ABSTRACT
Methodologies for evaluation of e-Health platforms are still lacking. We propose an e-Health evaluation framework that aims to provide a concise methodology for the evaluation of e-Health platforms under three main categories: usability evaluation, logistics evaluation and technical evaluation. For the scope of this paper, focus is given on defining techniques to enable the gathering of technical evaluation metrics using our proposed framework. Technical evaluation provides insight on the scalability, response time, functionality and performance of an e-Health platform. Using simulated patient data, preliminary evaluation experiments carried out on an existing e-Health platform, known as the Cloud4Health platform, proves the viability of our defined methodology for gathering of technical evaluation metrics. Experimental results showed up to 100 simulated patient's data could interact with the Cloud4Health platform concurrently with a maximum round-trip time latency value of 6.6 seconds for data samples to be processed. Future work aims to build upon this evaluation framework and develop techniques to gather both usability and logistics evaluation metrics when conducting evaluation of e-Health platforms.

KEYWORDS
e-Health evaluation, methodologies, platforms, simulated data, technical evaluation, evaluation metrics

1. INTRODUCTION
E-Health is the migration of healthcare, from traditional paper-based systems to electronic counter-parts (Haeyrinen et al., 2008). A prime example of e-Health is Electronic Health Records (EHR). Not only has EHRs removed the need for physical storage space, it also enables medical staff to look up a patient's medical record quickly and easily. The Commonwealth, an intergovernmental governmental organisation, in a document published by Rodrigues (2008), has acknowledged further benefits e-Health brings.

However, in the very same document, Rodrigues (2008) also highlights the many challenges e-Health must overcome before this technology is widely adoptable. Key challenges e-Health must overcome, from an organisational perspective, include governance, standardisation and cost (Rodrigues, 2008) whilst from a technical perspective we must consider the functionality, security and scalability of the technology (Pagliari, 2007; Henderson et al., 1999). From a public perspective, a key challenge for e-Health adaptation is in ensuring trust in this technology. Various reports on the inadequacy of e-Health technologies have emerged in UK news these past years. A prime example of this is in 2009, when the National Health Service (NHS) in England lost thousands of medical records (Savage, 2009) due to a lack of security in their computer systems. More recently, in July 2011, the NHS was once again put under the spotlight when computer criminals attempted to gain access to their systems that held patient medical records (Campbell, 2009). Hence, one can understand why patients may feel rather uneasy about medical facilities storing their personal data in an e-Health environment.

Numerous e-Health platforms have been proposed and developed to address the key challenges this technology faces including work by World Medical Card (2012), DOSSIA (2012), HealthVault (2011) Fan et al. (2011) and Rolim et al. (2010). The contribution this paper makes is in presenting the design of an e-Health evaluation framework that is capable of evaluating these platforms under the categories of usability, logistics and technical evaluation. The design of this e-Health evaluation framework aims to provide answer
for the question of whether or not an e-Health platform is capable of overcoming the key challenges it currently faces. For the scope of this paper we focus on methodologies for obtaining technical evaluation metrics.

The remainder of this paper is organised as follows. Section 2 discusses related works in regards to e-Health evaluation. Section 3 presents the current design of our e-Health evaluation framework. Section 4 provides methodologies for obtaining technical evaluation metrics along with design of preliminary experiments. Section 5 presents the results from conducting a preliminary technical evaluation on an existing e-Health platform. Finally, Section 6 provides a conclusion along with future work.

2. RELATED WORKS

Recent literature has shown that the primary method which e-Health evaluation takes place is live clinical trials. This, in essence, involves deploying an e-Health platform in a live clinical environment, e.g. hospital, and obtaining evaluation results from clinical staff and patients using the system on a day-to-day basis.

Research conducted by Cobb et al. (2005) present a live clinical trial of an e-Health system that provides support for patients wishing to quit smoking whilst Kessler et al. (2009) present a system for the delivery of psychotherapy via the internet. The work of Flynn et al. (2009) presents a third example of a live clinical trial in which they evaluate the effectiveness of an e-Health patient booking appointment service. This paper differs from all three authors’ work since the proposed evaluation framework considers e-Health from a high-level perspective, i.e. assessing the key challenges of e-Health from an organisational, public and technical perspective rather than how well the technology works for specific healthcare services, i.e. smoking cessation, therapy and appointment booking system.

In regards to evaluating e-Health from a high-level perspective, it has been found that although many propositions have been made on what should form part of an evaluation framework for testing e-Health platforms, little work has been presented on defining a clear methodology to conduct such an evaluation. Dansky et al. (2006) has noted “few blueprint for effect evaluation methodologies” (Dansky et al., 2006, p. 397) exist whilst the same backing can be provided in the work of Lilford et al. (2009) in which they find that no general consensus can be found in techniques used to evaluate e-Health platforms. Furthermore, although the work of Glasgow et al. (2007) highlights some of the key metrics which an e-Health evaluation should obtain, including appeal, use, cost and robustness, no fundamental answer on how such metrics should be measured is given.

The most concise evaluation methodology was found to be in the work of Catwell et al. (2009) in which the authors propose that evaluation of e-Health should take place from the very initial steps of designing an implementation. However, although they present workflow on how such an evaluation should take place, Catwell et al. (2009) do not provide specific metrics of evaluation. This research differs from the work of Catwell et al. (2009) in that the proposed evaluation framework is designed to evaluate already existing e-Health platforms rather than systems that are still in early stages of design along with presentation of specific metrics for evaluation. Our e-Health evaluation framework is presented in the next section of this paper.

3. E-HEALTH EVALUATION FRAMEWORK

3.1 Overview

Our e-Health evaluation framework is designed around three main categories of evaluation: Usability Evaluation, Logistics Evaluation and Technical Evaluation (Figure 1). These three categories are adapted from the work of Dansky et al. (2006) where they initially proposed a framework for the implementation of e-Health projects. It is proposed that the definition of these three main categories enables a concise evaluation of e-Health platforms to take place and is capable of answering the question of whether an implementation overcomes the key challenges this technology faces from an organisational, technical and public perspective.

In conducting evaluation of an e-Health platform with groups of clinical staff and patients, i.e. the public perspective, a key question to answer would be the ease-of-use in the system (Usability Evaluation). Having
a simple, understandable and intuitive system will most likely result in more trust in the e-Health platform in comparison with a complex system. An evaluation catered towards shareholders and directors of clinical environments, i.e. the organizational perspective, would more likely address questions such as the cost, upkeep and management of the system (Logistics Evaluation). Finally, for developers and IT administrators of e-Health platforms, i.e. the technical perspective, the key questions that should be answered in this category of evaluation is the functionality, security and scalability of the system (Technical Evaluation).

For the scope of this paper, the primary focus is on defining techniques for obtaining metrics under the category of Technical Evaluation. As our evaluation framework is still under further development, presentation of technical evaluation metrics aim to serve as a proof-of-concept of our design along with validity of this framework. Presentation of the two other categories of e-Health evaluation will be provided at a future date (see Section 6 for details). The next section discusses the design of technical evaluation metrics.

3.2 Technical Evaluation Design

As the name implies, Technical Evaluation focuses on assessing an e-Health platform from a technical standpoint therefore, many pre-existing evaluation metrics used in the testing of computer systems and networks can be used. The following is an outline of technical evaluation metrics which are currently defined as part of this framework:

- **CPU Utilisation**: relates to how much processing time is required for the upload and/or download of health care data along with general interaction with the e-Health platform. The measurement of this metrics provides us with an overview on whether the current hardware infrastructure is up to a sufficient standard for the hosting of a chosen e-Health platform. Results obtained via this metric are dependent on the processor(s) of the hosting platform. CPU utilisation is measured under the unit of percentage (%).

- **Packet Loss**: relates to the number of healthcare data samples that are lost or dropped during an evaluation. A low packet loss (preferably zero) is highly desirable, if not essential, for e-Health platforms since the key attribute in which all interactions take place around is healthcare data.

- **Upload / Download Time**: the duration of time taken for a healthcare data to be uploaded and retrieved from the e-Health platform respectively. Results obtained from this metric are dependent on the network interface cards (NIC) and protocols used by the e-Health platform under evaluation. Measured in units of ms (milliseconds).

- **Round-Trip Time (RTT)**: the duration of time taken for healthcare data to be uploaded to the e-Health platform, processed and then outputted to a clinical service, e.g. a patient monitoring system or EHR. As before, results from this metric are also dependent on NICs and the protocol used by the e-Health platform under evaluation. Measured in units of ms (milliseconds).

In order to obtain our defined technical evaluation metrics, it can be seen that some form of healthcare data must first interact with the e-Health platform under evaluation. We have chosen to use patient data for interaction with the e-Health platform under evaluation since this is the primary attribute which clinical environments revolve around. For the current scope of our framework, we have opted to use simulated patient data in place of real world patient data. Justification for the use of simulated data along with design of this software, named the Patient Simulator, is discussed in the next section.
3.3 Patient Simulator Design

To conduct a technical evaluation of an e-Health platform, and obtain results under the metrics defined in the previous section, patient data is required. To enable rapid evaluation of e-Health platforms to take place, we have chosen to simulate patient data instead of using real-world patient data. Simulating patient data also mitigates legal (Madsen et al., 1999) and ethical (Hardiker and Grant, 2011) barriers which real-world patient data imposes. In essence, the Patient Simulator is a software application that models and simulates virtual patient’s data and uses this data to interact with the e-Health platform under evaluation. The Patient Simulator was developed using Microsoft .NET C#.

The patient data we have chosen to simulate is known as the vital physiological signs of a patient. This includes the heart rate, blood pressure, temperature, respiratory rate and oxygen levels of a patient. In other words, these are the key attributes of a patient which would interact with an e-Health platform within a clinical environment. For the simulation of vital physiological signs, a Discrete Event-based Simulation (DES) method (Banks et al., 1984) is employed. DES refers to a simulation system in which variables only change at specific points in time (known as the time interval). A vital sign value is generated periodically using random normal distribution techniques (Martin, 1971). Table 1 presents the default mean and arbitrary standard deviation applied to all five vital physiological signs which are simulated in a patient by default.

<table>
<thead>
<tr>
<th>Vital Sign</th>
<th>Mean Value</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Rate</td>
<td>80 BPM</td>
<td>1.5</td>
</tr>
<tr>
<td>Blood Pressure</td>
<td>112 mmHG</td>
<td>2</td>
</tr>
<tr>
<td>Temperature</td>
<td>36.8 C</td>
<td>0.2</td>
</tr>
<tr>
<td>SpO2</td>
<td>97 %</td>
<td>0.5</td>
</tr>
<tr>
<td>Respiratory Rate</td>
<td>12 Breathes Per Minute</td>
<td>1</td>
</tr>
</tbody>
</table>

The mean value of blood pressure is based on the work of Pesola et al. (2001) in which they state that a normal systolic blood pressure is found to be 112 mmHg. From studies carried out by both Mackowiak et al. (1992) and Shoemaker (1996) the result of 36.8°C is applied for the mean body temperature. O’Driscoll et al. (2008) defines normal SpO2 as 96-98%, hence the average value of 97% is used. Finally, both Sherwood (2006) and Tortora and Derrickson (2008) agree that the mean respiration rate is found to be 12 breaths per minute.

In this paper, although variance and randomisation techniques were applied to the simulated patient data, the values generated were found not to be sufficiently realistic in comparison with real-world patient data. However, we propose the data simulated is still valid for experiments presented in this paper since we are evaluating the e-Health platform from a technical perspective, e.g. measuring response time and packet loss of arbitrary healthcare data samples, rather than evaluating the accuracy of clinical services, e.g. patient monitoring systems or EHRs. We outline technical evaluation experiments, conducted using simulated patient data, in the next section.

4. TECHNICAL EVALUATION EXPERIMENTS

4.1 Outline and Methodology

We present two sets of experiments in relation to technical evaluation for this paper. We have chosen to conduct the experiments on an existing e-Health implementation known as the Cloud4Health (Fan et al., 2011) platform (formally known as the DACAR project). Both experiments were conducted using simulated patient data generated on-the-fly at the start of each test. In each experiment, the only variation is the number of patients we choose to simulate. A grey-box testing approach was taken. The Cloud4Health project provided an API in order for the simulated data to interact with the platform but we do not consider the underlying hardware architecture or source code during experimentation. This helps provide a non-bias evaluation of the platform along with ensuring our defined evaluation metrics are not restricted to evaluating only one specific e-Health implementation. The scenario for each experiment is outlined as follows:
• **Experiment 1: Baseline Test** - The first test is very simplistic in nature. In simulating 100 samples of a single patient's data, the upload time was monitored for intervals of 0.5, 1 and 3 seconds. In other words, 100 samples of a single patient’s data was uploaded with three different time delays in order to evaluate whether any performance impact was found on the Cloud4Health platform based on how “talkative” the client, i.e. Patient Simulator, is. The primary aim of this experiment is establish a baseline result for how well the Cloud4Health platform handles a single patient's data being uploaded.

• **Experiment 2: RTT and CPU Utilisation** - The second experiment aimed to evaluate the RTT and CPU utilisation of the Cloud4Health platform. Up to 100 patient's data was uploaded concurrently. Each patient simulates 100 samples of data. The aim of this experiment is to assess the Cloud4Health platform's latency under a more realistic scenario which involves the input, processing and output of patient data.

To provide a brief summary, Experiment 1 measures the Technical Evaluation metric of Upload Time whilst Experiment 2 measures the metrics of RTT and CPU utilisation. Furthermore, we monitor packet loss in both experiments. The methodology for gathering these metrics is as follows:

• **CPU Utilisation**: A Microsoft Powershell script running directly on Cloud4Health server was used to obtain this metric. The script monitors the counter referred to as \processor\_total\% processor time (Edmead and Hinsberg, 2011). This counter returns the overall CPU utilisation of the server and the current value is logged to an output file every 1 second interval.

• **Packet Loss**: The total current samples stored in the Cloud4Health platform for each virtual patient was noted prior to the start of an experiment. The total samples simulated (100 per patient) is then subtracted from the current samples for each patient. A numeric value greater than 0 gives indication on the number of packets lost during the experiment.

• **Upload Time**: We gather upload time directly from the Patient Simulator application via the Stopwatch class (Microsoft, 2012) provided by .NET C#. An instance of the Stopwatch class is started upon uploading of data, and once the upload operation is complete, the Stopwatch is stopped. The elapsed time produced by the Stopwatch class results in the time taken to upload a single patient's data to the Cloud4Health platform.

• **RTT**: A push notification (Franklin and Zdonik, 1998) service was implemented on top of the Cloud4Health platform. Via the implementation of a Receiver client (acting as an example clinical service), the RTT metric is calculated based on subtracting the time stamp a packet was received (by the Receiver service) against the time stamp of when a patient data sample was sent to the e-Health platform (via the Patient Simulator). Strong time synchronization was achieved via an Active Directory server acting as time synchroniser (Smith, 2010) between the Patient Simulator and Receiver client machines.

5. RESULTS AND DISCUSSION

5.1 Experiment 1 - Baseline Test Results

Figure 2 provides the average upload time for 100 samples of a single virtual patient's data uploaded at intervals of 0.5, 1 and 3 seconds.

![Average Upload Time for 100 Samples](image)
No packet loss occurred during the running of this experiment. As the interval time decreases, the time taken to upload a single virtual patient’s data increases. Two conclusions can be made from this first experimental result: 1) talkativeness of a client, when upload a single patient’s data, only affects the Cloud4Health platform in a very minor manner and 2) upload times for a single patient’s data is exceptionally good with results of less than 58 ms when uploading data in intervals of 0.5 seconds.

5.2 Experiment 2 – Round-Trip Time and CPU Utilisation Results

As part of the second experiment, we present the technical evaluation results of RTT latency and CPU utilisation of the Cloud4Health platform. No packet loss occurred in the instance of running this experiment. Figure 3 shows the average RTT latency when simulating and uploading 20, 60 and 100 patient’s data whilst Figure 4 shows the CPU Utilisation results. Due to wide variance, Table 2 is also presented to give an overview of the minimum and maximum RTT latency values gathered during this experiment.

![Figure 3. Average round-trip time](image)

The results from the RTT graph show that in the case of simulating and uploading data for 20 virtual patients concurrently, the average RTT latency was very reasonable at 62.93 ms. In simulating 60 virtual patient’s data concurrently, the average RTT latency was found to be 64.80 ms – which is only a very minor increase in comparison with 20 patients. On the other hand, simulating 100 patients produced a significantly higher latency of 173.56 ms on average. Furthermore, with a maximum latency of 6674.21 ms and minimum latency of 51.11 ms, there is far wider range of variance in RTT latency when simulating 100 patients. This wide range of variance is an indication that a bottleneck may be present on the Cloud4Health platform.

![Figure 4. Average cpu utilisation](image)

CPU utilisation was the primary suspect in this increased variance with RTT latency but Figure 4 shows this is not the case. Even in the scenario of simulating 100 patients, the average CPU utilisation of the Cloud4Health platform was only slightly greater than 20%. Hence, it can be concluded from this experiment that although RTT latency values grow as the number of patients simulated increase there is currently no direct evidence to show that this has any correlation with CPU utilisation.

The experiment was stopped at 100 patients data simulated as we found that the Cloud4Health platform was unable to handle any higher volumes of patient data without unexpected errors. Although the experiment has shown that the Cloud4Health platform is generally responsive and uses very little CPU when processing
100 patient’s data concurrently, there is still room for improvement and optimisation to ensure this platform is capable of handling higher volumes of patient data.

Table 2. Minimum and maximum latency values

<table>
<thead>
<tr>
<th>Number of Patients</th>
<th>Min RTT Value (ms)</th>
<th>Max RTT Value (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>32.57</td>
<td>526.65</td>
</tr>
<tr>
<td>60</td>
<td>31.33</td>
<td>578.02</td>
</tr>
<tr>
<td>100</td>
<td>51.11</td>
<td>6674.21</td>
</tr>
</tbody>
</table>

6. CONCLUSION AND FUTURE WORK

We have proposed a framework which aims to evaluate e-Health platforms under three main categories: usability evaluation, logistics evaluation and technical evaluation. Methods for gathering of metrics related to technical evaluation are the focal point of this paper. Experiments carried out have demonstrated techniques that enable technical evaluation metrics, including upload time, CPU utilisation, packet loss and round-trip time, to be obtained from an e-Health platform. The primary test bed for conducting these experiments was the Cloud4Health platform and input of patient data was achieved via simulation in order to mitigate legal and ethical barriers imposed when attempting to use real-world patient data.

With capabilities of processing up to 100 patient's data concurrently with no packet loss, the experiments have shown that the Cloud4Health platform is very reliable in ensuring the safe arrival and storage of patient data in a very reasonable time (maximum of 6.6 seconds). However, there is also room for improvement for the scalability of the Cloud4Health platform since a live hospital and clinical environments will most likely have to cater for more than 100 patient's data being uploaded and download concurrently.

For future work, research is under way to define techniques to enable the gathering of metrics from the two other categories of evaluation this framework proposes: usability and logistics evaluation. Usability evaluation will involve a qualitative study on the general usability of a chosen e-Health platform from the perspective of the end-user, i.e. doctors, nurses and patients. The key goal of usability evaluation is to establish metrics that are capable of measuring a user’s perception when interacting with an e-Health platform and, via the analysis of the metrics obtained, determine whether the platform is capable of overcoming the challenge of public trust. We aim to conduct this study via interaction with real world clinical staff along with the completion of surveys and questionnaires.

Logistics evaluation will aim to provide techniques in order to obtain qualitative results on the general cost of using a chosen e-Health platform. This includes the cost involved in purchasing, hosting, maintaining and managing an e-Health platform from an organisational perspective. The Patient Simulator we have developed in this paper will be used for logistics evaluation to provide meaningful metrics such energy usage based on the number patient data interacting with an e-Health platform and hardware costs involved in storing increasing volume of patient data.

Finally, we also aim to improve upon techniques for gathering technical evaluation metrics from e-Health platforms. It is acknowledged that unexpected difficulties may arise in obtaining evaluation metrics from e-Health platforms other than the Cloud4Health platform, i.e. due to the use of different technologies. The grey-box testing approach taken in experiments carried out in this paper was the first step in ensuring our evaluation framework is interoperable with existing and future e-Health platforms. Future collaboration with the healthcare industry and adapting standard technologies will help further achieve this aim of interoperability. The end goal of this research is to provide an evaluation framework with a concise methodology to enable meaningful metrics to be obtained from any e-Health platform under test.

REFERENCES
