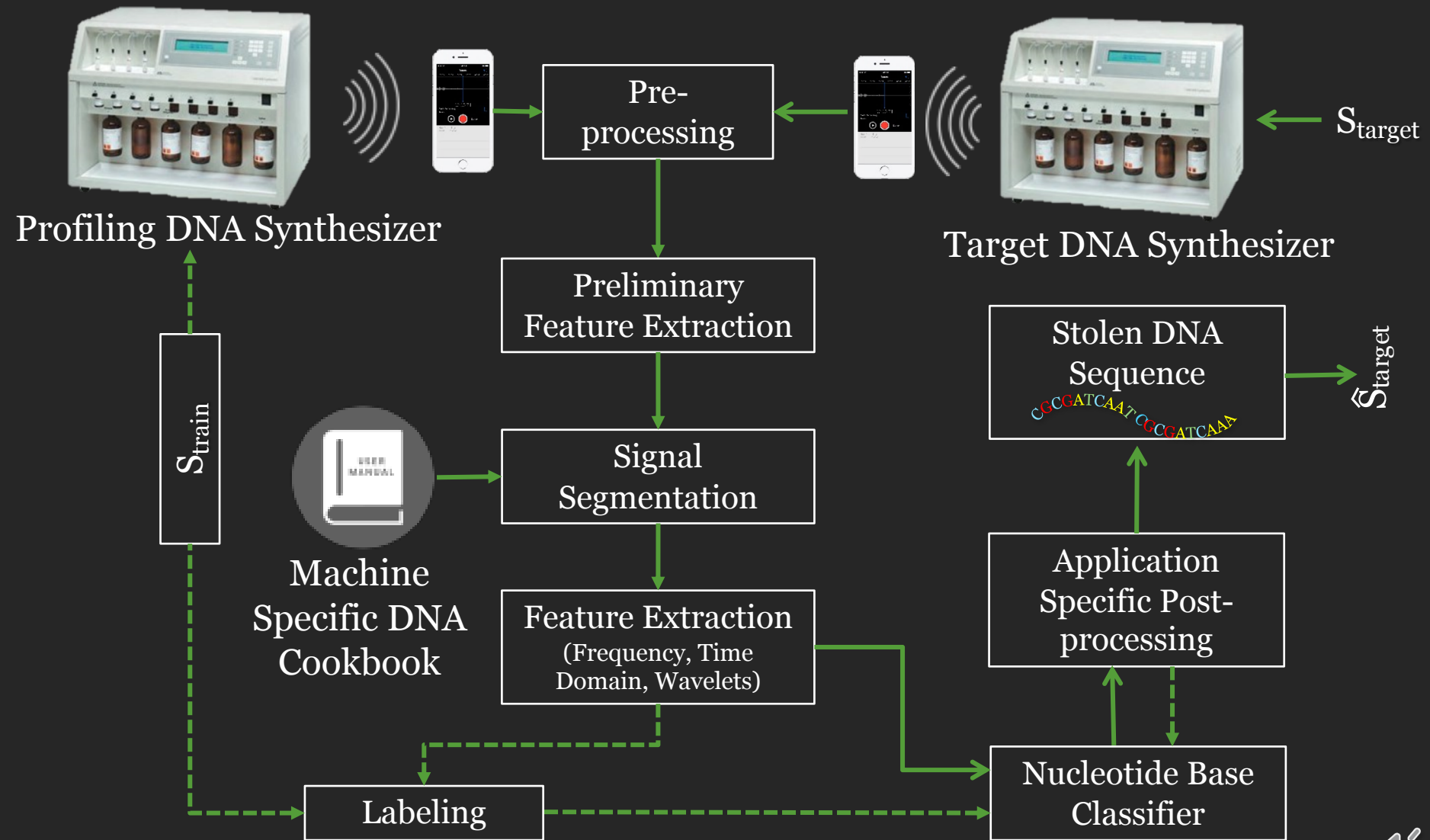
Adversary Model



DNA Synthesizer



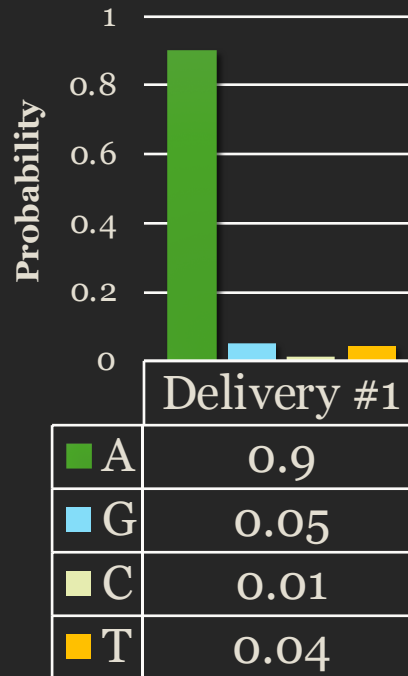
Attack Design



K-Best DNA Sequences: Motivation

- Post-processing stage detects a fault in classifier prediction: What is the next best sequence?

- Example



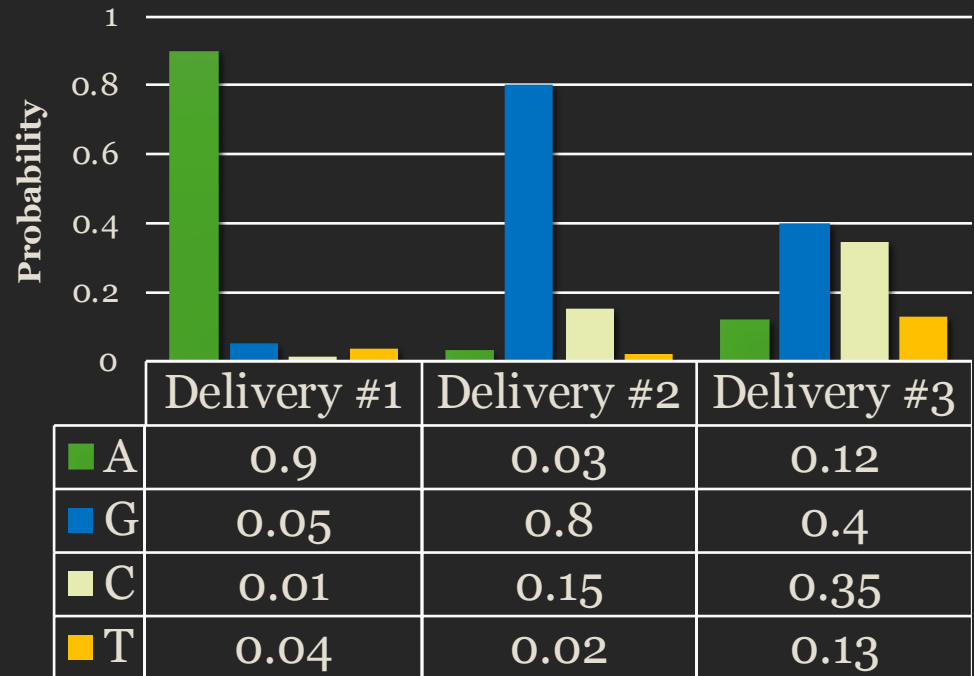
- First best prediction: **AGG**
- Second best prediction: **AGC**
- Third best prediction: **???**



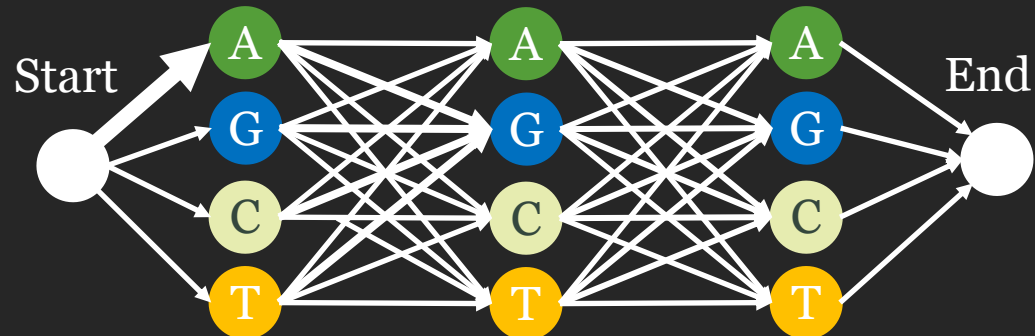
K-Best DNA Sequences: Algorithm

➤ DAG generation:

1. Add Start End nodes
2. Add 4 nodes per delivery
3. Fully connect consecutive 4 nodes by directed weighted edges corresponding to the destination nodes type probability



- ## ➤ Find the K longest paths
- [1] $O(n \log(n)+k)$



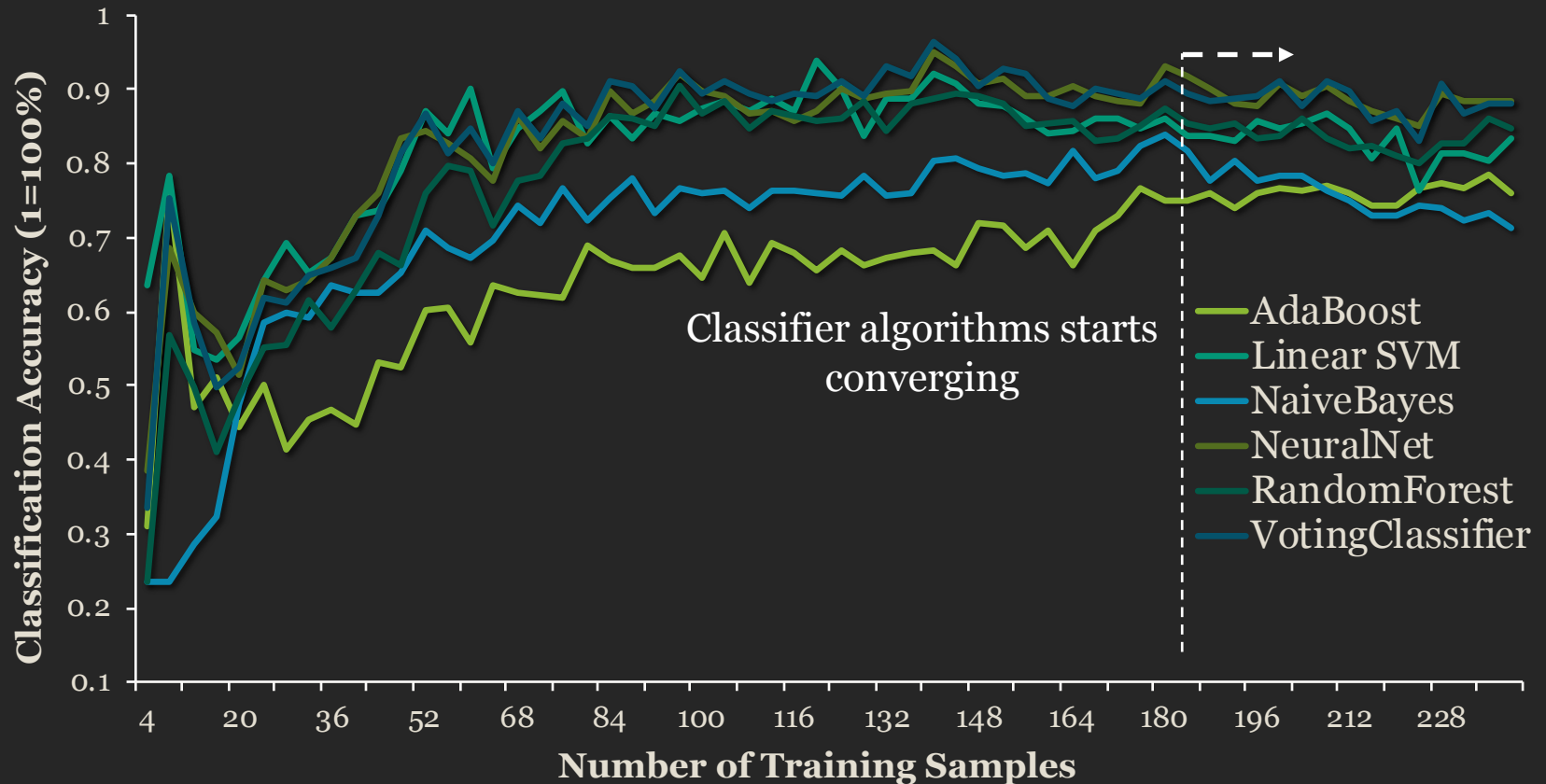
Experimental Setup



- Learning Possibility
- Noise Effect
- Distance Effect



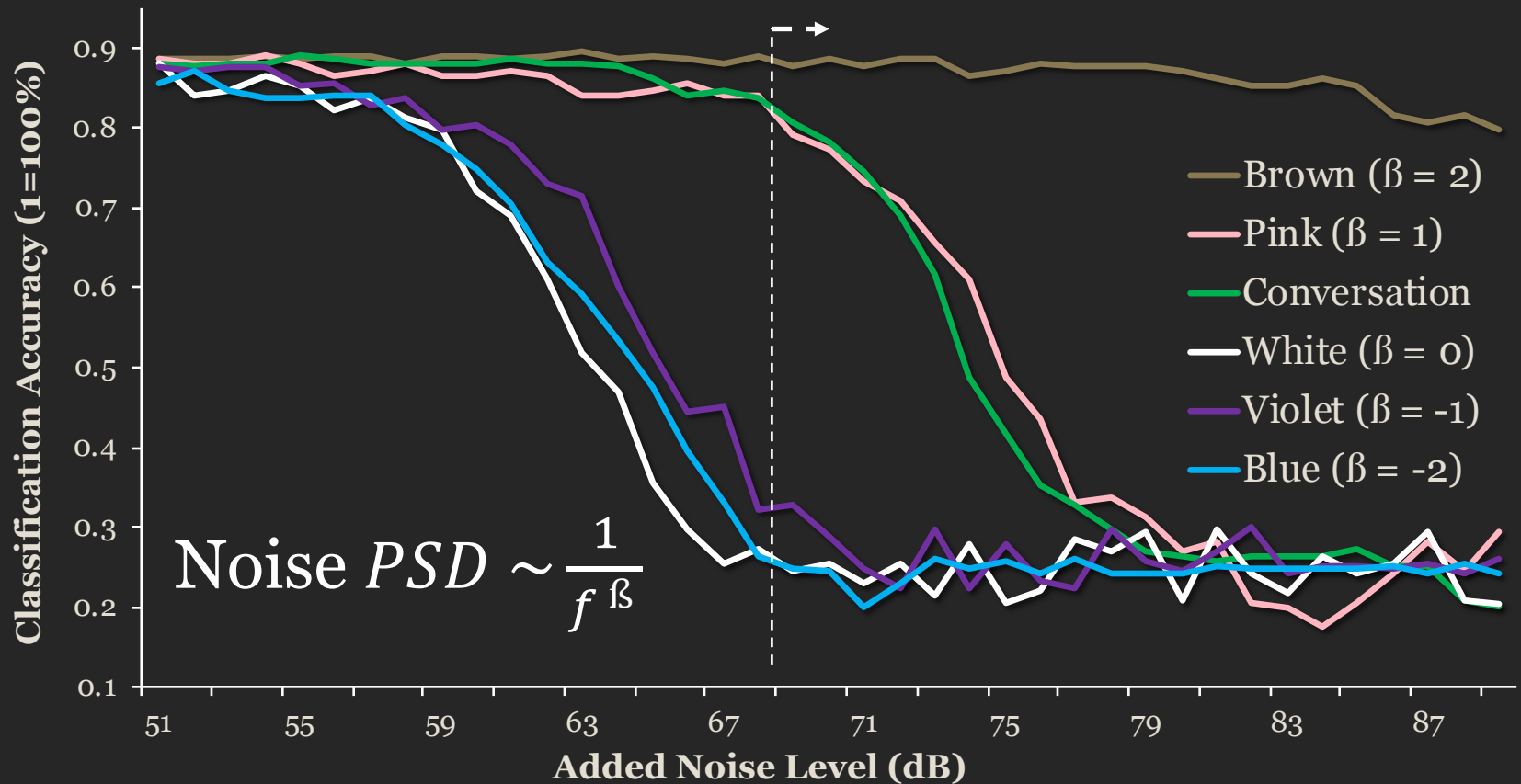
Results: Learning Curve



➤ Best classification accuracy: **88.07%**
(Voting Classifier)



Results: Noise Effect

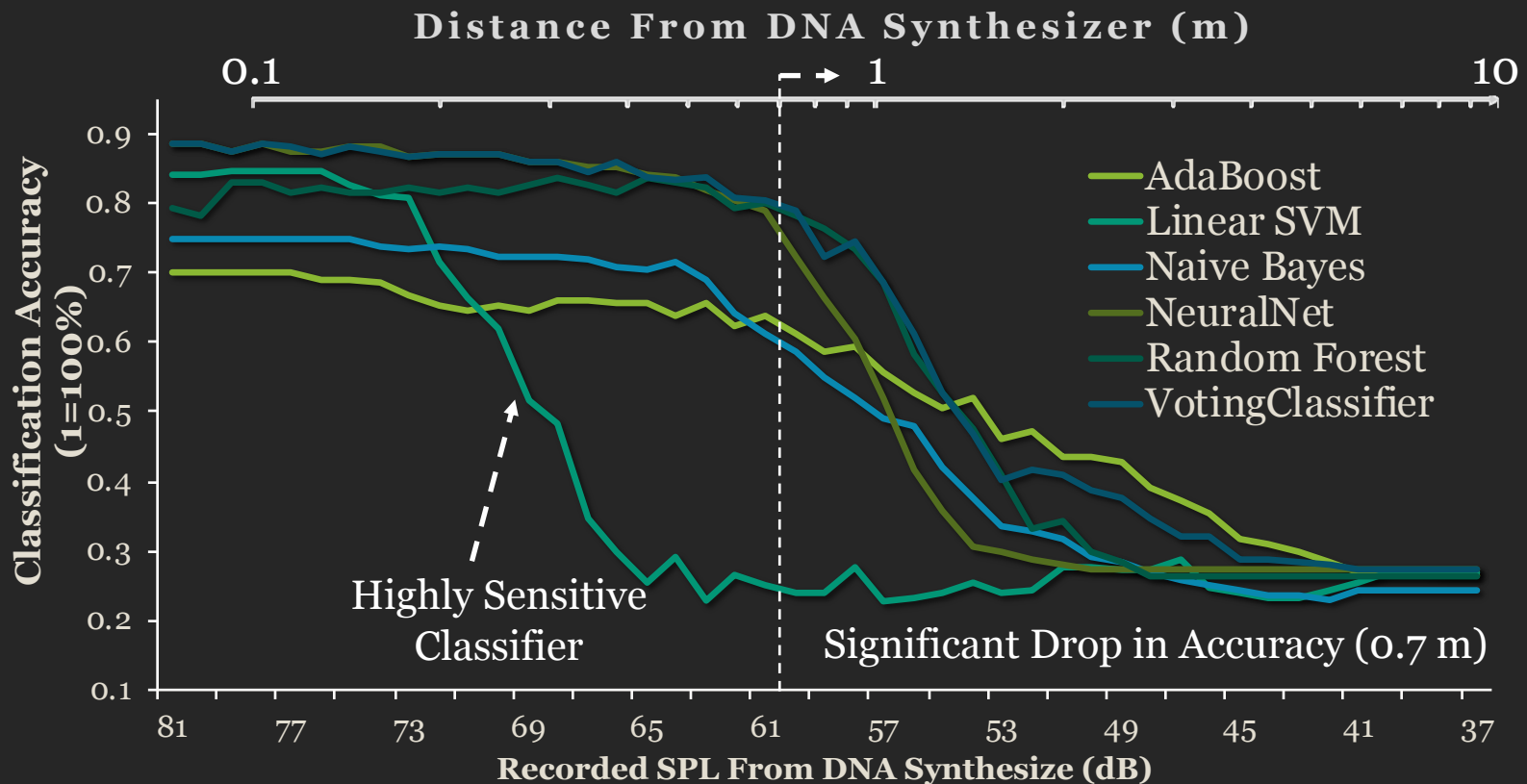


- Environmental noise does not effect the accuracy (Pink)
- High frequency noise blocks the attack better (Blue)



Results: Distance Effect

- Room noise level is neglectable for contact microphone
- Sound Pressure Level (SPL) : $L_2 = L_1 - |20\log(\frac{r_1}{r_2})|$
- Keep the room noise level constant, lower the SPL from DNA



Test Cases

To: Sina Faezi Cc: John Chaput; M. Al Faruque

Hi Sina:

Here is the original sequence I input to DNA synthesizer. The wrong predictions happened on T only which you conjectured as G, but there are six positions wrong.

DNA sequence: 5'-TGG CGA CAT TAT AAC CCG TCG GAT GAT CCG GGT CGT GTT CAC CTC-3'

Peptide sequence: WRHYNPSDDPGRVHL

On Thu, May 3, 2018 at 4:06 AM, Sina Faezi <sfaezi@uci.edu>

there are six positions wrong.

I noticed that since the machine starts synthesizing in the reversed order. If that is the case, here is the edited sequence:

TGG CGA CAT GAT AAC CCG TCG GAG GAT CCG GGG CGG GGG CAC CTC

WRHDNPSEDPGRGHL

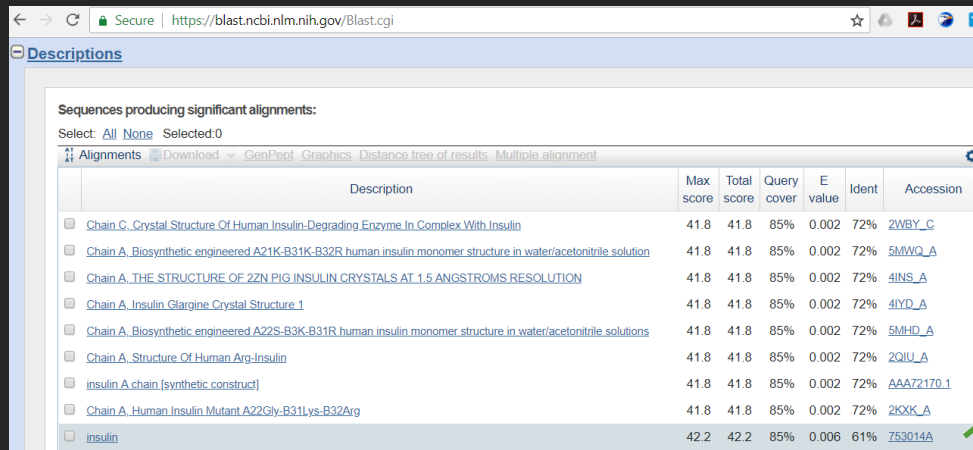
Kind Regards,
Sina Faezi

Graduate Student Researcher
<http://aicps.eng.uci.edu>



BLAST, Post processing

- Original Amino Sequence: GIVEACCTIICSLHELEDDCE



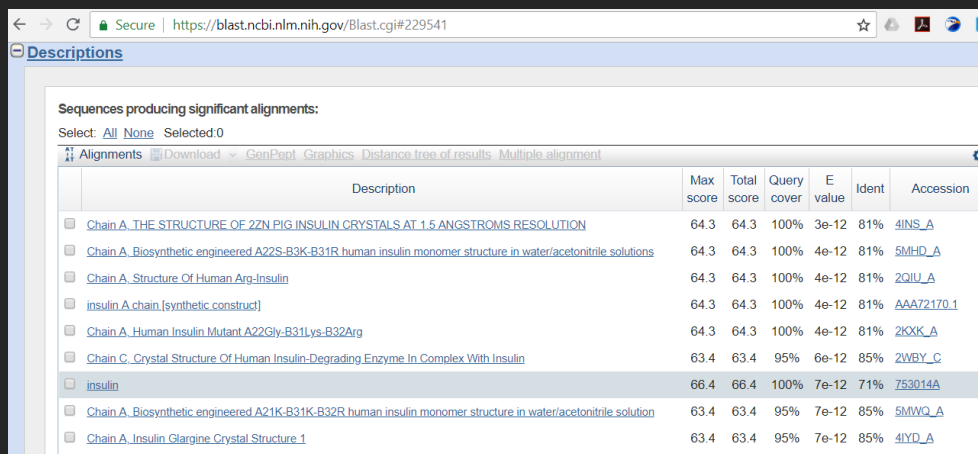
Sequences producing significant alignments:

Select: All None Selected: 0

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> Chain C. Crystal Structure Of Human Insulin-Degrading Enzyme In Complex With Insulin	41.8	41.8	85%	0.002	72%	ZWRB_C
<input type="checkbox"/> Chain A. Biosynthetic engineered A21K-B31K-B32R human insulin monomer structure in water/acetonitrile solution	41.8	41.8	85%	0.002	72%	5MWQ_A
<input type="checkbox"/> Chain A. THE STRUCTURE OF 2ZN PIG INSULIN CRYSTALS AT 1.5 ANGSTROMS RESOLUTION	41.8	41.8	85%	0.002	72%	4INS_A
<input type="checkbox"/> Chain A. Insulin Gargine Crystal Structure 1	41.8	41.8	85%	0.002	72%	4IYD_A
<input type="checkbox"/> Chain A. Biosynthetic engineered A22S-B3K-B31R human insulin monomer structure in water/acetonitrile solutions	41.8	41.8	85%	0.002	72%	5MHD_A
<input type="checkbox"/> Chain A. Structure Of Human Arg-Insulin	41.8	41.8	85%	0.002	72%	2QIU_A
<input type="checkbox"/> insulin A chain [synthetic construct]	41.8	41.8	85%	0.002	72%	AAA72170.1
<input type="checkbox"/> Chain A. Human Insulin Mutant A22Gly-B31Lys-B32Arg	41.8	41.8	85%	0.002	72%	2KXK_A
<input checked="" type="checkbox"/> insulin	42.2	42.2	85%	0.006	61%	753014A

Insulin

- Predicted Amino Sequence : GIVE**EC**CTS**IC**SL**Y**ELE**D**Y**C**D



Sequences producing significant alignments:

Select: All None Selected: 0

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> Chain A. THE STRUCTURE OF 2ZN PIG INSULIN CRYSTALS AT 1.5 ANGSTROMS RESOLUTION	64.3	64.3	100%	3e-12	81%	4INS_A
<input type="checkbox"/> Chain A. Biosynthetic engineered A22S-B3K-B31R human insulin monomer structure in water/acetonitrile solutions	64.3	64.3	100%	4e-12	81%	5MHD_A
<input type="checkbox"/> Chain A. Structure Of Human Arg-Insulin	64.3	64.3	100%	4e-12	81%	2QIU_A
<input type="checkbox"/> insulin A chain [synthetic construct]	64.3	64.3	100%	4e-12	81%	AAA72170.1
<input type="checkbox"/> Chain A. Human Insulin Mutant A22Gly-B31Lys-B32Arg	64.3	64.3	100%	4e-12	81%	2KXK_A
<input type="checkbox"/> Chain C. Crystal Structure Of Human Insulin-Degrading Enzyme In Complex With Insulin	63.4	63.4	95%	6e-12	85%	ZWRB_C
<input checked="" type="checkbox"/> insulin	66.4	66.4	100%	7e-12	71%	753014A
<input type="checkbox"/> Chain A. Biosynthetic engineered A21K-B31K-B32R human insulin monomer structure in water/acetonitrile solution	63.4	63.4	95%	7e-12	85%	5MWQ_A
<input type="checkbox"/> Chain A. Insulin Gargine Crystal Structure 1	63.4	63.4	95%	7e-12	85%	4IYD_A

Insulin



Countermeasures

- Secured structure
 - ❑ Symmetric component placement
 - ❑ Noise canceling material
- Artificial noise
- Delivery segment obfuscation
- Secured laboratory environment



Summary

- A cyber-physical attack on DNA synthesizers
- K-best sequences algorithm
- Real world attack scenario



Questions

Thank You!

Oligo-Snoop

A Non-Invasive Side Channel
Attack Against DNA Synthesis
Machines

