Abstract—Accurate flow cytometry analyses for disease diagnosis purposes requires powerful computational and storage resources that are rarely available in clinical settings. The emerging high-performance cloud computing technologies could potentially address the above-mentioned scalability challenge; however, potentially untrusted cloud infrastructures increases the security and privacy concerns significantly as the attackers may gain knowledge about the patient identity and medical information and affect the consequent course of treatment. In this paper, we present TSC, a trustworthy scalable cloud-based solution to provide remote cytometry analysis capabilities. TSC enables the medical laboratories to upload the acquired high-frequency raw measurements to the cloud for remote cytometry analysis with high-confidence data security guarantees. In particular, using fundamental cryptographic security solutions, such as the trusted platform module framework, TSC eliminates any possibility of unauthorized sensitive patient data exfiltration to untrusted parties, e.g., malicious or compromised cloud providers. Our evaluation results show that TSC effectively facilitates scalable and efficient disease diagnoses while preserving the patient privacy and treatment correctness.

I. INTRODUCTION

The holy grail of biomarker based disease diagnosis is advantageous not only because it enables early disease diagnosis, but also because it enables personalized medicine. The complete understanding of the patient’s condition at the molecular level will allow for the ability of prescribing a course of treatment which has a higher probability of being successful compared to making a prescription solely based on the patients outwardly symptoms. The vision of personalized medicine can be made a reality through the use of ubiquitous inexpensive detectors for monitoring biomarker levels.

One of the most commonly used biomarkers for disease diagnosis and monitoring is through obtaining cell counts. This includes both infectious disease and cancer. For example, one of the most commonly used methods for diagnosing and monitoring HIV patients is through obtaining CD4 cell counts. Other prevalent infectious diseases include tuberculosis and Malaria. An important test for diagnosing tuberculosis involves counting lymphocytes, whereas monocyte counts can offer important information for diagnosis of malaria. In the case of monitoring therapy for cancer, the counting of erythrocytes, leukocytes, neutrophils, eosinophils, basophils, lymphocytes, monocytes, and platelets all provide relevant information. Thus, it is absolutely necessary to develop a disposable diagnostic tool, which can rapidly and accurately count cells.

Cell counting can either be performed electrically using coulter counting or optically using flow cytometry. Often performing complete cell counts requires processing large volumes of test samples at high throughput. For example, to analyze a 10 ml blood sample for one hour would require sampling at least with a speed of 10 Msamples per second, generating Gbytes of data. Processing this data would require heavyweight digital filtering algorithms and also peak counting and analysis, all of which could be computationally expensive for a single embedded system. As a result, in scenarios such as this, secure transmission of the data to centralized servers, which would perform the necessary data processing, would be more efficient.

Recently, security in networked medical devices has become of concern both in terms of privacy of medical records, and also the threat of cyberattacks. In the context of cytometry, malicious altering of cell counts of patients can result in misdiagnosis and also can alter the course of therapy and treatment. In this paper, we present a software-based scheme for encryption of cytometry data to ensure secure transmission from a low cost ubiquitous biosensor to the cloud where post-processing of the data is rapidly performed. The cloud computational resources in TSC make use of cryptographical secure solutions such as trusted platform module (TPM) solutions to provide an efficient trustworthy execution and data processing environment for cytometry data analysis.

II. RELATED WORK

Conventional FACS is costly and does not meet the low cost and limited resource settings constraints, due its need for expensive labels and bulky optical equipment. As an alternative to lowering the optics cost, a promising strategy to reduce the instrumentation cost associated with cytometers is to move towards purely electrical detection, similar to what has been done previously with protein detection [5]. However, up until now, much effort has been made towards the development of low-cost bio-optical detection technologies, in particular by Ozcan et al. [2], [3], [6] Some promising approaches include the use of microfluidic channels integrated on miniaturized cameras, such as those integrated into cellphones, for ultra cheap cytometers [24]. Electrical detection can provide a more cost-effective solution to cytometry. The lock-in impedance measurement, which involves two electrode excitation and measurement system is the most common method for performing cytometry [3], [24]. In this method, one electrode is excited with an AC voltage, and the resulting current signal flowing from the second electrode is measured through subsequent
amplification, mixing (with a local oscillation signal), and low pass filtering stages. Hywel Morgan et al. demonstrated single cell dielectric spectroscopy using a microfabricated flow cytometer based on a multi-frequency lock-in technique [11], [22]. Saleh and Sohn [14]–[16], [21] used a four electrode sensing method for protein detection application. Other groups have improved both the impedance sensing scheme and throughput of the micro-fabricated Coulter counter. [25] Wu et al. [27] incorporated symmetric mirror channels in their sensing scheme. The use of symmetric mirror channels in the sensing scheme resulted in an improvement in the measured signal-to-noise ratio allowing for detection of 520 nm-diameter particles in a sensing pore of 50 × 16 × 20 μm. A four aperture design was also demonstrated [7], [8] capable of detecting and counting micron-sized particles through the corresponding sensing channels simultaneously. A high bandwidth RF probe was used to report a counting rate of 30 kHz in a single microfluidic channel [26]. Other interesting techniques such as the use of three-dimensional hydrodynamic focusing [18] have been proposed, creating a virtual narrow wall for maximizing detection sensitivity without risking the clogging of the channel. Recently demonstrated coulter counter applications include determination of the spermatozoa concentration in semen [19] as well as quantification of red blood cells in diluted whole blood using a microfluidic chip with polyelectrolytic gel electrodes (PGEs) [9]. In all of the above cases, massive amounts of data must be processed in order to attain accurate particle counts. In order to maintain the low cost of the cytometer, readout instrumentation, the implementation of the various data processing algorithms can be more rapidly performed on data processing servers, as opposed to performing the data analysis on an integrated embedded system.

On the other hand, over the past few years, there has been an increasing interest in remote secure computation solutions. Almost all the past efforts could be categorized into two groups. First, there are cryptographic methods [20] that make use of mathematical primitives to provide data processing capabilities over encrypted data in cloud without having to download the data locally. Boneh et al. [4] proposes a searchable encryption scheme using asymmetric cryptography. Using their proposed approach, the users could query for particular encrypted records via conjunctive range, subset, and comparison logical predicates. Shi et al. [20] later introduce multi-dimensional range querying framework over encrypted data that slightly improves the time complexity of [4] in particular cases. Second, there are system solutions for remote secure computation that focus on development of effective trust management engines. The most well-known recent methods make use of trusted platform module [12] in cloud infrastructures such the clients could remotely verify the execution environment and lack of unknown (and potentially malicious) processes, rather than protecting data itself. The major benefit of system solutions over the cryptographic secure computation techniques is that system techniques allow a lot more data processing and calculus functionalities whereas the cryptographic techniques facilitate particular functions such as search and addition.

III. THREAT MODEL

We describe our assumptions about the threat model against TSC. The cloud infrastructure used in TSC is assumed to be potentially untrusted that may be caused by either a malicious cloud provider [1], [23] or an attacker that manages to compromised the cloud infrastructure through exploitations of system vulnerabilities. In either case, to ensure the trustworthiness of cloud-based cytometry, TSC needs to deploy relevant security protection mechanisms. The trusted computing base [10] in TSC constitutes the cytometry data acquisition device that need to be trusted. This is practically a reasonable assumption as the device resides in the medical lab where the patient data is collected.

IV. TSC DESIGN

A. CYTOMETRY DATA ACQUISITION

In the context of protein biomarker detection, we proposed a decoupled architecture for protein capture, and impedance sensing resulting in at least 10x improvement in sensitivity compared to fluorescence based protein detection techniques. In this architecture, we inject test sample into the capture chamber, thus capturing target protein, then we incubate micron sized beads coated with secondary antibodies which bind to the proteins forming sandwich immunocomplexes. We wash off the loosely bound beads, and finally elute the specifically bound beads flowing them through the impedance cytometer detecting them one by one, thus quantifying protein abundance.

The cytometer was designed to have single bead sensitivity. We used a three electrode differential lock-in measurement (see Figure 1) architecture where the impedance is measured between electrodes A and B, and also between B and C, and the difference between the two is subtracted. An AC voltage is applied to electrode B, and electrodes A and C are tied to transimpedance amplifiers, the output of which feeds into a differential amplifier and then an amplitude demodulator. The resulting signal from each bead passing through is a signature negative and then positive peak. Figure
that transparently allows the users and applications to attest VMs whose fingerprints are verified to be intact. To restrict upload of the patient’s data to the cloud-based environment is proved to be untampered with by unauthorized parties, e.g., malicious cloud providers.

To prove the cloud integrity, TSC uses a remote attestation technique [17] that enables authorized remote users in the cytometry data acquisition laboratory to identify unauthorized modifications to their cloud-based execution environments, i.e., virtual machines (VMs). To that end, TSC connects to a trusted platform module (TPM) [13], a dedicated embedded microprocessor, that creates a hash key fingerprint of the hardware and software stack that bootstrapped on the cloud-based computer. Every TPM chip has a unique and secret cryptographic RSA key burnt in, and hence it is capable of performing platform authentication. After a secure connection between TSC and the cloud-based TPM is established, TSC authenticates the hardware and software stack running on the cloud. TSC follows a successful TPM-based cloud platform authentication by uploading the acquired cytometry patient data within the laboratory to the cloud. Needless to mention, the data are encoded using symmetric AES-512 encryption before the upload to prevent man-in-the-middle intrusions. Upon receipt of the data, the verified cloud platform decrypts the measurements and starts cytometry analyses such as couter counting and peak detection for disease diagnosis purposes. Consequently, TSC’s main objective from TPM-based remote attestation is to restrict upload of the patient’s data to the cloud-based VMs whose fingerprints are verified to be intact.

In our real-world implementations, we used an extended TPM for VMs on a cloud hypervisor, so-called vTPM [13], that transparently allows the users and applications to attest not only a cloud-based physical computer but also the software stack within the VMs running on that computer. Figure 2 shows our implementation setup.

V. Evaluations

To utilize vTPM for our VMs, we used a quad-core and TPM-enabled server with Ubuntu 12.4 operating system as the cloud-based computer. We installed a Xen 4.3 hypervisor with XSM/Flask security framework enabled. On Xen, we configured Dom0 as the default VM to manage other user-created VMs, so-called DomU. We installed a customized Linux kernel 3.7.1 for Dom0 that supports vTPM, and compiled a Linux kernel 3.7.9 as DomU that can be instantiated multiple times by the user. After successful creation of vTPM Manager, we built vTPM instances with their corresponding disk images. Finally, in the VM template of the guest VM, we associate the VM to a created vTPM. This enables the cloud user to attest the software stack on the VM before uploading and processing the patient’s data.

Processing of the collected patient’s data can be significantly expedited through parallelizing of the data processing. As mentioned earlier, we can utilize Cloud resources as a powerful and ubiquitous platform for parallel processing of the data. However, to address the security concerns involved in processing of the patient’s data on the Cloud, in this research, we have deployed virtual TPM (vTPM) [13] on the VMs provided by the Cloud providers. Using vTPM on the VMs assures that there has not been any malicious modification in the platform stack of the VMs. Nonetheless, vTPM imposes performance overhead to the VMs.

To evaluate the performance gain of parallel processing of the patient’s data on vTPM-enabled VMs, we set up a private Cloud platform. We also implemented a parallel version of our data processing to be able to execute the data processing on a multi-core architecture. Specifically, in this experiment, we have evaluated the performance of processing data on VMs with different number of cores (from one core to four cores) in two scenarios: without vTPM and with vTPM enabled on the VMs. The result of the experiment is illustrated in Figure 3.
We notice that, in general, as the number of cores increases, the time to process the patient’s data decreases. The nature of our collected data was in a way that the best parallelization performance is obtained when three cores are utilized for processing. In particular, we notice that the processing time drops from 11.30 seconds, when one core is utilized, to 7.4 seconds, when three cores are utilized. The reason for not getting further performance improvement is that the process is IO intensive (i.e., there are many Input/Output (IO) operations in the processing of the patient’s data) that operates as a bottleneck in the parallel processing of the data.

Additionally, we observe that in the scenario where vTPM was enabled on the VMs, the processing time increases slightly. The increase is due to the performance overhead imposed by the operation of vTPM. As demonstrated in Figure 2, the performance overhead is because of creating vTPM and vTPM manager virtual machines to enable a vTPM for a VM. However, as we can see in Figure 3, the performance loss is at most 0.2 seconds when two cores are utilized, which is insignificant.

VI. CONCLUSIONS

In this paper, we presented TSC, a trustworthy and scalable cytometry framework that enables high performance flow cytometry analysis for disease diagnosis purposes. TSC makes use of cloud computing infrastructures to satisfy high computation and storage requirements of coulter counting and peak detection procedures. To guarantee zero information leakage and patient privacy violations, TSC employs trusted platform module and information flow tracking solutions. Our evaluation results are very promising and show that TSC provide medical laboratories with an efficient and trustworthy environment for remote electronic patient data processing capabilities within a public cloud. Consequently, we conclude by noting that the processing of the patient’s data can be expedited by utilizing Cloud resources without any security concern for the patient’s data.

REFERENCES