This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2020.2994742, IEEE Transactions on Information Forensics and Security

Molecular Barcoding as a Defense against Benchtop Biochemical Attacks on DNA Fingerprinting and Information Forensics*

Mohamed Ibrahim, *Member, IEEE*, Tung-Che Liang, Kristin Scott, Krishnendu Chakrabarty, *Fellow, IEEE*, Ramesh Karri, *Fellow, IEEE*

***Abstract*—DNA fingerprinting can offer remarkable benefits, especially for point-of-care diagnostics, information forensics, and analysis. However, the pressure to drive down costs is likely to lead to cheap untrusted solutions and a multitude of unprecedented risks. These risks will especially emerge at the frontier between the cyberspace and DNA biology. To address these risks, we perform a forensic-security assessment of a typical DNA-fingerprinting flow. We demonstrate, for the first time, benchtop analysis of biochemical-level vulnerabilities in flows that are based on a standard quantification assay known as *polymerase chain reaction (PCR)*. After identifying potential vulnerabilities, we realize attacks using benchtop techniques to demonstrate their catastrophic impact on the outcome of the DNA fingerprinting. We also propose a countermeasure, in which DNA samples are each uniquely *barcoded* (using synthesized DNA molecules) in advance of PCR analysis, thus demonstrating the feasibility of our approach using benchtop techniques. We discuss how molecular barcoding could be utilized within a cyber-biological framework to improve DNA-fingerprinting security against a wide range of threats, including sample forgery. We also present a security analysis of the DNA barcoding mechanism from a molecular biology perspective.**

I. Introduction

DNA fingerprinting—the use of DNA profiling techniques for the identification of individuals—has had a dramatic impact in the court of law, intelligence-gathering, and counter-intelligence applications since its introduction in 1985 [1], [2], [3]. Over the years, DNA fingerprinting technologies have grown in sophistication to allow analysis on minute samples of DNA while recent advances in cyber-physical system (CPS) laboratory technologies, e.g., microfluidics-based assays, have promised to streamline and optimize the processing of DNA samples [4], [5], [6]. These advances have revolutionized not only information forensics but also DNA-processing flows in point-of-care clinical diagnostics, pathogen detection, and cancer research [7], [8].

The increasing complexity of contemporary DNA-fingerprinting flows has blurred the frontier between the cyber-space (computer-based or human-based control) and biology. Recent studies suggest that interactions across these spaces raise unprecedented security concerns, which create a whole new category of threats known as *cyberbiosecurity threats* [9]. In light of several instances of diverse cyberbiosecurity threats that have been discovered in recent years [10], [11], [12], it is imperative that the trust placed in DNA-fingerprinting flows—especially in information forensics laboratories—be thoroughly investigated.

Typical DNA-Fingerprinting Flow. In a traditional forensic DNA-fingerprinting flow, a sample (e.g., blood cells) is first collected, then DNA is extracted, amplified, detected, and either stored or destroyed [5]. The process of DNA amplification and quantification is performed using the polymerase chain reaction (PCR) assay [13]. The flow steps, shown in Figure 1, must be performed at a specially equipped forensics laboratory. According to a report presented in 2009, the regulation of forensics laboratories in the United States is fragmented, with only a handful of states requiring accreditation for DNA forensics [14]. Although a number of initiatives have been taken in the United States to strengthen forensic science [15], in 2019, not all states require accreditation for DNA forensics [16]. Even with accreditation and government oversight, cyberbiosecurity risks with the DNA forensics process have been identified, such as interception of shipments to compromise collected samples, tampering with computer-controlled processes to introduce stealthy discrepancies between the physical parameters of the process and the reported data, and altering secret bioinformatics databases to produce a misleading result [11], [17]. Motivations are varied and may include the obstruction of justice, activism, or sabotage of a biotechnology corporation, resulting in catastrophic financial and social impacts. Hence, the cyberbiosecurity risks of DNA fingerprinting need to be mitigated to prevent such activities.

Cyberbiosecurity Risks in DNA Fingerprinting. There

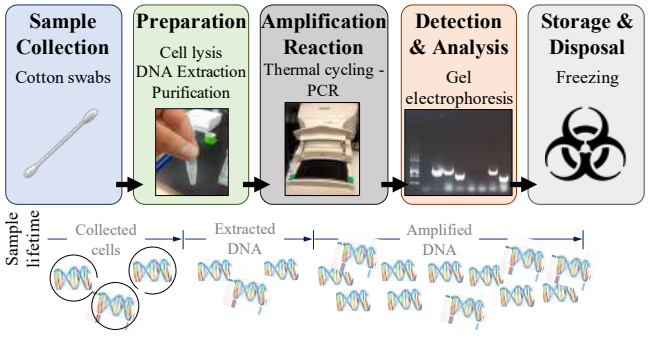


Fig. 1: Steps in a typical DNA-fingerprinting flow based on the PCR assay.

*This research was supported in part by the Army Research Office under grant number W911NF-17-1-0320 and the National Science Foundation under grant number CNS-183362.

Mohamed Ibrahim, Tung-Che Liang, and Krishnendu Chakrabarty are with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA (e-mail: mohamed.s.ibrahim@alumni.duke.edu; tung.che.liang@duke.edu; krish@duke.edu).

Kristin Scott is with the Department of Molecular Genetics and Microbiology, Duke University, Durham, NC, USA (e-mail: kristin.scott@duke.edu).

Ramesh Karri is with the Department of Electrical and Computer Engineering, New York University, New York, USA (e-mail: rkarri@nyu.edu).

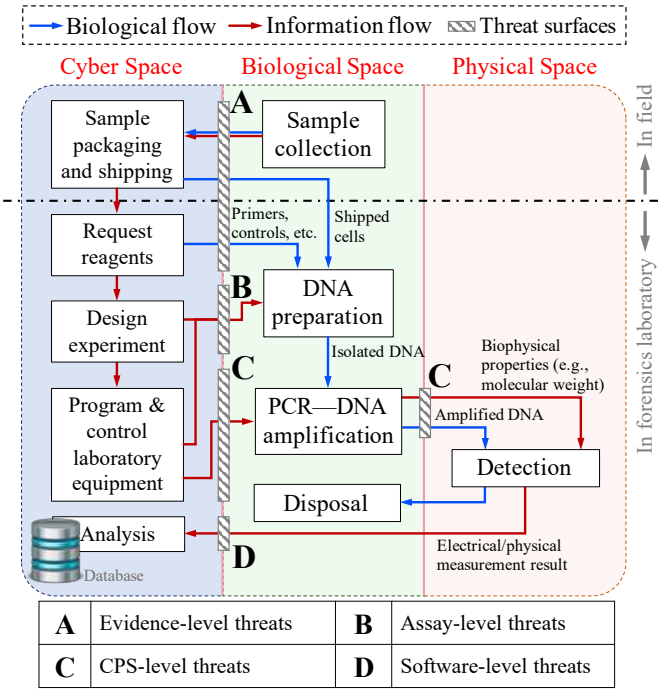


Fig. 2: A typical DNA-fingerprinting flow and related cyberbiosecurity threats.

is a broad set of the cyberbiosecurity threats in a DNA-fingerprinting flow, classified based on where an attack can take place within the flow of information. Based on the similar and clearly differentiated steps in typical DNA-fingerprinting flows, we identify three spaces: the *cyber space*, the *biological space*, and the *physical space*. Note that we separate the biological space and the physical space. The first example of a DNA-fingerprinting flow is shown in Fig. 1 where the flow steps can broadly be grouped into two main tasks: (1) biochemical processing, i.e., sample preparation and amplification, and (2) substrate detection and biosensing (sometimes referred to as DNA separation). Most DNA-fingerprinting flows utilize common assay procedures for biochemical processing, i.e., DNA extraction, purification, and PCR amplification. However, the task of substrate detection can vary among different flows, where different "sophisticated" physics-based methodologies and biosensing techniques are employed, and the goal is to achieve the highest detection throughput, accuracy, and speed. Examples of detection methodologies include benchtop techniques for gel electrophoresis [18] (described in this paper) and capillary electrophoresis [19], and also miniaturized techniques that use on-chip fluorescence sensors [20]. Our research suggests that the cyberbiosecurity threats can be classified into four main categories (see Figure 2): evidence-level threats, assay-level threats, cyber-physical system (CPS)-level threats, and software-level threats.

(1) *Evidence-level Threats:* Attacks on the integrity of DNA fingerprinting can occur even before DNA analysis begins. An attacker can tamper with the sample or the sample ID during sample packaging and shipping. DNA-fingerprinting flows assume that the samples under investigation were collected, labelled, and shipped in a trustworthy manner. That is, the samples have not been contaminated or were not exposed to environmental conditions that would lead to fouling of the samples or alteration in its intrinsic biological behavior. In a more adversarial setting, trust in the sample collection requires that the samples were neither deliberately replaced by a malicious actor nor tampered with harmful reagents. This trust assumption has been shown to be invalid [21]. The work in [22] also showed that by replacing biological cells in samples, a new form of bioterrorism can be introduced to Bio-NanoThings communication networks.

(2) *Assay-Level Threats:* Once the DNA samples are collected and shipped to a forensics laboratory, they must be processed and analyzed for comparison with reference samples or a DNA library. The instructions for carrying out the processing are specified as an assay that instructs when samples are mixed with reagents, heated, or split to obtain a desired result, such as in the PCR used for DNA amplification (Figure 2). Tampering with the assay can destroy precious samples that may be impossible to re-collect, or can lead to misleading and inconclusive results. Studies related to assay-level threats can be found in [23], [24], [25].

(3) *CPS-Level Threats:* The support systems that surround DNA fingerprinting are potential attack surfaces that must be evaluated. Programmable robotic arms or microfluidic biochips are supported by computer-based controllers and sensor feedback [26]. The signals transmitted from the controllers to the laboratory devices are vulnerable to attacks. The results obtained from integrated sensors—if any—must be transmitted electronically; they could potentially be tampered with. Detailed studies on security implications of CPS-level threats, especially in cyber-physical microfluidic laboratory-on-chip systems, have been reported [27], [28], [29].

(4) *Software-Level Threats:* Using the detection results, quantitative-analysis methods are used to identify genetic variants, e.g., phenotypes, compared with the genetic profiles of a large number of individuals. Most of these computational methods require access to big genomic repositories, potentially across several laboratories or sites, to improve the power of DNA-fingerprinting studies. However, Trojan attacks have hindered these large-scale studies by discouraging individuals and authorized laboratories from sharing their genomic data [30], [31]. Researchers have counteracted the impacts of such attacks by adapting modern cryptography mechanisms such as secure multiparty computation to large-scale genome-wide fingerprinting platforms [32].

Guidelines for Evidence-level Security. Security assessment of evidence-level threats has not gained attention despite their potential impacts on the end result. Study of these threats has thus far been considered only within cell-free DNA frameworks, e.g., DNA-sequencing frameworks [12]. In applications such as clinical diagnostics or forensic sciences, DNA samples can be obtained only after: (1) relevant human cells are collected, shipped, and processed, and (2) specific primer and control reagents are prepared (see Figure 2). The security implications of this early process has been overlooked, raising concerns about the effectiveness of the current cyberbiosecurity model.

As an example of a evidence-level threat, an attacker can fal-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2020.2994742, IEEE Transactions on Information Forensics and Security

sify blood and saliva samples containing DNA from a person other than the donor of these fluids [33]. Forged evidence can be realized using a strand of hair or drinking cup used by the target person. Forged DNA can then be extracted, amplified, and planted within the collected samples. The attacker may be either an intruder or a malicious technician in the forensics laboratory, and the attack does not require detailed knowledge of the forensic analysis platform.

To prevent this attack, crime-scene investigators can use utilities to securely "barcode" the collected evidence cells. In the barcoding routine, evidence cells are supplemented with specially designed DNA molecules. These molecules do not change the properties of the evidence cells, but they allow the evidence cells to have unique DNA signatures at specific regions of the DNA. These barcodes are assumed to be generated from a probabilistic distribution whose parameters are known to only authenticated persons at the front-end (molecular barcodes generation) and the back-end (deciphering barcoded samples at the forensics laboratory). A cyber-level key management scheme is also required to control the interactions among the participating parties.

Paper Contributions. Our research is the first attempt to consider evidence-level security implications of the DNA-fingerprinting flows. The key contributions of this paper are as follows.

- We demonstrate a benchtop experimental study that investigates evidence-level threats associated with DNA fingerprinting. We present a real-life demonstration of attacks along with the resulting outcomes.
- We introduce molecular barcoding to secure collected samples against evidence-level attacks.
- We describe a multi-space security framework that includes DNA barcoding (biological space) and data encryption (cyber space) to improve the security of DNA fingerprinting.
- We demonstrate a benchtop experimental study that implements the barcoding mechanism.
- We provide a security analysis for the multi-space security framework from a molecular biology perspective.

II. Background

In this section, we present an overview of biological and chemical processes related to DNA fingerprinting, in order to properly contextualize the central proposition of DNA barcoding.

A. DNA

Deoxyribonucleic acid (DNA) is a molecule that represents the genetic material in all living organisms. It is composed of a chain of structures known as *nucleotides* (also known as *bases*) whose order forms the identity of a DNA sequence. Each nucleotide can be one of four structures: adenine (**A**), cytosine (**C**), guanine (**G**), and thymine (**T**). This sequence forms a single-stranded DNA (ssDNA) that is chemically bonded with another complementary ssDNA to form a double-stranded DNA (dsDNA), which has a double helix structure. Figure 3 shows a linear representation of a dsDNA structure,

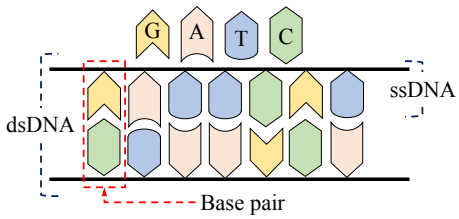


Fig. 3: A linear representation of ssDNA and dsDNA.

which is comprised of base (nucleotide) pairs. Often, dsDNA is just referred to as DNA for short.

B. DNA Fingerprinting and PCR

DNA fingerprinting is a laboratory method that is applied to collected cells to characterize an individual DNA sample either by size or specific nucleotides (**A**, **C**, **G**, or **T**). Most often, genomic regions of interest are targeted for analysis by either PCR or DNA sequencing. PCR is widely used and it enables information-forensics analysts to make numerous copies of isolated dsDNA molecules. A unique feature of PCR is that only dsDNA containing target sequence, referred to as an amplicon, are copied ("amplified"), with the aid of a specific set of chemical reagents as well as precisely controlled thermal cycling [34]. During amplification, the two strands (chains) of dsDNA are thermally separated. Next, under a lower temperature, each single strand is allowed to re-attach (anneal) to specially synthesized primers that have complementary nucleotide sequences. A DNA polymerase enzyme then extends the primer sequence by filling in the remainder of the complementary nucleotides, thus, constructing new copies of a target dsDNA. By repeating this process (i.e., thermal cycling and primer re-attachment), we can generate thousands of new copies of an amplicon.

Post DNA amplification, biophysical features, e.g., molecular weight, of these amplicons can be measured and the results can be used to distinguish between samples. Considering the forensic security of DNA-fingerprinting flows, these biophysical features can be key enablers for detecting evidence-level attacks or even securing these flows against such attacks, as discussed later in Section III.

C. DNA Preparation and Amplification

For each type of collected samples/cells, purified DNA is isolated to start DNA-fingerprinting experiments using PCR. PCR typically requires the preparation of three reactions and processing them in parallel through DNA amplification (Figure 1). One reaction represents PCR processing of the *sample under investigation* (SUI). A second reaction represents a reference *positive control* (PC), which is expected to produce visible amplicons at the end of the reaction. A third reaction represents a reference *negative control* (NC), which should not show any amplicons in the reaction result. PC is used to examine the efficiency of the reagents and the PCR procedure, whereas NC is used for calibration and for detecting carryover contamination. If amplicons are observed from the SUI reaction, then it is concluded that the given sample belongs to a cluster of sample types that exhibit biological activity at the locus of this particular amplicon. Likewise, if no amplicons

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2020.2994742, IEEE Transactions on Information Forensics and Security

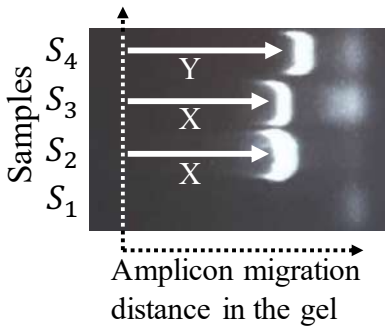


Fig. 4: Measurement of DNA amplification using gel electrophoresis.

are observed, i.e., similar to the NC pathway, then the sample belongs to a different cluster. Obviously, both PC and NC play a key role in the determination of the sample identity; therefore, they must be processed carefully.

D. DNA-Fingerprinting Measurement

Agarose gel electrophoresis is a routinely used method for separating protein or DNA molecules based on their molecular weight, which is proportional to the size of DNA strands [18]. Hence, PCR products (amplicons) are size-separated by the aid of an electric field where negatively charged DNA molecules migrate toward an anode (positive) pole. The shorter the amplicon (i.e., the lower the molecular weight), the further the sample will reach on the gel. Figure 4 shows the outcome of gel electrophoresis for four samples used in our study. Samples S_2 and S_3 are samples that exhibit amplicon generation through DNA amplification and the target amplicons contain a large number of base pairs (bp). Hence, the molecular-weight size markers (shown as white bands and named "DNA bands") associated with these samples indicate that a short distance (X) has been migrated. Sample S_4 also exhibits DNA amplification, but the resulting amplicons contain a smaller number of base pairs[1]. Therefore, the DNA band associated with S_4 indicates that S_4 has travelled a longer distance; see Y. Sample S_1 represents an NC pathway and it does not show a white band since no DNA amplification has occurred.

The process of DNA amplification is precise and it is sensitive to any modifications applied to the DNA-preparation process [35]. Therefore, the results we obtain from agarose gel electrophoresis can be analyzed to capture potential attacks on DNA-fingerprinting flows.

III. Evidence-level Attack Models

As described in Section I, DNA-preparation steps for PCR can be regarded as potential attack surfaces that can be exploited by a malicious adversary, who may be interested, for example, in tampering with the final results of a DNA analysis and information forensics. The tampering can result in the complete destruction of evidence (the true DNA-amplification profile), or modification of evidence such that it produces a misleading result. To demonstrate this capability, we carried out a benchtop experiment that executed DNA-amplification analysis on 6 different types of samples. Using these 6 samples, 4 parallel runs of DNA fingerprinting using PCR amplification were performed. The first run represents the golden, i.e., mainstream, implementation of the protocol, where all the protocol settings are adjusted normally. The remaining three runs represent malicious implementation of PCR, where DNA preparation was deliberately altered. We "simulated" three evidence-level attacks to demonstrate the following malicious behavior:

- *Attack 1:* An attack that causes all collected samples to maliciously report equivalent positive signals for DNA amplification; therefore, concealing the identity of the collected samples. We refer to this attack as *positive denial-of-service (PDoS) attack*.
- *Attack 2:* This attack causes all samples to maliciously report a negative value of DNA amplification (i.e., no DNA amplification occurs). This damages all the collected samples. We refer to this attack as *negative denial-of-service (NDoS) attack*.
- *Attack 3:* An attack that replaces or switches samples. This attack is *sample forgery*.

Attack 1 and Attack 2 completely destroy the evidence, whereas Attack 3 alters the evidence. Details of the benchtop experiment are presented below.

A. Benchtop Realization of Evidence-Level Attacks

We first describe the experiment setup for the 4 parallel PCR runs. Next, we explain the steps we used to implement the golden and the malicious DNA-fingerprinting processes. Figure 5 summarizes the steps described in this section and it aligns with the generic flow depicted in Figure 1.

Experiment Setup and Sample/Reagent Preparation. Having collected target DNA samples (Figure 5(a)), our implementation of a PCR run has three main steps: (1) sample and reagent preparation (Figure 5(b-c)); (2) PCR thermal-cycling and amplification (Figure 5(d)); (3) analysis using gel electrophoresis (Figure 5(e-f)). Hence, we first isolated DNA from the 6 DNA samples (via cell lysis and purification). Second, for each PCR run, we prepared a PCR mixture tube, also known as a master mix, that consists of 36 μL of buffer, 18 μL of deoxynucleotides (dNTPs), 36 μL of primers, 7.2 μL of $MgCL_2$, 82.8 μL of H_2O, and 3 μL of the Taq polymerase. Each quantity was pipetted carefully into the tube. Since our experiment has 4 parallel PCR runs, we prepared 4 PCR master mixes, as shown in Figure 5(b). Finally, to ensure complete diffusion of PCR reagents, PCR tubes were spun using a micro-centrifuge.

Golden vs. Malicious PCR. The 4 master mixes were clean and were prepared using the same quantities of reagents. To demonstrate the attacks described earlier, we altered the contents of master mix 2 to reflect Attack 1 (PDoS). To implement PDoS, we added a small quantity of PC (ura4+) to master mix 2. We altered master mix 3 to refect Attack 2 (NDoS). We implemented NDoS by adding ethylenediaminetetraacetic acid (EDTA) to master mix 3. Master mixes 1 and 4 were kept clean to implement the golden reactions and to implement Attack 3 (sample forgery), respectively[2]. The positive control

[1] A gene mutant with a smaller number of base pairs can be created from another by enabling enzymatic gene deletion.

[2] Sample-forgery attacks do not require alteration to the master mix.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2020.2994742, IEEE Transactions on Information Forensics and Security

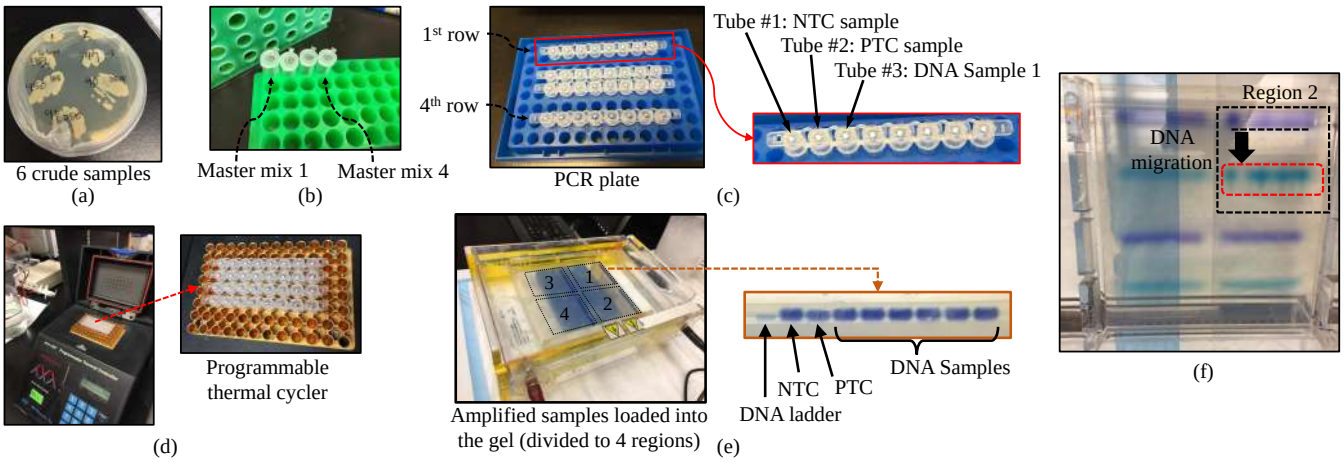


Fig. 5: Benchtop steps of DNA fingerprinting that involves 4 parallel PCR runs: (a) sample collection; (b) master-mix (reagent) preparation; (c) mixing samples and reagents; (d) thermal cycling for DNA amplification; (e) transfer of amplified samples to gel electrophoresis; (f) measurement of DNA fingerprints using gel electrophoresis.

(ura4+) is an endogenous gene that was previously selected as a region of interest of the genome. The negative control (EDTA) is a chelating agent that renders the Taq polymerase, which is necessary for DNA synthesis in the PCR reactions, inactive.

Next, we prepared a PCR plate that included all the chemicals needed for the 4 parallel runs as shown in Figure 5(c). We divided a set of empty tubes into 4 rows, where each row has 8 tubes. The reason for including 8 tubes (and not 6 tubes associated with the 6 DNA samples) was that two additional tubes per row were needed for PC and NC, as described in Section II.

The tubes in the first row on the plate (Figure 5(c)) were used for chemical reactions associated with the golden PCR assay. We pipetted 20 μL of PCR reagent in each tube. We added 1 μL of crude extract of each DNA sample to a separate tube, starting from tube #3 and ending at tube #8. To include a reference positive control (PC) and a reference negative control (NC) in the reaction, we pipetted 1 μL of distilled water (an NC solution that does not contain DNA) into tube #1 and 1 μL of ura4+ (a PC solution that is known to have the DNA of interest) into tube #2—we use tubes #1 and #2 for referencing. We repeated these steps to fill in the tubes in the second and the third rows. For the tubes of the second and third rows, we used master mixes 2 and 3. The tubes in the fourth row were filled with DNA samples that were randomly selected, triggering sample forgery.

After filling in all the tubes on the PCR plate, the tubes were sealed, placed in a balanced micro-centrifuge, and spun for a few seconds. Immediately after the spin, we placed the reaction tubes in the thermal cycler (Figure 5(d)), which had been programmed to perform DNA amplification by continuously raising and lowering the temperature of the PCR-plate content in discrete, pre-programmed steps. After the PCR program is finished, we transferred the PCR-plate contents into the gel electrophoresis apparatus, allowing the chemical solutions to migrate over the gel with the aid of an electric field; Figure 5(e) shows the transfer of 8×4 samples into the gel.

The gel was divided into four sections, and each section was loaded with samples from a certain row of tubes on the PCR plates. This approach allowed us to compare the outcomes of all the PCR runs. Also, in each section, we pipetted a specific chemical reagent, known as DNA ladder, which has fragments of DNA of different sizes, allowing us to benchmark the results of the PCR-related chemical solutions. Gel electrophoresis was run for an hour, then the gel block was transferred to a chamber that contains a source of UV light in order to visualize the DNA bands. Figure 5(f) shows a view of the gel after the electrophoresis process, viewing the migration of different PCR products.

B. Interpretation of Reaction Results

Figure 6 shows the results obtained by gel electrophoresis. The DNA-ladder plot has four regions: (1) the top-left region (Region 1) contains the DNA bands associated with the golden PCR reaction; (2) the top-right region (Region 2) represents the DNA bands based on Attack 1 (PDoS); (3) the bottom-left region (Region 3) shows the DNA bands for Attack 2 (NDoS); (4) the bottom-right region (Region 4) contains the DNA bands based on Attack 3 (sample switching). In all these sections, the result of the benchmark (the DNA ladder) is located at the leftmost column (column 1).

By analyzing Region 1 (trusted reaction), we observe that column 2 does not show any DNA bands; this result is expected since this column is associated with the NC tube where DNA amplification does not occur. In contrast, column 3 shows a DNA band, capturing the effect of DNA amplification within the PC tube. The remaining columns show various results of DNA amplification, depending on whether the target gene expresses at the target amplicon locus and also the length of the amplicon (number of base pairs). The DNA samples represented by columns 4 and 8 exhibit DNA amplification and they have amplicon mutants that are longer (contain more base pairs) compared to those in the DNA samples represented by columns 5 and 9.

In Region 2, we observe that all the columns exhibit the same high DNA-amplification profile, indicating that all the samples were deliberately manipulated to express the target gene. This behavior shows the impact of maliciously adding

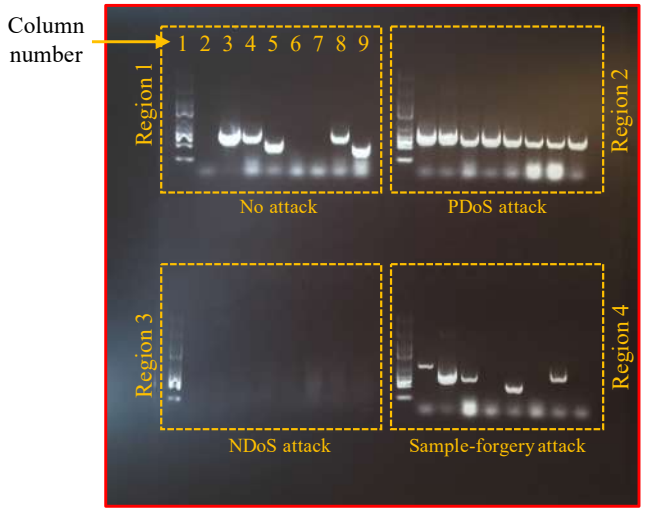


Fig. 6: Gel-electrophoresis results of the four DNA-fingerprinting runs (Region 1: golden DNA fingerprinting; Regions 2-4: malicious DNA fingerprinting).

PC reagent to all samples, causing a PDoS attack. This result is invalid especially because a white-band signal appears at column 1, which is supposed to suppress amplicons generation for the NC. As a result, this result cannot be used by an information forensics lab to distinguish DNA samples.

Similarly, in Region 3, no DNA amplification is reported at all samples since no DNA bands exist. This result indicates that all the samples, including the reference PC sample at column 3, were tampered with to suppress the expression of the target amplicons, causing NDoS attack. This result is also invalid and cannot be used to distinguish DNA samples. Note that both PDoS and NDoS attacks can be easily detected.

Finally, the different DNA bands observed in Region 4 indicate that no PDoS or NDoS attacks were launched. However, the profile of DNA bands shown in Region 4 is different from that of Region 1, indicating that DNA samples were likely switched or replaced with other samples, i.e., subjected to a sample-forgery attack. Note that this attack is hard to detect unless the switching action impacted either the PC or the NC samples, causing abnormal behavior at either column. For instance, in column 1 of Region 4, we observe a DNA band, meaning that DNA amplification has occurred. Since column 1 is associated with the NC sample, which is supposed to suppress DNA amplification, this observation is sufficient to prove that either the NC sample was tampered with/contaminated or the samples (including NC) have been switched.

In practice, Attack 3 (sample forgery) can be stealthy and hard to observe, especially if the attacker is aware of the locations of the NC and PC samples. Therefore, in Section IV, we present a benchtop study that provides an efficient countermeasure technique against Attack 3 based on DNA barcoding.

IV. Defense against Sample-Forgery Threats

In this section, we describe our approach for preventing sample-forgery attacks.

A. Molecular Barcoding of DNA Samples

We propose a defense that is inspired by a technique from microbiology called DNA barcoding [36]. DNA barcoding helps identify a sample similar to the optical identification of an object using a machine-readable barcode. Each sample is assigned one or a set of *unique* DNA barcodes that can be used to confirm the identity of the sample during downstream DNA-fingerprinting analysis. This technique can be implemented in the field (by collectors) and be applied once the cells are collected. Barcoded cells are shipped to an information forensics laboratory for regular PCR analysis.

At the information forensics laboratory, the embedded DNA barcode has to be amplified first then compared with a database to ensure that no sample-forgery attacks has occurred[3]. If the sequenced barcode matches the database, then the sample is considered genuine and it can be analyzed. On the other hand, if no barcode or an incorrect barcode is detected, then the associated sample is likely false and it has to be discarded. Figure 7 explains the above strategy using a schematic representation.

It is evident that molecular barcoding functions as a form of side-channel fingerprinting, which ensures that only genuine (trusted) samples can be used. The key principle that enables this security scheme is that each DNA barcode is initially designed with a unique sequence of nucleotides whose order forms the identity of a barcode. Recall from Section II that each nucleotide (or base) can be one of the following four options: adenine (**A**), cytosine (**C**), guanine (**G**), and thymine (**T**). This sequence forms a single-stranded DNA (ssDNA) that is chemically paired with another complementary ssDNA to form a double-stranded DNA (dsDNA).

Hence, to design a barcode of length L_b bp, the number of possible dsDNA structures n_{ds} is slightly lower than 4^{L_b}, if we exclude unrealistic combinations such as **AAAAA....** If $L_b = 280$ bp, which is a typical length of a small barcode, then $n_{ds} \approx 3.773962 \times 10^{168}$. Since each barcode has a specific nucleotide sequence, the barcode can be sequenced or detected (using PCR) at the down stream only using two specific primers (forward and reverse primers) that are tightly coupled with the nucleotide sequence. Hence, the number of primers is $2 \times 4^{L_{pmr}}$, where L_{pmr} is the primer length ($L_{pmr} = 21$ in our experiment). This large number ensures that launching a brute-force attack with the intent of identifying the sequence of a used barcode is impractical.

By analyzing the above characteristics of the molecular barcoding, we observe that the proposed biochemical routine can function as an effective defense mechanism only if the following conditions are satisfied:

1) Each barcode has a unique sequence of nucleotides, and so do the associated primers.
2) Only authorized collectors are allowed to barcode the samples. Only authorized analyzers can sequence/detect the barcodes before conducting DNA fingerprinting.
3) All the information related to the barcodes is transmitted securely between the collectors and analyzers.

A major challenge in the above conditions is that they apply to different spaces (cyber/information and biology spaces) and also different physical locations (field and forensics laboratory), putting the integrity of the molecular barcoding at risk. Hence, the routine for molecular barcoding needs to be

[3]Sample forgery can be performed by a man-in-the-middle (MITM) attacker who secretly relays and possibly alters transferred samples.

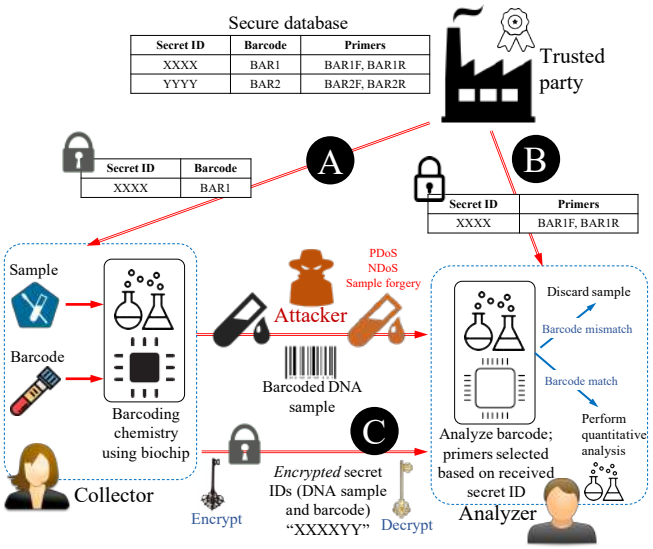


Fig. 7: Details of the multi-space security scheme.

coordinated with techniques for cyber-level security in order to enhance the integrity of the proposed defense. Such a coordination leads to a *multi-space* defense mechanism.

B. Multi-Space Defense Scheme

Figure 7 explains the steps in the proposed multi-space defense against sample-forgery. In the first step, a trusted party designs a set of barcodes with varying lengths and develops barcode-specific primers. These barcodes and the associated primers are registered in a secure database along with their secret identification numbers. Only authenticated users and collectors have access to this database.

Each of these classes has only a specific view of the database. In the second step, authenticated collectors obtain the barcodes and their associated secret identification numbers, whereas authenticated analyzers receive primers and their secret identification numbers. The identification of a barcode-specific primer must match the identification of the barcode. Also, such authenticated users are not required to have detailed knowledge of the barcoding sequences. The interactions described thus far, denoted by A and B in Figure 7, ensure that collectors and analyzers are individually trusted. Therefore, the process of molecular barcoding and detection can be trustworthy.

While collectors and analyzers can work independently off of the secret information and material obtained from the trusted party, they still need to interact because barcoded samples are prepared by collectors in-field and are delivered to analyzers in information forensics laboratories. A significant advantage of the proposed defense is that both types of users can interact securely and semi-anonymously. Therefore, in the third step, a collector communicates with an analyzer by sending two types of material: (1) a barcoded DNA sample (biological material), which encapsulates a secret molecular barcode obtained from the trusted party; (2) a public key-encrypted message (information material), denoted by C in Figure 7, which includes secret information about the barcode and the sample identification numbers. The use of public-key cryptography [37] ensures that analyzers can decrypt the message, using a private key, select the right type of primers based on the received identification message, and then perform barcode identification using PCR, as described in Section II. After completing the identification routine, the analyzer can verify the genuineness of all collected DNA samples.

This scheme does not prevent an adversary from intercepting an encrypted message and replacing it with an encrypted false message[4]. However, such an action is intrusive and is easy to detect since the end analyzer will not be able to identify the barcode, and the sample will eventually be discarded. Nevertheless, to avoid such a scenario, an end-to-end encryption protocol can be adopted to prevent eavesdroppers from accessing the encrypted message [38], [39].

By analyzing the above scheme along with the conditions described in the previous subsection, we show that this scheme can satisfy all the conditions.

V. Benchtop Demonstration and Analysis of Molecular Barcoding

In this section, we present our benchtop study implementing molecular barcoding for two DNA samples to demonstrate the effectiveness of the defense.

A. Specifications of Molecular Barcodes

Each sample is associated with a specific molecular barcode that has unique genomic characteristics (nucleotide sequence and fragment size). In our study, we use two molecular barcodes (gBlocks® Gene Fragments from Integrated DNA Technologies). The first barcode, denoted by BAR1, has 280 bp, and the second barcode, denoted by BAR2, has 190 bp. Table I shows the nucleotide sequence for BAR1 and BAR2. The difference in fragment size allows us to distinguish barcodes downstream after DNA amplification (Figure 4).

Each barcode is associated with two specific primers that are designed to anneal to a specific region in the barcode; one primer is designed to anneal at the 5'-end, called forward primer. The other primer, called reverse primer, is designed to anneal at the 3'-end of the barcode segment. The forward and reverse primers of BAR1 are denoted by BAR1F and BAR1R, respectively. The forward and reverse primers of BAR2 are denoted by BAR2F and BAR1R, respectively. The nucleotide sequence information for BAR1F, BAR1R, BAR2F, and BAR2R are shown in Table I.

B. Identification of Barcodes Using PCR

In order to detect the presence of a molecular barcode in a sample, we use traditional PCR to count the number of bp using agarose gel electrophoresis, as described in Sections II-B-II-D. By using the right type of primers in the PCR reaction, the barcode is amplified, allowing us to verify its presence in the associated reaction. Since BAR1 has a longer sequence of bp compared to BAR2, we expect that BAR2 amplicons migrate farther through the gel than BAR1 amplicons.

[4]A public key is known to everyone and it is tightly coupled with a private key, which is only known to an analyzer.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2020.2994742, IEEE Transactions on Information Forensics and Security

TABLE I: The sequence information for BAR1 and BAR2.

BAR1	5' - CAG TCA CTA TGG CGT GCT GCT AGC GCT ATA TGC GTT GAT GCA ATT TCT ATG CGC ACC CGT TCT CGG AGC ACT GTC CGA CCG CTT TGG CCG CCG CCC AGT CCT GCT CGC TTC GCT ACT TGG AGC CAC TAT CGA CTA CGC GAT CAT GGC GAC CAC ACC CGT CCT GTG GAT CCT CTA CGC CGG ACG CAT CGT GGC CGG CAT CAC CGG CGC CAC AGG TGC GGT TGC TGG CGC CTA TAT CGC CGA CAT CAC CGA TGG GGA AGA TCG GGC TCG CCA CTT CGG GCT C - 3'
BAR2	5' - GAC ATG AAG CTT TAA ATC AAT CTA AAG TAT ATA TGA GTA AAC TTG GTC TGA CAG TTA CCA ATG CTT AAT CAG TGA GGC ACC TAT CTC AGC GAT CTG TCT ATT TCG TTC ATC CAT AGT TGC CTG ACT CCC CGT CGT GTA GAT AAC TAC GAT ACG GGA GGG CTT ACC ATC TGG CCC CAG TGC TGC AAT GAT A - 3'
BAR1F	5' - CGC TAT ATG CGT TGA TGC AA - 3'
BAR1R	5' - AGA TGG TAA GCC CTC CCG TAT - 3'
BAR2F	5' - TGC TTA ATC AGT GAG GCA CCT - 3'
BAR2R	5' - AGA TGG TAA GCC CTC CCG TAT - 3'

TABLE II: Description of PCR reactions designed in the barcoding experiment.

Reaction	Tube	Sample 1	Sample 2	Barcodes BAR1	Barcodes BAR2	Primers BAR1F(R)	Primers BAR2F(R)
1	1			✓(1:10)		✓	
	2			✓(1:100)		✓	
	3			✓(1:1000)		✓	
2	1				✓(1:10)		✓
	2				✓(1:100)		✓
	3				✓(1:1000)		✓
3	1			✓(1:100)	✓(1:100)	✓	✓
4	1						
	2	✓		✓(1:100)			
	3		✓		✓(1:100)		
5	1						
	2	✓		✓(1:100)		✓	
	3		✓		✓(1:100)		✓
6	1						
	2	✓		✓(1:100)	✓(1:100)	✓	
	3		✓	✓(1:100)	✓(1:100)		✓
7	1						
	2	✓		✓(1:100)			✓
	3		✓		✓(1:100)	✓	

C. Experiment Procedure

The purpose of our experiment is twofold. First, we verify that BAR1 and BAR2 exhibit different characteristics (fragment size) that can be identified using gel electrophoresis. We achieve this objective without using actual DNA samples, i.e., only using barcodes. Second, we demonstrate that after a sample is barcoded, the amplification of the DNA barcode can occur *only if* the right primers are used in the reaction. This important demonstration, despite its technical challenges, confirms that molecular barcoding functions secure samples against forgery.

Experiment Design: To achieve the above goals, the experiment was designed by considering 7 parallel PCR reactions. We describe these reactions and their associated objectives in Table II and Table III, respectively.

Sample and Reagent Preparation: To start the experiment, we first prepared the DNA samples, molecular barcodes, and the reagents. We used two types of DNA samples, and they were both grown in yeast media so that they could be presented in the experiment in a liquid form (Figure 8(a)). Second, we

TABLE III: Objectives of PCR reactions.

Reaction	Objective
1	Verify PCR amplification of BAR1 and BAR2 at different dilutions. Verify the difference in fragment size.
2	
3	Verify the ability to use multiple "distinguishable" barcodes.
4	Verify that a barcode is not amplified if its specific primers are not used (the template is a barcoded sample).
5	Verify that a barcode is amplified if its specific primers are used (the template is a barcoded sample).
6	Verify that a barcode is amplified if its specific primers are used even if they are used along with other primers (the template is a barcoded sample).
7	Verify that a barcode is not amplified if the used primers are not the right ones (the template is a barcoded sample).

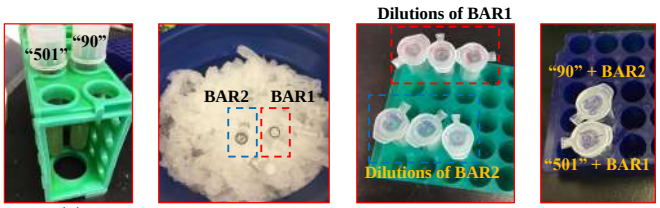


Fig. 8: Preparation of samples and reagents for barcoding.

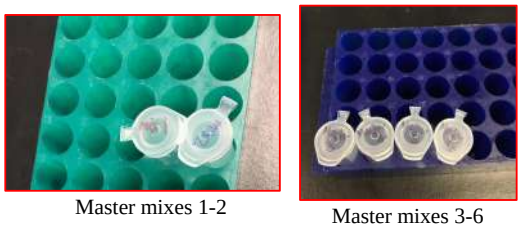


Fig. 9: Six tubes containing master mixes.

prepared the barcodes BAR1 and BAR2, shown in Figure 8(b), by serial dilution for each barcode. The dilutions we obtained for each barcode were 1:10, 1:100, and 1:1000 as shown in Figure 8(c). Third, we prepared barcoded samples by spiking the first DNA sample, labeled as 501, with BAR1 (1:100) and the second DNA sample, labeled as 90, with BAR2 (1:100), as shown in Figure 8(d). The diluted barcodes (Figure 8(c)) were used to implement the first and second PCR reactions, and the barcoded samples (Figure 8(d)) were used to implement the remaining reactions. We added only barcodes in Reaction 1 to Reaction 3; we added samples as well as barcodes in Reaction 4 to Reaction 7. We added sample-specific primers in Reaction 4 to Reaction 7.

Next, for each PCR reaction, we prepared a PCR master mix, which consists of primers (of DNA sample and/or barcodes), buffer, water, and Taq solution. The amounts of these components were adjusted based on the requirements of each PCR reaction. These components were pipetted into 6 master-mix tubes, as shown in Figure 9. Note that Reaction 3 is similar to Reaction 1 and Reaction 2, therefore no specific master mix was designed for Reaction 3.

Having prepared the samples and PCR master mixes, we loaded all the chemicals into 6 strips of empty tubes on a PCR plate; see Figure 10(a). Each 8-tube strip was associated with a PCR reaction, except the second strip which included Reaction 2 and Reaction 3. In each strip, we only used three out of eight tubes, as described in Table II.

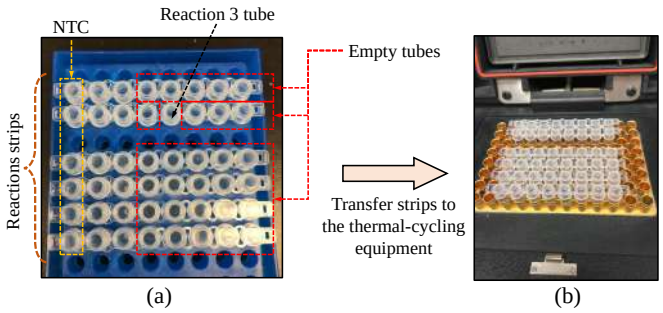


Fig. 10: (a) Preparation of the PCR plate. (b) Moving PCR strips to the thermal-cycling equipment.

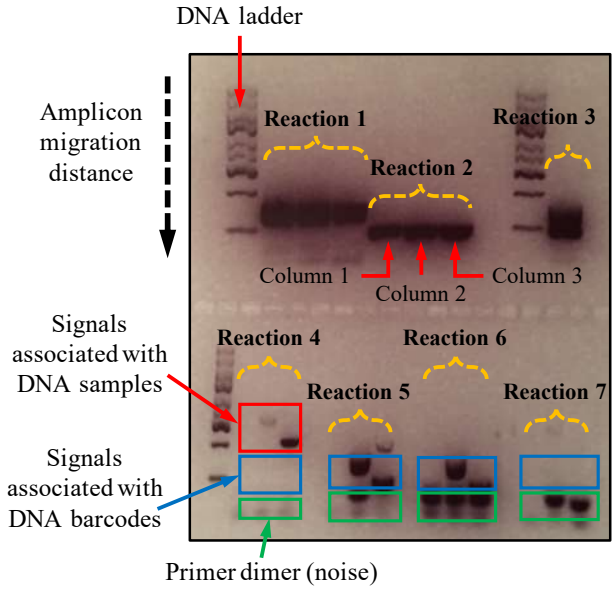


Fig. 11: Results of DNA amplification for the seven PCR reactions.

PCR Reaction and Gel Electrophoresis: Having filled in the assigned tubes on the PCR plate, the tubes were closed then spun for a few seconds using a micro-centrifuge. Next, we placed the strips in the thermal cycler and started the PCR reaction; see Figure 10(b). After the PCR program had finished, we transferred the PCR-plate contents into the gel electrophoresis apparatus, allowing the chemical solutions to migrate over the gel with the aid of an electric field. Gel electrophoresis was run for an hour, then the gel block was moved to a UV to a chamber that contains a source of UV light to visualize the DNA bands. Results interpretation is discussed next.

D. Interpretation of Results

Figure 11 shows the results obtained by gel electrophoresis. The plot is divided into 7 regions. Region 1 and Region 2 show the results of Reaction 1 and Reaction 2, respectively. In other words, they are associated with the serial dilution for BAR1 and BAR2, respectively. Region 3 shows the result for a combined BAR1/BAR2 reaction, i.e., Reaction 3. The results of Region 1, Region 2, and Region 3 are located at the top row.

Furthermore, the results shown in Region 4, Region 5, Region 6, and Region 7 (bottom row) are used to assess the barcoding performance in the presence of DNA samples, and they are associated with PCR Reaction 4, Reaction 5, Reaction 6, and Reaction 7, respectively. Each one of these regions contains 3 columns, where the leftmost column is related to the reference negative control, the middle and the rightmost columns are associated with DNA samples 501 and 90, respectively.

By comparing Region 1 and Region 2, we observe that the bands in Region 2 are located at a further distance compared with the bands in Region 1. This observation indicates that both BAR1 amplicons and BAR2 amplicons can be distinguished because of their different sizes, even if they are both included in a single tube (Region 3). The bands of BAR2 are located at a further distance since BAR2 (190 bp) has a shorter sequence of base pairs than BAR1 (280 bp).

The bands shown in Region 1, Region 2, and Region 3 are significantly bright, even after diluting the barcodes. This indicates that the copy number of DNA fragments that exists in BAR1 and BAR2, i.e., concentration, is significantly large. This result raises questions concerning the optimization of the copy numbers of a starting DNA sample and a molecular barcode when they co-exist in one tube. In other words, the DNA samples 501 and 90 produce amplicons of length 400 700 bp, whereas BAR1 and BAR2 produce amplicons of length 190 280 bp. With such a difference in molecular weights, a microliter volume of a DNA sample contains a lower copy number than a microliter volume of a barcode. Hence, by introducing an excessive number of barcode amplicons, the yield of the sample amplicons is significantly reduced. This is why we observe that the bands related to the sample amplicons are faint, as shown in Region 4, or they completely vanish, as shown in Regions 5-7. Discussion related to the optimization of copy numbers is presented later in this section.

In Region 4, no barcode bands are observed even though DNA barcodes were used in Reaction 4. This is because no barcode-specific primers were used in the reaction. Similarly, in Region 7, no barcodes are observed even though barcodes and primers were used. This is because we deliberately added the wrong primers—we used BAR1-specific primers when we added BAR2 and vice versa. These results indicate that the detection of a certain barcode is possible only if a specific set of primers are used.

In contrast, in Region 5, we observe the bands related to the barcodes as the right type of primers were used. In Region 6, we observe bands related to the barcodes since each tube contained all types of primers; a DNA barcode will bind with its specific primers during annealing.

Hence, despite the technical challenges related to the optimization of amplicon copy numbers, the above results show the effectiveness of the molecular-barcoding scheme in identifying samples and securing them against forgery. We carried out additional benchtop studies, based on serial dilution, to co-optimize DNA amplification for both DNA samples and barcodes.

E. Co-Optimization of DNA Samples and Barcodes Using Serial Dilution

To co-optimize DNA amplification for both DNA samples and barcodes, we performed two PCR experiments for a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2020.2994742, IEEE Transactions on Information Forensics and Security

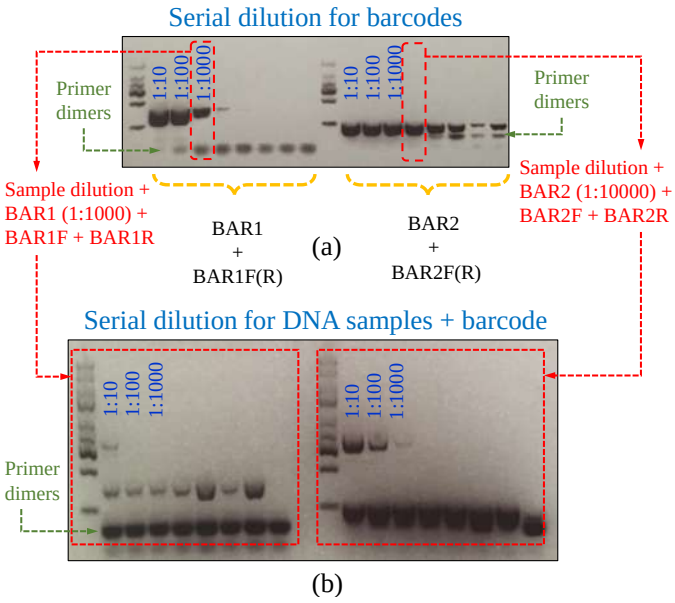


Fig. 12: Results of serial-dilution experiments that investigate the co-optimization of DNA amplification for DNA samples and molecular barcodes.

10-fold serial dilution of a genomic DNA sample and the barcodes BAR1 and BAR2. The goal of these experiments was to determine the optimum concentrations of the DNA sample and the barcodes such that the overall reaction yield is improved. In the first experiment, we prepared a serial dilution of BAR1 and BAR2 (similar to Reaction 1 and Reaction 2 in the previous experiment) followed by DNA amplification and gel-electrophoresis analysis. The outcome of this experiment is shown in Figure 12(a). The optimal dilution factors for BAR1 and BAR2 that result in clear bands with reduced primer dimers (noise) are 1:1000 and 1:10000, respectively.

Having determined the optimum dilution factors for the molecular barcodes, we conducted a second PCR experiment over a serial dilution of a multi-template setting, i.e., a DNA sample barcoded with (1:1000) BAR1 or (1:10000) BAR2. The result obtained from gel electrophoresis after running this experiment is shown in Figure 12(b). We observe that using dilution factors of (1:100) or lower for the DNA sample may degrade the sample yield, thus causing the bands associated with the sample amplicons to vanish. As a result, to co-optimize the DNA-amplification profile for the DNA sample and barcodes based on the above setting, the following conditions need to be fulfilled: (1) the volumes of the DNA sample and the barcode used in the same PCR reaction are equal; (2) by using (1:1000) BAR1 and (1:10000) BAR2, the DNA sample should not be diluted lower than 1:10 in order to get clear bands for both templates. This result is intuitive since the size of the sample amplicons that we used in the experiments is larger than the size of the barcode amplicons.

A different approach to co-optimize DNA amplification for the barcoded samples is to fix the dilution factors for both templates while examining the impact of using different volumes of DNA sample and the barcode. However, this approach is complex, and therefore, we did not consider it in this study.

VI. Security Analysis

In the threat model described in Section IV, an adversary can forge the sample when it is transferred from the collector to the analyzer. When the multi-space security scheme is used, it is difficult to forge samples during transit without being detected by the analyzer. This is because, the attacker must also forge the added barcodes. A forged barcode must satisfy two constraints: (1) it must use the primers as those for the legitimate barcode so that the forged barcode will be amplified using PCR in the analysis. (2) The forged barcode must have the same length as the legitimate barcode so that the gel-electrophoresis results of the forged barcode are the same as that of the legitimate barcode.

When the barcoded sample is transferred from the collector to the analyzer, an attacker can obtain (1) the encrypted message of the barcode, and (2) the biological sample that contains the barcode and the DNA sample (see Figure 7). One way the attacker can thwart the security scheme by breaking into the encrypted message. However, the message is encrypted using a standard public-key encryption algorithm such as RSA or ECC[5] [41], [42]. Since RSA is widely used and is a certified standard, e.g., IEEE P1316 [43], it is unlikely that the attacker can gain information from the RSA encrypted message.

Alternatively, the attacker may try to extract the barcode information from the biological material. To do so, the initial goal of the attacker is to guess the two primers of the barcode. The attacker can run PCR using the barcoded sample as well as the primers and then discover the length of the barcode. Here, we consider two classes of attackers: (1) a novice attacker with limited knowledge of biology, and (2) an attacker that is an expert in genomics.

The novice attacker who is not an expert in biology may randomly guess the sequence of the two primers. To successfully guess the sequence of a primer, the attacker needs to guess two factors: the primer length and the order of nucleotides. Let L_p be the event that the attacker guesses the right primer length and S be the event that the attacker guesses the right nucleotide order. The probability p_{pmr} that an attacker can successfully guess a primer is:

$$P_{pmr} = P(L_p) \cdot P(S|L_p) \tag{1}$$

Because the primer length ranges from 18 bp to 25 bp, $P(L_p) = \frac{1}{25-18+1} = \frac{1}{8}$. Each primer is composed of four different nucleotides. Thus, $P(S|L_p) = \frac{1}{4^l}$ where l is the length of the primer, and $P_{pmr} = \frac{1}{2^{3+2l}}$. In practice, the attacker will follow a "sampling without replacement" strategy. When the attacker fails to guess the primer in the first k attempts, the probability is adjusted to $P_{pmr}(k) = \frac{1}{2^{3+2l}-k}$. Because l is between 18 and 25, the probability of P_{pmr} and the probability of $P_{pmr}(k)$ are negligible and similar. The probability is always smaller than 1.8×10^{-12} for $l \geq 18$.

[5]RSA and ECC are vulnerable to quantum computer based attacks. In that case, one can consider using quantum-resistant cryptography that is currently being standardized by NIST [40]

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2020.2994742, IEEE Transactions on Information Forensics and Security

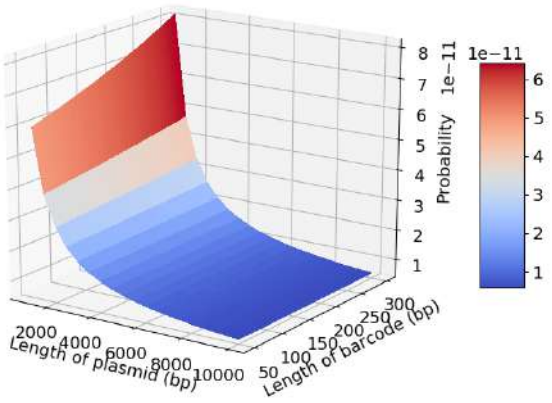


Fig. 13: The probability P_{bd} as a function of the barcode length and the plasmid size.

Example: In the experiment described in Section V, our primers are 21 bp long. The probability that the attacker can guess the two primers correctly is $P^2_{pmr} = (\frac{1}{2^{3+2\times 21}})^2 = \frac{1}{2^{90}}$.

An expert attacker, e.g., a biologist, may discover the barcode in a more sophisticated way. Since the security scheme is used to secure the human genome, the attacker may infer that the barcode design uses plasmid vectors that are widely used in molecular biology[6]. There are three factors that the attacker has to consider to guess the barcode. (1) The attacker needs to pick the right plasmid vector. According to a non-profit plasmid repository, there are over $70,000$ plasmid vectors that molecular biologists use [44]. Let PV be the event that the attacker picks the right plasmid vector. Therefore, $P(PV) = \frac{1}{70,000}$. (2) The attacker needs to guess the length of the barcode. A barcode can have 50–300 bp. The barcodes must be small enough so that they can be easily amplified using PCR on one hand and large enough to be visualized on the agarose gel. Assuming that L_b is the event that the attacker guesses the length of the barcode, $P(L_b) = \frac{1}{250+1} = \frac{1}{251}$. (3) The attacker needs to choose the right segment from the plasmid sequence. Let SG be the event that the attacker chooses the right segment of a plasmid vector. Assuming that the chosen plasmid sequence is of length l_{pld} and the barcode length is l_b, the probability $P(SG \mid PV \cap L_b) = \frac{1}{l_{pld}-l_b+1}$. As a result, the probability P_{bd} that the attacker discovers the barcode is described as:

$$\begin{aligned} P_{bd} &= P(PV) \cdot P(L_b) \cdot P(SG \mid PV \cap L_b) \\ &= \frac{1}{1.757 \times 10^7 \cdot (l_{pld} - l_b + 1)} \end{aligned} \tag{2}$$

Figure 13 shows the probability given plasmid length and a barcode length. For $l_{pld} > 1000$, the probability is smaller than 8×10^{-11}.

Example: In the experiment in Section V, we used a segment of 190 bp from the plasmid vector pBR332. Therefore, $P(L_b) = \frac{1}{251}$. The vector pBR332 was one of the first widely used cloning vectors in molecular biology, and it is $4,361$ bp long. Therefore, $P(SG \mid P \cap L_b) = \frac{1}{4361-190+1} = \frac{1}{4172}$.

[6]Plasmids are small DNA molecules that are physically separated from chromosomal DNA in cells. The plasmid sequences are unique in a way such that they are not found in human or mouse genomes.

TABLE IV: Probability that an attacker discovers a barcode that uses a specific plasmid with a barcode length choice.

Plasmid	Barcode Length (bp)		
	50	175	300
pUC18	2.16×10^{-11}	2.26×10^{-11}	2.38×10^{-11}
pBR322	1.32×10^{-11}	1.36×10^{-11}	1.40×10^{-11}
pTS36	7.33×10^{-12}	7.45×10^{-12}	7.57×10^{-12}
pAA31	6.43×10^{-12}	6.52×10^{-12}	6.61×10^{-12}
pBIN19	4.85×10^{-12}	4.91×10^{-12}	4.96×10^{-12}

The probability that the attacker can guess the barcode is $P_{bd} = \frac{1}{70,000} \times \frac{1}{251} \times \frac{1}{4172} \approx \frac{1}{7.33\times 10^{10}}$.

To examine how the forensic-security strength depends on the plasmid choice and barcode length, we ran a simulation based on several widely-used plasmid vectors, namely, pUC18, pBR322, pTS36, pAA31, and pBIN19. The sequence of the vectors can be found at the plasmid repository in [44]. The lengths of the plasmid vectors range from $2,696$ bp to $11,777$ bp. The probability that an attacker can discover a specific barcode is obtained using Equation (2). The length of the barcode is between 50 bp and 300 bp. Table IV shows the simulation results. No matter what type of plasmid vector is used to design the barcode, the probability that an attacker can discover the barcode is less than 2.5×10^{-11}.

Note that in the threat model described in Section IV, our molecular barcoding does not embed the fingerprints of the genuine sample in the barcode, i.e., it does not implement bio-PUFs. The proposed scheme only helps in protecting genuine samples but does not help in identifying forged ones. As shown in Fig. 7, molecular barcoding relies on the fact that the secret ID of the sample/barcode is transmitted via a trusted channel.

VII. Discussion

In this section, we present comparison with prior work. We also discuss the resilience of molecular barcoding against DNA sequencing-based attacks. Next, we discuss how to optimize molecular barcoding. Finally, we consider how the proposed scheme can be used in today's forensic workflows.

A. Comparison with Prior Work

An authentication assay has been proposed in information forensics to distinguish between a human DNA sample and DNA synthesized in the laboratory [45]. This work points out that DNA with any desired genetic profile can be synthesized *in vitro* and then used to replace the genuine DNA samples collected from crime scenes [46], [47]. The experimental results in [45] demonstrate that an authentication assay that monitors the methylation state of a DNA sample can distinguish between a human DNA sample and DNA synthesized in the laboratory. However, an authentication assay that monitors DNA methylation is susceptible to the attack model described in Section III. A stealthy attacker can replace (or swap) a DNA sample collected at the crime scene with a DNA sample isolated from an innocent person. Both DNA samples are methylated and thus, the authentication assay will deem the replacement sample to be authentic. Therefore, this defense can only thwart a subset of the sample-forgery.

Our defense can defend against many threats, including the attack described in [45]. When a PDoS attack is launched,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIFS.2020.2994742, IEEE Transactions on Information Forensics and Security

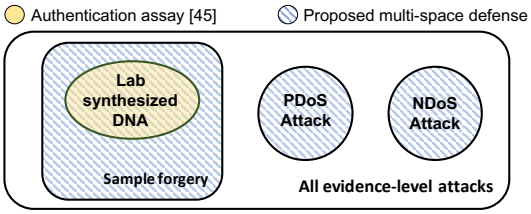


Fig. 14: Comparison between the state-of-art defense and our multi-space defense. The state-of-art defense can only detect lab-synthesized DNA forgery [45]; whereas, our defense can thwart all the attacks described in Section III.

the gel results will show an extra band when compared with a normal case. Therefore, the analyzer can easily detect such the attack. Similarly, when an NDoS attack is launched, the analyzer will not be able to see the band associated with the barcode in the gel results. Thus, the attack is detected. As shown in the experiments in Section V, our defense can detect the sample-forgery. Figure 14 shows the comparison between the work in [45] and our defense.

B. Sequencing Attacks for Molecular Barcoding

Recent advances in sequencing technology, such as whole-genome sequencing, have enabled scientists to sequence the whole genome of a sample without any prior knowledge [48]. However, obtaining the sequence of a barcode, which is in the mixture of crime-scene samples, may be impractical because the sequencing technique is time-consuming and expensive. Consider that, in order to obtain a barcode sequence, an attacker employs advanced sequencing technology on the solution acquired from the field (which contains the added barcode). This technology first breaks the sample genome as well as the barcode into small fragments, where each fragment has ~ 100 bps. Note that the genome of a human is approximately three billion bps. After sequencing, the attacker acquires the sequences of millions of fragments corresponding to the sample and the barcode. Therefore, the challenge of acquiring the barcode sequence lies in identifying the fragments from the barcode and assembling the fragments (in the right order) as the human genome and the barcode. Although it is technically possible for an expert in genomics with powerful software tools to understand the barcode sequence, the overall process is time-consuming and expensive. According to a report from NIH [49], the cost of sequencing the whole genome of a human is approximately $1,000 (USD) in 2019. To further increase the security strength of molecular barcoding, we can add more than one barcode in various lengths in the sample. After the advanced sequencing, the barcodes are cut into several fragments. Therefore, in order to discover the barcode sequences, the attacker first needs to reveal the number of barcodes used in the sample and the length of each barcode. As a result, it is impractical for an attacker to use this advanced sequencing technology to discover the barcodes.

C. Optimization for Molecular Barcoding

Molecular barcoding is a powerful technique that allows scientists to authenticate DNA samples *in vivo* even before genomic analysis starts. A major advantage of this technique is that designing robust molecular barcodes and relevant DNA primers that amplify them is easy. By using PCR, DNA primers anneal specifically to the barcode template based on its inherent DNA sequence and annealing temperatures. Thus, features of the barcode and paired primers can be optimized to improve forensic security. Our benchtop experiments showed that primer annealing depends on the barcode sequence and primer length, and any mismatches in these variables will reflect in the final result. This feature makes the barcoding immune to stealthy attacks, which maliciously change either environmental conditions of the barcoding or the genomic characteristics of the collected DNA sample, but without causing a denial of service.

Our benchtop results (Section V-E) showed that the multiplexing of these barcodes with isolated DNA samples may require systematic optimizations for some experimental settings to avoid interference of signals. This is due to the fact that optimum conditions for barcode detection using the PCR technique depend on the barcode/primer pair, as we did not observe robust amplicon signals for both the barcode and the genomic target in a single PCR reaction. Our ongoing research is focused on developing an optimization framework that adapts the experimental settings based on the barcode specifications. These settings include variables such as the optimal concentration values for the DNA sample and the barcode, and the annealing temperature profile of the PCR. This optimization framework will open the door for developing an automated, trustworthy DNA-fingerprinting system using microfluidics.

D. Integration in Today's Forensic Workflows

We consider two possible ways in which we can add the barcodes to the samples collected from a crime scene in forensic workflow. One possible way is that sample tubes have been pre-loaded with a barcode solution. The sample tubes are taken into the field, and then the crime-scene evidence is added directly to the tube with a sterilized swab. Alternatively, collection tubes could be prepared with lyophilized barcode(s), which would increase the stability of the barcode while in the field [50].

In this paper, we validated the presence/absence of added barcodes using gel electrophoresis. Our benchtop approach is meant to be simple and easy to understand and replicate, especially for security experts. The implemented concepts can be easily mapped to different technologies, e.g., gel electrophoresis, capillary electrophoresis (CE), or DNA sequencing, while maintaining the same security measures. As shown in Fig. 7, the security flow still requires molecular barcodes, corresponding primers, and the encrypted IDs. Today's forensic laboratories often use CE instruments to analyze short tandem repeat (STR) profiles for identification [51]. The main difference between using gel electrophoresis and using CE is that STR compares the presence/absence or number of repeats in a given sample to a control sample–or other samples-and asks whether the STR patterns match. Our barcodes are sequences that are not contained within the human (or mouse, Drosophila, yeast, etc) genomes and are thus "unique". Therefore, the barcode approach could be used with STR profiling,

in which barcode analysis is a first check to ascertain whether a sample has been tampered with (attacked or switched), and then once confirmed, STR analysis can be performed.

VIII. Conclusion

A major barrier toward a wide adoption of DNA-fingerprinting flows in criminal justice and bio-defense applications is the lack of trust and security measures. This paper takes the first steps in addressing the forensic-security implications of DNA-fingerprinting flows at the biochemistry level. We demonstrated a benchtop study that implements real-life evidence-level attacks. An effective countermeasure mechanism, based on molecular barcoding, is proposed and implemented. We present a forensic-security analysis of the molecular barcoding. The analytical results show that it is unlikely that an attacker, even an experienced biologist, can discover the molecular barcode. These results open the door to a new line of research on securing biomolecular systems.

References

[1] P. Gill, A. J. Jeffreys, and D. J. Werrett, "Forensic application of DNA 'fingerprints'," *Nature*, vol. 318, no. 6046, pp. 577–579, 1985.

[2] K. Norrgard, "Forensics, DNA fingerprinting, and CODIS," *Nature Education*, vol. 1, no. 1, p. 35, 2008.

[3] J. G. Shewale and R. H. Liu, *Forensic DNA Analysis: Current Practices and Emerging Technologies*. CRC Press, 2013.

[4] J. M. Bartlett and D. Stirling, "A short history of the polymerase chain reaction," *PCR protocols*, pp. 3–6, 2003.

[5] B. Bruijns, A. van Asten, R. Tiggelaar, and H. Gardeniers, "Microfluidic devices for forensic DNA analysis: A review," *Biosensors*, vol. 6, no. 3, p. 41, 2016.

[6] M. A. Jobling and P. Gill, "Encoded evidence: DNA in forensic analysis," *Nature Reviews Genetics*, vol. 5, no. 10, pp. 739–751, 2004.

[7] W. Jung, J. Han, J.-W. Choi, and C. H. Ahn, "Point-of-care testing (POCT) diagnostic systems using microfluidic lab-on-a-chip technologies," *Microelectronic Engineering*, vol. 132, pp. 46–57, 2015.

[8] K. Hsieh, B. S. Ferguson, M. Eisenstein, K. W. Plaxco, and H. T. Soh, "Integrated electrochemical microsystems for genetic detection of pathogens at the point of care," *Accounts of Chemical Research*, vol. 48, no. 4, pp. 911–920, 2015.

[9] R. S. Murch, W. K. So, W. G. Buchholz, S. Raman, and J. Peccoud, "Cyberbiosecurity: An emerging new discipline to help safeguard the bioeconomy," *Frontiers in Bioengineering and Biotechnology*, vol. 6, p. 39, 2018.

[10] E. Pauwels and A. Vidyarthi, "How Our Unhealthy Cybersecurity Infrastructure is Hurting Biotechnology," https://www.wilsoncenter.org/publication/how-our-unhealthy-cybersecurity-infrastructure-hurting-biotechnology, 2016, [Online; posted 29-March-2016].

[11] J. Peccoud, J. E. Gallegos, R. Murch, W. G. Buchholz, and S. Raman, "Cyberbiosecurity: From naive trust to risk awareness," *Trends in Biotechnology*, vol. 36, no. 1, pp. 4–7, 2018.

[12] P. Ney, K. Koscher, L. Organick, L. Ceze, and T. Kohno, "Computer security, privacy, and DNA sequencing: Compromising computers with synthesized DNA, privacy leaks, and more," in *USENIX Security Symposium*, 2017, pp. 765–779.

[13] P. Liu, X. Li, S. A. Greenspoon, J. R. Scherer, and R. A. Mathies, "Integrated DNA purification, PCR, sample cleanup, and capillary electrophoresis microchip for forensic human identification," *Lab on a Chip*, vol. 11, no. 6, pp. 1041–1048, 2011.

[14] N. R. Council, "Strengthening forensic science in the United States: A path forward," 2009.

[15] L. Wilson-Wilde, "The international development of forensic science standards—a review," *Forensic Science International*, vol. 288, pp. 1–9, 2018.

[16] M. Gamette and K. M. Reidy, "Raising the bar: Progress and future needs in forensic science," https://news.aafs.org/wp-content/uploads/2019/09/Matthew-Gamette-Science-Committee-Testimony-9-10-19-updated-9-13-19.pdf, 2019.

[17] K. G. Kozminski, "Biosecurity in the age of big data: a conversation with the FBI," *Molecular Biology of the Cell*, vol. 26, no. 22, pp. 3894–3897, 2015.

[18] P. Y. Lee, J. Costumbrado, C.-Y. Hsu, and Y. H. Kim, "Agarose gel electrophoresis for the separation of DNA fragments," *Journal of Visualized Experiments: JoVE*, no. 62, 2012.

[19] P. D. Grossman and J. C. Colburn, *Capillary electrophoresis: Theory and practice*. Academic Press, 2012.

[20] E. Y. Chan *et al.*, "Dna mapping using microfluidic stretching and single-molecule detection of fluorescent site-specific tags," *Genome Research*, vol. 14, no. 6, pp. 1137–1146, 2004.

[21] I. Pickrahn, G. Kreindl, E. Müller, B. Dunkelmann, W. Zahrer, J. Cemper-Kiesslich, and F. Neuhuber, "Contamination incidents in the pre-analytical phase of forensic dna analysis in Austria—Statistics of 17 years," *Forensic Science International: Genetics*, vol. 31, pp. 12–18, 2017.

[22] A. Giaretta, S. Balasubramaniam, and M. Conti, "Security vulnerabilities and countermeasures for target localization in bio-nanothings communication networks," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 4, pp. 665–676, 2015.

[23] S. S. Ali, M. Ibrahim, O. Sinanoglu, K. Chakrabarty, and R. Karri, "Microfluidic encryption of on-chip biochemical assays," in *IEEE Biomedical Circuits and Systems Conference*, 2016, pp. 152–155.

[24] S. Bhattacharjee, J. Tang, M. Ibrahim, K. Chakrabarty, and R. Karri, "Locking of biochemical assays for digital microfluidic biochips," in *IEEE European Test Symposium*, 2018, pp. 1–6.

[25] T.-C. Liang, M. Shayan, K. Chakrabarty, and R. Karri, "Secure assay execution on meda biochips to thwart attacks using real-time sensing," *ACM Transactions on Design Automation of Electronic Systems*, vol. 25, no. 2, pp. 1–25, 2020.

[26] Z. Zhong, Z. Li, K. Chakrabarty, T.-Y. Ho, and C.-Y. Lee, "Micro-electrode-dot-array digital microfluidic biochips: Technology, design automation, and test techniques," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 2, pp. 292–313, 2018.

[27] S. S. Ali, M. Ibrahim, O. Sinanoglu, K. Chakrabarty, and R. Karri, "Security assessment of cyberphysical digital microfluidic biochips," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 13, no. 3, pp. 445–458, 2016.

[28] J. Tang, M. Ibrahim, K. Chakrabarty, and R. Karri, "Towards secure and trustworthy cyberphysical microfluidic biochips," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2018.

[29] M. Shayan, S. Bhattacharjee, J. Tang, K. Chakrabarty, and R. Karri, "Bio-protocol watermarking on digital microfluidic biochips," *IEEE Transactions on Information Forensics and Security*, 2019.

[30] S. S. Shringarpure and C. D. Bustamante, "Privacy risks from genomic data-sharing beacons," *The American Journal of Human Genetics*, vol. 97, no. 5, pp. 631–646, 2015.

[31] M. A. Majumder, R. Cook-Deegan, and A. L. McGuire, "Beyond our borders? Public resistance to global genomic data sharing," *PLoS Biology*, vol. 14, no. 11, p. e2000206, 2016.

[32] H. Cho, D. J. Wu, and B. Berger, "Secure genome-wide association analysis using multiparty computation," *Nature Biotechnology*, vol. 36, no. 6, p. 547, 2018.

[33] K. Bolden, "DNA fabrication, a wake up call: The need to reevaluate the admissibility and reliability of DNA evidence," *Ga. St. UL Rev.*, vol. 27, pp. 409–1155, 2011.

[34] S. A. Bustin, *A-Z of Quantitative PCR, 1st Edition*. San Diego, CA: International University Line, 2004.

[35] R. Higuchi, C. Fockler, G. Dollinger, and R. Watson, "Kinetic PCR analysis: real-time monitoring of dna amplification reactions," *Bio/technology*, vol. 11, no. 9, p. 1026, 1993.

[36] P. D. Hebert, A. Cywinska, S. L. Ball, and J. R. DeWaard, "Biological identifications through DNA barcodes," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 270, no. 1512, pp. 313–321, 2003.

[37] W. Stallings, *Cryptography and Network Security: Principles and Practice*. Pearson Education India, 2003.

[38] K. Ermoshina, F. Musiani, and H. Halpin, "End-to-end encrypted messaging protocols: An overview," in *International Conference on Internet Science*, 2016, pp. 244–254.

[39] S. Baur, H. Boche, R. F. Schaefer, and H. V. Poor, "Secure storage capacity under rate constraints—continuity and super activation," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 959–970, 2019.

[40] L. Chen, L. Chen, S. Jordan, Y.-K. Liu, D. Moody, R. Peralta, R. Perlner, and D. Smith-Tone, *Report on post-quantum cryptography*. US Department of Commerce, National Institute of Standards and Technology, 2016.

[41] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Communications of the ACM*, vol. 21, no. 2, pp. 120–126, 1978.

[42] V. S. Miller, "Use of elliptic curves in cryptography," in *Conference on the theory and application of cryptographic techniques*. Springer, 1985, pp. 417–426.

[43] B. Kaliski and J. Staddon, "PKCS# 1: RSA cryptography specifications version 2.0," RFC 2437, October, Tech. Rep., 1998.

[44] J. Kamens *et al.*, "Addgene: nonprofit plasmid repository," https://www.addgene.org, 2019.

[45] D. Frumkin, A. Wasserstrom, A. Davidson, and A. Grafit, "Authentication of forensic DNA samples," *Forensic Science International: Genetics*, vol. 4, no. 2, pp. 95–103, 2010.

[46] R. S. Lasken and M. Egholm, "Whole genome amplification: abundant supplies of dna from precious samples or clinical specimens," *Trends in Biotechnology*, vol. 21, no. 12, pp. 531–535, 2003.

[47] S. Panelli, G. Damiani, L. Espen, G. Micheli, and V. Sgaramella, "Towards the analysis of the genomes of single cells: further characterisation of the multiple displacement amplification," *Gene*, vol. 372, pp. 1–7, 2006.

[48] M. Kuroda, *et al.*, "Whole genome sequencing of meticillin-resistant staphylococcus aureus," *The Lancet*, vol. 357, no. 9264, pp. 1225–1240, 2001.

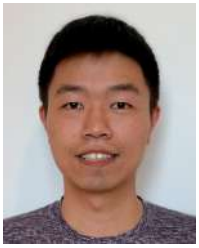
[49] W. Kris, "The cost of sequencing a human genome," https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost, 2019.

[50] J. L. Cleland *et al.*, "A specific molar ratio of stabilizer to protein is required for storage stability of a lyophilized monoclonal antibody," *Journal of Pharmaceutical Sciences*, vol. 90, no. 3, pp. 310–321, 2001.

[51] P. Liu *et al.*, "Integrated portable polymerase chain reaction-capillary electrophoresis microsystem for rapid forensic short tandem repeat typing," *Analytical Chemistry*, vol. 79, no. 5, pp. 1881–1889, 2007.

Mohamed Ibrahim received the B.S. and the M.Sc. degrees in electrical engineering from Ain Shams University in 2010 and 2013, respectively, and also the M.Sc. and the Ph.D. degrees in electrical and computer engineering from Duke University in 2017 and 2018, respectively. He is now a System-on-Chip Design Engineer at Intel Corporation, Santa Clara, CA. His current research interests include secure design automation, synthesis, and IoT connectivity of biomedical cyber-physical systems. During his Ph.D. time at Duke University, Dr. Ibrahim developed design and optimization methodologies for cyber-physical microfluidic biochip (CPMs) to support scalable biomolecular quantitative analysis, and he also co-started the research on the security, trust, and IoT connectivity of CPMs. He co-authored two books, contributed to three book chapters, and also published over 35 papers in premier journals, such as *Proceedings of the IEEE*, and refereed conference proceedings.

Dr. Ibrahim is the recipient of the 2018 Council of Graduate Schools/ProQuest Distinguished Dissertation Award in Mathematics, Physical Sciences, and Engineering; the 2019 Outstanding Dissertation Award from the Department of Electrical and Computer Engineering at Duke University; the Best Paper award at the 2017 IEEE/ACM DATE; the 2017 Postdoc Mobility award from the Technical University of Munich, Germany; two ACM conference travel awards from ACM-SIGBED in 2016 and ACM-SIGDA in 2017; and Duke Graduate School Fellowship in 2013. He served as a Technical Program Committee Member for IEEE/ACM DATE Conference (2019-), IEEE VLSID (2019), IEEE ISVLSI (2019-), ACM NanoCom (2019), IFIP/IEEE VLSI-SoC (2019), and as an expert reviewer for a large number of conferences and journals.



Tung-Che Liang received his B.S. degree in Electronics Engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2014, and the M.S.E degree from the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA, in 2020, where he is currently working toward the Ph.D. degree.

He was with Synopsys Inc., Hsinchu, Taiwan, as an R&D engineer. He was a yield & diagnosis intern at Intel, Santa Clara, CA, and a DFT intern at NVIDIA Inc., Santa Clara, CA. His current research interests include design automation and security for microfluidic systems.

Kristin Scott received the B.S. degree in Molecular Biology from Grove City College, Grove City, PA in 1992, and the Ph.D. degree in Biochemistry from the University of Iowa, Iowa City IA in 1998.

Since 2014, she has been with the Department of Molecular Genetics and Microbiology, Duke University, Durham, NC where she is currently an Assistant Professor. From 2008 to 2013 she was an Assistant Professor at the Duke Institute for Genome Sciences and Policy.

Krishnendu Chakrabarty received the B. Tech. degree from the Indian Institute of Technology, Kharagpur, in 1990, and the M.S.E. and Ph.D. degrees from the University of Michigan, Ann Arbor, in 1992 and 1995, respectively. He is now the John Cocke Distinguished Professor and Department Chair of Electrical and Computer Engineering (ECE), and Professor of Computer Science, at Duke University.

Prof. Chakrabarty is a recipient of the National Science Foundation CAREER award, the Office of Naval Research Young Investigator award, the Humboldt Research Award from the Alexander von Humboldt Foundation, Germany, the IEEE Transactions on CAD Donald O. Pederson Best Paper Award (2015), the ACM Transactions on Design Automation of Electronic Systems Best Paper Award (2017), multiple IBM Faculty Awards and HP Labs Open Innovation Research Awards, and over a dozen best paper awards at major conferences. He is also a recipient of the IEEE Computer Society Technical Achievement Award (2015), the IEEE Circuits and Systems Society Charles A. Desoer Technical Achievement Award (2017), the Semiconductor Research Corporation Technical Excellence Award (2018), and the IEEE Test Technology Technical Council Bob Madge Innovation Award (2018). He has been a Hans Fischer Senior Fellow at the Institute for Advanced Study, Technical University of Munich, Germany. He is a 2018 recipient of the Japan Society for the Promotion of Science (JSPS) Fellowship in the "Short Term S: Nobel Prize Level" category.

Prof. Chakrabarty's current research projects include: design-for-testability of integrated circuits and systems (especially 3D integration and system-on-chip); microfluidic biochips; hardware security; neuromorphic computing systems. He is a Fellow of ACM, a Fellow of IEEE, a Fellow of AAAS, and a Golden Core Member of the IEEE Computer Society. He was a Distinguished Visitor of the IEEE Computer Society (2005-2007, 2010-2012), a Distinguished Lecturer of the IEEE Circuits and Systems Society (2006-2007, 2012-2013), and an ACM Distinguished Speaker (2008-2016). Prof. Chakrabarty served as the Editor-in-Chief of *IEEE Design & Test of Computers* during 2010-2012, *ACM Journal on Emerging Technologies in Computing Systems* during 2010-2015, and *IEEE Transactions on VLSI Systems* during 2015-2018. He is an Associate Editor of *IEEE Transactions on Biomedical Circuits and Systems* and *ACM Transactions on Design Automation of Electronic Systems*.

Ramesh Karri is a Professor of Electrical and Computer Engineering at New York University. He co-directs the NYU Center for Cyber Security (http://cyber.nyu.edu). He co-founded the Trust-Hub (http://trust-hub.org) and organizes the Embedded Systems Challenge (https://csaw.engineering.nyu.edu/esc), the annual red team blue team event. Ramesh Karri has a Ph.D. in Computer Science and Engineering, from the University of California at San Diego and a B.E in ECE from Andhra University. His research and education activities in hardware cybersecurity include trustworthy integrated circuits, processors and cyber-physical systems; security-aware computer-aided design, test, verification, validation, and reliability; nano meets security; hardware security competitions, benchmarks and metrics; biochip security; additive manufacturing security. He has published over 275 articles in leading journals and conference proceedings.

Ramesh Karri' work in trustworthy hardware received best paper award nominations (ICCD 2015 and DFTS 2015), awards (ACM TODAES 2017, ITC 2014, CCS 2013, DFTS 2013 and VLSI Design 2012, ACM Student Research Competition at DAC 2012, ICCAD 2013, DAC 2014, ACM Grand Finals 2013, Kaspersky Challenge and Embedded Security Challenge). He received the Humboldt Fellowship and the National Science Foundation CAREER Award. He is a Fellow of the IEEE for his contributions to and leadership in Trustworthy Hardware. He is the Editor-in-Chief of ACM Journal of Emerging Technologies in Computing. Besides, he served/s as the Associate Editor of IEEE Transactions on Information Forensics and Security (2010-2014), IEEE Transactions on CAD (2014-), ACM Journal of Emerging Computing Technologies (2007-), ACM Transactions on Design Automation of Electronic Systems (2014-), IEEE Access (2015-), IEEE Transactions on Emerging Technologies in Computing (2015-), IEEE Design and Test (2015-) and IEEE Embedded Systems Letters (2016-). He served as an IEEE Computer Society Distinguished Visitor (2013-2015). He served on the Executive Committee of the IEEE/ACM Design Automation Conference leading the Security@DAC initiative (2014-2017). He has given keynotes, talks, and tutorials on Hardware Security and Trust.