
A

Appendix

Opcode Tables

In Chapter 2, you saw that Java bytecode is found in the code attribute part of the Java class file. Table A-1 lists all the possible Java bytecodes. The hex value for each opcode is shown along with the assembler-like opcode mnemonic.

Table A-1. *Java Bytecode-to-Opcode Mapping*

Opcode	Hex Value	Opcode Mnemonic
0	(0x00)	Nop
1	(0x01)	acnst_null
2	(0x02)	iconst_m1
3	(0x03)	iconst_0
4	(0x04)	iconst_1
5	(0x05)	iconst_2
6	(0x06)	iconst_3
7	(0x07)	iconst_4
8	(0x08)	iconst_5
9	(0x09)	lconst_0

Opcode	Hex Value	Opcode Mnemonic
10	(0x0a)	lconst_1
11	(0x0b)	fconst_0
12	(0x0c)	fconst_1
13	(0x0d)	fconst_2
14	(0x0e)	dconst_0
15	(0x0f)	dconst_1
16	(0x10)	bipush
17	(0x11)	sipush
18	(0x12)	ldc
19	(0x13)	ldc_w
20	(0x14)	ldc2_w
21	(0x15)	iload
22	(0x16)	lload
23	(0x17)	fload
24	(0x18)	dload
25	(0x19)	aload
26	(0x1a)	iload_0
27	(0x1b)	iload_1
28	(0x1c)	iload_2
29	(0x1d)	iload_3
30	(0x1e)	lload_0

31	(0x1f)	lload_1
32	(0x20)	lload_2
33	(0x21)	lload_3
34	(0x22)	fload_0
35	(0x23)	fload_1
36	(0x24)	fload_2
37	(0x25)	fload_3
38	(0x26)	dload_0
39	(0x27)	dload_1
40	(0x28)	dload_2
41	(0x29)	dload_3
42	(0x2a)	aload_0
43	(0x2b)	aload_1
44	(0x2c)	aload_2
45	(0x2d)	aload_3
46	(0x2e)	iaload
47	(0x2f)	laload
48	(0x30)	faload
49	(0x31)	daload
50	(0x32)	aaload
51	(0x33)	baload
52	(0x34)	caload

Opcode	Hex Value	Opcode Mnemonic
53	(0x35)	saload
54	(0x36)	istore
55	(0x37)	lstore
56	(0x38)	fstore
57	(0x39)	dstore
58	(0x3a)	astore
59	(0x3b)	istore_0
60	(0x3c)	istore_1
61	(0x3d)	istore_2
62	(0x3e)	istore_3
63	(0x3f)	lstore_0
64	(0x40)	lstore_1
65	(0x41)	lstore_2
66	(0x42)	lstore_3
67	(0x43)	fstore_0
68	(0x44)	fstore_1
69	(0x45)	fstore_2
70	(0x46)	fstore_3
71	(0x47)	dstore_0
72	(0x48)	dstore_1
73	(0x49)	dstore_2

74	(0x4a)	dstore_3
75	(0x4b)	astore_0
76	(0x4c)	astore_1
77	(0x4d)	astore_2
78	(0x4e)	astore_3
79	(0x4f)	iastore
80	(0x50)	lastore
81	(0x51)	fastore
82	(0x52)	dastore
83	(0x53)	aastore
84	(0x54)	bastore
85	(0x55)	castore
86	(0x56)	sastore
87	(0x57)	pop
88	(0x58)	pop2
89	(0x59)	dup
90	(0x5a)	dup_x1
91	(0x5b)	dup_x2
92	(0x5c)	dup2
93	(0x5d)	dup2_x1
94	(0x5e)	dup2_x2
95	(0x5f)	swap

Opcode	Hex Value	Opcode Mnemonic
96	(0x60)	iadd
97	(0x61)	ladd
98	(0x62)	fadd
99	(0x63)	dadd
100	(0x64)	isub
101	(0x65)	lsub
102	(0x66)	fsub
103	(0x67)	dsub
104	(0x68)	imul
105	(0x69)	lmul
106	(0x6a)	fmul
107	(0x6b)	dmul
108	(0x6c)	idiv
109	(0x6d)	ldiv
110	(0x6e)	fdiv
111	(0x6f)	ddiv
112	(0x70)	irem
113	(0x71)	lrem
114	(0x72)	frem
115	(0x73)	drem
116	(0x74)	ineg

117	(0x75)	lneg
118	(0x76)	fneg
119	(0x77)	dneg
120	(0x78)	ishl
121	(0x79)	lshl
122	(0x7a)	ishr
123	(0x7b)	lshr
124	(0x7c)	iushr
125	(0x7d)	lushr
126	(0x7e)	iand
127	(0x7f)	land
128	(0x80)	ior
129	(0x81)	lor
130	(0x82)	ixor
131	(0x83)	lxor
132	(0x84)	iinc
133	(0x85)	i2l
134	(0x86)	i2f
135	(0x87)	i2d
136	(0x88)	l2i
137	(0x89)	l2f
138	(0x8a)	l2d

Opcode	Hex Value	Opcode Mnemonic
139	(0x8b)	f2i
140	(0x8c)	f2l
141	(0x8d)	f2d
142	(0x8e)	d2i
143	(0x8f)	d2l
144	(0x90)	d2f
145	(0x91)	i2b
146	(0x92)	i2c
147	(0x93)	i2s
148	(0x94)	lcmp
149	(0x95)	fcmpl
150	(0x96)	fcmpg
151	(0x97)	dcmpl
152	(0x98)	dcmpg
153	(0x99)	ifeq
154	(0x9a)	ifne
155	(0x9b)	iflt
156	(0x9c)	ifge
157	(0x9d)	ifgt
158	(0x9e)	ifle
159	(0x9f)	if_icmpeq

160	(0xa0)	if_icmpne
161	(0xa1)	if_icmplt
162	(0xa2)	if_icmpge
163	(0xa3)	if_icmpgt
164	(0xa4)	if_icmple
165	(0xa5)	if_acmpeq
166	(0xa6)	if_acmpne
167	(0xa7)	goto
168	(0xa8)	jsr
169	(0xa9)	ret
170	(0xaa)	tableswitch
171	(0xab)	lookupswitch
172	(0xac)	ireturn
173	(0xad)	lreturn
174	(0xae)	freturn
175	(0xaf)	dreturn
176	(0xb0)	areturn
177	(0xb1)	return
178	(0xb2)	getstatic
179	(0xb3)	putstatic
180	(0xb4)	getfield
181	(0xb5)	putfield

Opcode	Hex Value	Opcode Mnemonic
182	(0xb6)	invokevirtual
183	(0xb7)	invokespecial
184	(0xb8)	invokestatic
185	(0xb9)	invokeinterface
186	(0xba)	invokedynamic
187	(0xbb)	new
188	(0xbc)	newarray
189	(0xbd)	anewarray
190	(0xbe)	arraylength
191	(0xbf)	athrow
192	(0xc0)	checkcast
193	(0xc1)	instanceof
194	(0xc2)	monitorenter
195	(0xc3)	monitorexit
196	(0xc4)	wide
197	(0xc5)	multianewarray
198	(0xc6)	ifnull
199	(0xc7)	ifnonnull
200	(0xc8)	goto_w
201	(0xc9)	jsr_w

Table A-2 lists all the possible Dalvik bytecodes encountered first in Chapter 3 and then throughout the book. The hex value for each opcode is shown along with the assembler-like opcode mnemonic.

Table A-2. *Dalvik Bytecode-to-Opcode Mapping*

Opcode	Hex Value	Opcode Mnemonic
1	(0x00)	Nop
2	(0x01)	move vx, vy
3	(0x02)	move/from16 vx, vy
4	(0x03)	move/16
5	(0x04)	move-wide
6	(0x05)	move-wide/from16 vx, vy
7	(0x06)	move-wide/16
8	(0x07)	move-object vx, vy
9	(0x08)	move-object/from16 vx, vy
10	(0x09)	move-object/16
11	(0x0A)	move-result vx
12	(0x0B)	move-result-wide vx
13	(0x0C)	move-result-object vx
14	(0x0D)	move-exception vx
15	(0x0E)	return-void
16	(0x0F)	return vx
17	(0x10)	return-wide vx
18	(0x11)	return-object vx
19	(0x12)	const/4 vx, lit4

Opcode	Hex Value	Opcode Mnemonic
20	(0x13)	const/16 vx, lit16
21	(0x14)	const vx, lit32
22	(0x15)	const/high16 v0, lit16
23	(0x16)	const-wide/16 vx, lit16
24	(0x17)	const-wide/32 vx, lit32
25	(0x18)	const-wide vx, lit64
26	(0x19)	const-wide/high16 vx, lit16
27	(0x1A)	const-string vx, string_id
28	(0x1B)	const-string-jumbo
29	(0x1C)	const-class vx, type_id
30	(0x1D)	monitor-enter vx
31	(0x1E)	monitor-exit
32	(0x1F)	check-cast vx, type_id
33	(0x20)	instance-of vx, vy, type_id
34	(0x21)	array-length vx, vy
35	(0x22)	new-instance vx, type
36	(0x23)	new-array vx, vy, type_id
37	(0x24)	filled-new-array {parameters}, type_id
38	(0x25)	filled-new-array-range {vx..vy}, type_id
39	(0x26)	fill-array-data vx, array_data_offset
40	(0x27)	throw vx

41	(0x28)	goto target
42	(0x29)	goto/16 target
43	(0x2A)	goto/32 target
44	(0x2B)	packed-switch vx, table
45	(0x2C)	sparse-switch vx, table
46	(0x2D)	cmpl-float
47	(0x2E)	cmpg-float vx, vy, vz
48	(0x2F)	cmpl-double vx, vy, vz
49	(0x30)	cmpg-double vx, vy, vz
50	(0x31)	cmp-long vx, vy, vz
51	(0x32)	if-eq vx, vy, target
52	(0x33)	if-ne vx, vy, target
53	(0x34)	if-lt vx, vy, target
54	(0x35)	if-ge vx, vy, target
55	(0x36)	if-gt vx, vy, target
56	(0x37)	if-le vx, vy, target
57	(0x38)	if-eqz vx, target
58	(0x39)	if-nez vx, target
59	(0x3A)	if-ltz vx, target
60	(0x3B)	if-gez vx, target
61	(0x3C)	if-gtz vx, target
62	(0x3D)	if-lez vx, target

Opcode	Hex Value	Opcode Mnemonic
63	(0x3E)	unused_3E
64	(0x3F)	unused_3F
65	(0x40)	unused_40
66	(0x41)	unused_41
67	(0x42)	unused_42
68	(0x43)	unused_43
69	(0x44)	aget vx, vy, vz
70	(0x45)	aget-wide vx, vy, vz
71	(0x46)	aget-object vx, vy, vz
72	(0x47)	aget-boolean vx, vy, vz
73	(0x48)	aget-byte vx, vy, vz
74	(0x49)	aget-char vx, vy, vz
75	(0x4A)	aget-short vx, vy, vz
76	(0x4B)	aput vx, vy, vz
77	(0x4C)	aput-wide vx, vy, vz
78	(0x4D)	aput-object vx, vy, vz
79	(0x4E)	aput-boolean vx, vy, vz
80	(0x4F)	aput-byte vx, vy, vz
81	(0x50)	aput-char vx, vy, vz
82	(0x51)	aput-short vx, vy, vz
83	(0x52)	iget vx, vy, field_id

84	(0x53)	iget-wide vx, vy, field_id
85	(0x54)	iget-object vx, vy, field_id
86	(0x55)	iget-boolean vx, vy, field_id
87	(0x56)	iget-byte vx, vy, field_id
88	(0x57)	iget-char vx, vy, field_id
89	(0x58)	iget-short vx, vy, field_id
90	(0x59)	iput vx, vy, field_id
91	(0x5A)	iput-wide vx, vy, field_id
92	(0x5B)	iput-object vx, vy, field_id
93	(0x5C)	iput-boolean vx, vy, field_id
94	(0x5D)	iput-byte vx, vy, field_id
95	(0x5E)	iput-char vx, vy, field_id
96	(0x5F)	iput-short vx, vy, field_id
97	(0x60)	sget vx, field_id
98	(0x61)	sget-wide vx, field_id
99	(0x62)	sget-object vx, field_id
100	(0x63)	sget-boolean vx, field_id
101	(0x64)	sget-byte vx, field_id
102	(0x65)	sget-char vx, field_id
103	(0x66)	sget-short vx, field_id
104	(0x67)	sput vx, field_id
105	(0x68)	sput-wide vx, field_id

Opcode	Hex Value	Opcode Mnemonic
106	(0x69)	sput-object vx, field_id
107	(0x6A)	sput-boolean vx, field_id
108	(0x6B)	sput-byte vx, field_id
109	(0x6C)	sput-char vx, field_id
110	(0x6D)	sput-short vx, field_id
111	(0x6E)	invoke-virtual { parameters }, methodtocall
112	(0x6F)	invoke-super {parameter}, methodtocall
113	(0x70)	invoke-direct { parameters }, methodtocall
114	(0x71)	invoke-static {parameters}, methodtocall
115	(0x72)	invoke-interface {parameters}, methodtocall
116	(0x73)	unused_73
117	(0x74)	invoke-virtual/range {vx..vy}, methodtocall
118	(0x75)	invoke-super/range
119	(0x76)	invoke-direct/range {vx..vy}, methodtocall
120	(0x77)	invoke-static/range {vx..vy}, methodtocall
121	(0x78)	invoke-interface-range
122	(0x79)	unused_79
123	(0x7A)	unused_7A
124	(0x7B)	neg-int vx, vy
125	(0x7C)	not-int vx, vy
126	(0x7D)	neg-long vx, vy

127	(0x7E)	not-long vx, vy
128	(0x7F)	neg-float vx, vy
129	(0x80)	neg-double vx, vy
130	(0x81)	int-to-long vx, vy
131	(0x82)	int-to-float vx, vy
132	(0x83)	int-to-double vx, vy
133	(0x84)	long-to-int vx, vy
134	(0x85)	long-to-float vx, vy
135	(0x86)	long-to-double vx, vy
136	(0x87)	float-to-int vx, vy
137	(0x88)	float-to-long vx, vy
138	(0x89)	float-to-double vx, vy
139	(0x8A)	double-to-int vx, vy
140	(0x8B)	double-to-long vx, vy
141	(0x8C)	double-to-float vx, vy
142	(0x8D)	int-to-byte vx, vy
143	(0x8E)	int-to-char vx, vy
144	(0x8F)	int-to-short vx, vy
145	(0x90)	add-int vx, vy, vz
146	(0x91)	sub-int vx, vy, vz
147	(0x92)	mul-int vx, vy, vz
148	(0x93)	div-int vx, vy, vz

Opcode	Hex Value	Opcode Mnemonic
149	(0x94)	rem-int vx, vy, vz
150	(0x95)	and-int vx, vy, vz
151	(0x96)	or-int vx, vy, vz
152	(0x97)	xor-int vx, vy, vz
153	(0x98)	shl-int vx, vy, vz
154	(0x99)	shr-int vx, vy, vz
155	(0x9A)	ushr-int vx, vy, vz
156	(0x9B)	add-long vx, vy, vz
157	(0x9C)	sub-long vx, vy, vz
158	(0x9D)	mul-long vx, vy, vz
159	(0x9E)	div-long vx, vy, vz
160	(0x9F)	rem-long vx, vy, vz
161	(0xA0)	and-long vx, vy, vz
162	(0xA1)	or-long vx, vy, vz
163	(0xA2)	xor-long vx, vy, vz
164	(0xA3)	shl-long vx, vy, vz
165	(0xA4)	shr-long vx, vy, vz
166	(0xA5)	ushr-long vx, vy, vz
167	(0xA6)	add-float vx, vy, vz
168	(0xA7)	sub-float vx, vy, vz
169	(0xA8)	mul-float vx, vy, vz

170	(0xA9)	div-float vx, vy, vz
171	(0xAA)	rem-float vx,vy,vz
172	(0xAB)	add-double vx, vy, vz
173	(0xAC)	sub-double vx, vy, vz
174	(0xAD)	mul-double vx, vy, vz
175	(0xAE)	div-double vx, vy, vz
176	(0xAF)	rem-double vx, vy, vz
177	(0xB0)	add-int/2addr vx, vy
178	(0xB1)	sub-int/2addr vx, vy
179	(0xB2)	mul-int/2addr vx, vy
180	(0xB3)	div-int/2addr vx, vy
181	(0xB4)	rem-int/2addr vx, vy
182	(0xB5)	and-int/2addr vx, vy
183	(0xB6)	or-int/2addr vx, vy
184	(0xB7)	xor-int/2addr vx, vy
185	(0xB8)	shl-int/2addr vx, vy
186	(0xB9)	shr-int/2addr vx, vy
187	(0xBA)	ushr-int/2addr vx, vy
188	(0xBB)	add-long/2addr vx, vy
189	(0xBC)	sub-long/2addr vx, vy
190	(0xBD)	mul-long/2addr vx, vy
191	(0xBE)	div-long/2addr vx, vy

Opcode	Hex Value	Opcode Mnemonic
192	(0xBF)	rem-long/2addr vx, vy
193	(0xC0)	and-long/2addr vx, vy
194	(0xC1)	or-long/2addr vx, vy
195	(0xC2)	xor-long/2addr vx, vy
196	(0xC3)	shl-long/2addr vx, vy
197	(0xC4)	shr-long/2addr vx, vy
198	(0xC5)	ushr-long/2addr vx, vy
199	(0xC6)	add-float/2addr vx, vy
200	(0xC7)	sub-float/2addr vx, vy
201	(0xC8)	mul-float/2addr vx, vy
202	(0xC9)	div-float/2addr vx, vy
203	(0xCA)	rem-float/2addr vx, vy
204	(0xCB)	add-double/2addr vx, vy
205	(0xCC)	sub-double/2addr vx, vy
206	(0xCD)	mul-double/2addr vx, vy
207	(0xCE)	div-double/2addr vx, vy
208	(0xCF)	rem-double/2addr vx, vy
209	(0xD0)	add-int/lit16 vx, vy, lit16
210	(0xD1)	sub-int/lit16 vx, vy, lit16
211	(0xD2)	mul-int/lit16 vx, vy, lit16
212	(0xD3)	div-int/lit16 vx, vy, lit16

213	(0xD4)	rem-int/lit16 vx, vy, lit16
214	(0xD5)	and-int/lit16 vx, vy, lit16
215	(0xD6)	or-int/lit16 vx, vy, lit16
216	(0xD7)	xor-int/lit16 vx, vy, lit16
217	(0xD8)	add-int/lit8 vx, vy, lit8
218	(0xD9)	sub-int/lit8 vx, vy, lit8
219	(0xDA)	mul-int/lit8 vx, vy, lit8
220	(0xDB)	div-int/lit8 vx, vy, lit8
221	(0xDC)	rem-int/lit8 vx, vy, lit8
222	(0xDD)	and-int/lit8 vx, vy, lit8
223	(0xDE)	or-int/lit8 vx, vy, lit8
224	(0xDF)	xor-int/lit8 vx, vy, lit8
225	(0xE0)	shl-int/lit8 vx, vy, lit8
226	(0xE1)	shr-int/lit8 vx, vy, lit8
227	(0xE2)	ushr-int/lit8 vx, vy, lit8
228	(0xE3)	unused_E3
229	(0xE4)	unused_E4
230	(0xE5)	unused_E5
231	(0xE6)	unused_E6
232	(0xE7)	unused_E7
233	(0xE8)	unused_E8
234	(0xE9)	unused_E9

Opcode	Hex Value	Opcode Mnemonic
235	(0xEA)	unused_EA
236	(0xEB)	unused_EB
237	(0xEC)	unused_EC
238	(0xED)	unused_ED
239	(0xEE)	execute-inline {parameters}, inline ID
240	(0xEF)	unused_EF
241	(0xF0)	invoke-direct-empty
242	(0xF1)	unused_F1
243	(0xF2)	iget-quick vx, vy, offset
244	(0xF3)	iget-wide-quick vx, vy, offset
245	(0xF4)	iget-object-quick vx, vy, offset
246	(0xF5)	iput-quick vx, vy, offset
247	(0xF6)	iput-wide-quick vx, vy, offset
248	(0xF7)	iput-object-quick vx, vy, offset
249	(0xF8)	invoke-virtual-quick {parameters}, vtable offset
250	(0xF9)	invoke-virtual-quick/range {parameter range}, vtable offset
251	(0xFA)	invoke-super-quick {parameters}, vtable offset
252	(0xFB)	invoke-super-quick/range {register range}, vtable offset
253	(0xFC)	unused_FC
254	(0xFD)	unused_FD

255	(0xFE)	unused_FE
256	(0xFF)	unused_FF

Index

■ A

Abstract syntax trees (AST), 171–172

Aggregation obfuscation, 131

cloning methods, 133

inlining and outlining methods,
132

interleaving methods, 132–133

loop transformations, 133

Android application, decompilers

APKs, 6

backend systems via web
services, 6

DEX file, 6

dexdump output, 7–8

DVM, 6

ProGuard, 6

reasons for vulnerable, 7

Android bytecode analysis

Casting.java, 198–201

Hello World application, 213–214

if statement, 218–220

Android Native Development Kit

(NDK), 17, 141–142

Android package file (APK), 93

adb pull Command, 101

AXMLPrinter2.jar Command, 102

Decoded AndroidManifest.xml,
102–103

decompilers, 6

apktool, 119

dex2jar, 118

Jad, 116–117

JD-GUI, 117–118

Mocha, 115

undx, 118

decompiling, 101

DEX file, 57

casting.class conversion, 59–
60

classes.dex file, 58

class file vs. DEX file, 58

unzipped, 58

downloading

backup tool, 94–95

description, 93–94

forums, 95

platform tools, 95–101

issues, 103

baksmali, 113–115

database schemas, 105

dedexer, 112–113

Dexdump, 109–112

disassemblers, 107

Dx, 109

fake apps, 106

hex editors, 107–108

HTML5/CSS, 106

web service keys and logins,
104–105

obfuscators

Crema, 143–144

DashO, 145–146

APK, obfuscators (*cont.*)
 JavaScript obfuscators, 146–149
 ProGuard, 144–145
 source code protection
 description, 119–120
 fingerprinting your code, 138–140
 native methods, 140–142
 non-obfuscation strategies, 142–143
 obfuscation, 121–138
 web services, 138
 writing two versions, 120–121
 unzipped, 102
 YUI compressor, 147–149
 zipped format, 101
Another Tool for Language Recognition (ANTLR)
 description, 166
 DexToXML, 167–168
 plug-in for Eclipse, 166–167
Apktool, 119

■ B

Ball
Baksmali, 113–115

■ C

Casting.java
 Android bytecode analysis, 198–201
 Casting.ddx, 197–198
 code, 196
 Java, 211–212
 parser, 201
 Casting.ddx Parser, 207–211
 for Loop Parser, 204–206
 Without Bytecode Parser, 201–203
 Without Pytecode, 203
Class file

access flags, 38
attributes and attributes count, 55
casting.class, 25
casting.java, 24
constant pool, 29
 count, 28
 cp_info Structure, 29
 field descriptors, 37
 for Casting.class, 30–36
 tags, 29
 Utf8 structure, 30
field attributes, 43–44
fields and field count, 43
 Casting.java field information, 41–42
 field access flag names and values, 42
 field_info data structure, 41
interfaces and interface count, 39–41
magic number, 27
method attributes
 Code Attribute, 48–49
 description, 48
 <init> method, 49–51
 init method attributes, 48
 main method, 51–55
methods and method count
 access flags, 46
 Casting.class method information, 44–46
 method_info structure, 44
 method name and descriptor
 constant-pool information, 47
minor and major version
 numbers, 28
parts, 26
struct, 26
superclass, 39
this class, 39
XML representation, 27

- Code fingerprinting, 16
- Computation obfuscation
 - dead or irrelevant code insertion, 127
 - extending loop conditions, 128
 - parallelizing code, 131
 - programming idioms removal, 130
 - reducible to non-reducible transformation, 128
 - redundant operands, 129
- Construction of Useful Parsers (CUP)
 - Decompiler.lex, 164
 - description, 160
 - example, 157
 - main method bytecode, 164
 - Parser.CUP, 162
 - Partial CUP Debug Output, 165
 - sections, 160
 - declaration, 161
 - grammar rules, 164–165
 - list of symbols and tokens, 162–164
 - user routines, 161–162
- Control obfuscations
 - aggregation, 131
 - cloning methods, 133
 - inlining and outlining methods, 132
 - interleaving methods, 132–133
 - loop transformations, 133
- classifications, 127
- computation
 - dead or irrelevant code insertion, 127
 - extending loop conditions, 128
 - parallelizing code, 131
 - programming idioms removal, 130
 - reducible to non-reducible transformation, 128
 - redundant operands, 129

- description, 127
- ordering, 134
 - reordering expressions, 134
 - reordering loops, 135

Copyright law, 13

CUP. *See* Construction of Useful Parsers (CUP)

■ D, E

Dalvik executable (DEX) file, 6

- APK file, 57
 - casting.class conversion, 59–60
 - classes.dex file, 58
 - class file vs. DEX file, 58
 - unzipped, 58
- class_defs section
 - access flags, 83–84
 - classes.dex, 82
 - description, 81
 - DexToXML, 83
 - struct, 82
- data section
 - class_data_item, 85–88
 - code_item, 88–92
 - description, 85
- DexToXML, 62
- DVM, 57
- field_ids section
 - classes.dex, 77
 - fields information, 78
 - struct, 76
- header section
 - checksum, 67
 - classes.dex, 65
 - description, 62
 - DexToXML Output, 65–66
 - Endian_tag, 67
 - fields, 63–65
 - Header_size, 67
 - magic number, 66
 - struct, 62–63

DEX file (*cont.*)

- method_ids section
 - classes.dex, 79
 - DexToXML, 79–81
 - methods, 81
 - struct, 78

- parts, 61

- proto_ids section
 - classes.dex, 74
 - data section, 76
 - description, 74
 - DexToXML, 75–76
 - struct, 74

- specification, 57

- string_ids section
 - classes.dex, 67–68
 - data Section, 71
 - DexToXML, 68–71

- struct, 61

- type_ids section, 71
 - classes.dex, 72
 - data Section, 73
 - DexToXML, 72–73

Dalvik virtual machine (DVM)

- APKs, 6
- backsmali, 113
- DEX file, 57

DashO, 247

- control flow option, 248
- description, 145
- GUI, 145, 248
- output
 - description, 249
 - JD-GUI output, 249–251
- protected code, 146
- wizard, 249

Data obfuscations

- aggregation, 136
 - array transformations, 137
 - class transformations, 137
 - merging scalar variables, 137
- classifications, 135
- ordering, 137

- storage and encoding
 - changing encoding, 135–136
 - splitting variables, 136
 - static to procedural data conversion, 136

Data section, DEX file

- class_data_item
 - classes.dex, 86
 - DexToXML, 87–88
 - encoded_field, 86
 - encoded_method, 86
 - static field and method information, 88
 - struct, 85
 - Uleb128, 85

- code_item
 - classes.dex, 89
 - DexToXML, 90–91
 - struct, 88

- description, 85

Dcc decompilers, 10

Decompilation, 17

Decompilers

Android application

- APKs, 6
- backend systems via web services, 6
- DEX file, 6
- Dexdump Output, 7–8
- DVM, 6
- ProGuard, 6
- reasons for vulnerable, 7

- apktool, 119

defining the problem

- casting bytecode, 152–153
- DexToSource parser, 154
- opcodes, 153

- description, 2

design

- parser design, 169–173
- theory, 152

- dex2jar, 118

- DexToSource, 151

- history of, 8
 - academic decompilers, 10
 - ALGOL, 9
 - dcc, 10
 - Hanpeter van Vliet and Mocha, 11
 - pirate software, 9
 - reverse-engineering techniques, 9
 - VB, 10–11
- implementation, 175
 - DexToSource (*see* DexToSource)
 - DexToXML (*see* DexToXML functions)
 - refactoring, 223–227
- Jad, 116–117
- JD-GUI, 117–118
- JVM
 - description, 3
 - Javap Output, 5
 - JDK, 3
 - reasons for vulnerable, 4
 - Simple Java Source Code, 4
 - specification, 4
- legal issues
 - ground rules, 12
 - protection laws, 12–14
- Mocha, 115
- moral issues, 15–16
- opcode
 - definition, 152
 - types, 153
- pirated software, 16–17
- protection
 - code fingerprinting, 16
 - encryption, 17
 - IPR protection schemes, 16
 - license agreements, 16
 - native code, 17
 - obfuscation, 16
 - pirated software, 16
 - schemes, 16
 - server-side code, 17
- tools
 - ANTLR, 165–168
 - compiler-compilers, 154
 - CUP, 160–165
 - JLex, 157–160
 - Lex and Yacc, 155–156
 - types, 154
- types, 2
- undx, 118
- virtual machine, 3
- Dedexer, 112–113
- dex.log file, 176
- Header of the Class, 176–177
- magic number
 - ANTLR Magic-Number Parser, 178–179
 - DexToXML ANTLR Grammar, 184–195
 - DexToXML Magic-Number Parser, 180
 - DexToXML.java, 181
 - header rule, 182
 - parsing output, 177–178
 - parsing rules, 180
 - Refactored DexToXML Header Grammar, 183–184
 - Refactored header_entry Rule, 183
 - tokenized, 179
- Dex2jar, 118
- Dexdump, 109–112
- DexToSource
 - Casting.java
 - Android bytecode analysis, 198–201
 - Casting.ddx, 197–198
 - code, 196
 - Java, 211–212
 - parser, 201–211
 - description, 151, 196
 - Hello World application

DexToSource, Hello World application
(*cont.*)

- Android bytecode analysis,
213–214

- Android screen, 212

- HelloWorld.ddx, 213

- Java, 216

- parser, 214–216

- if statement, 217

- Android bytecode analysis,
218–220

- escapeHTML.ddx, 217

- escapeHTML Method, 217

- Java, 223

- parser, 220–222

- parser, 154

DexToXML functions

- description, 176

- dex.log output parsing

- ANTLR, 178–179

- Header of the Class, 176–177

- magic number, 177–178, 180–
195

- rules, 179

Digital Millennium Copyright Act
(DMCA), 12

Disassemblers, 107

DVM. *See* Dalvik virtual machine
(DVM)

Dx command, 109

■ F, G

Fingerprinting your code

- description, 138

- digital-fingerprinting system
criteria, 140

■ H

Heap analysis tool (HAT), 22–23

Hexadecimal editors

- description, 107

- IDA, 108

Timebombed Trial App Code, 108

■ I

If statement

- Android bytecode analysis, 218–
220

- escapeHTML.ddx, 217

- escapeHTML Method, 217

- Java, 223

- parser, 220–222

Intellectual Property Rights (IPR)
protection schemes, 16

■ J, K

Jad decompilers, 116–117

Java

- Casting.java, 211–212

- decompiler, 171

- Hello World application, 216

- if statement, 223

JavaScript obfuscators, 146–149

Java virtual machine (JVM)

- block diagram, 22

- class file

- access flags, 38

- attributes and attributes

- count, 55

- casting.class, 25

- casting.java, 24

- constant pool count, 28–37

- field attributes, 43

- fields and field count, 41

- interfaces and interface

- count, 39–41

- magic number, 27

- method attributes, 48–55

- methods and method count,
44–48

- minor and major version

- numbers, 28

- parts, 26

- struct, 26

- superclass, 39
 - this class, 39
 - XML representation, 27
- description, 19–20
- design, 20–21
- Javap Output, 5
- JDK, 3
- reasons for vulnerable, 4
- simple java source code, 4
- simple stack machine
 - heap, 22–23
 - JVM stack, 24
 - method area, 23
 - parts, 21
 - PC registers, 23
- specification, 4, 20
- JD-GUI decompiler, 117–118
- JLex compiler
 - example, 157
 - sections
 - directives, 158–159
 - regular-expressions, 159–160
 - user code, 158

■ L

- Layout obfuscations
 - Crema-Protected Code, 125–126
 - description, 125
 - Operator Overloading, 126
- Lex and Yacc tool
 - description, 155
 - LALR(1) parser, 156
 - LL(k) parsers, 156
 - Sed and Awk, 156
 - tokens, 155

■ M, N

- Magic number, dex.log
 - ANTLR Magic-Number Parser, 178–179
 - DexToXML ANTLR Grammar, 184–195

- DexToXML Magic-Number Parser, 180
 - DexToXML.java, 181
 - header rule, 182
 - parsing output, 177–178
 - parsing rules, 180
 - Refactored DexToXML Header Grammar, 183–184
 - Refactored header_entry Rule, 183
 - tokenized, 179
- Mocha decompiler, 115
- Myths, Android, 230–231

■ O

- Obfuscation
 - case study, 230
 - code, 230
 - control, 127–135
 - data, 135–137
 - decompilers, 16
 - description, 122
 - JVM, 124
 - layout, 125–127
 - techniques, 138
 - transformations types, 122–124
- Obfuscators
 - Crema, 143–144
 - Dash0, 145–146
 - JavaScript obfuscators, 146–149
 - ProGuard, 144–145
- Opcode
 - definition, 152
 - types, 153
- Ordering obfuscation
 - reordering expressions, 134
 - reordering loops, 135

■ P, Q

- Parser design
 - Casting.java
 - Casting.ddx Parser, 207–211

- Parser design, Casting.java (*cont.*)
 - for Loop Parser, 204–206
 - Without Bytecode Parser, 201–203
 - Without Pytecode, 203
- Casting.smali Method, 173
- Hello World application, 214–216
- identifiers, 173
- if statement, 220–222
- integers, 173
- keywords, 173
- native format, 172
- strategy
 - AST, 171–172
 - benefits, 169
 - choice one, 171
 - choice three, 171–172
 - choice two, 171
 - disadvantages, 170
 - final decompiler design, 169
 - StringTemplates, 169
- token types, 173
- whitespace, 173
- Patent law, 13
- Platform tools, APK
 - description, 95
 - installation and usage, 99–101
 - rooting, 96–99
- Z4Root
 - disabling root, 99
 - installation, 96–97
 - temporary or permanent root, 98
- Program counter (PC) registers, 23
- ProGuard, 6, 231
 - configuration
 - default, 245
 - GUI, 246
 - proguard.cfg file, 245
 - debugging, 247
 - double-checking your work
 - Obfuscated t.java, 237–238

- obfuscated WordPress jar file, 243–244
- Original EscapeUtils.java Code, 234–236
- r.java Class, 238–243
- Unobfuscated
 - EscapeUtils.java, 236–237
- SDK output, 232–234
- Protection laws, decompilers
 - copyright, 13
 - description, 12
 - DMCA, 12
 - fair use, 12
 - Legal Protection of Computer Programs, 13
 - patents, 13
 - reverse engineering, 14

■ R

- Refactoring
 - opcode classifications, 223
 - refactored parser, 224–227
- Reverse engineering, 14
- Reverse-engineering techniques, 9

■ S

- Server-side code, 17
- Simple stack machine
 - heap, 22–23
 - JVM stack, 24
 - method area, 23
 - parts, 21
 - PC registers, 23

■ T

- Tools
 - backup tool, APK, 94–95
 - decompilers
 - ANTLR, 165–168
 - CUP, 160–165
 - JLex, 157–160

Lex and Yacc, 155–156

HAT, 22–23

platform, APK

description, 95

installation and usage, 99–101

rooting, 96–99

Z4Root, 96

■ U

Undx converter, 118

■ V

Visual Basic (VB), 10–11

■ W, X

Web services, APK, 138

■ Y

YUI compressor, 147–149

■ Z

Z4Root

disabling root, 99

installation, 96–97

temporary or permanent root, 98