新月プロトコル/1.0 ドラフト#1

新月プロジェクト

2005 年 7 月 11 日

1　はじめに

この文書は P2P 匿名掲示板-新月-のノード間のプロトコルについて述べる。新月の匿名掲示板ネットワークはノードにより構成される。ノード間でメッセージを交換して、相互に通信が行えるように、プロトコルが規定されている必要がある。新月のネットワークのノード間で交換されるメッセージは新月プロトコルに従う。この文書で述べる新月プトロコルのバージョンは 1.0 である。

2　目的

この文書の目的はノードが相互にメッセージを交換できるように、プロトコルを規定することである。そして、新月プロトコルの実装アプリケーションの開発ができる新月プロトコルの文書を提供することである。

3　新月プロトコルの歴史

新月プロトコル/1.0 が作成される以前も、新月プロトコルの文書は公開されていた。しかし、当時の文書は、すでに存在していた実装の shinGETsu の補足的な資料でしかなかった。新月の名がインターネット上でに広がるにつれて、新月プロトコルを元に実装された、様々な新月の実装が登場し始めた。各実装は一見、相互にネゴシエーションが行えていたと思えた。しかし、実装者ごとに新月プロトコルの解釈が異なっていたため、新月プロトコルを満たしきれていない実装が現れた。また、署名のシステムは文書化されていたものの不透明な点が数多くあった。そこで、これまで作成されてきた資料を元に、包括

的な新月プロトコルの文書を作成する必要があった。

4　新月のネットワーク

4.1　新月ネットワークの構成

新月のネットワークはノード間の接続によって構成される。図 1 は新月のネットワーク構成の概念図である。図中の点は、新月のネットワークのノードである。各々のノードはネットワーク上で一意な名前を持つ。この一意な名前をノード名という。

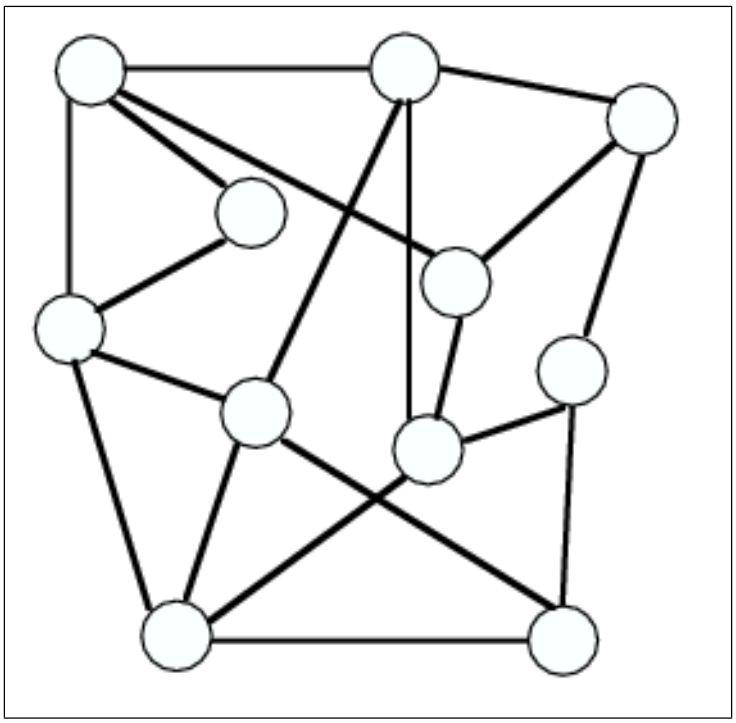

図 1　新月のネットワーク構成

ネットワーク上の各々ノードは、他のノードと通信を行うために他のノード名を保持する。ノード名を用いることにより、ネットワーク上のノードを一意に識別することができる。しかし、ノード間のノード名の相互の保持は保証しなくてよい。一方のノード A がもう一方のノード B の名前を保持していても、もう一方のノード B は、必ずしもノード A のノード名を保持する必要はない。

ノード間の通信は HTTP/1.1[1] の GET メソッドをベースに行う。ノード間のリクエスト・レスポンスのメッセージは GET メソッドの規約に従わなければならない。また、

ノード間のリクエスト・レスポンスのメッセージの文字エンコーディングは UTF-8 でなければならない。

新月のノード間の通信コマンドは、最終的に URL[2] として構築される。したがって、新月のノード間のコマンドは URL の仕様に従わなければならない。

4.2 ノード

新月プロトコルを実装しているプログラムをノードという。新月のネットワークに接続するプログラムは新月プロトコルを正確に実装しなければならない。プログラムの新月プロトコルの実装にあたっては、HTTP、URL の仕様も満たさなければならない。ノードはネットワーク上で一意の名前を持つ。これをノード名 (4.1) といった。ノード名の構成は URL の仕様に従わなければならない。

新月プロトコルは、ノード名の構成要素であるポート番号を指定しない。各々のノードはノード名に使用するポート番号を自由に設定することができる。ただし、ゲートウェイで公開する場合を除いて、んー土が使用する 1024 番より小さい数字のポート番号は、ウェルノウンポート番号に従う [3] べきである。

5 ノード間のプロトコルコマンド

ノード間の通信はプロトコルコマンドを交換することによって行う。プロトコルコマンドは URL のパス部分に、解釈可能な位置に付加する。次に、ノード間で通信を確立するためのプロトコルコマンドを示す。

/ping
: ノードは「PONG\n 相手ノードの IP アドレス」を返す。

/node
: ノードは接続しているノードから 1 つ選択し、そのノード名を返す。

/join/ノード名
: ノードは、相手ノードが指定するノード名が有効であることを /ping によって確かめ、自ノードのノード保持リストに加え、「WELCOME」または「WELCOME\n 別のノードのノード名」を返す。 相手ノードが指定するノード名が有効でないときはそれ以外のものを返す。 相手ノードのノード名は、接続しようとしている相手

ノード自身を示すものでなければならない。 相手ノードのノード名のホスト名は、省略することができる。 ノード名のパスは/を + に置き換えたものとする。 相手ノードにはノード保持リストに加えてもらい、レスポンスで指定したノードに接続することを期待する。

/bye/ノード名

ノードは相手ノードが指定するノード名が有効であることを /ping によって確かめ、自ノードのノード保持リストから削除し、 「BYEBYE」を返す。

/have/ファイル名

ファイル名のファイルを持っていれば「YES」、そうでなければ「NO」を返す。

/get/ファイル名/時刻引数

ファイル名で指定したファイルの、時刻引数を満たすレコードを返す。 時刻引数は次のうちどれかである。

「時刻」　指定した時刻のレコード
「-時刻」　指定した時刻以前のレコードすべて
「時刻-」　指定した時刻以降のレコードすべて
「時刻-時刻」　指定した時刻の間のレコードすべて
「時刻/識別子」　指定した時刻と識別子のレコード

/head/ファイル名/時刻引数

/get と同様のレスポンスである。しかし、各レコードの時刻と識別子のみを返す。

/have/ファイル名

ファイル名のファイルを持っていれば「YES」、そうでなければ「NO」を返す。

/update/ファイル名/時刻/識別子/ノード名

ノードにファイルが更新されたことを知らせる。 もしすでにファイルの更新に関する処理しているのなら何もしない。 自分の持っているファイルならば、それを更新したのち、 ノード名を自ノードのノード名に書き換えて周りのノードにも通知する。 そうでなければそのまま、送信されてきたプロトコルコマンドを接続して

いるノードに転送する。 ノード名のパスは/を + に置き換えたものとする。 /join と同様にホスト名を省略できる。

/

ノード固有のメッセージを表示する。 通信には使わないので、何を出力してもよい。

6 ファイル

6.1 ファイルの定義

新月プロトコルにおけるファイルとは、GET メソッドのレスポンスのメッセージボディのことを指す。ファイルはレコードの集合である。/ping、/have コマンドのレスポンスの内容もファイルが交換していると考える必要がある。

6.2 ファイル名

ファイル名は、半角英数字とアンダースコア (_) のみで構成する。ファイル名は prefix_basename という形式である。 prefix、basename に使えるのは半角英数字のみである。prefix はファイルの種類を表す。basename は、文字コードの UTF-8 で表現されるファイル名を、16 進数表現に変換した文字列である。16 進数表現に用いることができる文字は半角英字の A-F、半角数字のみである。

6.3 レコード

レコードとはファイルの 1 行のことである。レコードの形式は、次のように定義する。

タイムスタンプ<>識別子<>本文」

識別子は本文の MD5 値である。レコードの要素は、<>で区切る。タイムスタンプは、新月プロトコルで定義する基準時刻 (8) から経過した整数値の秒数を表す。

6.4 本文の書式

本文を構成する要素は「<>」で区切る。 本文は複数の名前付きフィールドで構成される。名前付きフィールドの形式は「名前:値」である。名前付きフィールドの名前に使えるのは半角英数字とアンダースコア (_) のみである。名前付きフィールドの名前は重複してはならない。 名前付きフィールドの stamp, id は予約されており、それぞれタイムスタンプ、識別子を表わす。本文での名前付きフィールドの出現順番は問わない。

名前付きフィールドのフィールドの値には「<」「>」が含まれてはならない。 例外として「<文字列>」(タグと呼ぶ) を含むことができる (文字列は長さが 1 文字以上であって、「<」「>」が含まれてはならない)。 どのようなタグが使えるのかはアプリケーションが定める。

7 プラグインとアプリケーション

ユーザに提供する機能をプラグインと呼ぶ。プラグインのうち、ファイルの書式を定義するプラグインを アプリケーションと呼ぶ。現在、アプリケーションには list、thread、note の 3 種類ある。

8 時刻

新月プロトコルで用いる時刻の基点は「1970 年 1 月 1 日午前 0 時」である。この時刻はグリニッジ標準時である。ある時刻は、基点時刻を基準として表される整数値の秒で表す。時間の単位は秒である。

9 ブラケットリンク

アプリケーションは、ファイルの内容をなんらかのデータ形式に変換し、結果を出力する場合において、本文にブラケットリンクがあるときは、文字列が指定するリンクを作成しなければならない。

アプリケーション type が存在するとする。標準的なブラケットリンクの形式は [[/type/文字列]] である。文字列の形式はアプリケーションが定義する。つまり、文字列の解釈はアプリケーションに委ねられる。/type は省略することができる。/type を省略したブラケットリンクの形式である [[文字列]] は、本文が記録されているファイル形

式の type が省略されていると考える。

現在、アプリケーションとして定義されている thread、note のブラケットリンクの作成例を次に示す。

[[すれっど]]
: 本文が記録されているファイル形式が thread なら、thread の「すれっど」へリンクする。

[[/list/りすと]]
: list の「りすと」へリンクする。

[[すれっど/7889e7db]]
: 本文が記録されているファイル形式が thread なら、thread「すれっど」の識別子「7889e7db」を持つレコードへリンクする。

[[/thread/すれっど/7889e7db]]
: thread「すれっど」の識別子「7889e7db」を持つレコードへリンクする。

10 署名

署名の暗号システムは、公開鍵暗号方式に従う。暗号化のアルゴリズムは RSA である。メッセージダイジェストのアルゴリズムは MD5 である。

10.1 文字列と多倍長整数間の変換アルゴリズム

文字列と多倍長整数間は、文字列の最下位 6 ビットから Base64 アルゴリズムの変換テーブルに準拠して変換する。文字列の先頭文字が整数の最下位 6 ビットに対応し、文字列の最後尾が最上位ビットに対応する。このとき、想定される文字列の字数が足りない時は、数値の末尾のビットに「0」を補充する。つまり、数値から文字変換時に文字列の末尾に A を付加する。

10.2 鍵生成アルゴリズム

素数テストにはミラーテストを使用する。最初の 10 個の素数に対して、テストを通過した擬素数を素数とみなす。公開乗数 e は 65537 とする。トリップ生成文字列から、生成する素数を p、q とする。トリップ生成文字列$key とはクライアントから取得する署名文字列である。

ステップ 1 ハッシュの計算

トリップ生成文字列を hashs とする。ここでの + 演算子は文字列の連結演算を意味する。MD5 の計算アルゴリズムを md5() と定義する。

$$hashs = md5(\$key) + md5(\$key +' pad1') + md5(\$key +' pad2') + md5(\$key +' pad3') \quad (1)$$

ステップ 2 ハッシュ文字列から p、q への変換

hashs の 0 から 27 バイト、28 から 63 バイトをそれぞれ p、q に代入する。このとき、リトルエンディアンで処理する。つまり、hashs の最上位バイトが p(q) の最下位バイトに対応するように処理する。

p については、

$$p.num[0] - p.num[6] \fallingdotseq hashs[0-27] \text{ (28 バイト)} \quad (2)$$

である。q については、

$$q.num[0] - q.num[8] \fallingdotseq hashs[28-63] \text{ (36 バイト)} \quad (3)$$

である。

ステップ 3 p、q の前処理

ステップ 2 で求めた p、q に対して、次のふたつの演算を行う。 p が 216 ビットで表される小さな因数になるのを防ぐために (A) の演算をする。 q が 280 ビットで表される小さな因数になるのを防ぐために (B) の演算をする。これにより、公開鍵 n が 496 ビットで表される数値を下回ることを抑制する。

(A). p.num[6] の下から 24 ビット目 (216 ビット目、第 215 ビット) を 1 にする。
(B). q.num[8] の下から 24 ビット目 (280 ビット目、第 279 ビット) を 1 にする。

ステップ 4 p、q を RSA に適する素数への変換

q を q 以上で最小の擬素数とする。 p を p 以上で最小の擬素数とする。次の関係が成立する p、q を求める。(p - 1)(q - 1) と e が互いに素で、かつ

$$t = 0\text{x}7743 \tag{4}$$
$$de \equiv 1 \ (mod \ (p-1)(q-1)) \tag{5}$$
$$n = pq \tag{6}$$

上記 3 式の関係が成立するとする。上記 3 式が成立する t、d、n に対し

$$t^{ed} \equiv t \ (mod \ n) \tag{7}$$

上記の関係を満たす p、q が得られるまで、 $p = p + 2$、$q = q + 2$ として (5) 式から、(7) 式が成立するまで繰り返す。

ステップ 5 鍵の命名

ステップ 4 で生成した n を公開鍵、d を秘密鍵と称する。文字列に変換する際は上記変換則 (10.1) を用いて 86 文字とする。

10.3 署名対象文字列

署名対象文字列を次のように定義する。

署名対象文字列 = field1:value<>field2:value2<>...

このフィールド書式は新月プロトコルの本文の書式に従う。フィールドの出現順序は target フィールドで指定された順番に従う。

10.4 署名アルゴリズム

RSA 暗号化 $m^d \equiv c \ (mod \ n)$ に使用する m を求める。署名対象ハッシュ文字列をとする。署名対象ハッシュ文字列の長さは 64 バイトである。署名対象ハッシュ文字列 Mes は

次のようにして求めることができる。

Mes[0-63] = md5(署名対象文字列) = md5(filed1:value1<>field2:value2<>...)

署名対象ハッシュ文字列 Mes の長さが 64 バイトに満たない場合、Mes を数値へ変換する時に、Mes の空きバイトには 0 を仮定する。 64 バイトを超える場合、超えた部分は無視する。文字列から数値への対応は次式で定義する。

$$m.num[0-15] \div Mes[0-63] \text{ (64 バイト)} \tag{8}$$

ただし、リトルエンディアンで処理する。つまり、Mes の最上位バイトは m の最下位バイトである。

参考文献

[1] IEFT:Hypertext Transfer Protocol – HTTP/1.1,
http://www.ietf.org/rfc/rfc2616.txt

[2] IEFT:Uniform Resource Locators (URL),
http://www.ietf.org/rfc/rfc1738.txt

[3] IANA:WELL KNOWN PORT NUMBERS,
http://www.iana.org/assignments/port-numbers