

Arabic Domain Names: Analysis & Solution

Khaled Alghathbar^{1,2}

¹Center of Excellence in Information Assurance

²Information Systems Dept.

College of Computer and Information Sciences

King Saud University, Riyadh

Kingdom of Saudi Arabia

kalghathbar@ksu.edu.sa

Abstract: In this paper we presented technical study and security assessment that may threaten the Arabic Domains in the new proposed multilingual domains (International domain Names IDN). We analyzed the Unicode Script (in particular the Arabic Scripts) and showed how vulnerable the proposed solution will be, then we presented suggestions for improved solution to be adapted to avoid future security risk when using the Arabic Domain names.

Keywords: Arabic Domain Names ·DNS ·IDN.

1. Introduction

The communication conducted between computers over the internet is known to be dependent on the digital address of the internet protocol (IP address) that comprises 32 bits expressed in the following form (192.168.1.2). Therefore, it is inevitable for a computer to exchange data with another over the internet to define the digital address of the latter. However, the internet users are not able to use these crude digits as it is difficult to memorize them, a domain name system (DNS) is built to utilize the names assigned to the computers shifting them into their equivalent digits to mitigate this difficulty and provide various advantages to the internet and its users.

Domain Name System (DNS) is defined as a decentralized (distributed) database over the internet that includes the information of domains, computers names, and their digital addresses inside each domain. The domain name system divides the network into separate parts that are locally administered and accessed over the network. Dependencies and response time are managed by means of replication and caching.

The domain name system is integral to Internet and networks as a whole; without DNS computers and websites can never be accessed through their names. When DNS is deactivated or sabotaged by attackers, the internet and networks users suffer serious damages such as directing them into fake website where their personal information e.g. usernames, passwords, and credit card numbers can be stolen. These damages also may result in the denial of accessing websites and other services provided over the internet or networks. It even can be aggravated to paralyze the whole internet to deny accessing websites using their names.

This research paper is concerned by presenting technical study and security assessment about the security threats that may endanger Arabic domain names in the newly proposed multilingual domains (International domain Names IDN). It also analyzes the Unicode Script (particularly the Arabic Scripts) to define how vulnerable the proposed solution will be, then an improved solution is suggested to avoid future security risks when using the Arabic domain names.

2. Characteristics of Arabic Language in the Context of Domain Name

Before discussing the topic of Arabic domain names, it is worthy to mention the main characteristics of Arabic language in relation to the context of domain names. In brief, these characteristics can be summed in the following:

- 1- Arabic language comprises 28 letter
(أ ب ت ث ج ح خ د ذ ر ز س ش ص ض ط ظ ع ف ق ك ل م ن ه و ي)
Some of these letters have various shapes e.g. the letter “Yaa”, it can have these forms (ي، ي، ي)
- 2- Unlike English, Arabic is written from right to left
- 3- Depending on its position in the same word, a letter can have different shapes without being altered e.g. the letter “Noon”, it can be shaped (ن، ن، ن) in the same word yet it is still the same letter.
- 4- Arabic language is diacritical i.e. there are marks to be put over or under the letter to give a meaning to the word such (حَسَن، حُسْن). However, these marks are not letters, they are just marks to define the right way of pronouncing the words difficult to read or understand.
- 5- For compounding words, a space is used in Arabic. En Dashes (-) are rarely used for compounding.
- 6- In the Arab region, there are two groups of digits. Both are written from the left to right with the same reference yet they have different codes in the Unicode table. The two groups are below:
 - Eastern numerals (currently called Indian numerals). They are the most widely used in the Arabic countries (٠, ١, ٢, ٣, ٤, ٥, ٦, ٧, ٨, ٩)
 - Moroccan numerals (currently called Arabic numerals). They are widely used in the northwest

African countries (0, 1, 2, 3, 4, 5, 6, 7, 8, 9)

- 7- Compared to other language, the use of abstractions like “UN”, “USA” are not common in Arabic.

The abovementioned points clarify the big difference between the Arabic domain names and the domain names in other languages. Therefore, it is inevitable to ensure that any multi-language domain names will meet the requirements of these characteristics and differences.

3. Study of the Multi-Language Domain Names Proposed Solution

A study is conducted to the multi-language domain names proposed solution focusing on the Arabic domain names. It should be considered that the proposed solution depends on the Requests for Comments (RFCs) summed as follows:

RFC3490: It defines the multi-language domain names specifying the framework of how to introduce them in programs and systems. This RFC is referred to as (IDNA) [6].

RFC3491: It specifies the subcategories of (RFC3454) that define the approved characters allowed to form a domain name besides the operations required to set up a domain name and comparing it with other texts in a clearly defined manner. This RFC is referred to as (Nameprep)[7].

RFC3492: It explains the operations and the special technique used to transform a multi-language domain name into its counterpart foreign letters and vice-versa. This RFC is referred to as (Puncode) [8].

RFC3454: It specifies the technique to set up a Unicode easy to be dealt with and compared to other strings to pinpoint the similarities and differences between them in a way that make such a technique an approved one. This RFC is referred to as (Stringprep) [5].

By the study of these characteristics and the proposed technical solution, some points turned to be cleared namely that no rules or exceptions have been dedicated to deal with the characteristics of neither the Arabic language or any other language (except for English represented in its letters) despite the great differences among the characteristics of these languages. The section below demonstrates the outcomes of investigating the proposed solution.

- 1- The approved solution does not involve linguistic issues or define the letters to be used by each language to facilitate its application
- 2- The solution requires altering all the programs that provide multi-language domain names so that they can be consistent with the approved solution instead of changing the entire infrastructure of the internet domain names
- 3- According to the RFCs, the proposed solution states the programs that do not support multi-language domain names e.g. (موقع.مصر) will directly deal with the counterpart of that name e.g. (xn--4gbrim.xn--wgbh1c)

- 4- It is possible to use any group of characters even if the domain name consists characters from different languages which is difficult to be implemented in languages that contradict in direction (left to right & right to left)

- 5- The solution implements significant rules that control its application besides other rules that dedicated to foreign letters e.g. shifting upper case letters into small case

- 6- The solution specifies no stipulations or controls to meet the requirements of Arabic language i.e. the solution adopts Arabic language as it does with foreign letters; its exceptions and rules. For example, the solution does not connect the forms of number spread in the Arab region so that they refer to the same symbol contrarily to English language where upper case and lower case letter are given the same symbols as explained below:

- The symbols assigned to “ABC.com” are the same symbols assigned to “AbC.CoM” i.e. they refer to the same single domain
- Unlikely, the symbols assigned to "مؤتمر-٢٠٠٦-الإسلامي" is different from the one assigned to "مؤتمر-2006-الإسلامي"

- 7- During the study, an assumption is discovered that may limit using any Arabic numerals with the multi-language domain names with “right-to-left” writing direction. The rule explains it is failure to fulfill the transfer process when the domain name starts with a writing direction different from its ending e.g. it is possible to write the following domain names (مؤتمر-٢٠٠٦-الوطني.السعودية ، ٩١١-للحوادث.قطر) yet it is not when writing the following (مؤتمر-٢٠٠٦.الإسلامي، اتصل-٩١١.قطر)

Feeble efforts are exerted to define the characters approved for writing Arabic domain names and their architecture [1]. However, all the concerned parties – especially the Arabic Academy which is subsidiary to The Arab League and the Arabic domain names registration centers – to conduct a detailed study about these characteristics to be issued as approved characteristics of the Arabic language so that they can be implemented by the approved solution or find another method that suits their implementation and support.

4. Unicode Characters: A Detailed Study

A detailed study is conducted about the Unicode characters (on which the multi-language domain names solution depends) especially the Arabic script section which comprises the letters of both Arabic language and other language such as Persian, Urdu, Pashto, Sindhi, and Kurdish besides various numerals and symbols.

The study is conducted to identify the characters and forms similar to the Arabic letters and numerals which is probable to misguide the user when using Arabic domain names. The study is conducted on the latest version of Unicode namely “Version 4.2” instead of “Version 3.2” because several new characters and symbols are added to the former besides keeping those of the latter. For effective

results, the study here is conducted both versions at the same time.

written in different direction from left to right and other languages that are written from right to left like Hebrew does not resemble Arabic letters.

The following table shows the Arabic letters (in their various forms), numerals, and similar other letters in “Arabic Script”. It is worthy mention that the study only is limited to the Arabic script of the Unicode since other languages are

No.	Character	Code	Description	Similar Character		Description
				Character	Code	
1	ء	0621	ARABIC LETTER HAMZA	◌ْ	0654	ARABIC HAMZA ABOVE
				◌ِ	0655	ARABIC HAMZA BELOW
				◌َ	0674	ARABIC LETTER HIGH HAMZA
				۶	06F6	EXTENDED ARABIC-INDIC DIGIT SIX
				۶	06FD	ARABIC SIGN SINDHI AMPERSAND
2	آ	0622	ARABIC LETTER ALEF WITH MADDA ABOVE	أ	0623	ARABIC LETTER ALEF WITH HAMZA ABOVE
				آ	0671	ARABIC LETTER ALEF WASLA
				أ	0672	ARABIC LETTER ALEF WITH WAVY HAMZA ABOVE
				آ	0675	ARABIC LETTER HIGH HAMZA ALEF
3	أ	0623	ARABIC LETTER ALEF WITH HAMZA ABOVE	أ	0622	ALEF WITH MADDA ABOVE
				آ	0671	ARABIC LETTER ALEF WASLA
				أ	0672	ARABIC LETTER ALEF WITH WAVY HAMZA ABOVE
				آ	0675	ARABIC LETTER HIGH HAMZA ALEF
4	ؤ	0624	ARABIC LETTER WAW WITH HAMZA ABOVE	ؤ	0676	ARABIC LETTER HIGH HAMZA WAW
				ؤ	0677	ARABIC LETTER U WITH HAMZA ABOVE
				ؤ	06C6	ARABIC LETTER OE
				ؤ	06C7	ARABIC LETTER U
				ؤ	06C8	ARABIC LETTER YU
				ؤ	06C9	ARABIC LETTER KIRGHIZ YU
				ؤ	06CA	ARABIC LETTER WAW WITH TWO DOTS ABOVE
				ؤ	06CB	ARABIC LETTER VE
				ؤ	06CF	ARABIC LETTER WAW WITH DOT ABOVE
5	إ	0625	ARABIC LETTER ALEF WITH HAMZA BELOW	إ	0673	ARABIC LETTER ALEF WITH WAVY HAMZA BELOW
6	ئ	0626	ARABIC LETTER YEH WITH HAMZA ABOVE	ئ	0678	ARABIC LETTER HIGH HAMZA YEH
				ئ	06CE	ARABIC LETTER YEH WITH SMALL V
7	ا	0627	ARABIC LETTER ALEF	ا	0661	ARABIC-INDIC DIGIT ONE
				ا	06F1	EXTENDED ARABIC-INDIC DIGIT ONE
8	ب	0628	ARABIC LETTER BEH	ب	066E	ARABIC LETTER DOTLESS BEH
				ب	067B	ARABIC LETTER BEEH
				ب	067E	ARABIC LETTER PEH
				ب	0680	ARABIC LETTER BEHEH
9	ة	0629	ARABIC LETTER TEH MARBUTA	ة	06C0	ARABIC LETTER HEH WITH YEH ABOVE
				ة	06C2	ARABIC LETTER HEH GOAL WITH HAMZA ABOVE
				ة	06C3	ARABIC LETTER TEH MARBUTA GOAL
10	ث	062A	ARABIC LETTER THE	ث	062B	ARABIC LETTER THEH
				ث	0679	ARABIC LETTER TTEH
				ث	067A	ARABIC LETTER TTEHEH

				ف	067C	ARABIC LETTER TEH WITH RING
				ف	067D	ARABIC LETTER TEH WITH THREE DOTS ABOVE DOWNWARDS
				ت	067F	ARABIC LETTER TEHEH
				ن	06BD	ARABIC LETTER NOON WITH THREE DOTS ABOVE
11	ث	062B	ARABIC LETTER THEH	ط	0679	ARABIC LETTER TTEH
				ط	067A	ARABIC LETTER TTEHEH
				ف	067C	ARABIC LETTER TEH WITH RING
				ف	067D	ARABIC LETTER TEH WITH THREE DOTS ABOVE DOWNWARDS
				ت	067F	ARABIC LETTER TEHEH
				ن	06BD	ARABIC LETTER NOON WITH THREE DOTS ABOVE
12	ج	062C	ARABIC LETTER JEEM	ج	0683	ARABIC LETTER NYEH
				ج	0684	ARABIC LETTER DYEH
				ج	0686	ARABIC LETTER TCHEH
				ج	0687	ARABIC LETTER TCHEHEH
				ج	06BF	ARABIC LETTER TCHEH WITH DOT ABOVE
				ج	06DA	ARABIC SMALL HIGH JEEM
13	ح	062D	ARABIC LETTER HAH	-	--	--
14	خ	062E	ARABIC LETTER KHAH	ح	0681	ARABIC LETTER HAH WITH HAMZA ABOVE
				ح	0682	ARABIC LETTER HAH WITH TWO DOTS VERTICAL ABOVE
				ح	0685	ARABIC LETTER HAH WITH THREE DOTS ABOVE
				ح	06BF	ARABIC LETTER TCHEH WITH DOT ABOVE
15	د	062F	ARABIC LETTER DAL	د	0689	ARABIC LETTER DAL WITH RING
				د	068A	ARABIC LETTER DAL WITH DOT BELOW
				د	068D	ARABIC LETTER DDAHAL
16	ذ	0630	ARABIC LETTER THAL	ذ	0688	ARABIC LETTER DDAL
				ذ	068B	ARABIC LETTER DAL WITH DOT BELOW AND SMALL TAH
				ذ	068C	ARABIC LETTER DAHAL
				ذ	068E	ARABIC LETTER DUL
				ذ	068F	ARABIC LETTER DAL WITH THREE DOTS ABOVE DOWNWARDS
				ذ	0690	ARABIC LETTER DAL WITH FOUR DOTS ABOVE
				ذ	06EE	ARABIC LETTER DAL WITH INVERTED V
17	ر	0631	ARABIC LETTER REH	ر	0693	ARABIC LETTER REH WITH RING
				ر	0694	ARABIC LETTER REH WITH DOT BELOW
				ر	0695	ARABIC LETTER REH WITH SMALL V BELOW
				ر	0696	ARABIC LETTER REH WITH DOT BELOW AND DOT ABOVE
18	ز	0632	ARABIC LETTER ZAIN	ز	0691	ARABIC LETTER RREH
				ز	0692	ARABIC LETTER REH WITH SMALL V
				ز	0696	ARABIC LETTER REH WITH DOT BELOW AND DOT ABOVE
				ز	0697	ARABIC LETTER REH WITH TWO DOTS ABOVE
				ز	0698	ARABIC LETTER JEH
				ز	0699	ARABIC LETTER REH WITH FOUR DOTS ABOVE
				ز	06EF	ARABIC LETTER REH WITH INVERTED V
19	س	0633	ARABIC LETTER SEEN	س	069A	ARABIC LETTER SEEN WITH DOT BELOW AND DOT ABOVE
				س	069B	ARABIC LETTER SEEN WITH THREE DOTS BELOW
				س	06DC	ARABIC SMALL HIGH SEEN
20	ش	0634	ARABIC LETTER SHEEN	ش	069A	ARABIC LETTER SEEN WITH DOT BELOW AND DOT ABOVE

				ثيٲ	069C	ARABIC LETTER SEEN WITH THREE DOTS BELOW AND THREE DOTS ABOVE
				ثيٲ	06FA	ARABIC LETTER SHEEN WITH DOT BELOW
21	ص	0635	ARABIC LETTER SAD	ص	069D	ARABIC LETTER SAD WITH TWO DOTS BELOW
22	ض	0636	ARABIC LETTER DAD	ض	069E	ARABIC LETTER SAD WITH THREE DOTS ABOVE
				ض	06FB	ARABIC LETTER DAD WITH DOT BELOW
23	ط	0637	ARABIC LETTER TAH	ط	0615	ARABIC SMALL HIGH TAH
24	ظ	0638	ARABIC LETTER ZAH	ظ	069F	ARABIC LETTER TAH WITH THREE DOTS ABOVE
25	ع	0639	ARABIC LETTER AIN	ع	060F	ARABIC SIGN MISRA
26	غ	063A	ARABIC LETTER GHAIN	غ	06A0	ARABIC LETTER AIN WITH THREE DOTS ABOVE
				غ	06FC	ARABIC LETTER GHAIN WITH DOT BELOW
27	ف	0641	ARABIC LETTER FEH	ف	060B	AFGHANI SIGN
				ف	06A1	ARABIC LETTER DOTLESS FEH
				ف	06A2	ARABIC LETTER FEH WITH DOT MOVED BELOW
				ف	06A3	ARABIC LETTER FEH WITH DOT BELOW
				ف	06A4	ARABIC LETTER VEH
				ف	06A5	ARABIC LETTER FEH WITH THREE DOTS BELOW
				ف	06A6	ARABIC LETTER PEHEH
				ف	06A7	ARABIC LETTER QAF WITH DOT ABOVE
28	ق	0642	ARABIC LETTER QAF	ق	066F	ARABIC LETTER DOTLESS QAF
				ق	06A4	ARABIC LETTER VEH
				ق	06A6	ARABIC LETTER PEHEH
				ق	06A7	ARABIC LETTER QAF WITH DOT ABOVE
				ق	06A8	ARABIC LETTER QAF WITH THREE DOTS ABOVE
29	ك	0643	ARABIC LETTER KAF	ك	06A9	ARABIC LETTER KEHEH
				ك	06AA	ARABIC LETTER SWASH KAF
				ك	06AB	ARABIC LETTER KAF WITH RING
				ك	06AC	ARABIC LETTER KAF WITH DOT ABOVE
				ك	06AD	ARABIC LETTER NG
				ك	06AE	ARABIC LETTER KAF WITH THREE DOTS BELOW
				ك	06AF	ARABIC LETTER GAF
				ك	06B0	ARABIC LETTER GAF WITH RING
				ك	06B1	ARABIC LETTER NGOEH
				ك	06B2	ARABIC LETTER GAF WITH TWO DOTS BELOW
				ك	06B3	ARABIC LETTER GUEH
				ك	06B4	ARABIC LETTER GAF WITH THREE DOTS ABOVE
30	ل	0644	ARABIC LETTER LAM	ل	06B5	ARABIC LETTER LAM WITH SMALL V
				ل	06B6	ARABIC LETTER LAM WITH DOT ABOVE
				ل	06B7	ARABIC LETTER LAM WITH THREE DOTS ABOVE
				ل	06B8	ARABIC LETTER LAM WITH THREE DOTS BELOW
31	م	0645	ARABIC LETTER MEEM	م	06D8	ARABIC SMALL HIGH MEEM INITIAL FORM
				م	06E2	ARABIC SMALL HIGH MEEM ISOLATED FORM
				م	06ED	ARABIC SMALL LOW MEEM
				م	06FE	ARABIC LETTER SINDHI POSTPOSITION MEN
32	ن	0646	ARABIC LETTER NOON	ن	06B9	ARABIC LETTER NOON WITH DOT BELOW
				ن	06BA	ARABIC LETTER NOON GHUNNA
				ن	06BB	ARABIC LETTER RNOON

				ن	06BC	ARABIC LETTER NOON WITH RING
				ن	06BD	ARABIC LETTER NOON WITH THREE DOTS ABOVE
				ن	06E8	ARABIC SMALL HIGH NOON
33	ه	0647	ARABIC LETTER HEH	و	0652	ARABIC SUKUN
				٥	0665	ARABIC-INDIC DIGIT FIVE
				ه	06BE	ARABIC LETTER HEH DOACHASHMEE
				ه	06C1	ARABIC LETTER HEH GOAL
				ه	06D5	ARABIC LETTER AE
				ه	06FF	ARABIC LETTER HEH WITH INVERTED V
34	و	0648	ARABIC LETTER WAW	و	064F	ARABIC DAMMA
				و	066F	ARABIC LETTER DOTLESS QAF
				و	06C4	ARABIC LETTER WAW WITH RING
				و	06C5	ARABIC LETTER KIRGHIZ OE
				و	06C6	ARABIC LETTER OE
				و	06C7	ARABIC LETTER U
				و	06C8	ARABIC LETTER YU
				و	06C9	ARABIC LETTER KIRGHIZ YU
				و	06CA	ARABIC LETTER WAW WITH TWO DOTS ABOVE
				و	06CB	ARABIC LETTER VE
				و	06CF	ARABIC LETTER WAW WITH DOT ABOVE
35	ى	0649	ARABIC LETTER ALEF MAKSURA	ى	06CC	ARABIC LETTER FARSI YEH
				ى	06CD	ARABIC LETTER YEH WITH TAIL
				ى	06CE	ARABIC LETTER YEH WITH SMALL V
				ى	06D2	ARABIC LETTER YEH BARREE
36	ي	064A	ARABIC LETTER YEH	ي	067B	ARABIC LETTER BEEH
				ي	067E	ARABIC LETTER PEH
				ي	0680	ARABIC LETTER BEHEH
				ي	06D0	ARABIC LETTER E
				ي	06D1	ARABIC LETTER YEH WITH THREE DOTS BELOW
				ي	06D2	ARABIC LETTER YEH BARREE
37	٠		ARABIC-INDIC DIGIT ZERO	٠	06F0	EXTENDED ARABIC-INDIC DIGIT ZERO
38	١		ARABIC-INDIC DIGIT ONE	١	0627	ARABIC LETTER ALEF
				١	06F1	EXTENDED ARABIC-INDIC DIGIT ONE
39	٢		ARABIC-INDIC DIGIT TWO	٢	06F2	EXTENDED ARABIC-INDIC DIGIT TWO
40	٣		ARABIC-INDIC DIGIT THREE	٣	06F3	EXTENDED ARABIC-INDIC DIGIT THREE
41	٤		ARABIC-INDIC DIGIT FOUR	--	--	--
42	٥		ARABIC-INDIC DIGIT FIVE	ه	0647	ARABIC LETTER HEH
				ه	06C1	ARABIC LETTER HEH GOAL
				ه	06D5	ARABIC LETTER AE
43	٦		ARABIC-INDIC DIGIT SIX	--	--	--
44	٧		ARABIC-INDIC DIGIT SEVEN	٧	06F7	EXTENDED ARABIC-INDIC DIGIT SEVEN
45	٨		ARABIC-INDIC DIGIT EIGHT	٨	06F8	EXTENDED ARABIC-INDIC DIGIT EIGHT
46	٩		ARABIC-INDIC DIGIT NINE	٩	06F9	EXTENDED ARABIC-INDIC DIGIT NINE

Table 1: Similar letters in Arabic Scripts

It is noticed that the first column represents the serial number of the character intended to be studied. The column is ordered according to the character's position in the Unicode (Arabic Script section). The 2nd, 3rd, and 4th columns represent the character's form, its code, and its description in order. The last three columns deals with the same features of the characters similar to the investigated one.

The technique used to define the similarity level depends on the character's shape approximation and the similarity of its parts; their numbers and position, taking into

consideration that some exceptions are made for dots since they might not be separate parts for being too small. For example, when searching for the characters similar to "أ" with *hamza*, first, the character's shape is analyzed into two parts. The main long stick-like vertical column and the small part represented by *hamza* over the main part. Then,

characters that are similar in the shape and the number of parts are searched. The outcomes are like "ا" and "آ" because

generally they similar in the number of parts yet there minor differences in the shape of *hamza*. However, the outcomes

do not include the character "ا" because it consists only one part despite the great similarity level it has in the main part.

5. Security Threats of Arabic Domain Names

5.1. Domain Homograph or Phishing

Similarity of Arabic domain names will be mainly discussed in that section, however, this threat should be known as an old one that endanger current domain names especially those of banks and payment transactions websites such as www.paypal.com. Similarities between characters like letters "L" "O" in the small case and numbers "1" "0"

in order are commonly used to deceive the user by producing domain names that look like original ones. Therefore, the pace should be quicker to activate multi-language domain names especially that possible solutions to avoid these threats are represented by spreading the technical awareness of the system user.

Concerning Arabic domain names, similarity may occur in the following:

A. Shape: The focus will be on the letters of the Arabic script since the other letters have different direction. A number of domain names can be registered by a group of people so that the end user finds it difficult to distinguish between them. For example, the domain name "المجلة الأولى" is different from "المجلة الأولى" because the latter replaced (1) instead of (ا). It is also different from "المجلة الأولى" using shortened 'alif' instead of Persian 'yaa' (U-06CC). Other potentials of similarities are possible as Table (1) explained in details.

B. Spelling Mistakes: It is common since a great amount of spelling and grammatical rules are not applied to the daily life especially among the public. For example, a user can write "اماره" instead of "إمارة" with a spelling error in hamza and bound *taa'*

C. System-supported Letters & Marks: The approved solution supports all the Unicode letters, marks, and symbols including the Arabic script. Therefore, users can register various domain names with differences that are too slight to be noticed even by the user. It can occur through:

- Vocalization marks: The domain name "حَسَن" is distinct from "حُسْن" and "حسن"
- Extension "—": The domain name "صالح" is distinct from "صالح" and "صَالِح"
- Other marks that are common in the Arabic script yet difficult to be noticed for their tiny size

D. Compounding: Words in Arabic are separated using a space yet the approved solution does not apply such a space. Therefore, users can utilize "—" instead of the space or compound the words together. For example, the domain name "مؤسسة الفلاني" is distinct from "مؤسسة الفلاني"

E. Using "0": The numeral "0" which is written in Arabic in the form of "٠" could be used to

misguide the users as if they are dealing with a sub-domain from the famous main domain. For example, the domain name "موقع بنك" is distinct from "موقع.بنك". The former represents a one-word complete domain name yet the latter represents a main domain "موقع" that have a sub-domain "بنك"

- F. Mixing Letters:** Letters can be mixed to look like a single letter. For example, the upper *hamza* can be modified to the letter 'alif' (ا) so that they may look like the letter "أ" to be used for registering a domain name
- G.** Combinations between the abovementioned points may result in various similarities among the domain names especially with long domain names

There are some factors that help spreading the threats:

- a) Users have lower awareness level that Arabic domain names may be similar besides the necessity of ensuring the domain name correct spelling i.e. not to depend on the unofficial sources like emails and SMS
- b) Domain name registration centers – especially foreign centers that are responsible for registering domain – still did not absorb the characteristics of Arabic language
- c) The current system supports all the Unicode character tables without any controls or stipulations that relate to Arabic language.
- d) Many other language shares the Arabic script with the Arabic language besides the close similarities among the letters of these languages
- e) Some fonts that are used to represent domain names in programs or operating systems may lead to higher similarities especially with small-size font lines.

However, various solutions are available to avoid these problems:

- a) Increasing the technical and security awareness among DNS users
- b) Altering the technical specifications of the solution or devising the technique to be mixed with this solution to avoid problems
- c) Classifying the Arabic script into subdivision according to the language e.g. Arabic section, Persian section, etc. Otherwise, removing other languages from the Arabic script so that they can be categorized in separate tables with controls and stipulations that suit each language
- d) Stop mixing letters from different languages i.e. using only one language to form the domain name e.g. not to mix Arabic and English letters to form a multi-language domain name
- e) Setting rules for registering Arabic domain names to limit the occurrence of problems such as adherence to the spelling and grammatical rules
- f) To save its intellectual property and rights, any entity when registering a domain name should register other domain names similar to the original one

5.2. The Difficulty of using multi-language domain names for critical services:

It is a temporary issue because in the forthcoming period it will be difficult to depend mainly and officially on the multi-language domain names. They should be used as a means supporting the current domain names until fulfilling the issues below:

- Complete support from Arab and foreign societies
- To develop the infrastructure so that all the famous DNS-based software and applications turn to support multi-language domain names such as:
 - Internet browsers such as MS IE
 - Email clients such as MS Outlook
 - File transfer programs such as WS_FTP
 - Others
- Spreading technical awareness about Arabic domain names among users
- Spreading the use of multi-language domain names generally and Arabic domain names specifically

This does not mean discouraging the use of Arabic domain names. It is a goal for which all Arab strive to be fulfilled as soon as possible. However, it is necessary to pinpoint the difficulty of applying critical services like bank websites and public service while depending only on multi-language domain names. An Arabic domain name can be established in addition to another foreign domain name to ensure the service availability even for users whose computers do not support multi-language domain names .

5.3. The Difficulty of Identifying the Contents of Log Files:

Log files are considered to be the most important source of investigating the historical data of an equipment, an operating system, or software in general. Systems and applications also known to be highly dependent on DNS to identify the names of computers and websites which are stored in log files. Storage techniques mostly use foreign letters (ASCII) to save files; a fact that makes it possible in the future to have log files containing names of equipments and websites stored using the "pubcode" of the multi-language domain name to make the log file appears like below:

```
May 30 12:10:55 server pure-ftpd: (?@127.0.0.1) [INFO]
New connection from xn--4gbrim.xn--mgb5a8an.xn--
ugb6blj
```

This log file refers the a message that a successful connection has been initiated from the computer named "موقع.نطاق.دولي" into the computer responsible for file transfer according to the date and time specified in the message. The computer name is obtained after using a software that transform the text "xn--4gbrim.xn--mgb5a8an.xn--ugb6blj" into "موقع.نطاق.دولي". The transformation process consumes time and efforts depending on the log file size and the necessity of reaching the data.

This problem is not limited to the Arabic domain names yet it covers all the multi-language domain names. However, it could be solved as below:

- The operating systems as well as the log files initiation and storage software should support an encryption/storage based on Unicode characters e.g. UTF-8 to directly save multi-language domain names in the log files.
- Developing some applications and tools that can read the log files then transform the name encrypted texts into the proper understandable multi-language domain names.

5.4. Big Organization should register Arabic domain names:

Registering Arabic domain names is expected to be officially started soon at the registration centers all over the Arab region. Therefore, it is inevitable for key entities, companies, and trademark owners to register these names before assigned to others who might bargain these names. It is also preferred to register Arabic domain names that are represent abbreviation or abstractions of their names.

The Arabic domain name registration centers should devise a way to protect property e.g. the owners of trademarks and commercial names can have a period of time to register their Arabic domain names before opening the registration to others. This period is known as Sunrise Period.

6. Conclusion

Using Arabic language to name website domains over the internet is a national demand especially because Arabic is the language of Quran. However, the efforts exerted to fulfill this demand are incomplete. Therefore, this research paper demonstrates the disadvantages and the threats that might result from using Arabic letters. It also proposes solutions and recommendations to avoid problems. It is a hope to proper and secure utilization of Arabic domain names.

7. References

- [1] Arabic Domain Names Registrars , www.arabic-domains.org.
- [2] P. Albitz and C. Liu (2001). DNS and BIND, 4th edition, O'Reilly & Associates.
- [3] A. Al-Zoman. (2004). Internet Domain Names: what do you know about them, First Edition.
- [4] S. Boran, "Running the BIND9 DNS Server securely", Boran , 2004, http://www.boran.com/security/sp/bind9_20010430.html
- [5] Working-Group , "RFC 3454: Preparation of Internationalized Strings ("stringprep")", The Internet Society, December 2002.

- [6] Working-Group , "RFC 3490: Internationalizing Domain Names in Applications (IDNA)", The Internet Society, March 2003
- [7] Working-Group , "RFC 3491: Nameprep: A Stringprep Profile for Internationalized Domain Names (IDN)", The Internet Society, March 2003
- [8] Working-Group , "RFC 3492: Punycode: A Bootstring encoding of Unicode for Internationalized Domain Names in Applications (IDNA)", The Internet Society, March 2003

Author Biographies



Khaled Alghathbar is the Director of Center of Excellence in Information Assurance and also on the faculty of Information Systems in King Saud University, Riyadh, Saudi Arabia. He received a PhD in Information Technology from George Mason University in 2004 and has several certifications in information systems security. His dissertation is in the area of incorporating access and flow control policies in requirements engineering. He developed a framework to formally verify access and flow control policies during the early stages of the development life cycle. He has published several papers in international journals, conferences and workshops.