**TECHNICAL ADVISORY GROUP ON
MACHINE READABLE TRAVEL DOCUMENTS**

Fifteenth Meeting

(Montreal, 17 to 21 May 2004)

Agenda Item 2: Report of the Document Content and Format Working Group
Agenda Item 2.1: Progress in updating Doc 9303 and proposed amendments

"e-Passport" INDICATOR – VERSION NUMBER & NEW CHECK CHARACTER SCHEME IN MACHINE READABLE TRAVEL DOCUMENTS

(Presented by the Document Content and Format Working Group (DCFWG))

1. INTRODUCTION

1.1 This working paper has arisen out of a request of the New Technologies Working Group (NTWG) to find a way to incorporate in the Machine Readable Zone (MRZ) an indicator that the Travel Document contains a contactless IC. This type of Travel Document will be referred to here as the "e-Passport", as the Machine Readable Passport (MRP) with contactless IC is the first example of this technology.

2. SUMMARY

2.1 The introduction of an e-Passport indicator in the MRZ introduces a serious backward compatibility problem with respect to installed border control reader equipment and systems. Whatever method is adopted, there will be difficulty. This working paper addresses this problem. It proposes a solution that, though not entirely backward compatible, conserves the essential structure of the MRZ. The solution has the additional merit of considerably improving the immediate and long term usability of the MRZ. The concepts were discussed at the DCFWG meeting (Basingstoke, 8 to 10 March 2004) and further refined by extensive discussion by email.

3. CONCEPTS

3.1 The e-Passport is the traditional MRP, essentially as described in Doc 9303, with the addition of Optional Expansion Capacity in the form of a contactless chip.

3.2 The contactless chip contains one or more biometrics and a digital signature (Public Key Infrastructure - PKI), encoded using the Logical Data Structure (LDS). By the means of the biometric the e-Passport is linked to the bearer, and by the means of the digital signature the document can be protected against forgery and tampering. These are two powerful concepts which make forgery and fraud considerably more difficult, if not impossible.

3.3 In order to combat unauthorized reading, several issuing countries have proposed that the chip be optionally protected by a password. This password is to be derived from the MRZ by traditional Optical Character Recognition (OCR) reading. This is known as "closed access" (as distinct from unlocked chips that would have "open access" and not require a password). The OCR readability is a critical issue here.

4. QUESTIONS POSED BY THE DCFWG

4.1 How can the presence of the chip be ascertained during the machine reading process? While the passport cover and data page can be marked with a humanly readable logo, usually this cannot be recognised by a machine reader. The NTWG has requested the DCFWG to consider an appropriate change to the MRZ to accomplish this function. Note that the chip itself cannot be relied on to perform this function as it may be inoperative, or may be locked ("closed access").

4.2 How can the MRZ reading process be made more accurate and reliable? Even if it is optional, if "closed access" is adopted by a large number of countries, then receiving countries will have to machine read the MRZ of the e-Passport in order to derive a key that can be used to unlock the closed chip. If, as has been reported, the error rate in machine reading is high in some countries, facilitation will suffer as manual entry must be used to compensate for this error rate.

5. PROPOSAL BY DCFWG – VERSION NUMBER AND NEW CHECK CHARACTER SCHEME

5.1 The DCFWG is proposing mandatory adoption of a new MRZ check character scheme and version number that will:

a) identify the MRP as an e-Passport;

b) overcome the limitations of the present check digit scheme;

c) make the OCR reading more secure; and

d) open the way for future improvements of the MRZ by introducing a version identifier.

5.2 It should be noted that any modification of the MRZ, whether minor (one character) or major (many characters), will result in the same major disruptive effect on existing machine readers and processing systems worldwide. However, for those countries introducing e-Passports, the effort in modifying the printing and reading of the MRZ is extremely small in comparison to the effort required to encode and read the contactless chip.

5.3 The objective is then to cause the least amount of disruption and to keep some measure of backward compatibility, especially for those countries that are planning to keep their present machine reading capability, that is, not planning to read e-Passports except for the MRZ as before.

5.4 The DCFWG considered and discarded two initial options:

a) to use the "PE" combination at the beginning of the first MRL (and other combinations for visas and MROTDs eg "VE" and "IE" respectively). This option has the advantage of preserving existing printing, reading and processing systems. However there are major disadvantages in that:

1) the use of the second letter to denote types of passports (eg "PD" for "diplomatic") will be lost; and

2) some countries are already using the "PE" combination for traditional MRPs. This will lead to great confusion if this option is adopted.

b) to replace the initial "P" at the beginning of the first MRL by the letter "E' (and other letter combinations for visas and MROTDs). The second letter can retain its significance as before, so for example "ED" could denote a diplomatic e-Passport. However, this option means that:

1) all existing readers and processing systems will need to be modified, and there is no measure of backward compatibility. This option would cause the maximum disruption for no gain, other than to denote e-Passports; and

2) it is also considered unsafe to rely on just one character to signal the presence of the chip, as forgers, desirous of hiding presence of the chip, will attack this character and try to change it back to that of a standard passport ("P").

5.5 In considering this major disruption to the world's reading and processing systems, the possibility of other solutions has been discussed that:

1) would not be as disruptive;

2) be potentially backward compatible; and

3) offer improved performance in terms of OCR-B readability.

5.6 Thus a third and better option is to replace the last five positions of the "Name" field with a new system of check characters and a version number. This involves shortening the "Name" field by 5 characters. It may also be desirable to further shorten the name field by a further 3 positions and insert 3 separators ("<<<") in order to clearly delimit the shortened "Name" field from the new "Check Character and Version Number" field.

5.7 The purpose of the version number is:

1) to identify the MRP as an e-Passport; and

2) to allow for future revisions if needed.

This first version number in all e-Passports would be the number "1" in position 42 of the upper MRL.

5.8 The purpose of the check characters is:

1) to improve the readability of the OCR; and

2) to further identify the MRP as an e-Passport, ie to remove the reliance on just one identifying character.

Candidates for simple but numerically efficient check character scheme have been identified that use four check characters (A-Z, 0-9 and <) and that provide a maximum of error detection and recovery for 88 characters. The

new check character scheme provides for more accurate reading of the MRZ as the scheme includes all data in the MRZ, not only selected fields as in the current scheme.

5.9 The proposed algorithm works like this:

a) every existing character is mapped to a corresponding number (eg 0..9 = 0..9; A = 10, B = 11 etc. until Z = 35; and < = 36);

b) all data fields are arranged in the defined sequence (eg Type, Country of Issue, Name, Version Number, Document Number, etc), including the existing 7-3-1 weighted check digits. This produces a row of 84 characters for the MRP case (because from the total number of 88 characters in the two lines the new four check characters must be deducted);

c) now calculate check character No.1 (CC1) by adding all numerical values corresponding to the 85 specific characters. The result is divided by 37 as many times as the integer remainder is bigger than 37 and the final remainder then is CC1;

d) calculate CC2 by the same procedure, while doing the sum the value for each character is multiplied by its position in the row;

e) calculate CC3 by the same procedure while doing the sum the first 37 character values in the row are multiplied by 1, the second 37 values are multiplied by 2 and the rest of them multiplied by 3;

f) calculate CC4 by a procedure [tba];

g) the check characters (CC1, CC2, CC3 and CC4) will range between 0 and 36 and each can be mapped back to a single character that is then inserted in the respective position in the MRZ (positions 40, 41, 43 and 44 of the upper MRL); and

h) the version number uses position 42 of the upper MRL in order to "split" the check characters into 2 groups of 2 characters each, thus preventing the spelling of a "nasty" word in any language.

5.10 The new check character scheme refers to the total MRZ rather than individual fields as before with the 7-3-1 weighting check digit scheme. Thus the presently unchecked "name" and other fields are included in the new scheme. By means of the mathematical function, these check characters can reliably identify up to 3 errors if the error locations are unknown, and correct the error if only one character position is corrupted. If the error locations are known two errors may be corrected. Note that the check digits used in the past only refer to their respective fields (document number, date of birth, date of expiry and the optional data field) and can usually identify errors, although sometimes two or more errors may cancel each other and not be identified. Error correction is possible only in certain circumstances. In general, the present 7-3-1 weighting check digit scheme is inefficient, and was partly adopted as it can be humanly calculated. The new check character scheme is more powerful and could still be humanly calculated although this is would be a lengthy task.

5.11 The essence of this third option is that the presence of the version number and the revised check character scheme will signify that the MRP is an e-Passport. The traditional 7-3-1 weighted check digits at positions 10, 20, 28, 43 and 44 of the lower MRL are left as before for maximum backward compatibility.

5.12 The new check character scheme will result in considerably improved OCR-B reading. This is vital if the MRZ is to be reliably read in order to derive a "key" or "password" in order to unlock a "closed" contactless chip.

5.13 This third option is considered to be the most backward compatible of all the options examined. Receiving countries with traditional machine readers and processing systems will be able to still read the MRZs of e-Passports. However, the new characters at the end of the name field may be identified as part of the name, causing a database mis-match. The use of 3 separators ("<<<") to separate the name field from the new field may, in most cases, prevent this problem. Countries should be urged to upgrade their traditional machine readers and systems by reprogramming or replacement to accommodate the new scheme.

5.14 Countries upgrading or installing new reading equipment and systems will be able to read both types of travel documents (ie traditional MRP or e-Passport). Accommodating the new check character scheme and the reduced name field length is essentially a programming exercise. This exercise represents a small fraction of the effort to introduce the reading of contactless ICs.

5.15 Resulting from the discussion in the DCFWG meeting, a similar alternative scheme was proposed but finally rejected. This alternative was to use the version number and new check character scheme and replace the existing 7-3-1 weighted check digits. The version number would be a "letter" (initially "A") and would replace the overall check digit at position 44 of the lower MRL. The new check characters would replace the existing check digits at positions 10, 20, 28 and 43 of the lower MRL. This alternative was discarded due to the following reasons:

a) existing readers would have a high chance of mis-reading as the existing check digit scheme would no longer compute, and the e-Passport would most likely be flagged as a forgery; and

b) inspectors would lose a valuable tool in detecting forged passports by not being able to calculate the check digits manually. Forgers often get these check digits incorrect. Of course, the same means of detection is available with the new check characters, but the new algorithm is much more difficult to manually compute.

5.16 Note that use of the optional data field of the MRZ (positions 29 to 43 of the lower MRL) for the check characters and version number is not possible as many issuing states use this field for other purposes.

6. ACTION BY THE TAG/MRTD

6.1 The DCFWG invites the TAG/MRTD to :

a) note the work and approach taken by the DCFWG on the identification of e-Passports using the MRZ, and the new check character scheme for the MRZ; and

b) recommend approval of the approach and work to date by the DCFWG, and approve the development of more precise technical specifications for inclusion in Doc 9303.

— END —