Execution of Provably Secure Assays on MEDA Biochips to Thwart Attacks*

Tung-Che Liang
Duke University
tung.che.liang@duke.edu

Mohammed Shayan
New York University
mos283@nyu.edu

Krishnendu Chakrabarty
Duke University
krishnendu.chakrabarty@duke.edu

Ramesh Karri
New York University
rkarri@nyu.edu

ABSTRACT

Digital microfluidic biochips (DMFBs) have emerged as a promising platform for DNA sequencing, clinical chemistry, and point-of-care diagnostics. Recent research has shown that DMFBs are susceptible to various types of malicious attacks. Defenses proposed thus far only offer probabilistic guarantees of security due to the limitation of on-chip sensor resources. A micro-electrode-dot-array (MEDA) biochip is a next-generation DMFB that enables the sensing of on-chip droplet locations, which are captured in the form of a droplet-location map. We propose a security mechanism that validates assay execution by reconstructing the sequencing graph (i.e., the assay specification) from the droplet-location maps and comparing it against the golden sequencing graph. We prove that there is a unique (one-to-one) mapping from the set of droplet-location maps (over the duration of the assay) to the set of possible sequencing graphs. Any deviation in the droplet-location maps due to an attack is detected by this countermeasure because the resulting derived sequencing graph is not isomorphic to the original sequencing graph. We highlight the strength of the security mechanism by simulating attacks on real-life bioassays.

1 INTRODUCTION

A digital microfluidic biochip (DMFB) is composed of a two-dimensional electrode array that manipulates discrete fluid droplets. When driven by a sequence of control voltages, the electrode array can perform fluidic operations such as dispensing, mixing, and splitting [8]. Because of the precise control over microfluidic operations, DMFBs are employed in lab-on-a-chip systems for biomolecular recognition, point-of-care diagnostics, and cell biology [9]. DMFBs have recently been commercialized; e.g., Illumina, a leading DNA sequencing company, has announced the use of DMFBs in their NeoPrep NGS Library Prep product [1]; USDA has approved the Baebies SEEKER DMFB platform as a high-throughput laboratory solution for screening diseases in a newborn child [2].

As the platforms are being adopted for safety-critical applications [11], security and trustworthiness of DMFBs have become the important focus of research. It has been shown that DMFBs are susceptible to attacks such as actuation tampering and miscalibration and that the attacks lead to disastrous assay outcomes [6, 7, 20]. A randomized checkpointing-based security method for DMFBs was presented in [20]. This method randomly checks areas on the DMFB platform using a CCD camera in the cyberphysical system to monitor assay execution. However, checkpoint-based validation is limited by real-time computing resources and memory needed for analyzing real-time data from a high-resolution camera. Therefore, the checkpointing-based validation can only provide probable security and not likely to inspire confidence in users of these DMFB systems.

The micro-electrode-dot-array (MEDA) architecture for DMFBs was introduced recently, and MEDA prototypes were fabricated using TSMC 0.35 μm CMOS technology [12, 15, 21]. A MEDA biochip is composed of an array of identical microfluidic unit components named *micro-electrode cells* (MCs). Each MC consists of a micro-electrode, a semiconductor control circuit, and a sensing module that enables real-time sensing of on-chip droplets. The sensor data specifies the location, the size, and the shape of the on-chip droplets. The real-time sensor data can be used to validate the execution of a bioassay [16]. From a security perspective, MEDA is promising as it overcomes the resource constraints of a traditional DMFB.

In this paper, we propose the first provable security solution for MEDA biochips by exploiting integrated droplet sensing. The proposed method can detect any operational attack and it does not require CCD cameras for detection. The key contributions of this paper are as follows:

- A security method that utilizes the information sensed from a MEDA biochip. This method automatically recognizes fluidic operations, and constructs the dependencies between these operations to compare against the "golden" bioassay.
- A formal proof that the proposed security solution can reconstruct the sequencing graph, i.e., the assay specification.
- Demonstration of result-altering attacks on the in-vitro glucose test assay [6] and denial-of-service attacks on the multiplexed in-vitro diagnostics on human physiological fluids [19]. Our results show that the proposed defense can thwart these attacks.

The remainder of this paper is organized as follows. Section 2 explains the MEDA architecture. Section 3 models threats and attacks on MEDA biochips. Section 4 presents details of the proposed security method and theoretical results. Section 5 showcases the working of the proposed defense against simulated attacks on real-life bioassays. Finally, conclusions are drawn in Section 6.

2 MEDA BIOCHIPS

In this section, we describe the MEDA biochip architecture and its working principle.

*This work 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.

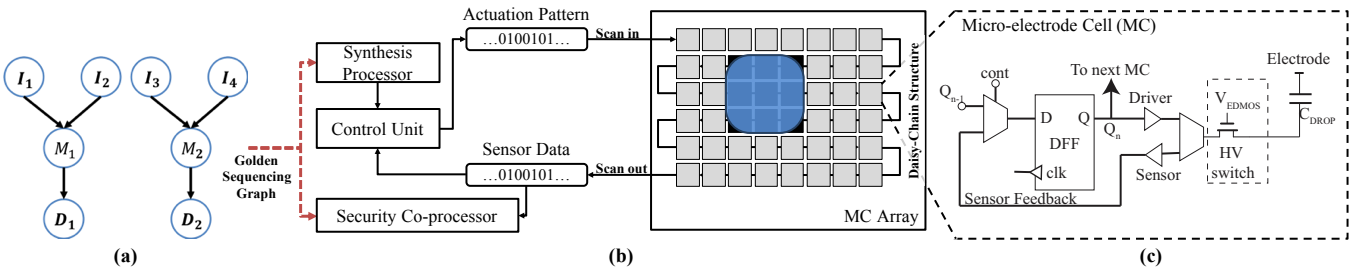


Figure 1: (a) An example of the given sequencing graph. The nodes I_i, M_i, and D_i represent dispensing, merging, and detecting operations. (b) The MEDA biochip architecture consists of a synthesis processor, a control unit, and a two-dimensional MC array. The proposed security mechanism is loaded on a security co-processor that is physically separate from the synthesis processor. (c) The circuit schematic of the sensor and control module of the micro-electrode cell. Q_n denotes the n^{th} cell in the daisy-chain structure.

2.1 MEDA Biochip Architecture

A MEDA biochip platform is composed of a processor, a control unit, and a two-dimensional MC array, as shown in Fig. 1(b). The MC integrates a micro-electrode, an activation circuit, and a sensing circuit, which is shown in Fig. 1(c). These circuits allow fine-grained droplet control and real time sensing [13]. The MCs are connected to form a daisy chain, and they can operate in two modes: sensing and actuation. The control bit for actuation is fed serially through the scan chain. In actuation mode, the electrode with logic '1' is actuated using high voltage. In sensing mode, the presence of a droplet is determined as follows. The capacitor formed by the lower and upper electrode plates is pre-charged. Then the discharge path is switched on. The discharge time depends on the electrode capacitance; the presence of a droplet increases the permittivity and hence the capacitance of the electrode. The flip-flop within the MC samples the capacitor node voltage. The clock is tuned such that the flip-flop captures '1' when a droplet is present and '0' otherwise. The detection results are shifted out as a vector specifying droplet location, and this vector is referred to as the *droplet-location map*.

2.2 MEDA Biochip Working Principle

To carry out bio-chemical assays on MEDA biochips, the given bioassay is interpreted as a sequencing graph that specifies the relationships of fluidic operations. An example of a sequencing graph is illustrated in Fig. 1(a). A synthesis tool that is loaded on the processor binds the operations to on-chip resources, generates an optimized schedule of these operations, and routes droplets on the biochip [17]. Based on the synthesis result, the actuation patterns for the micro-electrodes are translated by the control unit, and the actuation patterns are shifted to the MC array through the daisy-chain structure sequentially. After an actuation pattern is activated on the MC array, a location-map is scanned out to the control unit as a feedback of the scanned in actuation pattern [15]. An operation cycle includes the scan-in of an actuation sequence, the activation of the micro-electrodes, and the scan-out of a location-map. As a result, the scanned out droplet-location maps are cycle-by-cycle consecutive.

A provably security mechanism is proposed based on the consecutive droplet-location maps and the same sequencing graph for the synthesis processor. Details of this security mechanism will be explained in Section 4. To ensure that the control unit and the security mechanism are not simultaneously compromised, the security mechanism is installed in a separate monitoring co-processor, as shown in Fig. 1(b).

3 ATTACKS ON MEDA BIOCHIPS

In this section, we present the motivation for attacks and discuss the threat model in the context of MEDA biochips.

3.1 Threat Model

Motivations for attacking a MEDA biochip vary based on the target application. In point-of-care diagnostics, an attack may endanger patient health by ensuring an incorrect prescription based on compromised test results. To compete with rivals, companies may be interested in corporate sabotage by disrupting scientific experiments. Other motivations include terrorism, bypassing of pollution control, and revenge [6].

We assume that the attacker is able to change fluidic operations during synthesis. The attacker can be the CAD tool vendor; the synthesis tools can be tampered with to alter routing paths and module placements. Operations can also be inserted or deleted. Since remote control of experiments is now possible [4] [5], e.g., in environmental monitoring, an attacker can tamper with the assay by inserting malware in the synthesis processor or the control unit through the internet. These attackers do not have to understand the detailed structure of each biochip to compromise the outcome of the assays. Although we consider attacks at the synthesis level, we assume that the operators of MEDA biochips are trustworthy, i.e., the samples and reagents are loaded as expected. The MC array is also assumed to be functioning as intended, i.e., micro-electrode actuation and capacitance detection are not compromised. Additionally, we assume the integrity of the proposed security solution because it is loaded on a co-processor, which is physically separate from the control unit. Finally, the "golden" sequencing graph for the attack-detection method is assumed to be trustworthy.
Example: Fig. 2 shows an example of tampering with the synthesis result of the multiplexed in-vitro diagnosis. The lactate and glucose levels in human physiological fluids are mixed and tested. These tests are important for diagnosing diseases such as diabetes, myocardial infraction, congestive heart failure, and septicemia [14].

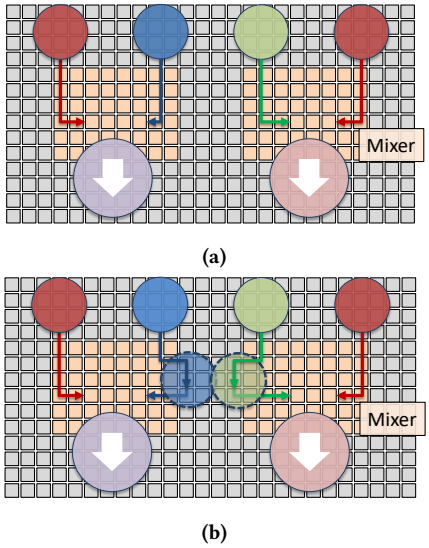


Figure 2: Malicious routing of droplets during the execution of a bioassay. (a) Originally, two pairs of droplets are mixed in separate mixers. (b) Two droplets are contaminated due to malicious routing.

The intended execution of multiplexed in-vitro diagnosis is shown in Fig. 2(a). When an adversary stealthily changes the routes during synthesis, as shown in Fig. 2(b), sample/reagent contamination may occur. If the compromised result is detected by an online error-recovery method [16] [22], the re-execution of the assay will be triggered. The re-execution may however fail due to the same attack, and another re-execution will then be launched. As a result, the repeated execution of the assay will lead to the wastage of precious samples and expensive reagents. On the other hand, if the compromised result is not detected by an online error-recovery method, the erroneous results will be deemed to be correct. The treatment prescribed on the basis of the misdiagnosis may jeopardize the health of the patient.

3.2 Attack Modeling

To alter the outcome of the assay, an attacker can modify some of the operations in the assay. The adversary can introduce spurious assay operations or delete critical operations during synthesis. For instance, if an aliquot (smaller) droplet is maliciously extracted from one droplet and merged with another droplet during execution, the victim droplet is contaminated [22]. The attacker can also alter an operation in a subtle manner. For example, an original 1:1 split operation may be replaced with a 3:1 split [18].

Example: Fig. 3 shows the original and modified sequencing graphs of the multiplexed in-vitro diagnosis. This attack is based on the misrouting example in Fig. 2, where a sample/reagent contamination is introduced during execution. Although the adversary only changes the transportation paths of two droplets, an unwanted merge operation and an unwanted split operation are stealthily added to the original execution. Therefore, the erroneous results may mislead prescription and further jeopardize the health of the patient.

A classification of attacks is presented in Table 1. We classify attacks according to the synthesis phase during which they are

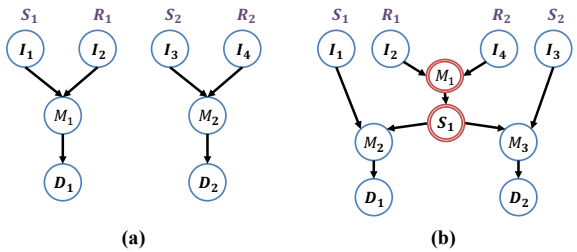


Figure 3: Contamination of the assay is modeled by the inserting of extra merge and split steps in the sequencing graph: (a) The golden sequencing graph; (b) The modified sequencing graph; S_1, S_2, R_1, and R_2 represent plasma, serum, glucose oxidase, and lactate oxidase, respectively; the nodes I_i, M_i, D_i, and S_i represent dispensing, merging, detecting, and splitting operations.

introduced, and map the related impacts to the input sequencing graph. The assay can be tampered with before synthesis commences. The addition (deletion) of operations in the given assay leads to extra (fewer) nodes in the sequencing graph. During synthesis, the constraints on completion or incubation times might be manipulated so that operations do not finish correctly. Two droplets might be routed too close to each other, causing unwanted merging of these droplets. Attackers can also change the synthesis output so that the assay fails, e.g., the example in Fig. 2 might result from a change in the synthesized route.

Table 1: Classification of Attacks.

Attacks before synthesis	**Modifications to the sequencing graph**
Insert extra operations	Insert extra nodes
Delete operations	Delete nodes
Alter operations	Replace nodes
Attacks during synthesis	**Possible Modifications**
Violate fluidic-distance design rule [13, 17, 19]	Insert extra merge and split steps
Attacks after synthesis	**Modifications to the sequential graph**
Overlap droplet routes	Insert extra merge and split steps
Replace input droplets of operations	Redirect input edges of nodes

4 ENSURING SECURITY BY RECONSTRUCTING THE SEQUENCING GRAPH

As a countermeasure to the threats described in Section 3, we propose a security mechanism based on the reconstruction of the sequencing graph from real-time sensor data (i.e., the droplet-location maps). The reconstructed graph is compared against the golden sequencing graph, and any deviation is flagged as an error.

Recall that the droplet-location maps are generated cycle-by-cycle according to the MEDA biochip design. As a result, the exploitation of these sensor data introduces negligible overhead. Since each droplet-location map provides information of on-chip droplets at a certain time, we claim that the overall operations of a given assay can be reconstructed using the consecutive droplet-location maps.

Input: A droplet-location map lm_t, the row length l
Output: A list of footprints FPs

```
FPs = {}
initial empty adjacency graph G_A
for a component b_i in lm_t do
    if (b_i = 1) then Add node e_i to G_A end if
end for
for a pair of nodes e_i and e_j in G_A do
    if (|i − j| = 1 or |i − j| = l) then
        Add an edge between e_i and e_j
    end if
end for
for a connected component g_c in G_A do
    fp = {e_i | e_i is a node of the component g_c}
    Add fp to FPs
end for
return FPs
```

Figure 4: Pseudocode for deriving footprints from a droplet-location map.

To establish the above claim, we present a theorem—we lead up to it step-by-step using lemmas. These lemmas reveal the one-to-one mapping between fluidic operations and consecutive (i.e., in adjacent clock cycles) droplet-location maps. The recognized operations can be used to reconstruct a sequencing graph that is isomorphic to the "golden" sequencing graph when there is no attack. The proofs of the lemmas and the theorem are presented in the appendix A. We first describe how each droplet-location map can be interpreted and we introduce the notation that will be used later.

4.1 Interpretation of Droplet-Location Maps

Our goal is to identify locations of droplets at any time step (clock cycle) from the scanned-out droplet-location map. Let the scanned out droplet-location map at time t be denoted by the vector lm_t. The length of this vector is the number of micro-electrodes on the platform. A component $b_i \in \{0, 1\}$ of lm_t indicates whether the micro-electrode e_i is under a droplet at time t, where e_i is the ith micro-electrode on the platform. To represent the locations of all droplets in a compact manner, a droplet-location map is transformed to an undirected graph called the *adjacency graph* $G_A = (V_A, E_A)$. Node that $V_A = \{e_i | b_i = 1,\ \forall i\}$. For any given $x, y \in V_A$, $(x, y) \in E_A$ if x and y are physically adjacent on the MEDA platform. Let the number of micro-electrodes in a row be l. Two micro-electrodes e_i and e_j are physically adjacent if either $|i - j| = 1$ or $|i - j| = l$. A connected component in G_A indicates a droplet's location, i.e., the droplet resides over these micro-electrodes in this component. We define the set of nodes of a connected component as a *footprint*; this represents a droplet's location. If a droplet d_x exists on the platform at time t, the set of nodes corresponding to this droplet's position is denoted by fp_t^x. The procedure for computing a droplet-location map is shown in Fig. 4.

Example: Fig. 5 shows two visualized droplet-location maps that are scanned out from a MEDA biochip at time $t-1$ and time t, respectively. A droplet-location map $lm_{t-1} = (b_1, ..., b_{50})$ can be obtained at time $t-1$, where only b_{12}, b_{13}, b_{18}, b_{19}, b_{22}, b_{23}, b_{28}, and b_{29} are equal to 1; the other elements of this vector are 0. The adjacency graph obtained from the droplet-location map lm_{t-1} is shown in Fig. 6. Nodes in this graph represent the micro-electrodes e_{12}, e_{13}, e_{18}, e_{19}, e_{22}, e_{23}, e_{28}, and e_{29}, and edges between nodes

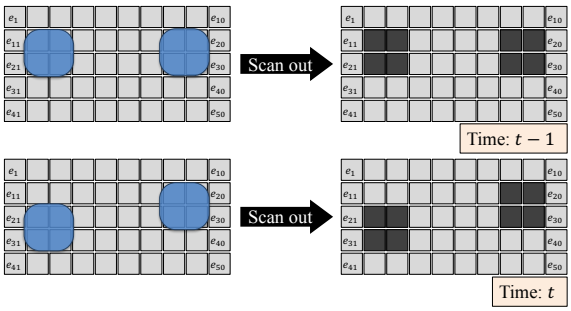


Figure 5: Droplets on the MEDA platform and the visualized droplet-location maps over two consecutive time slots.

indicate adjacency in the micro-electrode array. The two connected components indicate the locations of two droplets at time $t-1$. Consequently, two footprints fp_{t-1}^x and fp_{t-1}^y are obtained, where $fp_{t-1}^x = \{e_{12}, e_{13}, e_{22}, e_{23}\}$ and $fp_{t-1}^y = \{e_{18}, e_{19}, e_{28}, e_{29}\}$.

4.2 Recognition of Fluidic Operations

Using the droplet-location maps for two consecutive time slots that are derived from the MEDA platform, we are able to recognize the droplet operations. In the example of Fig. 5, two droplets are on the chip at times $t-1$ and t. Let the left droplet be d_x and the right droplet be d_y, and suppose d_x is moved downward at time t. Using the two droplet-location maps lm_{t-1} and lm_t, four footprints are obtained, i.e., as fp_{t-1}^x, fp_{t-1}^y, fp_t^x, and fp_t^y, where $fp_{t-1}^x = \{e_{12}, e_{13}, e_{22}, e_{23}\}$, $fp_{t-1}^y = \{e_{18}, e_{19}, e_{28}, e_{29}\}$, $fp_t^x = \{e_{22}, e_{23}, e_{32}, e_{33}\}$, and $fp_t^y = \{e_{18}, e_{19}, e_{28}, e_{29}\}$. By comparing the footprints at time $t-1$ and the footprints at time t, we find that $fp_{t-1}^x \cap fp_t^x \neq \emptyset$ and that $fp_{t-1}^y \cap fp_t^y \neq \emptyset$. Moreover, the overlapping footprints have the same number of electrodes, i.e., $|fp_{t-1}^x| = |fp_t^x|$ and $|fp_{t-1}^y| = |fp_t^y|$. This example shows that the transportation of any droplet can be inferred by examining its footprints.

Lemma 1. *A droplet d_x is transported on the MEDA platform between time slots $t-1$ and t if and only if $fp_{t-1}^x \cap fp_t^x \neq \emptyset$ and $|fp_{t-1}^x| = |fp_t^x|$.*

Furthermore, all basic fluidic operations can be inferred by examining two consecutive droplet-location maps.

Lemma 2. *A droplet d_x is dispensed on the MEDA platform between time slots $t-1$ and t if and only if $fp_t^x \neq \emptyset$ and $fp_{t-1}^x = \emptyset$.*

Lemma 3. *A droplet d_x is discarded from the MEDA platform between time slots $t-1$ and t if and only if $fp_{t-1}^x \neq \emptyset$ and $fp_t^x = \emptyset$.*

The above operations are related to only one droplet. Other droplet operations on MEDA biochips include two droplets, e.g., merging and splitting.

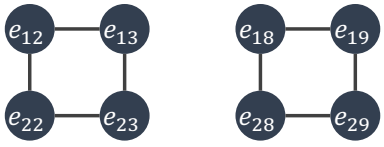


Figure 6: The adjacency graph at time t-1 for the scenario shown in Fig. 5.

Lemma 4. *A droplet d_x is split from a larger droplet d_y on the MEDA platform between time slots $t-1$ and t if and only if $fp_t^x \neq \emptyset$, $fp_{t-1}^y \neq \emptyset$, $fp_t^x \cap fp_{t-1}^y \neq \emptyset$, and $|fp_t^x| < |fp_{t-1}^y|$.*

Lemma 5. *A droplet d_x is obtained as a result of merging two droplets d_y and d_z on the MEDA platform between time slots $t-1$ and t if and only if $fp_t^x \neq \emptyset$, $fp_{t-1}^y \neq \emptyset$, $fp_{t-1}^z \neq \emptyset$, $fp_t^x \cap fp_{t-1}^y \neq \emptyset$, $fp_t^x \cap fp_{t-1}^z \neq \emptyset$, and $|fp_t^x| = |fp_{t-1}^y| + |fp_{t-1}^z|$.*

We assume that droplets remaining on the platform until the end of assay execution are used for detection. The detection operation is therefore recognized by the following lemma.

Lemma 6. *A droplet d_x is detected on the MEDA platform at time T (end of the assay) if and only if $fp_T^x \neq \emptyset$.*

Using the above lemmas, all the basic fluidic operations can be automatically inferred using location maps.

To increase the strength of the proposed defense, we also explicitly identify the type of reagent used in a dispensing operation. This can be accomplished by identifying the location of the dispensing port. Reagents are stored in separate reservoirs outside the MEDA platform, and they are dispensed through various channels to different locations on the platform to avoid contamination. A dispensing operation can be recognized by the footprint. Let the set of micro-electrodes where reagent A is dispensed on the platform be E_A. If a dispensing operation is recognized by Lemma 2 with a footprint fp_t^x, and $fp_t^x \cap E_A \neq \emptyset$, the dispensing operation must involve reagent A. Likewise, a splitting operation can also be specifically identified by the proposed method. Unlike a conventional DMFB that supports only a 1:1 splitting operation, a MEDA biochip enables unbalanced splitting operations, e.g., the 1:2 ratio. The proposed defense records the splitting ratio by determining the sizes of the resulting droplets. These details about the executed operations enable us to detect any deviations from the original assay.

4.3 Sequencing-Graph Construction

Since the droplet-location maps for an assay are obtained for consecutive time slots, the operations inferred from them are contiguous as well in time, i.e., each operation is related to a specifically identified previous operation and/or a similarly identified operation. Let the last droplet-location map of any assay execution be lm_T. The following lemma states that the droplet-location maps can be used to reconstruct the sequencing graph for an assay.

Lemma 7. *The sequencing graph can be reconstructed from a complete sequence of droplet-location maps from the start to the end of the assay, i.e., (lm_1, lm_2, ..., lm_T).*

If we denote fluidic operations as nodes and add edges between consecutive operations, a graph is generated with the dispensing, discarding, transporting, merging, and splitting operations. However, the golden sequencing graph does not include transportation operations because the specific droplet transportation paths do not affect the outcome of the assay (as long as the steps of the biochemical protocol are correctly followed). As a result, nodes corresponding to droplet transportation operations are omitted, and their parent nodes and child nodes are connected by edges inserted in a post-processing step. This step leads to a graph G_{re}, which we refer to as the *reconstructed graph.*

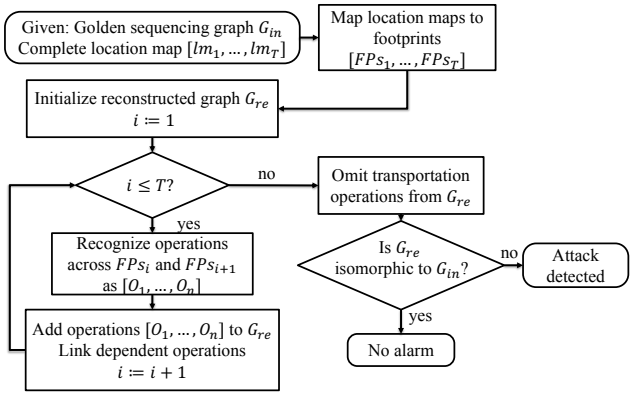


Figure 7: Control flow corresponding to the proposed defense.

If all operations from the given sequencing graph G_{in} are synthesized correctly on MEDA and not compromised by an attack, we can prove that G_{re} is isomorphic to G_{in}.

Theorem 1. *If a sequencing graph G_{in} is synthesized on a MEDA biochip and the biochip is not compromised by an attack, the sequencing graph G_{re} reconstructed from the droplet-location maps is isomorphic to G_{in}. If an attack compromises one or more operations, G_{re} is no longer isomorphic to G_{in}.*

4.4 Detection of Attacks

Theorem 1 shows that attacks can be detected if G_{re} is not isomorphic to G_{in}. Although the problem of checking for graph isomorphism is NP-complete [10], we are able to use it here because both G_{re} and G_{in} are relatively small in size. Recall the example in Fig. 2, where the contamination attack introduced two extra operations in G_{re} (Fig. 3). In this case, the number of nodes is different in G_{re} compared to G_{in}. Therefore, we can distinguish between the two graphs in linear time in this case by simply counting the number of nodes in the two graphs. The overall control flow of the proposed defense is shown in Fig. 7.

Note that the actuation patterns cannot be used to secure the execution of bioassays. Unlike sensor data, there is no one-to-one mapping between the synthesis result and the golden sequencing graph. For example, to transport a droplet on the biochip, several sequences of actuation can be applied. Besides, the actuation patterns for a droplet may not be consecutive in the cycle-by-cycle sense. When a droplet waits on the biochip for later operations, no actuation pattern is applied to this droplet.

5 SIMULATION RESULTS

We simulated two real-life assays and different attacks on a state-of-art MEDA platform with 60×30 micro-electrodes [15] to demonstrate the effectiveness of the proposed defense. The assays are synthesized using methods described in [17], and location maps are generated by simulating assay execution.

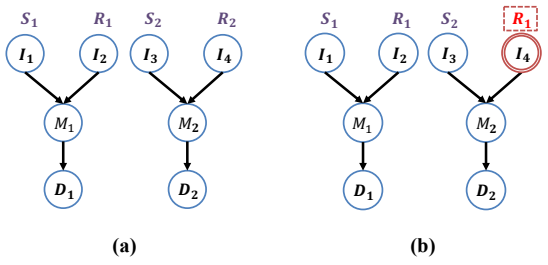


Figure 8: The swapping-reagent attack; S_1, S_2, R_1, and R_2 represent plasma, serum, glucose oxidase, and lactate oxidase, respectively; the nodes I_i, M_i, D_i, and S_i represent dispensing, merging, detecting, and splitting operations, respectively. (a) The golden graph. (b) The reconstructed graph.

5.1 Lactate, glucose, and pyruvate in-vitro test

We first study attacks on multiplexed in-vitro diagnostics of serum and plasma [19]. Glucose and lactate measurements are carried out in this procedure by mixing the samples from the patient, the glucose reagent, and the lactate reagent. The golden sequencing graph that represents the bioassay is shown in Fig. 8(a). Two attacks are inserted in the synthesis phase to tamper with the results: 1) reduce the distance constraint between droplets in order to cause contamination; 2) swap input droplets before mixing procedures.

5.1.1 Attack 1: This attack was introduced in Fig. 2; the minimum space constraint between glucose and lactate reagents is violated, and the paths of these droplets are maliciously routed to be too close. Thus, two droplets come in contact when they are transported to the corresponding mixers. The contaminated droplets cause the final detection results to be incorrect. Recall that the erroneous results may mislead prescription and jeopardize one's health. Our analysis method reconstructs the sequencing graph G_{re}; it contains extra operations (Fig. 3). The defense mechanism therefore provides an alarm because G_{in} is not isomorphic to G_{re}.

5.1.2 Attack 2: This attack stealthily swaps one of the dispense reagent with another for the mixing operation; hence, the detection outcome is compromised. Similar to Attack 1, the compromised detection may end up with disastrous consequences. Note that when the multiplexed in-vitro assay is executed on biochips, operations are executed in parallel to shorten the overall execution time. This attack is thus hard to detect because there are numerous on-chip droplets. The proposed method reconstructs the graph shown in Fig. 8. The defense procedure detects the attack; the analysis reveals that the reagent is different from that for G_{in}.

5.2 Glucose Test

We next demonstrate two other attacks on the glucose test assay. According to the 2017 National Diabetes Statistics Report from Centers for Diseases Control (CDC), 30.3 million people have diabetes in the United States [3]. Patients with diabetes must regularly monitor their blood-sugar level through glucose tests so as to inject the appropriate amount of insulin. If the test result is maliciously tampered to result in a higher glucose reading, higher doses of insulin will be injected, leading to hypoglycemia. On the other hand, if the test result is altered to result in a lower glucose level, less insulin will be injected, resulting in hyperglycemia.

To test the concentration of the glucose level in a patient's sample, serial dilution is used to generate the calibration curve [6]; see the blue solid line in Fig. 10. Given any sample, its glucose concentration is measured by comparing the absorbent reaction to the calibration curve. For example, if the absorbent-reaction rate of a sample is measured as 0.001, the concentration of this sample is interpolated as 100 mg/dL.

The golden sequencing graph of a glucose assay is shown in Fig. 9(a), where Chain 1 and Chain 2 are used for generating a series of different concentrations. These known concentrations are used to plot the calibration curve. In this example, a golden glucose solution with concentration of 600 mg/dL is diluted to get the series of calibrated concentrations. The diluted concentrations are 600 mg/dL, 300 mg/dL, 150 mg/dL, 75 mg/dL, and 0 mg/dL. These concentrations with their absorbent rates are measured to draw the golden calibration curve. The sample in Chain 3 is tested to get the absorbent reaction rate, and by interpolating with the golden curve, the concentration of the sample is determined as 94 mg/dL. We next consider two possible attacks on the assay: 1) alter the calibration curve on Chain 2; 2) tamper the concentration for Chain 3 with a waste droplet.

5.2.1 Attack 1: This attack changes the volume ratio of first split in Chain 2, and the concentrations of the child nodes are therefore changed accordingly; see Fig. 9(b). If the attack succeeds without being detected, the calibration curve is also altered, which is plotted as a red dot line in Fig. 10. The glucose concentration of the sample is mistakenly interpreted as 225 mg/dL, and the outcome for the patient may be hypoglycemia because of the insulin overdose. Fortunately, the proposed defense can easily detect this attack. The reconstructed graph is shown in Fig. 9(b), where the compromised splitting operation is doubly circled in red. The defense mechanism identifies the difference between G_{re} and G_{in} because the splitting ratios are specifically identified.

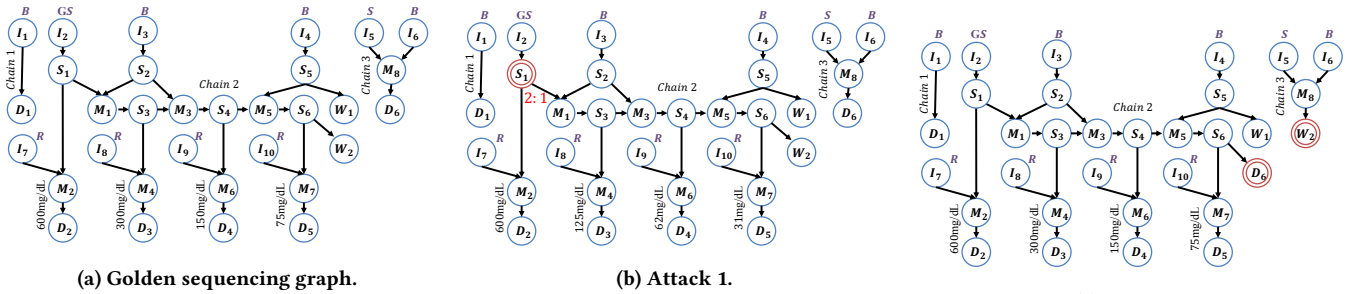
5.2.2 Attack 2: This stealthy attack uses a waste droplet in Chain 2 as the detection droplet, and then discards the valuable sample in Chain 3 as a waste droplet. The reconstructed graph is shown in Fig. 9(c), where the two swapped operations are doubly circled in red. In this case, the glucose level is severely underestimated. As a result, the patient remains in hyperglycemia, which may lead to headache, fatigue, blurred vision, or even diabetic coma. However, such an undesirable consequence can be avoided because the proposed defense is able to detect the attack by comparing G_{re} with G_{in}.

6 CONCLUSION

We have presented a lightweight, but effective, security solution for MEDA biochips. This approach can detect malicious attacks in the MEDA platform. We have shown that the provably secure assay execution can be guaranteed using this solution. Simulation results for two real-life assays demonstrate the effectiveness of this solution for thwarting attacks and preventing harmful consequences.

REFERENCES

[1] 2014. NeoPrep NFS Library Prep with Digital Microfluidics by Illumina. https://www.illumina.com/company/video-hub/F_Hks6OnSKM0.html



(a) Golden sequencing graph. (b) Attack 1. (c) Attack 2.

Figure 9: The golden glucose test assay and two malicious attacks. B, GS, and S are buffer, glucose solution and sample, respectively; the nodes I_i, M_i, D_i, W_i and S_i correspond to dispensing, merging, detection, the discarding of a waste droplet, and splitting operations, respectively. (a) The graphs for Chain 1 and Chain 2 are used for generating the calibration curve, and the concentration of the sample in Chain 3 is tested and interpolated with the calibration curve. (b) An attack is stealthily introduced in the second splitting operation. The splitting ratio is changed from $1:1$ to $2:1$. (c) An attack is inserted during high-level synthesis; the outcome droplet of Chain 3 is discarded and the waste droplet in Chain 2 is transported for detection.

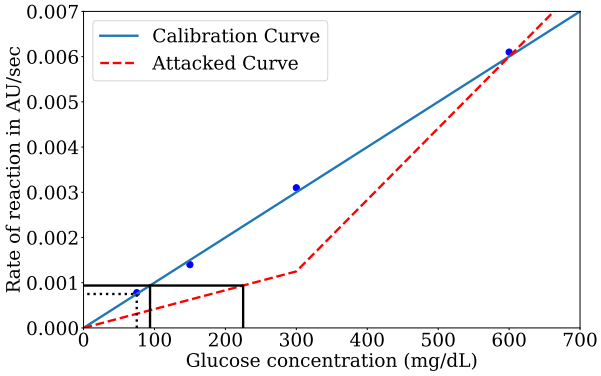


Figure 10: Golden glucose calibration curve and the attacked calibration curve.

[2] 2016. FDA advisors approve of Baebies SEEKER analyzer for newborns. https://www.baebies.com/fda-advisors-back-approval-baebies-seeker-analyzer-newborns/

[3] 2017. National Diabetes Statistics Report. https://www.cdc.gov/diabetes/data/statistics/statistics-report.html.

[4] 2018. Automated cell and molecular biology laboratory. https://www.transcriptic.com/.

[5] 2018. Emerald Cloud Lab. https://www.emeraldcloudlab.com/.

[6] Sk Subidh Ali et al. 2016. Security assessment of cyberphysical digital microfluidic biochips. *IEEE/ACM Trans. CBB* 13, 3 (2016), 445–458.

[7] Sk Subidh Ali et al. 2016. Supply-Chain Security of Digital Microfluidic Biochips. *Computer* 49, 8 (2016), 36–43.

[8] Kihwan Choi et al. 2012. Digital microfluidics. *Annual Review of Analytical Chemistry* 5 (2012), 413–440.

[9] Wei-Lung Chou et al. 2015. Recent advances in applications of droplet microfluidics. *Micromachines* 6, 9 (2015), 1249–1271.

[10] Luigi P Cordella et al. 2004. A (sub) graph isomorphism algorithm for matching large graphs. *IEEE trans. PAMI* 26, 10 (2004), 1367–1372.

[11] Richard B Fair et al. 2007. Chemical and biological applications of digital-microfluidic devices. *IEEE Design & Test of Computers* 24, 1 (2007).

[12] Yingchieh Ho et al. 2016. Design of a micro-electrode cell for programmable lab-on-CMOS platform. In *Proc. ISCAS*. 2871–2874.

[13] Oliver Keszocze et al. 2017. Exact routing for micro-electrode-dot-array digital microfluidic biochips. In *Proc. ASP-DAC*. 708–713.

[14] Ryoji Kurita et al. 2002. Microfluidic device integrated with pre-reactor and dual enzyme-modified microelectrodes for monitoring in vivo glucose and lactate. *Sensors and Actuators B: Chemical* 87, 2 (2002), 296–303.

[15] Kelvin Yi-Tse Lai et al. 2015. An intelligent digital microfluidic processor for biomedical detection. *Journal of Signal Processing Systems* 78, 1 (2015), 85–93.

[16] Zipeng Li et al. 2016. Error recovery in a micro-electrode-dot-array digital microfluidic biochip. In *Proc. ICCAD*. 1–8.

[17] Zipeng Li et al. 2017. Droplet size-aware high-level synthesis for micro-electrode-dot-array digital microfluidic biochips. *IEEE Trans. on bioCAS* 11, 3 (2017), 612–626.

[18] Zipeng Li et al. 2017. Sample preparation on micro-electrode-dot-array digital microfluidic biochips. In *Proc. ISVLSI*. 146–151.

[19] Fei Su, William Hwang, and Krishnendu Chakrabarty. 2006. Droplet routing in the synthesis of digital microfluidic biochips. In *Proc. DATE*, Vol. 1. 1–6.

[20] Jack Tang et al. 2017. Secure Randomized Checkpointing for Digital Microfluidic Biochips. *IEEE Trans. CAD* (2017).

[21] Gary Wang, Daniel Teng, and S-K Fan. 2011. Digital microfluidic operations on micro-electrode dot array architecture. *IET Nanobiotechnology* 5, 4 (2011), 152–160.

[22] Zhanwei Zhong, Zipeng Li, and Krishnendu Chakrabarty. 2017. Adaptive error recovery in Micro-electrode-dot-array biochips based on droplet-aliquot operations and predictive analysis. In *Proc. ICCAD*. 615–622.

A APPENDIX

We present proofs for the lemmas and theorems stated in the paper. To help readers better understand the proofs, we list our notation in Table 2.

Table 2: Notation used for the proofs.

Symbols	Description
G_{in}	Input sequencing graph
G_{re}	Reconstructed sequencing graph
n_i	ith node in sequencing graph
I_i	ith dispensing operation in sequencing graph
S_i	ith splitting operation in sequencing graph
M_i	ith merging operation in sequencing graph
W_i	ith waste-droplet discarding operation in sequencing graph
D_i	ith detection for droplets in sequencing graph
lm_t	Location map as a Boolean-value vector generated at time t
t	Time scale in range $0 < t \leq T,\ t \in \mathbb{N}$
T	End time of assay execution, $T \in \mathbb{N}$
d_x	A droplet x on a given MEDA biochip
fp_t^x	Set of electrodes that represents the location of a droplet d_x at time t
U	Universal set of all the micro-electrodes in the MEDA platform

Lemma A.1: If a droplet d_x exists on the chip at time slots $t-1$ and t, $fp_t^x \cap fp_{t-1}^y = \emptyset$ for any other droplet d_y.

Proof. Let the footprint of droplet d_x at time $t-1$ be fp_{t-1}^x, and the droplet closest to d_x at time $t-1$ be d_y. Due to the need for a safe distance between droplets [17], the minimal distance between fp_{t-1}^x and fp_{t-1}^y is four micro-electrodes. For fp_t^x and fp_{t-1}^y to overlap, the droplet d_x has to move pass through at least four micro-electrode in a single cycle. According to [15], the length of one actuation cycle is 110 ms, and the length of one micro-electrode is 37 µm. This implies that d_x must move at a velocity of at least $(4 \times 37)/110 = 1.35$ mm/s. This velocity is well above the upper limit 1 mm/s [17]. Therefore, a droplet cannot overlap with other droplet's footprint from the previous cycle. □

Lemma A.2: A droplet d_x exists on a MEDA biochip at times $t-1$ and t if and only if $fp_t^x \neq \emptyset$, $fp_{t-1}^x \neq \emptyset$, and $fp_t^x \cap fp_{t-1}^x \neq \emptyset$.

Proof. ($\Rightarrow$) Suppose a droplet d_x exists at $t-1$ and t. Thus $fp_t^x \neq \emptyset$ and $fp_{t-1}^x \neq \emptyset$.The proof of $fp_t^x \cap fp_{t-1}^x \neq \emptyset$ is obvious from that fact that a droplet must overlap adjacent micro-electrodes to ensure movement. Next we prove the "if" part. A non-empty footprint fp_{t-1}^x indicates that a droplet d_x exists at time $t-1$. Likewise, a non-empty fp_t^x indicates that a droplet d_x exists at time t. Assume that these two droplets are not the same, and let the droplet at $t-1$ be $d_{\bar{x}}$. Since $fp_t^x \cap fp_{t-1}^x \neq \emptyset$, d_x must move to the previous location of droplet $d_{\bar{x}}$ at time t. This contradicts Lemma A.1. Hence, the two droplets d_x and $d_{\bar{x}}$ are actually the same. □

Lemma 2 (restated): A droplet d_x is dispensed on the MEDA platform between time slots $t-1$ and t if and only if $fp_t^x \neq \emptyset$ and $fp_{t-1}^x = \emptyset$.

Proof. ($\Rightarrow$) The droplet d_x does not appear on the platform until time t. As a result, the footprints of d_x before time $t-1$ are empty sets, and the footprints of droplet d_x after time t are not empty set.
($\Leftarrow$) From Lemma A.2 we know if fp_{t-1}^x or fp_t^x is an empty set, d_x must not exists at both $t-1$ and t. We also know that $fp_t^x \neq \emptyset$ indicates that d_x exists on the platform at time t. Thus d_x must not exist on the platform at time $t-1$. Hence, the droplet is dispensed to the platform at time t. □

Lemma 3 (restated): A droplet d_x is discarded from the MEDA platform between time slots $t-1$ and t if and only if $fp_{t-1}^x \neq \emptyset$ and $fp_t^x = \emptyset$.

Proof. ($\Rightarrow$) The droplet d_x exists on the platform until time $t-1$, implying $fp_{t-1}^x \neq \emptyset$ and $fp_t^x = \emptyset$.
($\Leftarrow$) Based on Lemma A.2, we know if fp_{t-1}^x or fp_t^x is an empty set, d_x must not exists at both $t-1$ and t. Since $fp_{t-1}^x \neq \emptyset$ indicates that d_x exists on the platform at time $t-1$, d_x must not exist on the platform at t. Hence, we conclude that d_x is discarded to waste reservoirs at t. □

Lemma 1 (restated): A droplet d_x is transported on the MEDA platform between time slots $t-1$ and t if and only if $fp_{t-1}^x \cap fp_t^x \neq \emptyset$ and $|fp_{t-1}^x| = |fp_t^x|$.

Proof. ($\Rightarrow$) Since d_x exists at $t-1$ and t, $fp_{t-1}^x \cap fp_t^x \neq \emptyset$ by Lemma A.2. The volume of a droplet equals to the droplet's bottom area times the gap between upper and lower plates. Since the volume of d_x and the gap between plates do not change over time, the bottom area of d_x remains the same, i.e., $|fp_t^x| = |fp_{t-1}^x|$.
($\Leftarrow$) By Lemma A.2, we know that these two footprints indicate the same droplet d_x. Besides, $fp_{t-1}^x \neq \emptyset$ indicates the area where d_x locates at time $t-1$, and so does $fp_t^x \neq \emptyset$. As a result, the droplet is transported at time t. □

Lemma 4 (restated): A droplet d_x is split from a larger droplet d_y on the MEDA platform between time slots $t-1$ and t if and only if $fp_t^x \neq \emptyset$, $fp_{t-1}^y \neq \emptyset$, $fp_t^x \cap fp_{t-1}^y \neq \emptyset$, and $|fp_t^x| < |fp_{t-1}^y|$.

Proof. ($\Rightarrow$) Assume that we divide the droplet d_y into two smaller droplets at time $t-1$, and let one of them be $d_{\bar{x}}$, which has the same volume of the droplet d_x and $fp_{t-1}^{\bar{x}} \cap fp_t^x \neq \emptyset$. According to Lemma 1, $fp_{t-1}^{\bar{x}} \cap fp_t^x \neq \emptyset$ and $|fp_{t-1}^{\bar{x}}| = |fp_t^x|$. We also know that $fp_{t-1}^{\bar{x}} \subset fp_{t-1}^y$ because the droplet $d_{\bar{x}}$ is part of the droplet d_y. As a result, $fp_t^x \cap fp_{t-1}^y \neq \emptyset$ and $|fp_{t-1}^y| > |fp_t^x|$.
($\Leftarrow$) Let $fp_{t-1}^{\bar{x}} \subset fp_{t-1}^y$, where $fp_{t-1}^{\bar{x}} \cap fp_t^x \neq \emptyset$ and $|fp_{t-1}^{\bar{x}}| = |fp_t^x|$. From the Lemma 1, we can infer that a droplet d_x is transported from area above $fp_{t-1}^{\bar{x}}$ to area above fp_t^x at time t. Since $fp_{t-1}^{\bar{x}} \subset fp_{t-1}^y$, the droplet d_x was part of the droplet d_y at time $t-1$. Hence, d_x is extracted from the droplet d_y. □

Lemma 5 (restated): A droplet d_x is obtained as a result of merging two droplets d_y and d_z on the MEDA platform between time slots $t-1$ and t if and only if $fp_t^x \neq \emptyset$, $fp_{t-1}^y \neq \emptyset$, $fp_{t-1}^z \neq \emptyset$, $fp_t^x \cap fp_{t-1}^y \neq \emptyset$, $fp_t^x \cap fp_{t-1}^z \neq \emptyset$, and $|fp_t^x| = |fp_{t-1}^y| + |fp_{t-1}^z|$.

Proof. ($\Rightarrow$) Because d_x exisits at t, $fp_t^x \neq \emptyset$. The droplet d_x can be viewed as two droplets $d_{\bar{y}}$ and $d_{\bar{z}}$ connecting together, where the volume of $d_{\bar{y}}$ equals to the volume of droplet d_y and the volume of $d_{\bar{z}}$ equals to the volume of droplet d_z. These two droplets are transported as d_x and d_y at t, respectively. According to Lemma 1, $fp_t^{\bar{y}} \cap fp_{t-1}^y \neq \emptyset$, $fp_t^{\bar{z}} \cap fp_{t-1}^z \neq \emptyset$, $|fp_t^{\bar{y}}| = |fp_{t-1}^y|$, and $|fp_t^{\bar{z}}| = |fp_{t-1}^z|$. Since droplets $d_{\bar{y}}$ and $d_{\bar{z}}$ are two parts of the droplet d_x, $fp_t^x = fp_t^{\bar{y}} \cup fp_t^{\bar{z}}$. Hence, $fp_t^x \cap fp_{t-1}^y \neq \emptyset$, $fp_t^x \cap fp_{t-1}^z \neq \emptyset$, and $|fp_t^x| = |fp_{t-1}^y| + |fp_{t-1}^z|$.
($\Leftarrow$) fp_t^x can be partitioned into two footprints $fp_t^{\bar{y}}$ and $fp_t^{\bar{z}}$, where $|fp_t^{\bar{y}}| = |fp_{t-1}^y|$, $fp_t^{\bar{z}}| = |fp_{t-1}^z|$, $fp_t^{\bar{y}} \cap fp_{t-1}^y \neq \emptyset$, and $fp_t^{\bar{z}} \cap fp_{t-1}^z \neq \emptyset$. According to Lemma 1, the droplet d_y is transported to the area above $fp_t^{\bar{y}}$, and the droplet d_z is transported to the area above $fp_t^{\bar{z}}$ at time t. Since two footprints $fp_t^{\bar{y}}$ and $fp_t^{\bar{z}}$ are subsets of the footprint fp_t^x, two droplets d_y and d_z are actually two parts of the droplet d_x at time t, i.e., they merge as the droplet d_x at time t. □

Lemma 6 (restated): A droplet d_x is detected on the MEDA platform at time T (end of the assay) if and only if $fp_T^x \neq \emptyset$.

Proof. The proof is obvious from the assumption that on-chip droplets at the end of the assay are for detection. □

Lemma 7 (restated): The sequencing graph can be reconstructed from a complete sequence of droplet-location maps from the start to the end of the assay, i.e., $(lm_1, lm_2, ..., lm_T)$.

Proof. Let fp_t^x be any footprint obtained at t, where $1 < t \leq T$, and $fp_t^x \subset U$. The operation of d_x at t can be inferred from fp_{t-1}^x. Because location maps are consecutive, the dependencies between the contiguous operations are also acquired. Hence, a sequencing graph can be reconstructed from the complete sequence of location maps, and G_{re} can be obtained by omitting transportation operations from this graph. □

Theorem 1 (restated): If a sequencing graph G_{in} is synthesized on a MEDA biochip and the biochip is not compromised by an attack, the sequencing graph G_{re} reconstructed from the droplet-location maps is isomorphic to G_{in}. If an attack compromises one or more operations, G_{re} is no longer isomorphic to G_{in}.

Proof. Let n_i be ith node in G_{in} where $n_i \in \{I_i, D_i, S_i, M_i, W_i\}$. After the assay is synthesized on a MEDA biochip, there exists a footprint fp_t^i for operation n_i at time t. The operation n_i in G_{re} is reconstructed from fp_t^i based on Lemma 7. Since node n_i can represent any operation in G_{in}, G_{in} and G_{re} must be isomorphic.

Based on the attack classification in Table I from the paper, two kinds of modification on G_{in} are 1) node change and 2) edge redirection. In the first scenario, assume n_i in G_{in} is compromised by an attack, and n_i is replaced with $\bar{n_i}$. After the assay is synthesized on a MEDA biochip, there exists a footprint fp_t^i for operation $\bar{n_i}$. The operation $\bar{n_i}$ is reconstructed in G_{re} from fp_t^i based on Lemma 7. Since $n_i \neq \bar{n_i}$, G_{re} is not isomorphic to G_{in}. In the second scenario, assume there is an edge from n_i to n_j in G_{in}, and the edge is redirected from n_k to n_j due to an attack. After the compromised assay is synthesized on a MEDA biochip, there exists footprints for operations of n_k, n_j, and transportation from n_k to n_j. These operations are reconstructed in G_{re} based on Lemma 7, and there is an edge connecting n_k to n_j. Because the edge from n_k to n_j is not the same as the edge from n_i to n_j, G_{re} is not isomorphic to G_{in}. □