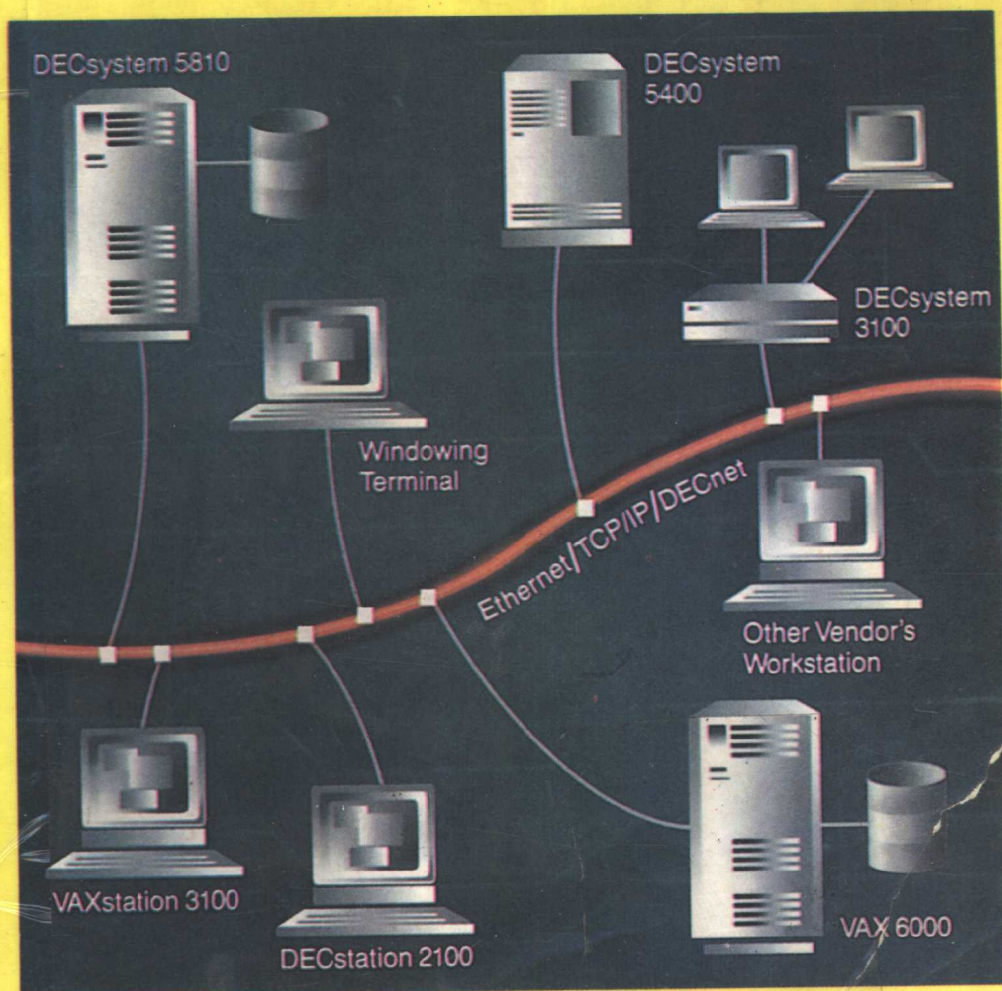


网络互联协议 TCP/IP 详解

王劲松 编



科学技术文献出版社

北京科海培训中心系列教材

网络互联协议 TCP/IP 详解

王劲松 编

科学技术文献出版社

(京)新登字 130 号

内 容 简 介

计算机网络是计算机科学的一个重要分支,而 TCP/IP 协议是一组非常著名的网络通信协议。本书系统地分析了组成 TCP/IP 协议的各个协议的细节,论述了这些协议的实现方法,并重点阐述了协议软件的本质。

全书共分为二十一章,分别剖析了 TCP、SNMP、NET、ARP、UDP、RIP、ICMP 和其他一些协议。各章后面备有供进一步研究的资料和练习。书末有两个附录,分别介绍了过程调用的交叉引用和用于程序代码的 Xinu 过程和函数。本书可作为高等院校计算机专业的高年级本科生和研究生有关课程的教材,也可以供广大计算机软件工作者以及有关技术人员学习参考。

网络互联协议 TCP/IP 详解

王劲松 编

科学技术文献出版社出版

(北京复兴路 15 号 邮政编码 100038)

北京艺辉胶印厂印刷

新华书店北京发行所发行 各地新华书店经营

*

787×1092 毫米 16 开本 29 印张 705 千字

1993 年 5 月第 1 版 1993 年 5 月第 1 次印刷

印数:1—5000 册

科技新书目:

ISBN 7-5023-2009-1/TP·104

定价:26.00 元

前言

•TCP/IP(即“传输控制协议与网际协议”的缩写)是一组非常著名的网际通信协议,特别是近几年来愈来愈引起世人的关注。在美国和其他一些西方国家,TCP/IP 协议已成为建立计算机局域网、广域网以及互联网的重要通信协议。在我们国家,目前已有不少部门采用或即将采用 TCP/IP 协议建立广域网或互联网,有的局域网也将 TCP/IP 协议作为多机种间互相通信的共同协议。

TCP/IP 协议是一套完整的协议簇。除了本身的 TCP(传输控制协议)和 IP(网际协议)之外,还包括其它多种协议,其中有工具性协议、管理性协议以及应用性协议等等。本书将分析各个协议的细节,论述这些协议的实现方法,并重点阐述协议软件的本质。

诚然,关于各个协议的官方规格说明以及有关其实现和使用的讨论,在请求注释文件(RFC)中均有论述。尽管有些 RFC 文件让初学者有些难以理解,但是这些文件仍是详细信息的权威性资料,本书并不想简单地重复所有这些信息。更为重要的是,RFC 包括了各种协议,并留下了许多没有解决的、有关协议交互的问题。例如,路由选择信息协议(RIP)说明了一个网关是如何在其 IP 路由选择表中安排路由的,以及网关是如何将该表中的路由通知其它网关的,RIP 亦说明了必须超时和撤销的路由。但是,RFC 文件中并未说明 RIP 和其它协议的交互作用,于是,便产生了问题:“路由暂停对表中不是由 RIP 安排的路由有何影响?”此外,人们还须考虑这样一个问题:“RIP 的修改应当取代由网络管理人员手工安排的路由吗?”

为有助于说明协议间的交互,并确保我们的解决方法相互一致,我们设计并建立了一个可用系统作为本书的中心例子。该系统提供了 TCP/IP 协议簇中的大多数协议,其中包括:TCP、IP、ICMP、UDP、ARP、RIP 以及 SNMP。此外,还提供了一个接口服务(finger service)的客户和服务器的例子。由于本书给出了每个协议的程序代码,因此,读者可以研究其实现方法并理解其内部结构。最为重要的是,由于示例系统将这些协议软件集成为一个可用的软件,因而,读者可以很清楚地理解这些协议间的交互。

例子中的程序代码力图遵循协议标准,并采用最新的思想。例如,书中的 TCP 代码包括傻窗口避免(silly window avoidance)、Jacobson—Karels 慢起动及避免信息拥挤的优化,这些特性是在通常的商业实现中所没有的。然而,我们应充分认识到商业界并不总是遵循这些公布的标准,我们力图使系统适应特定环境的应用。例如,程序代码中有一个配置参数,该参数允许程序代码使用网际标准或 BSD—UNIX 系统来解释 TCP 的紧急数据指针(urgent data pointer)。

本书既可作为网络方面的高级教程,也可用作研究生的教材。作为大学教程,应将重点放在前几章,并省略有关 SNMP 和 RIP 的章节。研究生们将在本书有关 TCP 的章节中发现最有趣的、最富有挑战性的概念。自适应性重发(adaptive retransmission)及有关高性能的启发性方法尤为重要,应当给予特别的关注。在本书的练习中提供了一些备选的实现方法和概括;这些练习并不只是对文中所述实现方法的机械式重复,因此,学生们需要在课外进行大胆的尝试以解决这些练习。

对于涉及网络互联和 TCP/IP 协议的技术人员,或关心网络技术方面的进展的技术人员,也是一本非常不错的参考书。

编 译 者
1993 年 2 月·北京

目 录

前 言

第一章 绪论和概述	1
1.1 TCP/IP 协议	1
1.2 必须了解的细节	1
1.3 协议间交互的复杂性	1
1.4 本书所采用的方法	2
1.5 研究程序代码的重要性	2
1.6 Xinu 操作系统	3
1.7 本书其它部分的组织	3
1.8 本章小结	3
供进一步研究的资料	4
第二章 TCP/IP 软件在操作系统中的结构	5
2.1 简介	5
2.2 进程的概念	5
2.3 进程的优先级	6
2.4 进程的通讯	6
2.5 交互进程通讯	8
2.5.1 端口	9
2.5.2 消息传递	10
2.6 设备驱动程序、输入和输出	10
2.7 网络输入和中断	11
2.8 传送报文分组到高层协议	12
2.9 从 IP 到传输协议传送数据报文	12
2.9.1 将输入的数据报文传送到 TCP	13
2.9.2 将输入的数据报文传送到 UDP	13
2.10 到应用程序的传送	14
2.11 输出信息流	14
2.12 从 TCP 通过 IP 到网络输出	15
2.13 UDP 输出	16
2.14 本章小结	16

供进一步研究的资料	19
练习	19
第三章 网络接口层	20
3.1 简介	20
3.2 网络接口的抽象概念	20
3.2.1 接口结构	21
3.2.2 统计使用情况	23
3.3 接口的逻辑状态	23
3.4 本地主机接口	23
3.5 缓冲区管理	24
3.5.1 大缓冲区的解决方法	25
3.5.2 链表的解决方法(mbufs)	25
3.5.3 示例中的解决方法	25
3.5.4 其它缓冲区的问题	26
3.6 多路分用输入报文分组	26
3.7 本章小结	28
供进一步研究的资料	28
练习	28
第四章 地址的发现和赋值(ARP)	29
4.1 简介	29
4.2 ARP 软件概念的组织结构	29
4.3 ARP 设计的示例	29
4.4 ARP 高速缓冲存储器的数据结构	31
4.5 ARP 输出处理	33
4.5.1 搜索 ARP 高速缓冲存储器	33
4.5.2 广播一个 ARP 请求	34
4.5.3 ARP 输出过程	35
4.6 ARP 输入处理	37
4.6.1 向表格中增添已分解的条目	37
4.6.2 发送正在等待的报文分组	38
4.6.3 ARP 输入过程	38
4.7 ARP 高速缓冲存储器的管理	40
4.7.1 分配高速缓冲存储器条目	41
4.7.2 周期性的高速缓冲存储器维护	42
4.7.3 解除已排队的报文分组的分配	43
4.8 ARP 初始化	44

4.9	ARP 配置参数	45
4.10	本章小结	45
	供进一步研究的资料	45
	练习	46
第五章	IP: 全程软件组织结构	47
5.1	简介	47
5.2	中心转换	47
5.3	IP 软件设计	47
5.4	IP 软件的组织结构和数据报文流	48
5.4.1	选择输入数据报文的策略	48
5.4.2	允许 IP 进程阻塞(block)	50
5.4.3	IP 使用的常数定义	54
5.4.4	检查和的计算	56
5.4.5	处理定向广播	56
5.4.6	识别广播地址	58
5.5	IP 首部中的字节排序	59
5.6	向 IP 发送数据报文	60
5.6.1	发送在本地产生的数据报文	60
5.6.2	发送输入数据报文	62
5.7	表格的维护	62
5.8	本章小结	63
	供进一步研究的资料	64
	练习	64
第六章	IP: 路由选择表和路由选择算法	65
6.1	简介	65
6.2	路由的维护和查找	65
6.3	路由选择表的组织结构	65
6.4	路由选择表的数据结构	66
6.5	路由的来源和持久性(persistence)	68
6.6	为数据报文选择路由	68
6.6.1	实用过程	68
6.6.2	获得路由	72
6.6.3	数据结构初始化	73
6.7	周期性路由表的维护	74
6.7.1	增加路由	75
6.7.2	删除路由	78

6.8	IP 选项处理	80
6.9	本章小结	81
	供进一步研究的资料	82
	练习	82
第七章	IP:分段和重构.....	83
7.1	简介	83
7.2	分段数据报文	83
7.2.1	分段片断	83
7.3	分段的实现	84
7.3.1	发送一个片段	86
7.3.2	拷贝数据报文的首部	87
7.4	数据报文重构	88
7.4.1	数据结构	88
7.4.2	互斥	89
7.4.3	向表中增加片段	89
7.4.4	在溢出过程中放弃	91
7.4.5	测试完整的数据报文	92
7.4.6	用片段建立数据报文	94
7.5	片段表的维护	95
7.6	初始化	97
7.7	本章小结	98
	供进一步研究的资料	98
	练习	98
第八章	IP:错误处理(ICMP)	100
8.1	简介	100
8.2	ICMP 信息格式	100
8.3	ICMP 信息的实现	100
8.4	处理输入的 ICMP 信息	102
8.5	控制 ICMP 重定向信息	105
8.6	设置子网络掩码	107
8.7	为 ICMP 报文分组选择报源地址	108
8.8	产生 ICMP 错误信息	109
8.9	避免有关错误的错误	111
8.10	为 ICMP 分配缓冲器	112
8.11	ICMP 信息的数据部分	114
8.12	产生 ICMP 重定向信息	115

8.13	本章小结	117
	供进一步研究的资料	117
	练习	117
第九章	UDP: 用户数据报文	118
9.1	简介	118
9.2	UDP 端口和多路分用	118
9.2.1	用于两两通讯的端口	118
9.2.2	用于多对一通讯的端口	119
9.2.3	操作模式	119
9.2.4	多路分用的错综(subtle)结果	120
9.3	UDP(用户数据报文)	121
9.3.1	UDP 的说明	121
9.3.2	输入数据报文队列说明	123
9.3.3	把 UDP 端口号映射到队列中	124
9.3.4	分配自由的队列	125
9.3.5	变换成为/来自网络字节的顺序	125
9.3.6	处理到达的数据报文	126
9.3.7	UDP 检查和的计算	128
9.4	UDP 输出处理	129
9.4.1	发送 UDP 数据报文	130
9.5	本章小结	132
	供进一步研究的资料	132
	练习	132
第十章	TCP: 数据结构和输入处理	133
10.1	简介	133
10.2	TCP 软件概述	133
10.3	发送控制块	133
10.4	TCP 数据段格式	138
10.5	序列空间比较	139
10.6	TCP 有限状态机器	140
10.7	状态转换的示例	141
10.8	有限状态机器的说明	142
10.9	TCB 的分配和初始化	143
10.9.1	分配一个 TCB	143
10.9.2	解除 TCB 的分配	144
10.10	有限状态机器的实现	145

10.11	处理输入数据段	146
10.11.1	把一个 TCP 首部变换为本地字节顺序	148
10.11.2	计算 TCP 检查和	149
10.11.3	为数据段寻找 TCB	150
10.11.4	检查数据段的有效性	151
10.11.5	为当前状态选择进程	152
10.12	本章小结	154
	供进一步研究的资料	154
	练习	154

第十一章 TCP:有限状态机器的实现 155

11.1	简介	155
11.2	CLOSED 状态处理	155
11.3	降级关闭(Graceful Shutdown)	155
11.4	关闭后的时间延迟	156
11.5	TIME-WAIT 状态处理	157
11.6	CLOSING 状态处理	158
11.7	FIN-WAIT-2 状态处理	159
11.8	FIN-WAIT-1 状态处理	160
11.9	CLOSE-WAIT 状态处理	161
11.10	LAST-ACK 状态处理	162
11.11	ESTABLISHED 状态处理	163
11.12	数据段中的数据处理的	165
11.13	保持接收到的八位位组的轨迹(track)	167
11.14	终止一个 TCP 连接	170
11.15	建立一个 TCP 连接	171
11.16	初始化 TCB	171
11.17	SYN-SENT 状态处理	172
11.18	SYN-RECEIVED 状态处理	173
11.19	LISTEN 状态的处理	176
11.20	为一个新 TCB 初始化窗口变量	177
11.21	本章小结	178
	供进一步研究的资料	179
	练习	179

第十二章 输出处理 180

12.1	简介	180
12.2	控制 TCP 输出的复杂性	180

12.3	四个 TCP 输出状态	181
12.4	作为一个过程的 TCP 输出	181
12.5	TCP 输出信息	182
12.6	编制输出状态和 TCB 编号的程序代码	182
12.7	TCP 输出过程的实现	182
12.8	互斥	184
12.9	IDLE 状态的实现	184
12.10	PERSIST 状态的实现	184
12.11	TRANSMIT 状态的实现	185
12.12	RETRANSMIT 状态的实现	187
12.13	发送一个数据段	187
12.14	计算 TCP 数据长度	190
12.15	计算序列的计数	192
12.16	其它 TCP 过程	192
12.16.1	发送 RESET	193
12.16.2	转换到网络字节顺序	194
12.16.3	在输出缓冲区等待空间	195
12.16.4	激活等待 TCB 的进程	196
12.16.5	选择初始序列号	197
12.17	本章小结	198
	供进一步研究的资料	198
	练习	198

第十三章 TCP: 定时器管理 199

13.1	简介	199
13.2	定时事件的一般数据结构	199
13.3	TCP 事件的数据结构	200
13.4	定时器、事件和消息	200
13.5	TCP 定时器进程	201
13.6	删除 TCP 定时器事件	203
13.7	删除一个 TCB 的所有事件	204
13.8	决定一个事件的剩余时间	205
13.9	插入一个 TCP 定时器事件	206
13.10	无延迟时启动 TCP 输出	207
13.11	本章小结	208
	供进一步研究的资料	208
	练习	209

第十四章 流控制和自适应重发 210

14.1	简介	210
14.2	自适应重发的困难	210
14.3	协调自适应重发	210
14.4	重发定时器和补偿	211
14.4.1	Karn 算法	211
14.4.2	重发输出状态处理	211
14.5	基于窗口的流控制	213
14.5.1	傻窗口综合症(Silly Window Synchome)	213
14.5.2	接收者一侧的傻窗口避免	213
14.5.3	零窗口之后的优化性能	214
14.5.4	调整发送者的窗口	215
14.6	最大数据段尺寸的计算	216
14.6.1	发送者的最大数据段尺寸	216
14.6.2	选项处理	217
14.6.3	通告输入最大数据段	219
14.7	拥挤的避免和控制	220
14.7.1	倍增递减	220
14.8	慢启动和拥挤避免	220
14.8.1	慢启动	220
14.8.2	阈值点之后的较慢递增	221
14.8.3	拥挤窗口增加的实现	221
14.9	往返估算和超时	222
14.9.1	快速平均修正算法	223
14.9.2	处理输入的应答信号	224
14.9.3	为窗口外的数据产生应答信号	226
14.10	其它的解释和方法	227
14.11	本章小结	227
	供进一步研究的资料	227
	练习	228

第十五章 TCP:紧急数据处理和推进功能 229

15.1	简介	229
15.2	区外信号发送	229
15.3	紧急数据	229
15.4	解释标准	230
15.4.1	通知应用程序	231

15.4.2	多个并行应用程序	231
15.5	紧急模式数据的数据结构	232
15.6	从输入数据段中提取紧急数据	232
15.7	序列号、键和表的次序	234
15.8	紧急数据队列操作例行程序	234
15.8.1	紧急队列节点的初始化	236
15.8.2	解除紧急队列节点的分配	236
15.9	读紧急数据	237
15.10	收集空位	238
15.11	跳过空位	239
15.12	从 TCP 中读取数据	240
15.13	发送紧急数据	242
15.14	记录紧急数据消息的位置	242
15.14.1	可能的方案	243
15.14.2	给紧急数据输出表中增加一个节点	243
15.15	传送紧急数据	244
15.16	TCP 的推进功能	244
15.17	用失序传送解释推进	245
15.18	输入中推进的实现	245
15.19	本章小结	246
	供进一步研究的资料	247
	练习	247
第十六章	插座级接口	248
16.1	简介	248
16.2	设备的接口	248
16.2.1	单字节 I/O	249
16.2.2	非传送函数的扩展	249
16.3	作为设备的 TCP 连接	250
16.4	TCP 客户程序的例子	250
16.5	TCP 服务器程序的例子	251
16.6	TCP 主设备的实现	253
16.6.1	TCP 主设备的 open 函数	253
16.6.2	形成消极的 TCP 连接	254
16.6.3	形成积极的 TCP 连接	255
16.6.4	分配未用的本地端口	256
16.6.5	完成积极的连接	258
16.6.6	控制 TCP 主设备	259

16.7	TCP 从设备的实现	260
16.7.1	从 TCP 从设备输入数据	260
16.7.2	TCP 从设备上的单字节输入	262
16.7.3	通过 TCP 从设备输出	263
16.7.4	关闭 TCP 连接	265
16.7.5	TCP 从设备的控制操作	266
16.7.6	从消极设备接受连接	267
16.7.7	改变听队列的尺寸	268
16.7.8	从从设备获得统计数据	268
16.7.9	设置或清除 TCP 选项	270
16.8	从设备的初始化	271
16.9	本章小结	272
	供进一步研究的资料	273
	练习	273

第十七章 RIP: 主动路由传播和被动获取 274

17.1	简介	274
17.2	主动和被动模式的参与者	274
17.3	基本的 RIP 算法和代价度量	275
17.4	不稳定性和解决办法	275
17.4.1	计数到无穷大	275
17.4.2	网关失效(Crash)和路由超时	276
17.4.3	分割前景(Split horizon)	276
17.4.4	抑制反向	277
17.4.5	使用抑制反向的路由超时	277
17.4.6	触发更新(Triggered Update)	277
17.4.7	防止广播风暴(Broadcast Storms)的随机化	278
17.5	消息类型	278
17.6	协议特征	279
17.7	RIP 的实现	280
17.7.1	两种实现方式	280
17.7.2	说明	280
17.7.3	输出的概念组织	282
17.8	主要的 RIP 进程	282
17.8.1	必须为 0 的域必须为 0	284
17.8.2	处理输入响应	285
17.8.3	在更新中锁定	286
17.8.4	校验地址	286

17.9	响应输入请求	287
17.10	产生更新信息	288
17.11	初始化更新消息的副本	289
17.11.1	向更新消息副本增加路由	290
17.11.2	计算要通告的尺度	292
17.11.3	为 RIP 消息分配数据报文	292
17.12	产生周期性的 RIP 输出	293
17.13	RIP 的局限性	294
17.14	本章小结	295
	供进一步研究的资料	295
	练习	295
第十八章 SNMP: MIB 变量、表示和赋值		296
18.1	简介	296
18.2	服务器的组织和名称的映射	296
18.3	MIB 变量	297
18.3.1	表格中的域	298
18.4	MIB 变量名	298
18.4.1	名称的数字表达	298
18.5	名称中的字典顺序	299
18.6	前缀去除	299
18.7	与 MIB 变量有关的操作	300
18.8	表格的名称	300
18.9	名称层次结构中的概念线索	301
18.10	MIB 变量的数据结构	302
18.10.1	使用单独的函数完成操作(threading)	303
18.11	快速查找的数据结构	303
18.12	散列表的实现	304
18.13	MIB 赋值的规格	305
18.14	用于赋值中的内部变量	309
18.15	散列表查找	310
18.16	SNMP 结构和常量	312
18.17	ASN.1 表达式操作	315
18.17.1	长度的表示	316
18.17.2	将整数转换为 ASN.1 形式	318
18.17.3	将对象 id 转换为 ASN.1 形式	319
18.17.4	转换数值的通用例程	322
18.18	本章小结	324

供进一步研究的资料.....	324
练习.....	324
第十九章 SNMP:客户和服务	326
19.1 简介	326
19.2 服务器中的数据表示方法	326
19.3 服务器的实现	326
19.4 分析 SNMP 消息	328
19.5 转换赋值表中的 ASN.1 名称	332
19.6 分解查询	333
19.7 解释 get_next 操作	334
19.8 间接操作的应用	335
19.9 表格的间接操作	337
19.10 产生一个反向的回答消息	338
19.11 从内部形式转换到 ASN.1	341
19.12 服务器使用的实用函数	342
19.13 SNMP 客户的实现	342
19.14 变量的初始化	344
19.15 本章小结	347
供进一步研究的资料.....	347
练习.....	347
第二十章 SNMP:表格访问函数	348
20.1 简介	348
20.2 表格的访问	348
20.3 表格的对象标识符	348
20.4 地址条目表格函数	349
20.4.1 地址条目表格的 get 操作	350
20.4.2 地址条目表格的 get_first 操作	351
20.4.3 地址条目表格的 get_next 操作	352
20.4.4 地址条目表格中的增量搜索	354
20.5 地址转换表格函数	354
20.5.1 地址转换表格的 get 操作	356
20.5.2 地址转换表格的 get_first 操作	357
20.5.3 地址转换表格的 get_next 操作	358
20.5.4 地址条目表格中的增量搜索	359
20.5.5 混沌顺序	361
20.5.6 地址转换表格的 set 操作	361