Web Transcript Verification Using Check Digit as Secure Number

Muhamad Rais Abd Halim, Hairulnizam Mahdin, Siti Salwani Yaacob
Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia.
rais@uthm.edu.my

***Abstract*—The case of fraudulent on academic qualification document is increasing due to the advancement of editing technology. This makes document forgery easy to be done. It has becoming a worldwide problem to verify the academic documents including the degree's certificate and academic transcript. It is frightening situation when people without the right qualification working as professional in the area that will harm the society. Manual verification can be time consuming and not practical to be implemented. Therefore, it is paramount to have a web-based solution that can function all day long to perform this task. In this paper, dual check digit approach is proposed to verify the authenticity of an academic transcript. Three check digit methods which are Universal Product Code Mod 10 (UPC Mod 10), International Standard Book Number (ISBN Mod 11), and Luhn Mod 10 have been applied to test their reliability in carrying this task. The simulation results show that the combination of ISBN-11 and UPC Mod 10 performed very well to validate the authenticity of the transcripts.**

***Index Terms*—Check Digit; Transcript Verification; UPC Mod 10; ISBN Mod 11; Luhn Mod 10.**

I. INTRODUCTION

The cases of forgery on academic documents such as degree's certificate and academic transcripts are increasing nowadays [1]. One of the factors is the advancement of inexpensive hardware and software that enable people to produce document that is highly resemble with the original. This is actually a frightening scenario because there is chance that we will deal with unqualified personnel that used fake certificate and transcript to work in a particular field. There are many approaches that have been proposed to check the authenticity of degree certificate [2–5] but there is not much work has been done in verifying academic transcript. Academic transcript carry details information on the subjects that have been taken with the grades scored. It also includes the points earned based on the grades that will categorized the degree whether it is distinction or general. Some of the jobs require not only degree but also specify minimum pointer for their recruit. Therefore, it is need to verify the authenticity of an academic transcript. In this paper we proposed an approach to verify academic transcript documents based on check digit method. Simulation results show that the proposed approach is more reliable than the original check digit method including Universal Product Code (UPC) [6,7], International Standard Book Number (ISBN-11) [8], and Luhn [9] when tested on different error cases. This paper is organized as follows: Section 1 covered the introduction of this paper. Section 2 covered the previous research and techniques used in related works. Section 3 is about the methodology used in this research. In Section 4, this paper discuss on the simulation results and performance analysis of the proposed method. The conclusion of the research is presented in Section 5.

II. RELATED WORKS

Albrecht [10] outlined that the increase in document forgery by fraudulent individuals has brought about a dire need to develop methods that protects the integrity of documents, both hardcopy and softcopy documents. To provide security to the information contained in a document, mechanisms and tools for prevention and detection need to be used. This section covers methods and techniques to perform integrity control. The methods include digital signatures, watermarking and check digit that used to detect document forgeries.

A. Watermarking

Watermarking is a mechanism of embedding an imperceptible signal or signature, which is visible or invisible to protect the integrity of intellectual property rights of digital media such as text, audio, video and images [11]. Watermarking algorithm is made up an embedding algorithm; embeds the signature on the document and an extraction algorithm; decodes its content and detects the signature. Bamatraf [12] describes watermarking as the process of hiding or embedding an imperceptible signal or signature into a given document and argues that the watermark embedded into a document needs to be robust enough to withstand alterations caused by fraudulent individuals or malevolent attacks. This method has been used to protect the authenticity especially text [13], images [14] and document [15].

W. Diffie and M.E. Hellman introduced digital signature in the mid-seventies [16]. Digital signature is an asymmetric cryptographic mechanism used for demonstrating the authenticity of a document. In other word, digital signature is a mathematical technique used to validate the authenticity and integrity of a message, software or digital document. Digital Signature is used to sign intellectual properties such as information transactions and also software distribution purposes. A unique code that acts as a signature can be attached to message by sender. The encryption of message with the sender's private key and the hash of the message are taking to form the signature. The recipient unpacks received message, and use the same hashing function to the signed and encrypted message then compared to the decrypted signature.

B. Digital Signature

Digital Signature can be classified into two processes. First process is signing and encryption with four step; generates a unique code using hashing, encrypted using private key of the sender, packed together message signature and the public key and encrypted packed unit using receipt's public key to form signed and encrypted message. Second process is decryption and verification with five steps; decrypt received using the receiver's private key to form a packed unit of message, unpacked into plain text message, apply same hashing function that used by sender, decrypted by using the received public key of the sender and message digest computed from the plain message received by the recipient. Signed and encrypted data or message can only be decrypted using the correct Private Key of the recipient thus ensuring the privacy.

C. Common Errors

According to the empirical investigations by Gallian [21], when receiving input from humans, there are common errors occurrences of sequence of digits, the most common transmission errors are the following

Table 1
Common errors occurrence

No.	Error Type	Form
1.	Single error	a → b
2.	Adjacent transposition error	ab → ba
3.	Twin error	aa → bb
4.	Jump transposition	abc → cba
5.	Jump twin error	aca → bcb
6.	Phonetic errors	60 → 16 ("sixty" to "sixteen")

Single digit error occurs when only a single digit in the whole string of numbers or alphabets is typed wrongly. Single transposition error occurs when the n^{th} and the $(n+1)^{th}$ digits are replaced by the $(n+1)^{th}$ and the *nth* respectively. Twin error occurs when two consecutive identical digits are incorrectly typed as two other identical digits. Jump transposition error occurs when the n^{th} and the $(n+2)^{th}$ digits are replaced by the $(n+2)^{th}$ and the n^{th} respectively. Jump twin error occurs when the n^{th} and $(n+2)^{th}$ digits are identical and are mistyped as two other identical digits or alphabets. Phonetic Error occurs when there are two have similar pronunciations in some languages. An example of such errors is pronunciation number 16 (sixteen) and 60 (sixty) in English.

D. Error Detection

Error detection involves adding a check digit to the digit or alphanumeric and verification. Check digits generated and usually placed behind the digit or alphanumeric. To confirm 'digit or alphanumeric' error-free, check digit verification performed. This section will discuss some of the check digit generation method such as Luhn 10 method, UPC Modulus 10 method and ISBN Modulus 11 method.

E. Check Digit

According to [17], check digit is a numeric value that has been calculated mathematically and added to data to ensure that original data has not been altered or an incorrect but valid value submitted. The first notable check digit algorithm was developed by an IBM scientist Hans Peter Luhn and it is specified in ISO/IEC 7812-1 [18]. Check digit algorithm has been widely used in various applications such as to verify the bank International Bank Account Number (IBAN) numbers and International Standard Book Number (ISBN) numbers for books [12].

There are number of methods based on check digit including UPC modulus 10 [7], ISBN modulus 11 [8] and Luhn 10 [9]. The following subsection will present the existing methods in check digit.

a. UPC Modulus 10

UPC Modulus 10 is used to protect the integrity Universal Product Code (UPC) that contains thirteen digits. The first twelve digits represent the product code and last digit is check digit. The first digit identifies a broad category of product type, the next five digits identify the manufacturer, the next five the product, and the last is the check digit.

Definition 1:
The check digit α_{13} for the UPC number $\alpha_1, \alpha_2 ... \alpha_{12}$ is chosen to satisfy the condition, where Weight $W_1, W_2, W_3 .... W_{12}$ is denoted by weighing factor (3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1).

$$X = ((\alpha_1 * W_1) + (\alpha_2 * W_2) + ... + (\alpha_{12} * W_{12})) \quad (1)$$

$$X = \sum_{i=1}^{n=12} (\alpha i * Wi) \quad (2)$$

$$\text{UPC Check Digit: } 10 - (X \% 10) \quad (3)$$

b. ISBN Modulus 11

Every published book will have a unique identification number based on International Standard Book Number system (ISBN). It was introduced by Gordon Foster and WH Smith in 1966 [8]. ISBN later published by the International Standard Organization (ISO) for standardization in 1970 and being known as the ISO 2108.

The check digit α_{10} for the ISBN number $\alpha_1, \alpha_2 ... \alpha_9$ is chosen to satisfy the condition, where Weight $W_1, W_2, W_3 .... W_9$ is denoted by weighing factor (10,9,8,7,6,5,4,3,2).

$$X = ((\alpha_1 * W_1) + (\alpha_2 * W_2) + ... + (\alpha_{12} * W_{12})) \quad (3)$$

$$X = \sum_{i=1}^{n=9} (\alpha i * Wi) \quad (4)$$

$$\text{ISBN Check Digit: } 11 - (X \% 11) \quad (5)$$

c. Luhn 10 Method

This method is also known as modulo 10 or mod 10 algorithm. It is used to validate identification number such as credit card numbers and IMEI number. It is created by IBM scientist Peter Luhn [9].

Definition 2
The check digit α7 for the IBM number is; α1, α2... α7, where α8 is the last digit.

F. Literature Summary

All method detects 100% for single errors. ISBN Modulus 11 and UPC method detects 100% error for single transposition but Luhn 10 method detects 0%. Siddharth [22] mentions that only ISBN Modulus 11 can detect jump transposition with 100% rate compared 0% for Luhn 10 and UPC method. UPC method is best error detect for phonetic error with 100% success rate proportionate ISBN Modulus 11 with 90% and Luhn 10 with 90%. ISBN Modulus 11 and Luhn 10 success rate for jump twin error is 100% better than UPC with 88.9% success rate. Comparison for total accuracy, highest success rate is ISBN Modulus 11 with 91.21% followed by UPC with 79.67% and Luhn 10 with 63.15%. Table 2 below gives a summary of the success rates of three modern conventional check digit methods.

Table 2
Check digit methods success rates based on [21] and [22]

Errors	ISBN Modulus 11	Luhn 10	UPC
Single error	100%	100%	100%
Single transposition	100%	0%	100%
Jump transposition	100%	0%	0%
Twin error	88.89%	88.89%	88.89%
Phonetic error	90%	90%	100%
Jump twin	100%	100%	88.9%
Total accuracy	91.21%	63.15%	79.67%

The performance of each method on different errors is summarized in Table 2. Based on this result it can summarize that the Luhn 10 and UPC unable to detect jump transposition error. The Luhn 10 also does not detect single error. The ISBN Modulus 11 method does detect all error with lowest score of 88.89% for twin error. Luhn 10 is the least effective and ISBN Modulus 11 method is the most effective for detecting errors.

In the case of academic transcript fraudulent, check digit method is the most suitable method to be adapted due to its modest requirements and simplicity to be implemented using web services. In this research, we combine methods in check digit and test it performance under various errors cases.

III. METHODOLOGY

To perform the experiments, the attributes that commonly available in the academic transcripts are identified. The attributes are to be used to determine the security number. Here are the attributes contained that commonly used in academic transcript:

1. Name
2. Programme of study
3. Faculty
4. Student matric number
5. Session of admission
6. Year of graduation
7. Identity card number
8. Credits transferred
9. Final CPA
10. Semester
11. Session
12. CPA every semester
13. GPA every semester
14. Course codes
15. Course names
16. Credit hours
17. Grade point
18. Grades
19. Semester credits
20. Total accumulated credits

There are 20 attributes contained in one's academic transcript. However not all of these attributes are used in creating security number.

A. Secure Number Generation

This section will discuss the process of generating secure number. Attributes that are required in the generation number is secure Semester, Session, Student Matric Number and GPA. Figure 3.1 below gives an illustration of process security number generation.

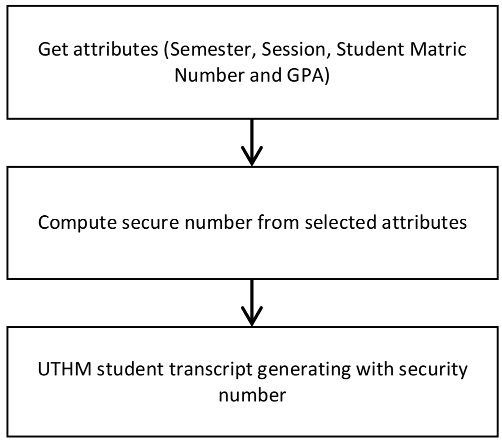

Figure 1: Generating UTHM Student transcripts with security number

Figure 1 describe how the process to generate the academic transcript's security number. Four attributes from UTHM student transcript required for secure number generation. The attribute is Semester, Session, Student ID and GPA. After getting these attributes, algorithm is applied to compute security number. This secure number will be print together into the student academic transcripts. The algorithm is the combination of check digit methods which is UPC Modulus 10 and ISBN Modulus 11. The process is shown in Algorithm 1.

Algorithm 1: Dual Digit Check Digit

INPUT: Semester, Session, StudentID, GPA
BEGIN
1. Generate first check digit
2. Generate second check digit
3. Print first check digit and second check digit
4. Generate student transcript
5. ENDFOR

END DFH

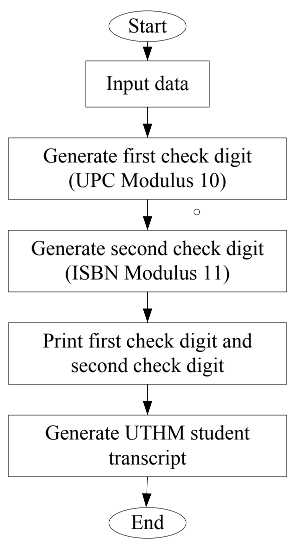

Figure 2: Dual digit check digit flowchart

IV. Performance Analysis

This section discussed simulation result of existing and proposed check digit method to generating secure number. Universal Product Code (UPC-10), Luhn-10 and International Book Standard Number (ISBN-10) are chosen as prominent check digit. A proposed method will be compared with these three methods of their efficiency to detect an error.

A. Simulation Result for Single Error

Single error occurs when only a single digit in the Dual Check Digit whole string of numbers is typed wrongly. Table 2 shows comparisons of the efficiency of the proposed against three other prominent check digit methods in error detecting. UPC modulus 10, ISBN modulus 11, Luhn 10 and proposed method detect 100% of single digit error. This shows that all method is suitable for generating a UTHM student transcript secure number.

Table 2
Simulation result for single error

Method	Method Efficiency
UPC Modulus 10	100%
ISBN Modulus 11	100%
Luhn 10	100%
Dual Check Digit	100%

B. Simulation Result for Single Transportation Error

A single transposition error is simple error of data entry. Transposition errors occur when switching the position of two adjacent digits (transposed). In other word this type of error occurs when the ***nth*** and the ***(n + 1)th*** digits are replaced by the ***(n + 1)th*** and the ***nth*** respectively. This error discussed twice with difference digit position. Table 3 shows the transpose of first and second digit. Table 4 shows the transpose of second and third digit.

Table 3
Test result for single transposition error 1

Method	Method Efficiency
UPC Modulus 10	89.91%
ISBN Modulus 11	100%
Luhn 10	9.83%
Propose Method	100%

Table 3 above shows that UPC modulus 10 detect 89.91% of transposition error, ISBN modulus 11 and Dual Check Digit detect 100% of transposition error and Luhn 10 detect only 9.83% of transposition error.

Table 4
Test result for single transposition error 2

Method	Method Efficiency
UPC Modulus 10	89.90%
ISBN Modulus 11	100%
Luhn 10	10.19%
Dual Check Digit	100%

Table 4 above shows that UPC modulus 10 detect 89.90% of transposition error, ISBN modulus 11 and Dual Check Digit detect 100% of transposition error and Luhn 10 detect only 10.19% of transposition error.

C. Simulation Result isbnfor Jump Transportation Error

A jump transposition error is a switching of two non-adjacent digits, for example "123" instead of "321". Number "1" is switching with number "2". In other word by Hemangi Rane (2015), jump transposition error occur when the ith and $(i+2)^{th}$ digit are replaced by $(i+2)^{th}$ and i^{th} respectively .

Table 5 shows the simulation result for jump transposition error. UPC modulus 10 detects 0% jump transposition errors. ISBN modulus 11 detects 100% errors while Luhn 10 detects 0% errors. The proposed method detects 100% of jump transposition errors.

Table 5
Test result for jump transposition error

Method	Method Efficiency
UPC Modulus 10	0%
ISBN Modulus 11	100%
Luhn 10	0%
Dual Check Digit	100%

D. Simulation Result for Twin Error

A twin error occurs when two identical side-by-side digits change to a difference pair of identical digits. For example, if we have “44” in the number, somebody might somehow change it to “22”. Table 6 shows the simulation results of proposed method against three (3) other prominent check digit methods in error detecting. UPC modulus 10, Luhn 10 and Dual Check Digit, each detects 100% of twin errors. ISBN modulus 11 detects 92.17% of twin errors respectively.

Table 6
Twin error test result

Method	Method Efficiency
UPC Modulus 10	100%
ISBN Modulus 11	92.17%
Luhn 10	100%
Dual Check Digit	100%

E. Simulation Result for Jump Twin Error

Table 7 shows the simulation results of proposed method. UPC modulus 10 method detects 96.79% jump twin errors. The ISBN modulus 11 detects 100% jump twin errors. A Luhn 10 method detects 96.79% jump twin error, and Dual Check Digit detects 100% jump twin errors.

Table 7
Test result for Jump twin error

Method	Method Efficiency
UPC Modulus 10	96.79%
ISBN Modulus 11	100%
Luhn 10	96.79%
Dual Check Digit	100%

F. Simulation Result for Jump Twin Error

Table 8 shows the simulation results of proposed method against thre other prominent check digit methods in error detecting UPC modulus 10, Luhn 10 and Dual Check Digit, each detects 100% of twin errors. ISBN modulus 11 detects 98% of twin errors respectively.

Table 8
Test result for Phonetic error

Method	Method Efficiency
UPC Modulus 10	100%
ISBN Modulus 11	98%
Luhn 10	100%
Dual Check Digit	100%

G. *Analysis Result*

Dual Check Digit has compared against prominent method in several errors. The errors involved are single error, single transposition error, jump transposition error, twin error, jump twin error and phonetic error. Table 9 shows the analysis results of proposed method against three (3) other prominent check digit methods in error detecting.

Table 9
Average test results

Method	Dual Check Digit (%)	ISBN Modulus 11 (%)	UPC Mod 10 (%)	Luhn 10 (%)
Single Error	100	100	100	100
Single Transposition Error 1	100	100	89.91	9.83
Single Transposition Error 2	100	100	89.9	10.19
Jump Transposition Error	100	100	0	0
Twin Error	100	92.17	10	100
Jump Twin Error	100	100	96.79	96.79
Phonetic Error	100	98	100	100
Average Method Efficiency	100	98.6	82.37	59.54

According to the table 5.9 above, our proposed method has an average total accuracy of 100%. The current UPC method have an average total 95.36%, ISBN and Luhn have an average total accuracy of 86.75% and 56.75% respectively. Thus, based on the test results, we can argue that our proposed method is relatively more accurate in error detection compared to the other three methods.

V. CONCLUSIONS

This paper has proposed method based on analyzed various methods that can be used to implement verification on transcripts via web services. Three methods including Universal Product Code (UPC-10), International Standard Book Number (ISBN-10), and Luhn Mod 10 are compared. From the experiments, it shows that the combination from two of these methods which is UPC-10 and ISBN-11 produced a very variable method to validate the authenticity of the academic transcript. It also can be applied to other document such as the degree certificate by using suitable parameter.

ACKNOWLEDGEMENT

The authors would like to thank for the support given to this research by Ministry of Education (MOE) and Universiti Tun Hussein Onn Malaysia (UTHM), under ERGS grant Vot: E054.

REFERENCES

[1] Kassim C. K. H. C. K., Nasir N. E. M., and Ahmad S., 2015. Academic Dishonesty of Accounting Students at Higher Learning Institutions, *Mediterr. J. Soc. Sci*. 6 (4):702.

[2] Du Toit P. H., 2012. Using action research as process for sustaining knowledge production: A case study of a higher education qualification for academics. *South Afr. J. High*. Educ. 26(6):1216–1233.

[3] Rathbun G. andTurner N., 2012. Authenticity in academic development: the myth of neutrality. *Int. J. Acad. Dev*. 17(3):231–242.

[4] GARWE E. C., 2015. Qualification, Award and Recognition Fraud in Higher Education in Zimbabwe. *J. Stud. Educ.* 5(2)119–135.

[5] Singhal A. and Pavithr R. S., 2015. Degree Certificate Authentication using QR Code and Smartphone, *Int. J. Comput. Appl.* 120:(16).

[6] Grover A. and Berghel H., 2011. A survey of RFID deployment and security issues. *J. Inf. Process. Syst.* 7(4):561–580.

[7] Pindar Z. A., Jamel S., and Disina A. H., 2015. An Enhanced UPC Mod 10 Check Digit Method. *Int. J. Knowl. Based Comput. Syst.* 2 :(1).

[8] Kamaku W., Mwathi C., and Kivunge B., 2012. Limitations in the Convectional ISBN-10 Code. Am. *Int. J. Contemp. Res.* 2(2):553–569.

[9] Luhn, H. P., 1960.Computer For Verifying Numbers. US Patent 2950048.

[10] Albrecht C., Albrecht C., and Tzafrir S., 2011. How to protect and minimize consumer risk to identity theft. J. Financ. Crime. 18(4):405–414.

[11] Singh P. and Chadha R. S., 2013. A survey of digital watermarking techniques, applications and attacks, I*nt. J. Eng. Innov. Technol. IJEIT*. 2 (9):165–175.

[12] Bamatraf R. I. and Salleh M. N. M., 2011. A new Digital Watermarking using LSB and Inverse bit. *J. Comput*. 3

[13] Kuang Q. and, Xu X. 2011. A New Zero-Watermarking Scheme Based on Features Extraction for Authentication of Text. J. Converg. *Inf. Technol.* 6(11):155–165.

[14] Kester Q., Nana L., Pascu A. C., Gire S., Eghan J. M., and Quaynor N. N., 2014. A Hybrid Cryptographic and Digital Watermarking Technique for Securing Digital Images based on a Generated Symmetric Key. *Int. J. Comput. Appl.* 94(19).

[15] Fernando M.-R. and Manuel F.-M. J., 2012. SWAD: Secure Watermarking for Authentication of Scanned Documents. J. Commun. *Comput.* 9:1226–1230.

[16] Diffie W. and Hellman M. E., 1976. New directions in cryptography. I*nf. Theory IEEE Trans.* On. 22 (6):644–654.

[17] Ludvigsson J. F., Otterblad-Olausson P., Pettersson B. U., and Ekbom A. 2009. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. Eur. *J. Epidemiol.* 24 (11):659–667.

[18] Luhn H. P. 2008. Key word-in-context index for technical literature (kwic index). *Am. Doc.* 11(4):288–295.

[19] Simkin M. G. Five Data Validation Cases. *J. Inf. Syst. Educ.* 19(3):271.

[20] Hussein K. W., Sani N. F. M., Mahmod R., and Abdullah M. T., 2013. Enhance Luhn Algorithm for Validation of Credit Cards Numbers. Int. J. Comput. Sci. Mob. Comput. 2 (7):262–272.

[21] Gallian J. A., 1996. Error detection methods. ACM Comput. Surv. CSUR. 28(3):504–517.

[22] Siddharth A., 2012. Error detection in numeric codes. Reson.-Heidelb. 17(7):653.