International Journal of Cyber-Security and Digital Forensics (IJCSDF) 7(1): 1-9
The Society of Digital Information and Wireless Communications (SDIWC), 2018 ISSN: 2305-0012

A Mathematical Model for Resolving Minimum Password Length Controversy

[1]S. Agholor, [2]A. S. Sodiya and [3]D. O. Aborisade

[1]Department of Computer Science,
Federal College of Education, Abeokuta, Nigeria

[2,3]Department of Computer Science,
Federal University of Agriculture, Abeokuta, Nigeria

[1]sagholor@fce-abeokuta.edu.ng

[2]sinaronke@yahoo.co.uk, [3]daaborisade@funaab.edu.ng

ABSTRACT

Information Security has become one of the most pressing issues facing businesses in today's competitive e-commerce that is driven by online transactions. User authentication serves as the first line of defence against security breaches, which predominantly uses passwords. There have been growing controversies as per the minimum length of a password required to make the password withstand guessing and hacking attacks. For example, a password can receive a rating as "strong" with only six characters on Facebook but not on Gmail where it must have at least eight characters. There is, therefore, the urgent need to address these minimum password length controversies in view of its negative consequences on the security of the end-users' web accounts. In this paper, a mathematical model for determining minimum password length was developed. A combination of entropy formula and the bit strength threshold were used in developing the mathematical model, which was implemented using PHP. This was tested and a table of minimum password length needed for different character sets was generated. It is hoped that software developers as well as web account owners will find the table useful.

KEYWORDS

Bit Strength Threshold, Brute-Force Attacks, Entropy Model, Minimum Password Length, Randomness

1 INTRODUCTION

A password is a character or sequence of characters used to determine that a device user requesting access to a system is really that particular user [1]. The main authentication mechanism employed in millions of computer installations and websites is passwords [2]. According to [3] and [4], password will continue to maintain its predominance as a form of authentication even in the face of new developments in authentication systems such as biometric. This seems to remain unchanged in the foreseeable future [3, 5, 6]. This is so because of its easiness, cost effectiveness, simplicity and familiarity to all users [5, 7]. In addition, it provides adequate security to the end-users' online accounts, although there are some concerns about weak passwords which can lead to weak security [7]. However, findings from separate studies conducted by [8] and [9] showed that the weakness is not within the password authentication itself, but the choice of the passwords by the end-users.

Since passwords are the first line of defence in web accounts, there is the need to educate end-users on how to create good passwords. Towards this end, many systems administrators as well as software developers provide advice to their users on how to construct a good password. This advice comes in form of password policies, feedback mechanism, among others. One of these password policies is the minimum password length. However, there have been growing controversies on the minimum password length that can give the needed protection of end-users' online accounts. These conflicting minimum password lengths bring confusion to the end-users with far reaching negative consequences on the security of the end-users' web accounts. This calls for urgent scientific solution. To the best of our knowledge, this is the first scientific attempt at resolving the minimum password length controversy.

The remainder of this paper is organized into five sections namely: brief history of password development, related work, methodology, implementation and results. Finally, recommendations and conclusion were discussed.

2 BRIEF HISTORY OF PASSWORD DEVELOPMENT

Password, which was formerly called watchword in the olden days, has been used since ancient times, especially in the Roman military. In those days, a sentry, that is, a soldier stationed at a place, especially at a gate to prevent the passage of unauthorized persons, would challenge those wishing to enter an area or approaching it to supply a watchword and would only allow such a person or group of persons to pass if they knew the watchword [10]. In modern times, usernames and passwords whose implementation were pioneered by Unix system are commonly used by people during a login process that controls access to protected devices. It was initially implemented on a simple text password system whereby account passwords were stored verbatim in a file [11]. Even though the file was protected from casual reading and writing, a privileged or skillful user could gain access to the file, thereby compromising all end-users' passwords. As a result of this drawback, [12] proposed securing each password with encryption before storing them during account creation. During login at a later time, the system would encrypt the entered password, compare the result to the stored encrypted password for that user, and grant access if they matched. Otherwise, access is denied. The encryption, in this regards added a layer of security to the stored passwords.

However, if the encryption key is compromised or broken, all passwords would also be compromised. To remedy this vulnerability, [13] and [14] proposed passing the entered password through a one-way hashing function, instead of using encryption. Thus, an attacker who compromises a file of hashed passwords will not be able to obtain the plaintext passwords without performing a password guessing attack, whereby the attacker chooses and hashes a candidate password, and compares the result to each of the hashes contained in the password file. Any hash that matches the hash in the password file is the actual password for the corresponding account.

From its introduction by Unix system, the study of [5] considered the four decades of research work on passwords and then classified passwords into two generations, first and second generation. The first generation considered mainly two attributes, that is, security and usability, while the second generation considered three attributes, that is, security, usability and deployability of the password the trade-off between them.

3 RELATED WORK

End-users of today's computer systems manage a large number of online accounts that require passwords. Each authentication system has different rules or policies for which passwords are acceptable and which passwords are not. Some passwords must be eight characters; some must be over eight characters; some must contain multiple classes of characters; some cannot accept certain characters [15].

An analysis of minimum and maximum password lengths of twenty-three selected websites that have very high global ranking in terms of popularity was carried out by [16]. The result of the work, which is presented in table 1, showed that there is no uniformity in the web community about the length of passwords.

Table 1. Minimum and Maximum Password Length of some selected Websites

SN	Website	Minimum Characters	Maximum Characters
1	Adobe	6	100
2	Amazon	6	128
3	Apple	8	32
4	Ebay	6	64
5	Evernote	6	64
6	Facebook	6	2001+
7	Flipkart	4	2001+
8	Google	8	100
9	Hotmail	8	16
10	IEEE	8	64
11	Linkedin	6	16
12	Lycos	6	20
13	Oracle	8	80
14	Outlook	8	16
15	PayPal	8	20
16	Pinterest	6	2001+

17	Rediff	6	12
18	Springer	6	60
19	Stackoverflow	8	2001+
20	Twitter	6	2001+
21	Wikipedia	1	2001+
22	Wordpress	6	50
23	Yahoo	8	32

Source: Culled from [16]

From table 1, the minimum password length ranges from one to eight. This clearly showed that there is no consistency and common agreement on the minimum and maximum password length required for creating a password that can give basic protection. Further findings by [16] revealed that all the websites decide their minimum and maximum characters to be allowed in the password field according to their own wish. In other words, there is no scientific approach for determining minimum password length among the websites studied.

In a related work, [2] conducted a study of password policies of five different websites. The result of this study which is presented in table 2, showed a minimum password length of six to eight characters.

Table 2. Minimum Password Length of some Websites

Name	Minimum Password Length
Gmail	8 characters minimum length
You Tube	8 characters minimum length
Facebook	6 characters minimum length
MS Live	8 characters minimum length
Yahoo	6 characters minimum length

Source: From [2]

Again, results from table 2 showed non-uniformity in the minimum password length. Furthermore, there is no scientific approach used in arriving at this minimum password length.

A study of password policies of seventy-four different websites which included top, high and medium traffic websites, universities, banks, brokerages and government websites was conducted by [17]. The result of the study is presented in table 3.

Table 3. Minimum Password Length across 74 Websites

SN	Website	Minimum Password Length
1	Google	8
2	Facebook	6
3	Yahoo!	6
4	Youtube	6
5	AOL	8
6	Live	6
7	Wikipedia	1
8	eBay	6
9	Amazon	6
10	Ask	6
11	Weather	6
12	Answers	1
13	Myspace	6
14	Craigslist	6
15	Adobe	6
16	Capitalone.com	8
17	rockyou.com	8
18	typepad.com	6
19	overstock.com	5
20	latimes.com	6
21	intuit.com	6
22	cbssports.com	4
23	wowwiki.com	1
24	virginia.edu	6
25	pgatour.com	1
26	hollywood.com	1
27	occupied.com	4
28	istockphoto.com	5
29	highschoolsports.net	1
30	Fidelity	6
31	Vanguard	8
32	Schwab	6
33	WellsFargo	6
34	BoA	8
35	JP Morgan Chase	7
36	Citibank	6
37	PayPal	8
38	US Bank	8
39	Ohio State U	8
40	Arizona State U	8
41	U. of Florida	8
42	U. of Minn.	6
43	U. of Texas	8
44	U. Central Florida	8
45	Michigan State	8
46	Texas A & M	6
47	U. South Florida	6
48	Penn. State U	8
49	MIT	6
50	Stanford	8
51	UC Berkeley	8
52	CMU	8
53	UTUC	8
54	Cornell	7
55	Princeton	8
56	U. of Washington	8
57	Georgia Tech.	8
58	irs.gov	8

59	usps.com	8
60	nih.gov	8
61	ca.gov	8
62	ed.gov	8
63	noaa.gov	12
64	weather.gov	12
65	census.gov	8
66	ssa.gov	7
67	nasa.gov	12
68	U. of Phoenix	7
69	Columbia	6
70	Northwestern	6
71	VA	8
72	USAJobs	8
73	TreasuryDirect	8
74	Twitter	6

Source: Culled from [17]

From table 3, the result of the study showed a minimum password length that varies from one character to twelve characters. As in the case of other studies highlighted above, there is no scientific method of arriving at this minimum password length.

Closely related to the work of [17] but with a wider coverage was the work of [18] that performed a study of how passwords are handled in one hundred and fifty different websites. This study was a large collection of password policies and practices of one hundred and fifty websites. The results of this study showed that 123 websites representing 82% imposed a minimum password length, while 27 websites representing 18% do not restrict end-users to minimum password length. By inference, it follows that 18% allow minimum password length of 1 character. Again, among the websites with minimum password length the result showed no uniformity as 78 websites (52%) accept minimum password length of 6 characters, followed by 21 and 15 websites (14% and 10%) with a minimum password length of 4 and 5 characters respectively. This result is consistent with the finding of [19], though with a smaller sample. The findings from the two studies showed that there is no uniformity in minimum password length.

In a similar work, [20] presented a detailed structural analysis of username segment in e-mail addresses of MBA students in academic institutions offering MBA courses in Gujarat State of India. Their findings showed that the institution tend to design the username segment of their e-mail addresses by choosing words or combination of words from specific categories. The paper also highlights the use of special characters, digits and random words in designing the usernames. Other studies similar to this work are the works of [21], [22] and [23]. These works focused on textual analysis of digits used for designing Yahoo-Group identifiers, classification of character usage in unique addresses employed for accessing Yahoo-Group service and analysis of usage of four-digit year number in designing Yahoo-Group identifiers respectively. The results of these studies showed an interesting user behavior in the use of digits, character and four-digit year in designing Yahoo-Group identifiers respectively. There is no uniformity.

From the findings of the related work, it is clear that there is no consensus on the minimum password length. This has far reaching negative implication on the security of the end-users' web accounts, hence the need for a scientific method of determining minimum password length that will be acceptable to all concern and withstand online and offline attacks. This is the goal of this paper.

4 METHODOLOGY

4.1 The Model Development Process

As discussed in the related work section, there have been controversies on the minimum password length requirement. To resolve these controversies, we adopted a combination of entropy formula and the bit strength threshold in developing the model.

4.1.1 The Entropy Formula

The entropy of a given password according to [24], [15] and [17] is given by:

$$H = Len * (Log(C))/(Log(2)) \quad (1)$$

where:
H is the computed entropy of the password;
Len is the length of the password; and
C is the character set that make up the password.

Using equation (1), the entropy per character for various character pools is given in table 4.

Table 4. Entropy per Character for various Character pools

SN	Character Set	Entropy
1	Numeric	3.222
2	Lowercase	4.701
3	Uppercase	4.701
4	Symbols (Special Characters)	5.045
5	Numeric+Lowercase	5.170
6	Numeric+Uppercase	5.170
7	Numeric+Symbols	5.427
8	Lowercase+Uppercase	5.701
9	Numeric+Lowercase+Uppercase	5.955
10	Numeric+Lowercase+Symbols	6.109
11	Numeric+Lowercase+Uppercase+ Symbols	6.570

From table 4, it showed that a combination of numeric, lowercase, uppercase and symbols to create a password gives the highest entropy, hence the highest security. This entropy is the ideal situation where the end-user makes the password very random. This ideal situation is only obtainable from machine-generated password, hence we shall develop the entropy formula for both machine-generated passwords and human-generated passwords.

4.1.1.1 The Entropy Formula of a Machine-Generated Password

In this case, equation (1) holds. Thus, if we take Len1 for the minimum password length for the machine-generated password, then equation (1) becomes:

$$H = Len1*(Log (C))/(1.1*Log(2)) \quad (2) \text{ and}$$

$$Len1 = (H*Log(2))/(Log(C)) \quad (3)$$

where:
Len1 is the minimum password length for the machine-generated password;
H is the minimum entropy needed to withstand brute-force/guessing attacks.
C is the character set.

4.1.1.2 The Entropy Formula of a Human-Generated Password

Since human-generated password is just slightly random, equations (2) and (3) can not apply. According to [25], when an end-user generates a password using either lowercase characters only or uppercase characters only, the randomness of the password is estimated to be between 2 and 3 bits per character as against the machine-generated 4.7 bits. This showed that when an end-user chooses a password, about 36% (1.7 bits) to 57% (2.7 bits) of its randomness is lost when compared to machine generated passwords.

From the foregoing and using the lower bound which has more frequency (67%), we modify equation (1) by introducing a penalty of 1.1 bits which is 67% of 1.7 bits. The penalty introduced is to accommodate the randomness lost in a human chosen password. Thus, if we take Len2 as the length of the human-generated password, then equation (1) becomes:

$$H = Len2*(Log (C))/(1.1*Log(2)) \quad (4) \text{ and}$$

$$Len2 = (H*1.1*Log(2))/(Log(C)) \quad (5)$$

where:
Len2 is the minimum password length of the human-generated password;
H is the minimum entropy needed to withstand brute-force/guessing attacks.
C is the character set.
1.1 is the penalty introduced to accommodate the randomness lost.

4.1.2 The Bit Strength Threshold

The bit strength threshold is used to determine the minimum entropy required to withstand brute-force/guessing attacks. It should be noted that some basic benchmarks have been established for brute-force searches in the context of attempting to find keys used in encryption. The problem is not the same since these approaches involve astronomical numbers of trials, but the results are suggestive for password choice.

The following historical background will lead us in finding answer to the minimum entropy required to withstand brute-force attacks. The 56-bit DES encryption was broken by [26] in less than 3 days using specially designed hardware. In a related development, the 64-bit key was cracked by [27] in 1757 days or 4 years,

9 months and 23 days. Furthermore, [28] started work on a 72-bit key in 2002. The progress made so far is shown in Figure 1.

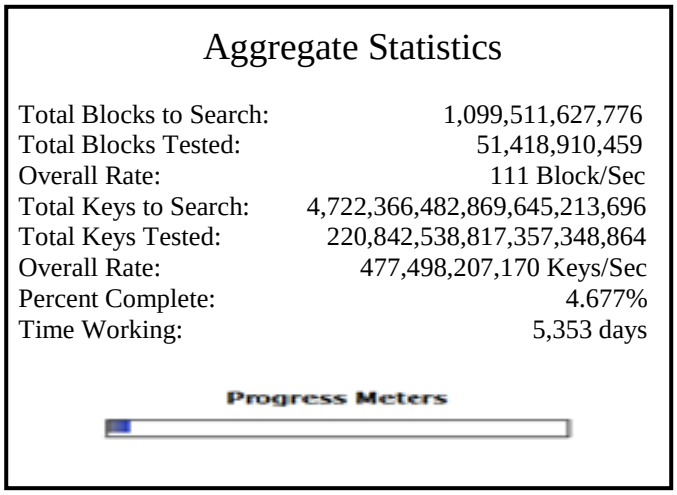

Figure 1. RC5-72/Overall Project Stats as at 29/07/2017

Source: www.stats.distributed.net/project.php?project_id=8

From Figure 1, a total of 5,353 days or 14.7 years has been used with 4.677% completion. Based on the above result, [28] estimated that cracking a 72-bit key using current hardware will take about 45, 579 days or 124.8 years.

As a result of the above examples, there can be no exact answer to the somewhat different problems of the password strength required to resist brute-force attacks in practice. However, [29] recommended 80-bit for the most secured passwords. In this paper, we adopted the recommendation of [29] as the bit strength threshold.

4.2 The Proposed Model for Determining the Minimum Password Length

The goal of this work is to develop a model that will take care of both machine-generated passwords and human-generated passwords. To achieve this, we made use of the recommended bit strength threshold of 80-bit, and a combination of equations (3) and (5). Thus, the model for computing the minimum password length is given by equation (6):

$$Len=1/2((80*Log(2))/(Log(C))+80*1.1*Log(2))/(Log(C))) \quad (6)$$ and

$$Len=1/2\{(Log(2)/Log(C))*(168)\} \quad (7)$$

where:
Len is the minimum password length;
80 is the bit strength threshold, that is, the minimum entropy needed to withstand brute-force/guessing attacks according to [29].
C is the character set.
1.1 is the penalty introduced to accommodate the randomness lost if it is a human-generated password.

Equation (6) is the developed Model.

4.3 Algorithm of the Model

The Model algorithm is presented below.
Let C be each character set
Let C1..C11 be numeric, lowercase.. num+lowcase+uppcase+special characters
Select Choice (C1…C11)
If (choice = C1) then
C=C1
CALL MODEL
endif
If (choice = C2) then
C=C2
CALL MODEL
endif
If (choice = C3) then
C=C3
CALL MODEL
endif
If (choice = C4) then
C=C4
CALL MODEL
endif
If (choice = C5) then
C=C5
CALL MODEL
endif
If (choice = C6) then
C=C6
CALL MODEL
endif
If (choice = C7) then
C=C7
CALL MODEL
endif
If (choice = C8) then
C=C8
CALL MODEL
endif
If (choice = C9) then

```
C=C9
CALL MODEL
endif
If (choice = C10) then
C=C10
CALL MODEL
endif
If (choice = C11) then
C=C11
CALL MODEL
endif
END
PROCEDURE MODEL
Len=1/2((Log(2)/Log(C))*(168))
OUTPUT Len
ENDPROC
```

5 IMPLEMENTATION AND RESULTS

The model was implemented using PHP. The result obtained when different character set was supplied to the program is presented in table 5.

Table 5. Summary of Minimum Password Length for various Character Pools.

Character Set	Minimum Password Length
Numeric only	25
Lowercase only	18
Uppercase only	18
Symbols (Special Characters) only	17
Numeric+Lowercase	16
Numeric+Uppercase	16
Numeric+Symbols	15
Lowercase+Uppercase	15
Numeric+Lowercase+ Uppercase	14
Numeric+Lowercase+ Symbols	14
Numeric+Lowercase+ Uppercase+Symbols	13

From table 5, the minimum password length when all character sets are used in creating a password is 13. It is 25 if the character set is numeric only. Interestingly, a combination of numeric and lowercase to form a password has the same minimum password length as when a combination of numeric and uppercase is used to create the password, which in both cases the minimum password lengths are 16. Other minimum password lengths are 18 when lowercase only or uppercase only is used in creating a password; it is 14 when either, numeric, lowercase and uppercase or numeric, lowercase and symbols are used in forming the password. It is 15 when either numeric and symbols are used in creating the password or lowercase and uppercase are used in creating the password.

The first implication of the study is that the authentication scheme such as the ATM currently being used in Nigeria that requires only 4 digits is not healthy in driving a cashless economy of the country as this will increase guessing and other related attacks.

The second implication is that users should be educated on the need to make their passwords as random as possible and to adhere to the minimum password length in line with table 5 on all web accounts.

6 RECOMMENDATIONS AND CONCLUSION

6.1 Recommendations

(1) Existing authentication schemes that require only 4 digits as password should adjust immediately to the minimum password length outlined in table 5.

(2) Future authentication schemes developers should enforce minimum password length in line with table 5 and also enforce password randomness.

6.2 Conclusion

This work has tried to put an end to the controversies surrounding minimum password length by using a scientific method in arriving at the minimum password length of different character sets. Developers are encouraged to make reference to table 5 and/or equation (7) when developing password authentication scheme.

REFERENCES

1. Agholor, S., Sodiya, A. S., Akinwale, A. T. Adeniran, O. J., Aborisade, D. O.: A Preferential Analysis of Existing Password Managers from End-Users' View Point. In: International Journal of Cyber-Security and Digital Forensics, vol. 5, no. 4, pp. 187-196 (201).

2. Agholor, S., Sodiya, A. S.: An Assessment of Feedback Mechanism of some Selected Websites towards Improved End-Users' Password. In: Proc. of 11th International Conference of the Nigeria Computer Society, Iloko-Ijesa, pp. 44-49, (2013).

3. Andreas, S.: Influencing User Password Choice Through Peer Pressure. An Unpubblished M.Sc. thesis submitted to the Department of Electrical and Computer Engineering, The University of British Columbia, pp. 1-120, (2011).

4. Bonneau, J., Herley, C., Van Oorschot, P. C.: The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes. In: Proc. of IEEE Symposium on Security and Privacy, pp. 553-567, (2012).

5. Gayathiri, C.: Text Password Survey: Transition from First Generation to Second Generation, retrieved on 13/10/2017 from www.blogs.ubc.ca/.../ (2013).

6. Herley, C., Van Oorschot, P.: A Research Agenda Acknowledging the Persistence of Passwords. In: IEEE Security & Privacy Magazine, vol. 10, no. 1, pp. 28-36, (2012).

7. Forget, A.: A World with Many Authentication Schemes. An Unpublished Ph.D. thesis submitted to the Faculty of Graduate and Postdoctoral Affairs, School of Computer Science, Carleton University, Ottawa, Ontario, Canada, pp. 1-244, (2012).

8. Gaw, S., Felten, E. W.: Password Management Strategies for Online Accounts. In: Proc. of the 4th SOUPS, Pittsburgh, PA, USA, pp. 1-12, (2006).

9. Saheen, G. A., Jaber, A. A., Hamami, A. A.: The Effect of Weight Factors Characters on Password Selection. In: World of Computer Science and Information Journal, vol. 3, no. 6, pp. 110-113, (2013)

10. Wikipedia.: Password defined, retrieved on 28/07/2017 from www.wikipedia.org/wiki/password (2017).

11. Morris, R., Thompson, K.: Password Security: A case History. In: communications of the ACM, vol. 22, no.11, pp. 594-597, (1979).

12. Wilkes, A.: Time-Sharing Computer Systems. In: American Elsevier, pp. 1-8, (1968).

13. Evans, A., Kantrowitz, W., Weiss, E.: A User Authentication Scheme not requiring Secrecy in Computer. In: communications of the ACM, vol. 17, no. 8, pp. 1-6, (1974).

14. Lamport, L.: Password Authentication with Insecure Communication. In: Communications of the ACM, vol. 24, no. 11, pp. 22-30, (1981).

15. Kuo, C., Romsnosky, S., Cranor, L. F.: Human Selection of Mnemonic Phrase-based Passwords. In: Proc. of SOUPS, New York, Ny, USA, pp. 67-78, (2006).

16. Saini, J. R.: Analysis of minimum and Maximum Character Bounds of Password Lengths of Globally Ranked Websites. In: International Journal of Advanced Networking Applications, pp. 1-5, (2015).

17. Florencio, D., Herley, C.: Where Do Security Policies Come From? In: Proc. of SOUPS, New York, NY, USA, pp. 1-14, (2010).

18. Bonneau, J., Preibusch, S.: The Pasword Thicket: technical and market failures in human authentication on the web. In: The Ninth Workshop on the Economics of Information Security, pp. 1-49, (2010).

19. Furnell, S.: An assessment of websites practices. In: Computers and Security, vol. 26, nos. 7 & 8, pp. 445-451, (2007).

20. Saini, J. R., Desai, A. A.: Structural Analysis of Usernames Segment in e-mail Addresses of MBA Institutes of Gujarat State of India. In: International Journal of Human and Social Sciences, vol. 5, no. 6, pp. 356-360, (2010).

21. Saini, J. R., Desai, A. A.: A Textual Analysis of Digits used for Designing Yahoo-Group Identifiers. In: IUP Journal of Information Technology, vol. 6, no. 2, pp. 34-42, (2010).

22. Saini, J. R., Desai, A. A.: A Classification of Character Usage in Unique Addresses Employed for Assessing Yahoo!Group Service. In: Karpagan Journal of Computer Science, vol. 12, no. 1, pp. 233-240, (2011).

23. Saini, J. R., Desai, A. A.: An Analysis of usage of Four-digit Year Number in Designing Yahoo-Group Identifiers. In: ADIT Journal of Engineering, vol. 8, no. 1, pp. 22-27, (2011).

24. Shannon, C. E.: Prediction and Entropy of Printed English. In: Bell Systems Technical Journal, vol. 30, pp. 50-64, (1951).

25. Eastlake, D., Schiller, J., Crocker, S.: Randomness Requirements for Security, retrieved on 30/11/2012 on www.datatracker.letf.org/doc/rfc4086 (2006).

26. Electrical Front Foundation.: Cracking DES: Secrets of Encryption Research, Wiretap Politics & Chip Design. In: San Francisco, CA, 1ed, pp. 1-280, (1998).

27. Edwards, M.: 64-bit RC5 Algorithm finally Cracked, retrieved on 30/07/2017 from www.m.windowsitpro.com (2002).

28. Project RC5 retrieved on 30/07/2017 from www.distributed.net (2002).

29. NIST.: Special Publication 800-63. Electronic Authentication Guideline.: In: Technical Report, National Institute of Standards and Technology (NIST), Computer Security Division, Gaithersburg, USA, pp. 1-45, (2006).