Groupe de travail Réseau
Request for Comments : 2938
RFC mise à jour : 2533
Catégorie : En cours de normalisation

G. Klyne, Content Technologies
L. Masinter, AT&T
septembre 2000
Traduction Claude Brière de L'Isle

Identification des caractéristiques de support composites

Résumé
Dans la RFC2533 est présenté un format d'expression pour décrire les capacités de caractéristiques des supports comme une combinaison de simples étiquettes de caractéristiques de supports.

Le présent document décrit un format abrégé pour un ensemble de supports composites, sur la base d'un hachage de l'expression des caractéristiques qui décrivent ce composite.

Table des Matières

1. Introduction

Dans "Syntaxe pour décrire les ensembles de caractéristiques des supports" [1], un format d'expression est présenté pour décrire les capacités des caractéristiques de support comme une combinaison de simples étiquettes de caractéristiques de supports [2].

Le présent document propose un format abrégé pour un ensemble composite de caractéristiques de supports, fondé sur un hachage de l'expression des caractéristiques qui décrivent ce composite.

Le présent mémoire étend la syntaxe d'expression décrite dans la RFC2533 [1] en s'appuyant sur elle, et il est supposé que le lecteur est familier avec l'interprétation des expressions d'ensembles de caractéristiques décrites ici.

1.1 Organisation du document

La Section 2 établit les fondements et les objectifs des références d'ensembles de caractéristiques.

La Section 3 présente une syntaxe pour les références d'ensembles de caractéristiques, et décrit comment elles se rapportent aux expressions d'ensembles de caractéristiques.

1.2 Terminologie et conventions du document

Un certain nombre de termes et autres conventions du document sont définies dans ce paragraphe, qui sont utlisés avec une signification spécifique dans le présent mémoire. Les termes sont donnés dans l'ordre alphabétique.

déréférencer
Acte de remplacer une référence d'ensemble de caractéristiques par l'expression correspondante de l'ensemble de caractéristiques. Aussi appelé "résolution".

ensemble de caractéristiques
ensemble des caractéristiques de support décrites par une assertion de caractéristique de support, comme décrite dans "Syntaxe pour la description des ensembles de caractéristiques de supports" [1]. (Voir dans ce texte une définition plus formelle de ce terme.)

étiquette d'ensemble de caractéristiques
nom qui se conforme à la syntaxe d'une étiquette de caractéristique [2] qui est utilisée pour noter un ensemble de caractéristiques plutôt qu'une seule caractéristique.

expression d'ensemble de caractéristiques
chaîne qui décrit un certain ensemble de caractéristiques, formulée conformément aux règles de "Syntaxe pour la description des ensembles de caractéristiques de supports" [1] (et eventuellement étendues par d'autres spécifications).

référence d'ensemble de caractéristiques
brève construction qui fait référence à un certain ensemble de caractéristiques. (Voir aussi à "déréference".)

résolution
(voir "déréference").

La présente spécification utilise la notation syntaxique et les conventions décrites dans la RFC2234, "BNF augmenté pour les spécifications de syntaxe : ABNF" [3].

Note : Des commentaires comme celui-ci apportent des informations non essentielles supplémentaires sur les raisons qui sous-tendent le présent document. De telles informations ne sont pas nécessaires pour la construction d'une mise en œuvre conforme, mais peuvent aider ceux qui souhaitent avoir une compréhension plus approfondie des concepts.

2. Motivation et objectifs

La gamme des capacités de caractéristiques de supports d'un système de traitement de messages peut être assez large, et l'expression de l'ensemble correspondant de caractéristiques [1] peut atteindre une taille significative.

On a déterminé une exigence pour permettre d'identifier des ensembles récurrents de caractéristiques par une seule valeur de référence, qui peut être combinée avec d'autres éléments dans une expression d'ensemble de caractéristiques. On prévoit que des mécanismes seront fournis pour permettre au receveur de telles références d'ensembles de caractéristiques de découvrir l'expression d'ensemble de caractéristiques correspondante, mais un tel mécanisme sort du domaine d'application de la présente spécification.

Donc, les objectifs de la présente proposition sont :

- de fournir une forme abrégée pour se référer à une expression arbitraire d'ensemble de caractéristiques.
- la signification (c'est-à-dire, l'expression d'ensemble de caractéristiques correspondante) d'une référence d'ensemble de caractéristiques devrait être indépendante de tout mécanisme particulier qui pourrait être utilisé pour le déréférencer.
- être capable de vérifier si une expression d'ensemble de caractéristiques donnée correspond à une référence d'ensemble de caractéristiques sans avoir à effectuer une opération explicite de déréférencement (c'est-à-dire, sans générer de trafic réseau supplémentaire).

- pour les traitements du protocole qui se conforment à la RFC2533 [1], être capable de traiter intelligemment une référence d'ensemble de caractéristiques sans connaissance explicite de sa signification (c'est-à-dire, l'introduction de références d'ensemble de caractéristiques ne devrait pas interrompre les traitements de l'expression des caractéristiques existantes). En fait, l'interprétation et les règles de traitement applicables de la RFC2533 [1] s'appliquent également aux expressions qui contiennent des références d'ensembles de caractéristiques.

Note : Cette proposition n'essaye pas de régler les problèmes de "chevauchement" ou de "par défaut". (Lorsque un ensemble de caractéristiques peut être référencé et modifié de façon sélective.)

Parmi les circonstances dans lesquelles cette forme abrégé peut être utilisée, on citera :

- une expression de caractéristique de support qui contient une sous expression répétée. Si la sous expression est assez grande, de l'espace peut être économisé en l'écrivant toute une fois, puis en utilisant la forme abrégée pour s'y référer.
- une capacité qui est normalement une gamme d'appareils, comme une certaine classe de télécopieurs où un grand nombre d'étiquettes de caractéristiques sont impliquées, mais seul un petit nombre d'ensembles de caractéristiques courants. Si le receveur comprend, ou peut découvrir, que certaines abréviation représentent un certain ensemble de caractéristiques, la taille de l'expression des caractéristiques peut être réduite en utilisant les abréviations.

Si les abréviations d'ensemble de caractéristique sont utilisées de cette façon, il se peut qu'elles puissent être interprétées par une simple recherche sur un tableau plutôt que par l'analyse complète de l'expression de la caractéristique. (Rendre cela utile en pratique va dépendre d'une réalisation appropriée des sous ensembles de caractéristiques.)

Des exemples d'un tel usage sont donnés à la section 4 du présent mémoire.

Le présent mémoire ne spécifie pas comment un programme qui reçoit une abréviation d'ensemble de caractéristiques devrait découvrir l'expression d'ensemble de caractéristiques correspondant : voir le paragraphe 3.2.

3. Représentation de caractéristique composite

La présente spécification repose sur deux idées centrales :

- l'utilisation de prédicats auxiliaires (introduits dans la RFC2533 [1]) pour former la base d'un identifiant d'un ensemble de caractéristiques,
- l'utilisation d'un jeton fondé sur une fonction de hachage calculé sur l'expression de l'ensemble de caractéristiques référencé.

Une raison clé d'utiliser une fonction de hachage pour générer un identifiant est de définir un espace de nom global sans exiger d'autorité de dénomination centrale. De nouvelles étiquettes d'ensemble de caractéristiques peuvent être introduites par toute partie qui suit les règles de formulation appropriées, sans référence à une autorité centralisée.

Des services de résolution locaux peuvent être nécessaires pour transposer les étiquettes d'ensemble de caractéristiques en leurs expressions correspondantes d'ensembles de caractéristiques, mais ils ne sont à même de changer la signification d'aucune d'entre elles. Un service de résolution qui manquerait à retourner l'expression correcte sera détecté par une application appelante, qui devrait rejeter toute valeur incorrecte fournie.

Note : Lorsque on utilise une référence à un ensemble de caractéristiques, sa signification est définie par la substitution de l'expression de caractéristique référencée à l'expression de référence. Lorsque toutes les références ont ainsi été remplacées, le résultat est interprété comme une expression de caractéristique normale.

En particulier, si la référence à une expression de caractéristique contient des étiquettes de caractéristiques qui sont aussi contraintes par l'expression de référence, les contraintes sont interprétées selon la RFC2533 [1], sans considération de leur origine. Par exemple, (en utilisant la notation introduite ci-dessous) :

(& (pix-x=100) (pix-y≤300) (h.SBB5REAOMHC09CP2GM4V07PQP0))

où (h.SBB5REAOMHC09CP2GM4V07PQP0) se résout en :

(& (pix-x≤200) (pix-y≤150))

donne un résultat équivalent à :

(& (pix-x=100) (pix-y≤150))

3.1 Format de référence hachée d'ensemble de caractéristiques

La présente spécification introduit une forme particulière de nom de prédicat auxiliaire avec la syntaxe suivante :

```
fname          = "h." 1*BASE32DIGIT
BASE32DIGIT    = DIGIT
```

/ "A" / "B" / "C" / "D" / "E" / "F" / "G" / "H"
/ "I" / "J" / "K" / "L" / "M" / "N" / "O" / "P"
/ "Q" / "R" / "S" / "T" / "U" / "V"

La séquence de chiffres en base-32 représente la valeur d'une fonction de hachage calculée sur l'expression d'ensemble de caractéristiques correspondante (voir aux paragraphes suivants). Noter que la syntaxe ci-dessus permet les lettres majuscules ou minuscules pour les chiffres en base-32 (selon la RFC2234 [3]).

Donc, au sein d'une expression d'ensemble de caractéristiques, une référence d'ensemble de caractéristiques hachée aura la forme suivante :

(h.123456789abcdefghijklmnopq)

3.1.1 Calcul de la valeur du hachage

La valeur du hachage est calculée en utilisant l'algorithme MD5 [6] sur le texte de l'expression d'ensemble de caractéristiques référencée sous réserve de certaines normalisations. L'expression de la caractéristique doit se conformer à la syntaxe donnés pour "filter" dans la RFC2533 [1]:

filter = "(" filtercomp ")" *(";" parameter)

Les étapes du calcul d'une valeur de hachage sont :

1. Normalisation des espaces : toutes les espaces, CR, LF, TAB et tous les autres caractères de contrôle de disposition qui peuvent être incorporés dans la chaîne d'expression de la caractéristique, autres que ceux contenus entre des chaîne entre guillemets sont retirés (ou ignorés pour les besoins du calcul de la valeur du hachage).

2. Normalisation de la casse : toutes les lettres en minuscules dans l'expression de caractéristique, autres que celles contenues dans des chaînes entre guillemets, sont converties en majuscules. C'est à dire que les caractères qui ne sont pas entre des guillements avec les valeurs US-ASCII de 97 à 122 (en décimal) sont changé en les caractères correspondants dans la gamme de 65 à 90.

3. Clacul du hachage : l'algorithme MD5, décrit dans la RFC1321 [6], est appliqué à la chaîne d'expression de caractéristiques normalisée (représentée par une séquence d'octets contenant des codes de caractères US-ASCII ; voir aussi la section 5).

Le résultat obtenu à l'étape 3 est une valeur de 128 bits (16 octets) qui est convertie en une représentation en base-32 pour former la référence de l'ensemble de caractéristiques.

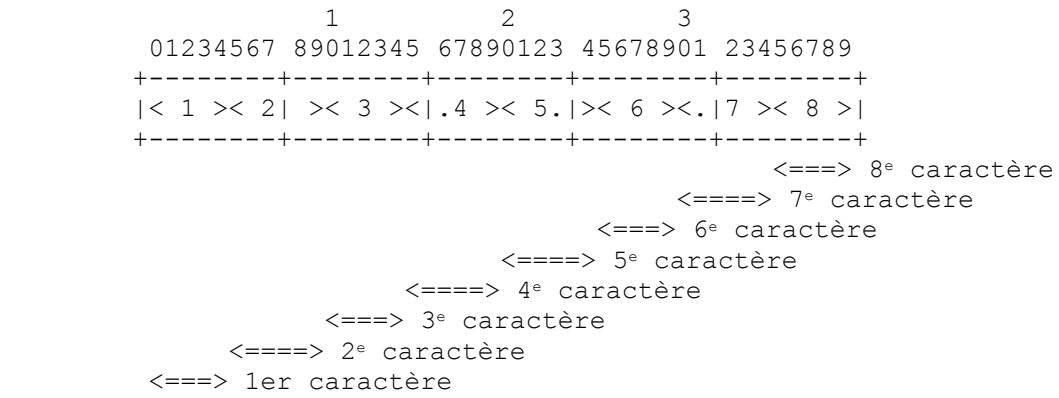
Note : Dans certaines circonstances, le retrait de TOUTES les espaces peut résulter en une chaîne invalide d'expression de caractéristiques. Cela ne devrait pas être un problème car cela n'est dait que pour les besoins du calcul de la valeur du hachage, et des expressions significativement différentes de caractéristiques sont supposées différer d'autres façons que par les espaces.

Note : La normalisation de la casse est réputée appropriee car la confontation des étiquettes et jetons de caractéristiques est insensible à la casse.

3.1.2 Représentation de valeur en base 32

La RFC1321 [6] décrit comment calculer une valeur de hachage MD5 qui est une séquence de 16 octets. Elle doit alors être codée comme une valeur en base-32, qui est une séquence de caactères de chiffres en base-32.

Chaque caractére successifs dans une valeur en base-32 représente cinq bits successif de la séquence d'octets sous-jacente. Donc, chaque groupe de huit caractères représente une séquence de cinq octets (40 bits):



La valeur (c'est-à-dire, la séquence de bits) représentée par chaque caractère de chiffre en base-32, est indiquée par le tableau suivant :

"0" 0	"A" 10	"K" 20	"U" 30
"1" 1	"B" 11	"L" 21	"V" 31
"2" 2	"C" 12	"M" 22	
"3" 3	"D" 13	"N" 23	
"4" 4	"E" 14	"O" 24	
"5" 5	"F" 15	"P" 25	
"6" 6	"G" 16	"Q" 26	
"7" 7	"H" 17	"R" 27	
"8" 8	"I" 18	"S" 28	
"9" 9	"J" 19	"T" 29	

Lors du codage d'une valeur en base-32, chaque groupe complet de 5 octets est représenté par une séquence de 8 caractères indiquée ci-dessus. Si un groupe de moins de 5 octets reste après cela, ils sont codés en utilisant autant de caractères supplémentaires que nécessaire: 1, 2, 3 ou 4 octets sont codés respectivement par 2, 4, 5 ou 7 caractères. Tous les bits en trop représentés par des caractères de chiffre en base-32 sont mis à zéro.

Lors du décodage d'une valeur en base-32, la transposition inverse est appliquée : chaqque groupe complet de 8 caractères code une séquence de 5 octets. Un groupe final de 2, 4, 5 ou 7 caractères code respectivement une séquence de 1, 2, 3 ou 4 octets. Tous les bits en trop représentés par le groupe final de caractères sont éliminés.

Donc, pour une valeur de hachage MD5 de 128 bits (16 octets) les quinze premiers octets sont codés comme 24 caractères de chiffres de base-32, et l'octet final est codé par deux caractères.

Note : La représentation en Base64 (selon MIME [4]) serait plus compacte (21 caractères plutôt que 26 pour la valeur de hachage MD5 de 128 bits) mais un nom de prédicat auxiliaire est défini (par [1]) comme ayant la même syntaxe qu'une étiquette de caractéristique, et les règles de correspondance d'étiquette de caractéristiques (selon [2]) déclarent que la correspondance d'étiquette de caractéristique est insensible à la casse.
La représentation en Base36 a été examinée (c'est-à-dire, en utilisant toutes les lettres "A"-"Z") mais n'a pas été utilisée parce que cela exigerait des opérations de multiplication et division d'une précision élevée pour coder et décoder les valeurs de hachage.

3.2 Résolution des identifiants d'ensemble de caractéristique

Le présent mémoire ne rend obligatoire aucun mécanisme particulier pour déréférencer un identifiant d'ensemble de caractéristiques. Il est estimé que des mécanismes spécifiques de déréférencement seront spécifiés pour toute application ou protocole qui les utilisera.

Les paragraphes qui suivent décrivent des façons dont les informations de déréférencement d'ensemble de caractéristiques peuvent être incorporées dans une expression d'ensemble de caractéristique. Elles se fondent sur des définitions de prédicats auxiliaires au sein d'une clause "où" [1].

Lorsque une référence d'ensemble de caractéristiques hachée est utilisée, la conformité aux règles de hachage prend le pas sur toute autre détermination de l'expression de caractéristique. Toute expression obtenue ne peut cependant pas être substituée à la référence fondée sur le hachage sauf si elle donne la valeur correcte du hachage.

3.2.1 Protocole d'interrogation

Un protocole qui fournit des interrogations de type question/réponse (par exemple, HTTP, LDAP, etc.) peut être établi pour fournir un service de résolution.

Donc, une interrogation à un serveur, associée à cette capacité pourrait être effectuée sur un identifiant d'ensemble de caractéristiques. La réponse retournée serait une expression CONNEG ; par exemple,

```
    (h.SBB5REAOMHC09CP2GM4V07PQP0)
où
    (h.SBB5REAOMHC09CP2GM4V07PQP0) :- (& (pix-x≤200) (pix-y≤150) )
fin
```

ou juste :

```
    (& (pix-x≤200) (pix-y≤150) )
```

Ce résultat serait combiné avec l'expression originale pour obtenir un résultat n'incluant pas le prédicat fondé sur le hachage.

Ce processus pourait être encore amélioré en utilisant les mécanismes de résolution d'URN (par exemple, ., DNS NAPTR [10]) pour découvrir le protocole et le serveur de résolution.

3.2.2 Détails d'ensemble de caractéristiques en ligne

Dans ce cas, une référence est résoute en incluant sa définition en ligne dans une expression.

L'expression d'ensemble de caractéristiques associée à une valeur de référence peut être spécifiée directement dans une clause "où", en utilisant la syntaxe de définition de prédicat auxiliaire [1] ; par exemple,

```
    (& (dpi=100) (h.SBB5REAOMHC09CP2GM4V07PQP0) )
où
    (h.SBB5REAOMHC09CP2GM4V07PQP0) :- (& (pix-x≤200) (pix-y≤150) )
fin
```

Cette forme pourrait être utilisée sur demande (et le mécanisme de demande est alors défini par le protocole d'application invocateur) ou lorsque le générateur estime que le receveur pourrait ne pas comprendre la référence.

C'est une erreur que l'expression de caractéristique en ligne ne donne pas la valeur du hachage contenue dans le nom du prédicat auxiliaire.

Note : Vue isolément, la valeur de ce format ne semble pas évidente, en ce que la forme (h.xxx) du prédicat auxiliaire pourrait être remplacée par n'importe quel nom arbitraire. On prévoit que cette forme pourrait être utilisée comme réponse passe-partout dans une séquence le long de lignes de :

```
A> Les capacités sont :
    (& (dpi=100) (h.SBB5REAOMHC09CP2GM4V07PQP0) )
B> Je ne comprend pas :
    (h.SBB5REAOMHC09CP2GM4V07PQP0)
A> Les capacités sont :
    (& (dpi=100) (h.SBB5REAOMHC09CP2GM4V07PQP0) )
  où
    (h.SBB5REAOMHC09CP2GM4V07PQP0) :- (& (pix-x≤200) (pix-y≤150) )
  fin
```

4. Exemples

Les exemples d'expressions d'ensembles de caractéristique ci-après contiennent des références d'ensembles de caractéristiques :

```
    (& (dpi=100) (h.SBB5REAOMHC09CP2GM4V07PQP0) )

    (& (dpi=100) (h.SBB5REAOMHC09CP2GM4V07PQP0) )
```

```
où
   (h.SBB5REAOMHC09CP2GM4V07PQP0) :-
      (& (pix-x<=200) (pix-y<=150) )
fin

   (h.QGEOPMCF02P09QC016CEPU22FO)
où
   (h.QGEOPMCF02P09QC016CEPU22FO) :-
      (| (& (ua-media=continuous) (dpi=200) (dpi-xyratio=200/100)
      (color=Binary) (paper-size=B4) (image-coding=MH) )
   (& (ua-media=continuous) (dpi=200) (dpi-xyratio=200/100)
      (color=Binary) (paper-size=B4) (image-coding=MR) )
   (& (ua-media=stationery) (dpi=300) (dpi-xyratio=1)
      (color=Binary) (paper-size=A4) (image-coding=JBIG) )
   (& (ua-media=transparency) (dpi=300) (dpi-xyratio=1)
      (color=Binary) (paper-size=A4) (image-coding=JBIG) ) )
fin
```

Les exemples suivants se fondent sur les travaux sur la télécopie Internet, et montrent comment un hachage de caractéristiques pourrait être utilisé pour exprimer les caractéristiques les plus courantes. Une forme du système de télécopie Internet dont on s'attend à ce qu'elle soit assez commune est ce qu'on appelle le système en "mode simple", dont les capacités sont déscrites par l'expression de caractéristiques suivante :

```
(& (image-file-structure=TIFF-minimal)
   (MRC-mode=0)
   (color=Binary)
   (image-coding=MH) (MRC-mode=0)
   (| (& (dpi=204) (dpi-xyratio=[204/98,204/196]) )
      (& (dpi=200) (dpi-xyratio=[200/100,1]) ) )
   (size-x<=2150/254)
   (paper-size=A4)
   (ua-media=stationery) )
```

Cela pourrait être exprimé par l'identifiant d'ensemble de caractéristiques fondé sur le hachage suivant :

```
   (h.MSB955PVIRT1QOHET9AJT5JM3O)
```

L'exemple suivant décrit les capacités d'un système de télécopie couleur Internet. Noter qu'un certain nombre de valeurs de caractéristiques sont applicables en commun avec '(color=grey)' et '(color=full)':

```
(& (image-file-structure=TIFF)
   (MRC-mode=0)
   (| (& (color=Binary)
      (image-coding=[MH,MR,MMR])
      (| (& (dpi=204) (dpi-xyratio=[204/98,204/196]) )
         (& (dpi=200) (dpi-xyratio=[200/100,1]) )
         (& (dpi=300) (dpi-xyratio=1) ) ) )
   (& (color=grey)
      (image-coding=JPEG)
      (image-coding-constraint=JPEG-T4E)
      (color-levels<=256)
      (color-space=CIELAB)
      (color-illuminant=D50)
      (CIELAB-L-min>=0)
      (CIELAB-L-max<=100)
      (dpi=[100,200,300]) (dpi-xyratio=1) )
   (& (color=full)
      (image-coding=JPEG)
      (image-coding-constraint=JPEG-T4E)
      (color-subsampling=["1:1:1","4:1:1"])
      (color-levels<=16777216)
      (color-space=CIELAB)
      (color-illuminant=D50)
      (CIELAB-L-min>=0)
```

```
        (CIELAB-L-max<=100)
        (CIELAB-a-min>=-85)
        (CIELAB-a-max<=85)
        (CIELAB-b-min>=-75)
        (CIELAB-b-max<=125)
        (dpi=[100,200,300]) (dpi-xyratio=1) ) )
    (size-x<=2150/254)
    (paper-size=[letter,A4,B4]) )
    (ua-media=stationery) )
```

Séparer les capacités communes donne :

```
(& (image-file-structure=TIFF)
    (MRC-mode=0)
    (| (& (color=Binary)
        (image-coding=[MH,MR,MMR])
        (| (& (dpi=204) (dpi-xyratio=[204/98,204/196]) )
            (& (dpi=200) (dpi-xyratio=[200/100,1]) )
            (& (dpi=300) (dpi-xyratio=1) ) ) )
        (& (color=grey)
            (color-levels<=256)
            (h.QVSEM8V2LMJ8VOR7V682J7079O) )
        (& (color=full)
            (color-subsampling=["1:1:1","4:1:1"])
            (color-levels<=16777216)
            (CIELAB-a-min>=-85)
            (CIELAB-a-max<=85)
            (CIELAB-b-min>=-75)
            (CIELAB-b-max<=125)
            (h.QVSEM8V2LMJ8VOR7V682J7079O) ) )
    (size-x<=2150/254)
    (paper-size=[letter,A4,B4]) )
    (ua-media=stationery) )
où
(h.QVSEM8V2LMJ8VOR7V682J7079O) :-
    (& (image-coding=JPEG)
        (image-coding-constraint=JPEG-T4E)
        (color-space=CIELAB)
        (color-illuminant=D50)
        (CIELAB-L-min>=0)
        (CIELAB-L-max<=100)
        (dpi=[100,200,300]) (dpi-xyratio=1) )
fin
```

5. Considérations d'internationalisation

Les expressions d'ensembles de caractéristiques et les chaînes d'URI sont actuellement définies pour ne comporter que des caractères tirés du répertoire US-ASCII [1], [5] ; dans ces circonstances, la présente spécification n'est pas impactée par les considérations d'internationalisation (autres que celles déjà applicables aux URI [5]).

Mais, si de futures révisions de la syntaxe d'ensemble de caractéristiques permettent des caractères non US-ASCII (par exemple, entre guillemets) une représentation canonique devra alors être définie pour les besoins du calcul des valeurs de hachage. Un choix pourrait être d'utiliser une représentation équivalente à UTF-8 comme base du calcul du hachage d'ensemble de caractéristiques. Un autre choix pourrait être de laisser cela à la responsabilité du protocole d'application (mais cela pourrait conduire à des ensembles de caractéristiques non interopérables entre des protocoles différents).

Une autre solution concevable est de mettre en majuscules l'expression de caractéristiques en préparation pour le calcul d'une valeur de hachage. Cela ne s'applique pas au contenu des chaînes de sorte qu'il est vraisemblable que cela ne pose pas de probème. Mais si des changements interviennent qui permettent des caractères non US-ASCII dans les étiquettes de caractères ou dans les chaînes de jetons, il faudra veiller à définir comment effectuer correctement la conversion de casse.

6. Considérations pour la sécurité

Pour la plus grande partie, les considérations de sécurité sont les mêmes que celles qui s'appliquent en général à l'identification de capacités dans [1], [2], [9].

Une considération qui pourrait éventuellement s'ajouter est que l'utilisation d'un identifiant d'ensemble de caractéristiques spécifique peut révéler plus d'informations sur un système que ce qui est nécessaire pour une simple transaction.

7. Remerciements

Nos idées ont été précisées par les discussions préalables avec Martin Duerst, Al Gilman et Ted Hardie. D'utiles suggestions d'amélioration ont été fournies par Maurizio Codogno.

8. Références

[1] G. Klyne, "Syntaxe de description des ensembles de caractéristiques des supports", RFC2533, mars 1999. (*MàJ par* RFC2738, RFC2938) *(P.S.)*

[2] K. Holtman, A. Mutz, T. Hardie, "Procédure d'enregistrement d'étiquette de caractéristique de support", RFC2506, mars 1999. (BCP0031)

[3] J. Luciani et autres, "Protocole de synchronisation d'antémémoire de serveur (SCSP) - NBMA", RFC2334, avril 1998.

[4] N. Freed et N. Borenstein, "Extensions de messagerie Internet multi-objets (MIME) Partie 1 : Format des corps de message Internet", RFC2045, novembre 1996. *(D. S., MàJ par 2184, 2231, 5335.)*

[5] T. Berners-Lee, R. Fielding et L. Masinter, "Identifiants de ressource uniformes (URI) : Syntaxe générique", RFC2396, août 1998. (*Obsolète, voir* RFC3986)

[6] R. Rivest, "Algorithme de résumé de message MD5", RFC1321, avril 1992. *(Information)*

[7] "Applied Cryptography" Bruce Schneier John Wiley and Sons, 1996 (seconde édition) ISBN 0-471-12845-7 (fichier) ISBN 0-471-11709-9 (papier)

[8] G. Klyne, "Cadre de négociation de contenu indépendant du protocole", RFC2703, septembre 1999. *(Information)*

[9] "Numerical Recipes" William H Press, Brian P Flannery, Saul A Teukolski and William T Vetterling, Cambridge University Press (1986) ISBN 0 521 30811 9 (L'approximation de la fonction Gamma est présentée au chapitre 6 "Special Functions". Il y a eu plusieurs éditions ultérieures de ce livre, de sorte que la référence du chapitre peut changer.)

[10] R. Daniel, M. Mealling, "Résolution des identifiants de ressource uniformes avec le système des noms de domaines", RFC2168, juin 1997. (*Obsolète, voir* RFC3401, RFC3402, RFC3403, RFC3404) (*MàJ par* RFC2915) *(Expérimentale)*

[11] Code source Java de l'algorithme de confrontation d'ensembles de caractéristiques, avec option de calcul de hachage d'ensemble de caractéristiques. Lien sur <http://www.imc.org/ietf-medfree/>

9. Adresse des auteurs

Graham Klyne
Content Technologies Ltd.
1220 Parkview,
Arlington Business Park
Theale
Reading, RG7 4SA
United Kingdom
téléphone : +44 118 930 1300
Fax : +44 118 930 1301
mél : GK@ACM.ORG

Larry Masinter
AT&T Labs
75 Willow Road
Menlo Park, CA 94025
USA
téléphone : +1-650-463-7059
mél : LMM@acm.org
http://larry.masinter.net

10. Appendice A Le paradoxe de l'anniversaire

Note : Toute cette section constitue un commentaire, et n'affecte en aucune façon la spécification des références d'ensemble de caractéristiques.

L'utilisation d'une valeur de hachage pour représenter un ensemble de caractéristiques arbitraire se fonde sur l'hypothèse qu'il n'y a pas deux ensembles distincts de caractéristiques qui vont donner la même valeur de hachage.

Il y a une faible possibilité, mais non nulle, que deux differents ensembles de caractéristiques donnent la même valeur de hachage.

On suppose que la fonction de hachage de 128 bits répartit les valeurs de hachage pour les ensembles de caractéristiques, même ceux qui ont de très petites différences, de façon aléatoire et également à travers la gamme des 2^128 (approximativement 3*10^38) valeurs possibles. C'est une propriété fondamentale d'un bon algorithme de résumé comme MD5. Donc, les chances que deux expressions distinctes d'ensemble de caractéristiques donnent le même hachage sont inférieures à 10^38. C'est négligeable en comparaison de, par exemple, la probabilité qu'un système receveur subisse une défaillance après avoir reçu des données conformes à un ensemble négocié de caractéristiques.

Mais lorsque le nombre d'ensembles distincts de caractéristiques en circulation augmente, la probabilité de répéter une valeur de hachage augmente de façon surprenante. Ce fait est illustré par le "paradoxe de l'anniversaire" : étant donné une collection aléatoire de seulement 23 personnes, il y a plus d'une chance sur deux qu'il existe une paire de personnes qui ont le même jour anniversaire. Ce sujet est exposé plus en détails aux paragraphes 7.4 et 7.5 de l'ouvrage de Bruce Schneier "Applied Cryptography" [7].

Le tableau ci-dessous montre les probabilités du "paradoxe de l'anniversaire" qu'au moins une paire d'ensembles de caractéristiques ait la même valeur de hachage pour différents nombres d'ensembles de caractéristiques utilisés.

Nombre d'ensembles de caractéristiques utilisés	Probabilité de deux ensembles avec la même valeur de hachage
1	
2	E-39
10	E-37
1E3	E-33
1E6	E-27
1E9	E-21
1E12	E-15
1E15	E-9
1E18	E-3

Les calculs de probabilités ci-dessus sont approximatifs, étant effectués à l'aide de logarithmes d'approximation d'une fonction Gamma par Lanczos [9]. La formule de probabilité est P=1-(m!/((m-n)! m^n)), où "m" est le nombre total de valeurs de hachage possibles (2^128) et "n" est le nombre d'ensembles de caractéristiques utilisés.

Si les expressions originales d'ensembles de caractéristiques sont générées manuellement, ou seulement en réponse à un processus contraint manuellement, le nombre total d'ensembles de caractéristiques en circulation va vraisemblablement rester très faible par rapport au nombre total de valeurs de hachage possibles.

Le résultat de tout cela est que en supposant que les ensembles de caractéristiques sont générés manuellement, même en prenant en compte l'effet du paradoxe de l'anniversaire, la probabilité d'identifier de façon incorrecte un ensemble de caractéristiques en utilisant une valeur de hachage est toujours négligeable comparée aux autres modes de défaillance possibles.

11. Déclaration de droits de reproduction

Remerciement
Le financement de la fonction d'édition des RFC est actuellement fourni par l'Internet Society.