Preserving Content Integrity of Digital Holy Quran: Survey and Open Challenges

SAQIB HAKAK1, AMIRRUDIN KAMSIN1, OMAR TAYAN2, MOHD. YAMANI IDNA IDRIS1, ABDULLAH GANI1, AND SABER ZERDOUMI1

1Faculty of Computer Science and Information Technology Department of Computer System and Technology, University of Malaya, Kuala Lumpur, Malaysia
2Computer Engineering Department and NOOR Research Center, College of Computer Science and Engineering, Taibah University, Madinah, Saudi Arabia.

Corresponding author: Saqib Hakak (saqibhakak@gmail.com)

This work was supported by the University of Malaya through the Faculty of Computer Science and Information Technology under Grant RP003A 14HNE and Taibah University-NOOR Research IT Center Al-Madinah Al-Munawwarah, Saudi Arabia under Grant NRC1-126B, 2015-2016.

ABSTRACT In recent years, a new trend has come up, which is that of reading the digital Quran online. This text was revealed more than 1400 years ago in the Arabic language and has been protected from all possible ways of distortion until today. Unfortunately, driven by the desire to make profit or gain publicity, fraudsters have started modifying certain Quranic verses. These alterations are misleading many people who are thus deprived of the original and accurate message of the Holy Quran. This paper focuses on systematically analyzing and categorizing existing research related to preserving and verifying the content integrity of the Quran. This paper further assesses these existing studies in terms of their evaluation parameters and findings. We find that the existing studies can be classified according to their format and methods, i.e., the online formats in which the Quranic content is available, methods employed to protect the Quranic content from modification, and last methods of verification. This paper concludes with the issue of future challenges and their possible solutions.

INDEX TERMS Quran, content integrity, hadith validation, Quran verse authentication, tampering, authentication.

I. INTRODUCTION

The major challenge faced by researchers today is that the ever-growing range of digital media vastly exceeds the latter’s ability to put proper authentication or integrity mechanisms in place. The past years have witnessed a, massive increase in the use of digital content accessible through the internet which has dramatically increased cases of copyright violations and this raised the issue of integrity, authenticity of digital content and data vulnerability [1], [2]. Due to the same reason, more and more research is being undertaken in the area of data integrity, authentication and security. The statistics are shown in Fig 1 taken from the web of a science database.

The excessive reliability on the internet and the increase in users has further exaggerated the problem of integrity and authenticity. According to World Internet statistics [3], the number of internet users is increasing fivefold as shown in Fig 2. With this trend of increase in internet users, the rate of publishing sensitive digital content online is also on the rise. There is lot of sensitive digital content available online which can be accessed and downloaded from different sources, such as religious websites, social media websites and other online blogs.

By sensitive content is meant here that the content constitutes material of utmost importance which requires protection of confidentiality, integrity or availability. This sensitive content may appear in the form of text, image, audio or video. As sensitive content is also considered the Holy Quran as the most authentic and unaltered religious text of all times. It constitutes the duty of every Muslim to protect its authenticity and integrity [4]. Furthermore, the majority of Muslims today use the internet for online education on religious precepts, Quran recitation and memorization, banking and socializing [5], [6]. The Quranic text or content is available in the form of simple text or images on numerous websites. However, there is no proper way or procedure to ascertain the authenticity or integrity of this sensitive published content. Besides, there is no converging trend where attempt has been made to unify various approaches for the purpose of authentication in the area of sensitive content like Digital