



# (12) 发明专利申请

(10) 申请公布号 CN 115021762 A

(43) 申请公布日 2022. 09. 06

(21) 申请号 202210672376.1

(22) 申请日 2022.06.15

(71) 申请人 道和邦(广州)电子信息科技有限公司

地址 510440 广东省广州市白云区嘉禾街  
鹤龙二路96号粤旺大厦C栋305室

(72) 发明人 不公告发明人

(51) Int.Cl.

H03M 7/30 (2006.01)

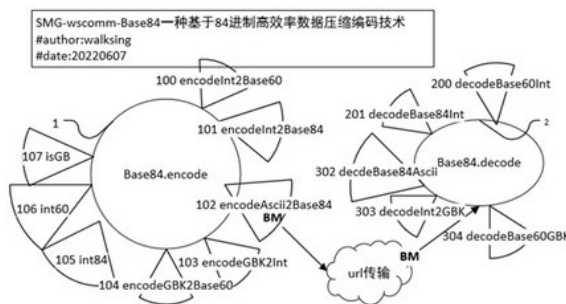
权利要求书3页 说明书5页 附图3页

## (54) 发明名称

一种基于84进制高效率数据压缩编码算法

## (57) 摘要

一种基于84进制高效率数据压缩编码算法：该算法以10进制数字为每俩位截取压缩为1位目标码，不足1位采用进位可变长编码，且对ascii字符编码支持混合编码，具有智能检测ASCII码中数字,GBK;经过测算其运行效率比传统Base64高约37倍;纯数字数据压缩率接近50%;本发明主要用于动态令牌加密编码传输;以及航天航空领域的大量观测大数进行压缩,结合幻数加密算法(参同期专利) MagicNumberEncrypt可实现高安全性,高可靠性信息保密压缩传输;高倍空间压缩编码,高效率加密解密技术为超算力输出,超分布式存储输出必放一异彩。



1. 一种基于10进制数字值为基数的编码:对其数值有选择的进行60进制, 84进制进行重新编码实施的技术方案,其特征如下:

encodeInt2Base84,encodeInt2Base60,encodeAscii2Base84

解码依次对应:decodeBase84Int,decodeBase60Int,decodeBase84Ascii

1.0 -控制符 1.1 位于最后一位代表奇数

1.2 位于除1.1则表示高进位:for encodeInt2Base84,解码dx+=84;

1.3 位于除1.1则表示低进位:for encodeAscii2Base84解码dx;

1.4.0 \*/控制符;

1.4.1 \*控制符 \*标识为整数数据如果以/结尾则标识>127的字符为GBK编码

xx为GBK编码5位数字/汉字,如果原gbk中文为奇数位,则自动填充\0' 填充偶数位;

1.4.2 \*xx.即ASCII数字 ,for encodeAscii2Base84 方案,当且仅当ascii码混合字母与数字,其数字必须连续满足大于5位时,该模块才启用encodeInt2Base60数字压缩技术,

1.4.1,1.4.2 编码自左起依次取2位,取整数值->dx,如果奇数末位依据1.1 补- ;

1.5.0 84进制参MT84码表规则(参表MT84码表),60进制参MT60表;

1.5.1 丢弃ASCII0-31 控制码;

1.5.2 设x为ascii 字符变量;dx=ASCII(x) 序号dx-=32;

1.5.3 低进位 变换dx<43则 buf -;buf MT84[dx];

1.5.4 非进位 变换dx-=43;buf MT84[dx];

1.6 url传输+ ->\*\* 服务端逆替换 空格 或\*\*->+再解码,=进行urldecode;

1.7 解码根据MT84与ascii 码表逆转换;

1.8 基于1.4.1,1.4.2 规则编码为encodeInt2Base60 ref MT60码表{0-9 A-Y,a-y}共60 位,z为base60 左进位,Z 结束控制符,后跟1位数字标识解码多余用于控制数据还原;

1.9 因base84 自动判断整数包括体内用base60编码,故虽然同一符号在不通编码规则不同,同时设计编码规避重复;Base84 含zZ./,但这些符号在base60不存在.故混合编码只是局部嵌入,不会打乱;

1.9.1 一个ascii2base84编码可容纳多个int2Base60码块,遇到不同块调不同解码程序进行解码;

1.10 GBK编码以5位数字,不足前补0,解码除掉前缀0,取ascii码;

1.11 ascii码只是变换位置不会减少编码位数,对int字符串会压缩编码;

1.12 当int码,GBK码长小于2,返回NoN 非数字;

1.13 本发明数字压缩编码精要,采用数字10进制,每俩位一组进行分隔进行编码,解码

1.13.1 基于1.13 俩位数字最大99 ,base60编码,其进位最大1以z代替。

2.基于Base84.encodeInt2Base84 ,实现的一种高压缩比的数字码压缩技术,其压缩率最大可接近50%,其特征如下:

2.1 基于1.0,1.5.1,1.5.2,1.5.3 ,可知单个ascii符号最多可表示俩位10进制数值范围[0,83],而大于83且<=99的概论只有16%,当相邻俩位10进制数大于83,经过编码用-

MT84[dx%84]标识,解码是z替换为84进制的高位,逆运算dx+=84即可还原信息,即这时用-加余码组成字母2位标识原2位10进制数字;也就是说只有16%的概率需要用可变长2位,其余84%的概率是1位字符代替俩位信息;在编码数据时,每2位数字通过换算大概率用1位字符标识,这是本发明以接近50%的压缩比基础;

2.2 其产生的编码字符,由MT84表字符决定,除+字符在url 转发需要转义\*\*,其它字符原封发送;当然在解码前需要进行urldecode 下较为可靠。

3. 一种基于1可实施的高效,高压缩比的时间数字订单压缩技术,其特征如下:

3.1 基于1.8 encodeInt2Base60,MT60表,可实现优雅符号密文输出,该技术方案主要用于密文需要存储在数据库,或对外接口服务,长数字订单就是例子;如果订单符号发送一批奇怪符号,大概率会被蹩脚程序员当作乱码删除,引发不必要的损失;该压缩率有60%的概率为1:2压缩,有40%的概率1:1不压缩;实际测算基于时间订单日期除年外均在60以内数字,其实该算法最早实现就是针对时间数字进行压缩的,这样概率提升到90%是1:2压缩,即50%的压缩;除时间数字外是60%的概率1:2压缩;

3.2 因为编码、解码过程尽可能一个循环流程下来操作;避免不必要的循环;故是高效率的。

4. 一种可智能检测数字可变长,可复合拼装包裹体的编码压缩技术,其特征如下:

4.1 基于1.4.0,1.4.1,1.4.2 可知,在encodeAscii2Base84过程,编码模块,会智能识别数字连续满超过5位才启用压缩数字技术,因为数字太短不足以提现其技术优势,比如ip地址不可以被解析压缩,只能按ascii码编码更快捷;但对应ascii字符串中长串数字该模块会智能启用压缩技术编入包裹体,其特征是密文含\*xxxx.,对GBK中文则以\*xxx/为包裹体,当有连续中文会合在一起解析直到结束。

5. 一种可用于url传递,可嵌入Cookie编码技术,其特征如下:

5.1 基于1.50,1.6,1.7,1.8可知编码表不含”’ & 等url敏感字符,其设计之初就是高效的url传递‘;’为cookie分隔符,’&’为url分隔符;’+’会被转义’\*\*’。

6. 基于4一种较Base64位编码更高效,编码速度更快的编码,其特征如下:

6.1实测基于4base84.encodeAscii2Base84 实例为用户真实属性数据明文数据1422,明文含中英文,符号数字混合字符:

Base84.encodeAscii2Base84算法对比Base64算法数据对比见下:

算法	明文长	1422	编码长	数据压缩率	编码耗时(纳秒)	base84:x效比	解码比
Base84	1508	106%	1080999	1:1	1:1		
X1:Base64	1928	136%	4205437	3.9:1	9.8:1		
X2:Hex	2888	203%	2523338	2.3:1	0.35:1		

据此测试数据不难发现,Base84数据压缩率103%最优,接近1:1(对ascii码不压缩,但如果ascii字符串含有大量连续的数字会压缩),效率是Base64的3.9倍,是Hex的2.3倍;无论从压缩率,还是压缩效率都是成倍提高,其中解码比base64提高9.8倍;

6.2 基于2 Base84.encodeInt2Base84

测试

sn= 01615201309015370011537001615201309015370016152013090153700161520130

明文长:68

base84:sn:B90Be-GP1AL1:B90Be-GP1AQPUNJB1:B90Be,36,99680 ns

base64:sn:MDE2MTUyMDEzMDkwMTUzNzAwMTE1MzcwMDE2MTUyMDEzMDkwMTUzNzAwMTYxNTIwMTMwOTAxNTM3MDAxNjE1MjAxMzA=,92,3728560 ns

算法对比Base64算法数据对比见下:

算法:明文长68	编码长	数据压缩率	编码耗时(纳秒)	base84:x效比	解码比
Base84	36	53%	99680	1:1	1:1
X1:Base64	92	135%	3728560	37:1	33:1
X2:Hex	136	200%	234852	2.4:1	0.64:1

据此测试数据不难发现,Base84对纯数字数据压缩率53%最优,接近50%,效率是Base64的37倍,是Hex的2.4倍;无论从压缩率,还是压缩效率都是成倍提高,其中解码是base64的33倍,(注:测试环境win7+jdk1.78,eclipse)。

## 一种基于84进制高效率数据压缩编码算法

### 技术领域

[0001] 本发明是为解决电子商务中对订单号压缩需求,传统订单号通常以时间为序列15位数字精巧到秒可杜绝网站多用户并发重单问题,故订单号不能太短,太短则大用户环境下会产生重复订单,使得业务逻辑溢出漏洞;用户在同一批次购物车可选择多家商家不同产品同时下单,每商家产生一个订单号,多批次订单号在一次购物结束进行清算会产生一个总单号进行提交,完成支付;此过程为方便查询数据有俩种设计方案:

方案A多批次商家订单放在一张表关联总支付订单号,

方案B 一个购物总订单号关联多商家订单中间逗号分隔;

A:一次购物可容纳商家的数量不限,但在经常查询订单会遍历查所购买所有支付商家的关系表,进行冗余遍历查询,效率低下.;

B:可容量商家数量字段空间有限字长不超过250,查询方便一次关联即可查处所有支付商家,最多可容纳 $250 / (15+1) = 15$ 个不同类商品订单;而基于本算法 $\text{encodeInt2Base60}/\text{encodeInt2Base84}$ 可实现订单号保留原信息同时,将数字长度压缩率接近50%.由此产生可容量订单数 $250 / (15/2+1+1) = 27$ 个不同类订单号存储,效率提升接近45%.同时免除数据库冗余遍历查询,在保障订单不重基础上,效率双倍提升;27个订单存储基本满足用户需求.订单号太短不行,太短容易引发并发冲突;太长亦不行,订单长限定:15位经过压缩算法保持数据唯一性,减少空间存储提升效率.

正是基于本需求需要研发高压缩比数字压缩编码技术,速度快,无损压缩,本发明为一种新型数据压缩编码技术,该压缩技术利用ascii 84个字符码,为输出码,以数字串每俩位截取转整数值为基数进行base60,base84编码压缩,理论数据压缩率(编码后数据长度与编码前数据长度比)极限可接近50%,Base84:Base64(开源)对比,编码效率提升29:1,解密效率提升:33:1,数据压缩率:50%:

本项发明,为SMG-VME可迭代分布式操作系统,SMG-VME-AFS可迭代分布式存储系统延伸的电子商务商业价值,继而引发大数字压缩编码压缩/解压缩技术,尤其适合于航天航空数据压缩传输,本发明设计之处就考虑适合网络间传输.

[0002] 本项发明,为SMG-VME-AFS可迭代分布式存储,下底层编码存储技术可替代base64编码,本发明包含订单类编码base60压缩/解压技术;ascii 编码,base84类编码。

[0003] 本项发明,本发明可以用于并行计算,分布式计算,实现超计算,超存储电子商务领域重要且广泛使用的底层数据编码技术,且目标是设计高效高倍数据压缩加密技术.

专利审查:

截至2022-06-06网络搜索暂无同案例

### 背景技术

[0004] 本发明主要应用在工业互联网,万物智能互联,分布式计算,万物互联数据高效率高倍信息传递,信息存储,其产生背景源于订单号的发生,严格控制在15位并采用高倍压缩编码技术转存数据,并用于订单支付,但其应用不限于此,可更多应用于航天航空大数据类

项压缩编码,提升效率.

## 附图说明

[0005] 图1是Base84组件功能图;

```

1Base84.encode
100 encodeInt2Base60
101 encodeInt2Base84
102 encodeAscii2Base84
103 encodeGBK2int
104 encodeGBK2Base60
105 int84
106 int60
107 isGB
2 Base84.decode 解码
200 decodeBase60Int
201 decodeBase84Int
202 decodeBase84Ascii
203 decodeInt2GBK
204 decodeBase60GBK

```

图2是encodeInt2Base60 编码;

调用逻辑

输入10001 S=intString->10002 判断奇数为真最后字符填100021 Z->10003  
for循环s 依次移动取2位数字转整数dx->1000301 判断dx>60则1000302 buf z标识进位  
1,解码需要加dx+60-> 1000303 dx%=60操作->buf MT60[dx];

->1000305 读取\*xxx. \*xxxZ将其存入mbuf ,此为包裹体不再编码->10003051  
buf mbuf ->输出buf编码结束

图3是encodeInt2Base84 编码;

调用逻辑

输入10101 S=intString->10102 判断奇数为真最后字符填101021 ‘-’->10103  
for循环s 依次移动取2位数字转整数dx->1010301 判断dx>84则1010302 buf ‘-’标识进  
位1,解码需要加dx+84-> 1010303 dx%=84操作->buf MT84[dx];

->1010305 读取\*xxx. \*xxxZ将其存入mbuf ,此为包裹体不再编码->10003051  
buf mbuf ->输出buf编码结束;

图4 是MT84 ascii字符表;

图5是MT60 ascii字符表;

图6是encodeAscii2Base84流程图;

调用逻辑:

102 ascii2Base84编码->1021 识别数字连续且大于5位,否则按1023ascii字符  
码处理->10211 编码后嵌入\*xx.包裹体->1022识别GBK码即,dx>127的字符码 ->1023 识

别ascii字符码->10231 取ascii数字码dx并encodeInt2Base84 buf->1022,1021 均为数字,启动数字压缩模块1024 压缩编码->1022 为GBK中文编码压缩结束->10221 插入\*xxx/GBK包裹体

->1025 encodeInt2Base60 压缩数字码产生1026 MT60[dx]字符码 buf之

->10231 输出1027 MT84字符码 buf之->返回buf编码结束

-----

发明要义:

1.一种基于10进制数字值为基数的编码:对其数值有选择的进行60进制,84进制进行重新编码实施的技术方案,其特征如下:

encodeInt2Base84,encodeInt2Base60,encodeAscii2Base84

解码依次对应:decodeBase84Int,decodeBase60Int,decodeBase84Ascii

1.0 -控制符 1.1 位于最后一位代表奇数

1.2位于除1.1则表示高进位:for encodeInt2Base84,解码dx+=84;

1.3位于除1.1则表示低进位:for encodeAscii2Base84解码dx;

1.4.0\*./控制符;

1.4.1\*控制符 \*标识为整数数据如果以/结尾则标识>127的字符为GBK编码

xx为GBK编码5位数字/汉字,如果原gbk中文为奇数位,则自动填充\0' 填充偶数位;

1.4.2 \*xx.即ASCII数字 ,for encodeAscii2Base84 方案,当且仅当ascii码混合字母与数字,其数字必须连续满足大于5位时,该模块才启用encodeInt2Base60数字压缩技术,

1.4.1,1.4.2 编码自左起依次取2位,取整数值->dx,如果奇数末位依据1.1 补-

;

1.5.084进制参MT84码表规则(参表MT84码表),60进制参MT60表;

1.5.1 丢弃ASCII0-31 控制码;

1.5.2 设x为ascii 字符变量;dx=ASCII(x) 序号dx-=32;

1.5.3 低进位 变换dx<43则 buf -;buf MT84[dx];

1.5.4 非进位 变换dx-=43;buf MT84[dx];

1.6 url传输+ ->\*\* 服务端逆替换 空格或\*\*->+再解码,=进行urldecode;

1.7 解码根据MT84与ascii 码表逆转换;

1.8 基于1.4.1,1.4.2 规则编码为encodeInt2Base60 ref MT60码表{0-9 A-Y, a-y}共60 位,z为base60 左进位,Z结束控制符,后跟1位数字标识解码多余用于控制数据还原;

1.9 因base84 自动判断整数包括体内用base60编码,故虽然同一符号在不通编码规则不同,同时设计编码规避重复;Base84 含zZ./,但这些符号在base60不存在.故混合编码只是局部嵌入,不会打乱;

1.9.1 一个ascii2base84编码可容纳多个int2Base60码块,遇到不同块调不同解码程序进行解码;

1.10 GBK编码以5位数字,不足前补0,解码除掉前缀0,取ascii码;

1.11 ascii码只是变换位置不会减少编码位数,对int字符串会压缩编码;

1.12 当int码,GBK码长小于2,返回NoN 非数字;

1.13 本发明数字压缩编码精要,采用数字10进制,每俩位一组进行分隔进行编码,解码

1.13.1 基于1.13 俩位数字最大99 ,base60编码,其进位最大1以z代替。

[0006] 基于Base84.encodeInt2Base84 ,实现的一种高压缩比的数字码压缩技术,其压缩率最大可接近50%,其特征如下:

2.1 基于1.0,1.5.1,1.5.2,1.5.3 ,可知单个ascii符号最多可表示俩位10进制数值范围[0,83],而大于83且<=99的概论只有16%,当相邻俩位10进制数大于83,经过编码用-MT84[dx%84]标识,解码是z替换为84进制的高位,逆运算dx+=84即可还原信息,即这时用-加余码组成字母2位标识原2位10进制数字;也就是说只有16%的概率需要用可变长2位,其余84%的概率是1位字符代替俩位信息;在编码数据时,每2位数字通过换算大概率用1位字符标识,这是本发明以接近50%的压缩比基础;

2.2 其产生的编码字符,由MT84表字符决定,除+字符在url 转发需要转义\*\*,其它字符原封发送;当然在解码前需要进行urldecode 下较为可靠。

[0007] 一种基于1可实施的高效,高压缩比的时间数字订单压缩技术,其特征如下:

3.1 基于1.8 encodeInt2Base60,MT60表,可实现优雅符号密文输出,该技术方案主要用于密文需要存储在数据库,或对外接口服务,长数字订单就是例子;如果订单符号发送一批奇怪符号,大概率会被蹩脚程序员当作乱码删除,引发不必要的损失;该压缩率有60%的概论为1:2压缩,有40%的概率1:1不压缩;实际测算基于时间订单日期除年外均在60以内数字,其实该算法最早实现就是针对时间数字进行压缩的,这样概率提升到90%是1:2压缩,即50%的压缩;除时间数字外是60%的概率1:2压缩;

3.2 因为编码、解码过程尽可能一个循环流程下来操作;避免不必要的循环;故是高效率的。

[0008] 一种可智能检测数字可变长,可复合拼装包裹体的编码压缩技术,其特征如下:

4.1 基于1.4.0,1.4.1,1.4.2 可知,在encodeAscii2Base84过程,编码模块,会智能识别数字连续满超过5位才启用压缩数字技术,因为数字太短不足以提现其技术优势,比如ip地址不可以被解析压缩,只能按ascii码编码更快捷;但对应ascii字符串中长串数字该模块会智能启用压缩技术编入包裹体,其特征是密文含\*xxxxx.,对GBK中文则以\*xxx/为包裹体,当有连续中文会合在一起解析直到结束。

[0009] 一种可用于url传递,可嵌入Cookie编码技术,其特征如下:

5.1 基于1.50,1.6,1.7,1.8 可知编码表不含” & ” 等url敏感字符,其设计之初就是高效的url传递 ‘;’ 为cookie分隔符,’ & ’ 为url分隔符;’ + ’ 会被转义’ \*\*’ 。

[0010] 基于4一种较Base64位编码更高效,编码速度更快的编码,其特征如下:

6.1实测基于4base84.encodeAscii2Base84 实例为用户真实属性数据明文数据1422,明文含中英文,符号数字混合字符:

Base84.encodeAscii2Base84算法对比Base64算法数据对比见下:

算法:	明文长	1422	编码长	数据压缩率	编码耗时(纳秒)	base84:x	效比	解码比
	Base84	1508	106%	1080999	1:1	1:1		



X1:Base64	1928	136%	4205437	3.9:1	9.8:1
X2:Hex	2888	203%	2523338	2.3:1	0.35:1

据此测试数据不难发现,Base84数据压缩率103%最优,接近1:1(对ascii码不压缩,但如果ascii字符串含有大量连续的数字会压缩),效率是Base64的3.9倍,是Hex的2.3倍;无论从压缩率,还是压缩效率都是成倍提高,其中解码比base64提高9.8倍;

## 6.2 基于2 Base84.encodeInt2Base84

### 测试

sn= 0161520130901537001153700161520130901537001615201309015370016152

0130

明文长:68

base84:sn:B90Be-GP1AL1:B90Be-GP1AQPUNJB1:B90Be,36,99680 ns

base64:sn:MDE2MTUyMDEzMDkwMTUzNzAwMTE1MzcwMDE2MTUyMDEzMDkwMTUzNzAwMTYxNTIwMTMwOTAxNTM3MDAxNjE1MjAxMzA=,92,3728560 ns

算法对比Base64算法数据对比见下:

算法:明文长68	编码长	数据压缩率	编码耗时(纳秒)	base84:x效比解	码比
Base84	36	53%	99680	1:1	1:1
X1:Base64	92	135%	3728560	37:1	33:1
X2:Hex	136	200%	234852	2.4:1	0.64:1

据此测试数据不难发现,Base84对纯数字数据压缩率53%最优,接近50%,效率是Base64的37倍,是Hex的2.4倍;无论从压缩率,还是压缩效率都是成倍提高,其中解码是base64的33倍,(注:测试环境win7+jdk1.78,eclipse)。

[0011] 现实意义:

一种基于84进制高效率数据压缩编码算法:该发明可完全代替传统的base64 进制,该算法以10进制数字为每俩位截取压缩为1位目标码,不足1位采用进 位可变长编码,且对ascii字符编码支持混合编码,具有智能检测ASCII码中数 字,GBK;理论上任何ascii码,GBK均可转为10进制模式再通过本发明的数 字压缩编码进行重编码;经过测算其运行效率比传统Base64高约37倍;纯数字 数据压缩率接近50%;本发明主要用于动态令牌加密编码传输;以及航天航空领 域的大量观测大数进行压缩,结合幻数加密算法(参同期专利)MagicNumberEncrypt可实现高安全性,高可靠性信息保密压缩传输;高倍空间 压缩编码,高效率加密解密技术为超算力输出,超分布式存储输出必放一异彩。

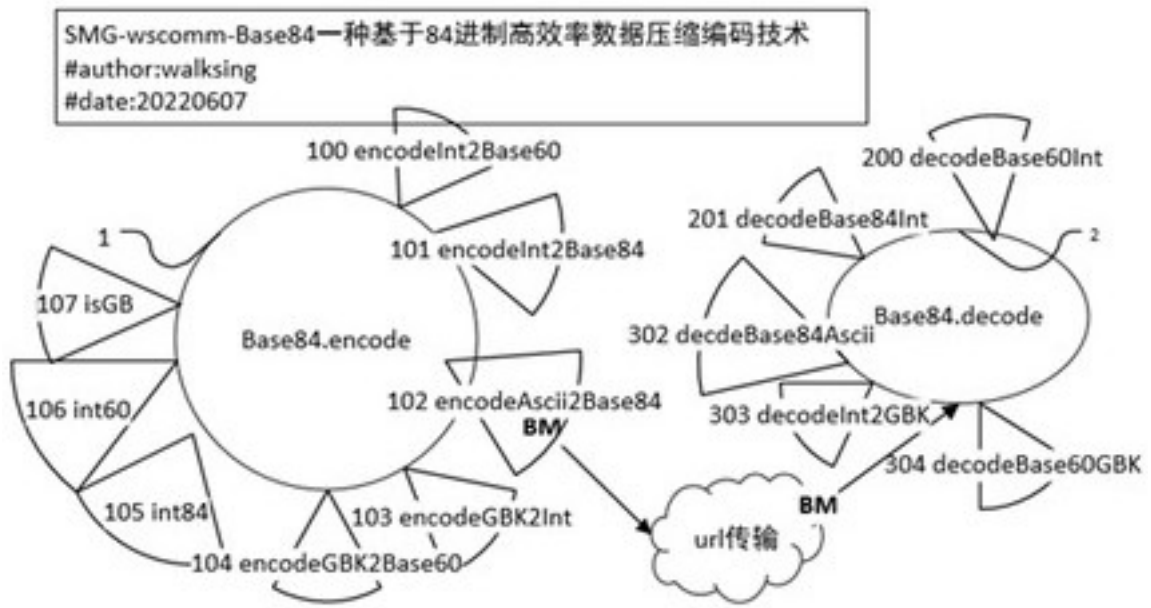


图1

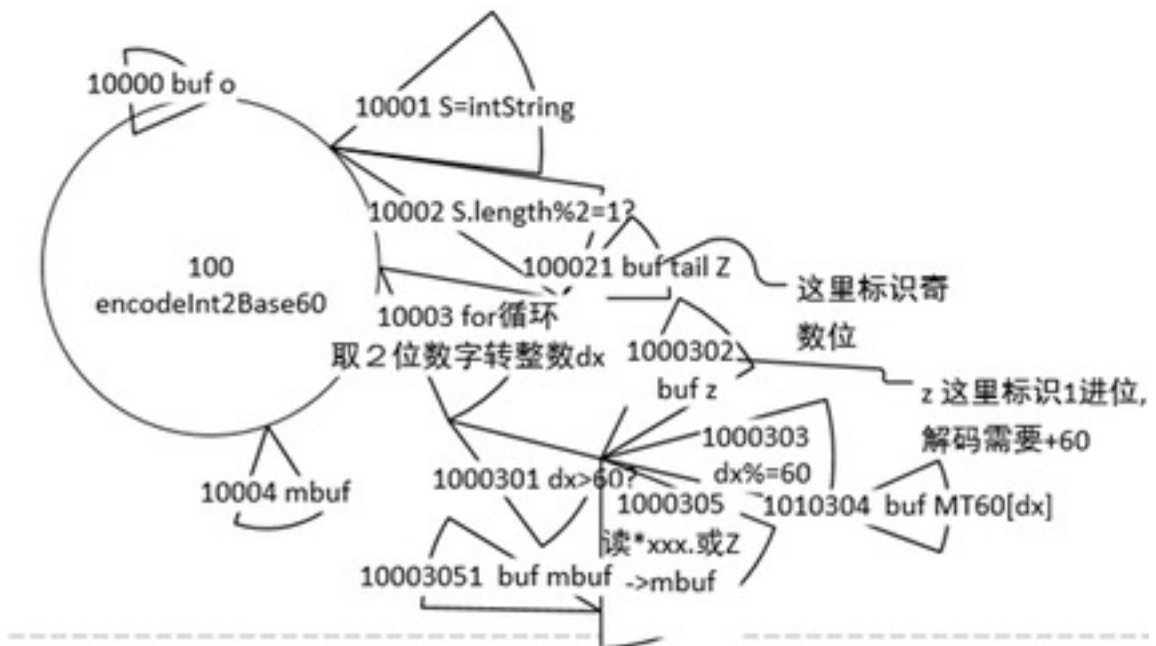


图2

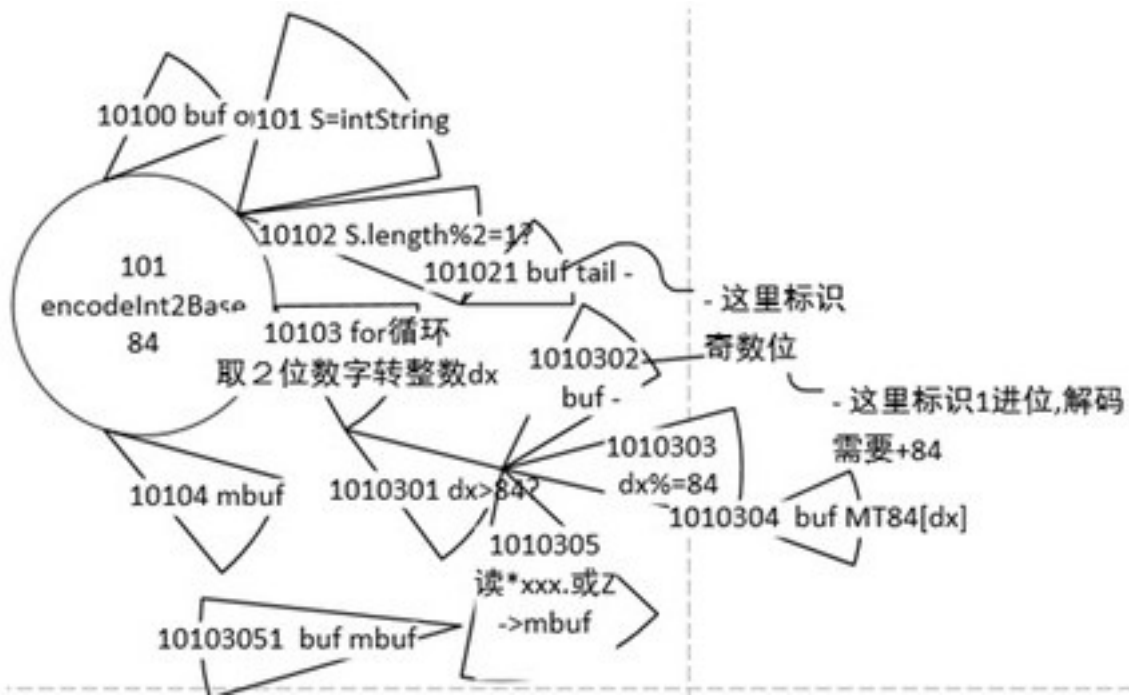


图3

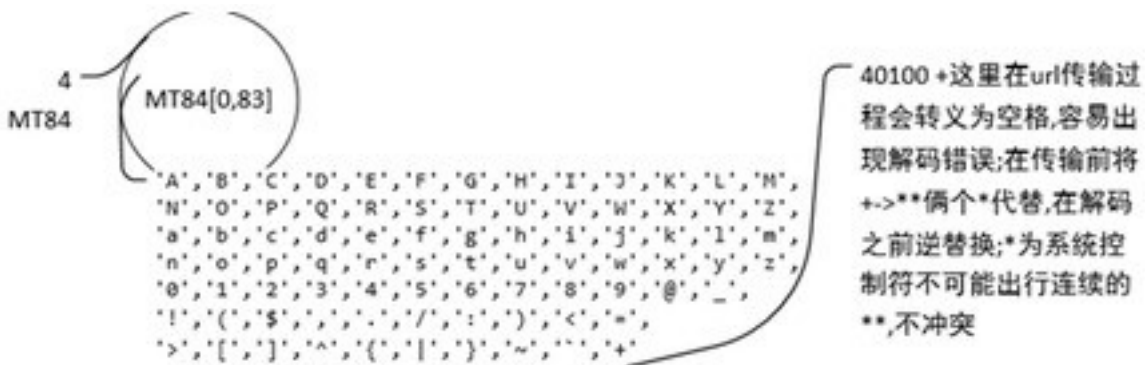


图4

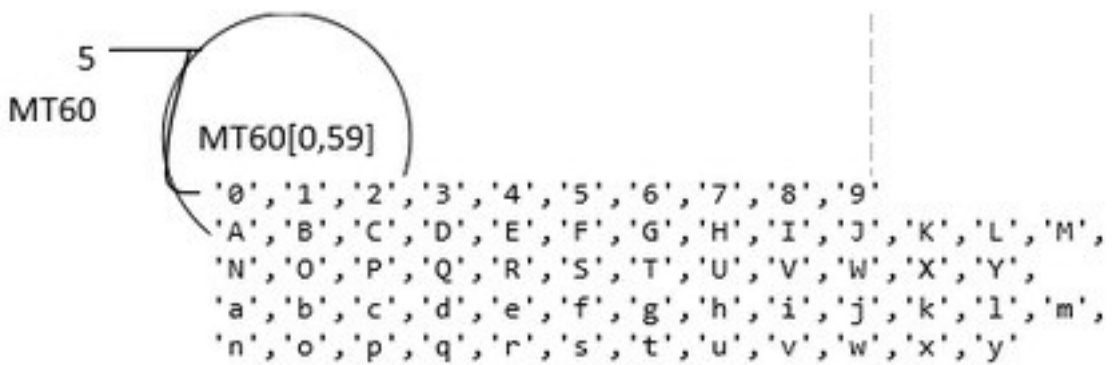


图5

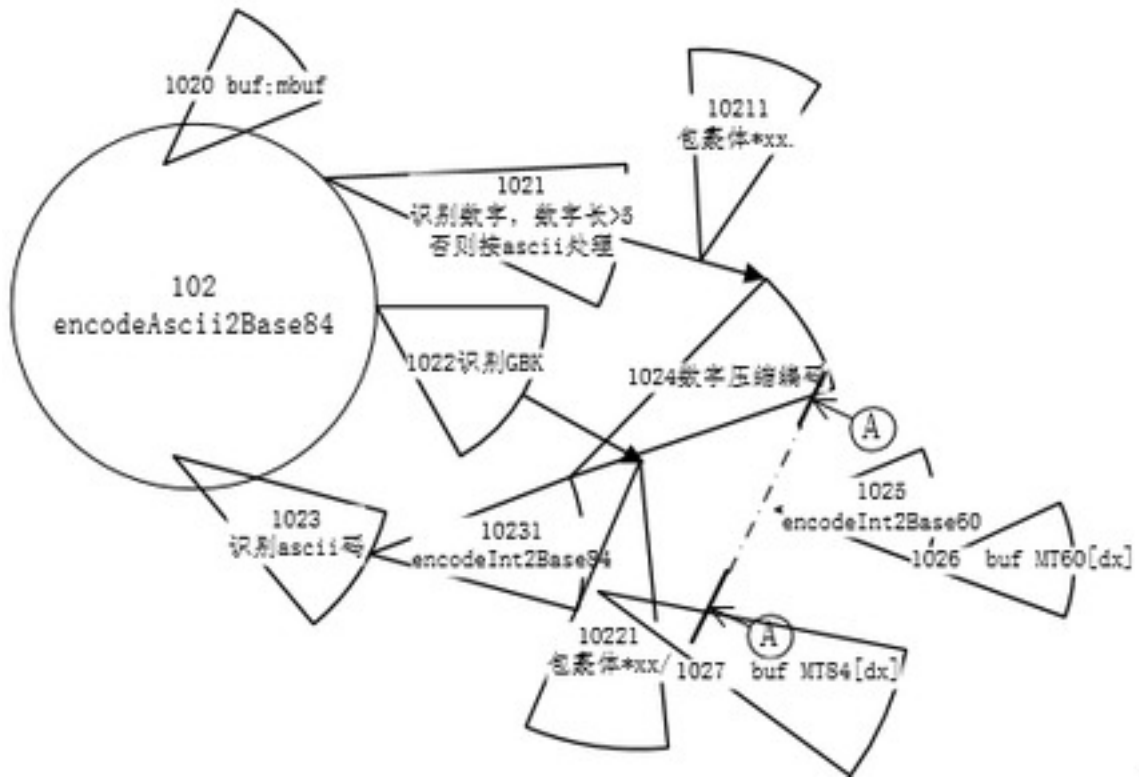


图6