

Index

Numerics

32-/64-bit architecture support, effect on application performance, 68

A

accelerator compression architectures

per-packet compression, 282–283
session-based compression, 284–286

accelerator TCP proxy functionality, 267–269

accelerators, 9

functions of, 112

message prediction, 116
object caching, 112–114
read ahead, 114
WAFS, 116
write-behind, 115

logical integration, 118–121

physical integration, 117

service architecture, 121

nontransparent accelerators, 122
transparent accelerators, 122, 125

technology overview, 106

compression, 111
data suppression, 110–112
WAN optimization, 107–109

ACL administration, centralized control, 191–192

Acme Data Corporation scenario, 292–302

acquiring content, 202–203

content size considerations, 205–206

cookie-based, 204

department-managed portals, 207

distribution models

Direct Fetch, 207–210
distributed hierarchy, 210–215

method of, selecting, 174–178

origin server content placement, 204–205

advanced congestion avoidance algorithms, overcoming TCP packet loss, 261

advanced TCP implementations, 261–263

BIC-TCP, 266

High-Speed TCP, 264–265

S-TCP, 265

Almost Write Inc. scenario, 314–319

application acceleration, 9

application chatter, 4

Application layer (OSI model), 30

application protocols

effect on application performance

CIFS, 51–52

FTP, 54–55

HTTP, 52–54

MAPI, 56–57

NFS, 55–56

RPC, 56–57

effect on latency, CIFS, 42–43

applications

latency, reducing, 5–6

managing in distributed workforce, 3–4

performance-limiting factors, 24–25

32-/64-bit architecture support, 68

application protocols, 51–57

bandwidth, 25–31

cache capacity, 67

CPU, 67

disk storage, 70–73

dual-core support, 67

file systems, 74–79

front side bus speed, 67

hyper-threading, 68

latency, 33–43

network stability, 57–58, 61

NICs, 79–80

operating systems, 61–67
RAM, 69–70
throughput, 45, 48–50
targeting for accelerators, 10
testing, 4

application-specific caching, 129, 140

advantages of, 129–130
cache validation, 130, 133
CIFS, 134–137
content freshness, 130, 133
HTTP, 142

IMS response, 142–143

HTTPS, 143–144

SSL, 145–146

TLS, 146

message prediction, 157–159
multiplexing, 159–160
opportunistic locks, 138
pipelining, 140
streaming media, 147–149
write-back optimization, 141

application-specific read-ahead, 154–157

asynchronous replication, 15

authentication

centralized control, 189–191
of requests for prepositioned content,
200–201

B

bandwidth

cost of, 238–239
effect on application performance, 25–31
fairness, 259

RTT fairness, 263

managing on CDNs, 183–184

bandwidth discovery, 246–247

TCP congestion avoidance, 249–250
TCP slow start, 247–248

batch opportunistic locks, 137

BDP (bandwidth delay product), 49, 255

**BIC-TCP (Binary Increase Congestion TCP),
261, 266, 348**

bidirectional compression library, 287

branch accelerators, 155

business continuance, 17–18

byte range requests, 151

C

C3 Technology LLC scenario, 302–307

**cache capacity, effect on application
performance, 67**

caching, 128

application-specific, 129, 140

advantages of, 129–130

cache validation, 130, 133

CIFS, 134–137

content freshness, 130, 133

HTTP, 142–143

HTTPS, 143–146

message prediction, 157–159

multiplexing, 159–160

opportunistic locks, 138

pipelining, 140

streaming media, 147–149

write-back optimization, 141

calculating

BDP, 49, 255

content delivery times, 234

CDNs (content delivery networks), 163–165

bandwidth, sharing, 183–184

centralized control, 188

ACL administration, 191–192

*authentication and authorization,
189–191*

SNMP, 192

streaming media, 189

- centralized edge management, 193–194
- centralized monitoring, SNMP, 192–193
- components of, 168–172
- content acquisition, 202–203
 - content size considerations*, 205–206
 - cookie-based content acquisition*, 204
 - method of, selecting*, 174–178
 - origin server content placement*, 204–205
- costs, managing, 180
- department-managed portals, 207
- desktop management suites, 184
 - management functions, combining*, 185–187
 - storage requirements, establishing*, 187
- distribution models
 - Direct Fetch*, 207–210
 - distributed hierarchy*, 210–215
- evolution of, 164–165
- general content storage, 232
- managing, 173–175
 - across multiple platforms*, 178–179
- proxy modes, 230–231
- software-based, 224–226
 - streaming media, executive demand*, 228–230
 - native protocol playback*, 226–228
- streaming media storage, 233–234
- time-of-day distribution, 215–217
 - multicast distribution*, 218–223
 - unicast distribution*, 218
- usage planning, 181–183
- centralized management, 8**
- checksum, 242**
- checksum verification, 38**
- CIFS (Common Internet File System), 9, 45, 134–137, 194–195**
 - effect on application performance, 51–52
 - opportunistic lock, 136–138
 - read-ahead optimization, 155
- classification, 94**
- codebook compression, 110**
- Command Grill Corporation scenario, 308–314**
- compression, 111**
 - accelerator compression architectures
 - per-packet compression*, 282–283
 - session-based compression*, 284–286
 - data suppression, overcoming TCP link capacity limitations, 274–275, 281
 - directionality, 287
 - TCP link capacity limitations, overcoming, 271
 - techniques, leveraging, 239
 - traditional data compression, overcoming TCP link capacity limitations, 274
- compression domain, 274**
- congestion avoidance, 46**
 - TCP, 249–250
- connection establishment, TCP, 242**
- connectionless transport protocols, 25**
- connection-oriented services, 241**
- connections (TCP), mice, 251–253**
- content, repositioning, 165–166**
- content acquisition, 202–203**
 - content size considerations, 205–206
 - cookie-based, 204
 - department-managed portals, 207
 - distribution models
 - Direct Fetch*, 207–210
 - distributed hierarchy*, 210–215
 - method of, selecting, 174–178
 - origin server content placement, 204–205
- content crawling, 175–176**
- content delivery times, calculating, 234**
- content distribution scenario, 166–168**
- content request read-ahead, 157**
- content-serving protocols**
 - CIFS, 194–195
 - FTP, 197
 - HTTP, 195–196
 - HTTPS, 196–197
 - RSTP, 197–199
 - TFTP, 199
- cookie-based content acquisition, 204**
- core accelerators, 155**
- cost of bandwidth, comparing LANs and WANs, 238–239**
- CPU, effect on application performance, 67**
- cwnd (congestion window), 47**

D

- DAS (direct-attached storage), 70**
- data link layer (OSI model), 30**
- data pattern matching, 278**
- data protection, 15–17**
- data replication, 15**

data suppression, 110–112
 directionality, 287
 TCP link capacity limitations, overcoming, 274–275, 281

database applications, web-based, 149–154

DDR SDRAM, 69

decoding data, 278–280

delay, 33
 forwarding delay, 42
 processing delay, 41
 propagation delay, 41
 serialization delay, 41

department-managed portals, 207

DiffServ, 98–99

Direct Fetch, 207–210

directionality, 287

disaster recovery, 17–18

disk storage, effect on application performance, 70–72
 Fiber Channel, 73
 SAS, 73

distributed environments
 server components, 12–13
 services, 11

distributed hierarchy distribution model, 210–215

distributed servers
 data, protecting, 6–7
 managing, 6

distribution models
 Direct Fetch, 207–210
 distributed hierarchy, 210–215

dual-core support, effect on application performance, 67

dynamic message prediction, 116

E

elephants, 251

encoding data, 278

encrypted content distribution, 223–224

enterprise applications, optimizing, 154

exclusive opportunistic locks, 137

F

fairness, 259
 RTT fairness, 263

FEC (forward error correction), 223
 TCP packet loss, overcoming, 260

Fiber Channel, effect on application performance, 73

FIFO queuing, 101

file systems, effect on application performance, 74
 HP UX, 76
 Linux, 76–79
 Microsoft Windows Server 2003, 75
 Solaris, 74–75

fill-the-pipe optimization, 257
 scalable TCP implementation method, 258
 window scaling method, 257

flow control, 39

forwarding delays, 42

front side bus speed, effect on application performance, 67

FTP, 197
 effect on application performance, 54–55

G-H

general content storage, 232

guaranteed delivery, 242–244

hard QoS, 97

hardware, effect on application performance
 32-/64-bit architecture support, 68
 cache capacity, 67
 CPU, 67
 disk storage, 70–73
 dual-core support, 67
 file systems, 74–79
 front side bus speed, 67
 hyper-threading, 68
 NICs, 79–80
 RAM, 69–70

hierarchical data pattern matching, 275

HP-UX, effect on application performance, 65–66

HS-TCP (High-Speed TCP), 264–265

HTTP, 195–196

- application-specific caching, 142
 - IMS response, 142–143*
- content crawling, 175–176
- effect on application performance, 52–54

HTTPS, 196–197

- application-specific caching, 143–144
 - SSL, 145–146*
 - TLS, 146*

hyper-threading, effect on application performance, 68**I****IANA (Internet Assigned Numbers Authority)**

- port assignments, 323–325, 331
- registered ports, 331–345

IBM operating systems, effect on application performance, 66–67**identifying**

- need for accelerators, 10
- network components, 58
- non-network components, 58

IMS response (HTTP), 142–143**IntServ, 97–98****IP flows, 86****J-K-L****journaling, 76–77****LANs, cost of bandwidth, 238–239****Large Initial Windows, 253–254****latency**

- effect on application performance, 33–41
- of application protocols, effect on
 - application performance, 42–43
- of applications, reducing, 5–6
- transport protocols, effect on, 40

LBFS (Low Bandwidth File System), 348**Level II opportunistic locks, 137****leveraging compression techniques, 239****LFNs (long and fat networks), 48, 251, 255**

- fill-the-pipe optimization, 257
 - scalable TCP implementation method, 258*
 - windows scaling method, 257*

link capacity limitations of TCP, overcoming, 270

- with compression, 271
- with data suppression, 274–275, 281
- with traditional data compression, 274

live streaming, 147, 199–200**logical integration of accelerators into network, 118–121****M****managing**

- applications in distributed workforce, 3–4
- CDNs , 173–175
 - bandwidth, 183–184*
 - costs, 180*
 - usage planning, 181–183*
- distributed servers, 6

MAPI, effect on application performance, 56–57**message prediction, 116, 157–159****message validity signature, 278****messages**

- TCP ACK, 35, 38
- TCP SYN, 35

mice connections, 251

- TCP slow start, overcoming performance
 - limitations, 252–253

Microsoft Internet Explorer, 5**Microsoft Outlook, 5****Microsoft TechNet, 348****Microsoft Windows, effect on application performance, 61**

- NetFMA, 63
- receive-side scaling, 63
- TCP Chimney Offload, 62

monitoring

- CDNs, centralized control, 192–193
- network devices
 - SNMP, 59*
 - syslog, 60*

multicast content distribution, 218–220, 223

- unusable multicast address ranges, 220–222

multiplexing, 159–160

N

NBAR (Network Based Application Recognition), viewing network utilization, 90–92

NetDMA, effect on Microsoft Windows performance, 63

NetFlow, viewing network utilization, 84–87, 90

NetFlow collector, 87, 90

network devices, monitoring
with SNMP, 59
with syslog, 60

network interception, 118

Network layer (OSI model), 30

network stability, effect on application performance, 57–58, 61

network utilization, viewing
with NBAR, 90–92
with NetFlow, 84–87, 90

NFS (Network File System), 45
effect on application performance, 55–56

NICs, effect on application performance, 79–80

nonhierarchical data pattern matching, 275

nonstrict content coherency, 132–133

nontransparent accelerators, 122

O

object caching, 112–114

one-way latency, 33

operating systems, effect on application performance, 61
HP-UX, 65–66
IBM, 66–67
Microsoft Windows, 61–63
Red Hat Linux, 64–65
Sun Microsystems, 63–64

opportunistic locks, 136–138

origin server content placement, 204–205

OSI reference model, 29

oversubscription, 27

P

packet classification, 95–96

packet loss (TCP), overcoming, 259–261

packet marking
DiffServ, 98–99
IntServ, 97–98

pass-through authentication, 190

PBR (Policy-Based Routing), 119

peer-to-peer sharing, 94

performance
factors affecting applications in WAN environments, 24–25
32-/64-bit architecture support, 68
application protocols, 51–57
bandwidth, 25–31
cache capacity, 67
CPU, 67
disk storage, 70–73
dual-core support, 67
file systems, 74–79
front side bus speed, 67
hyper-threading, 68
latency, 33–43
network stability, 57–58, 61
NICs, 79–80
operating systems, 61–67
RAM, 69–70
throughput, 45, 48–50
limitations of TCP slow start, overcoming, 252–253
vendor-specific test results, obtaining, 60

per-packet compression, 282–283

persistent compression, 284

physical inline, 117

physical integration of accelerators into network, 117

physical layer (OSI model), 30

pipelining, 140, 159–160

policing, 99

port assignments, IANA, 323–331

post-queuing, 95
optimization, 105–106

prediction, 116

prepositioned content, 113, 157
authenticating requests for, 200–201

pre-queuing operations, 95–96
packet marking, 97–99
traffic policing, 99

presentation layer (OSI model), 30

priority queuing, 102

proactive acceleration, 128

processing delays, 41

propagation delay, 41
protecting data on distributed servers, 6–7
proxy functionality, accelerator TCP proxy, 267–269
PSH flag (TCP), 244

Q

QoS, 92
 behavioral model, 95
 packet classification, 95–96
 post-queuing optimization, 105–106
 pre-queuing operations, 96
 packet marking, 97–99
 traffic policing, 99
 soft QoS, 98
queuing, 95, 100
 FIFO, 101
 post-queuing optimization, 105–106
 priority queuing, 102
 traffic shaping, 104
 WFQ, 103–104

R

RAID (Redundant Array of Inexpensive Disks), 71
RAM, effect on application performance, 69–70
range requests, 151
rate limiting, 104
rate-based transmission versus large initial windows, 253–254
reactive acceleration, 128
read ahead, 114, 154
read-ahead acceleration, 155–157
receive-side scaling, effect on Microsoft Windows performance, 63
Red Hat Linux, effect on application performance, 64–65
reducing application latency, 5–6
redundancy planning, 60–61
registered ports, 331–345
Reiser4 file system, 79
replication, 15
retransmission management (TCP), 245–246
RFC (Requests for Comments), 348

roundtrip latency, 33
RPC (Remote Procedure Call), 45
 effect on application performance, 56–57
RSTP (Real-Time Streaming Protocol), 197–199

S

SACK (selective acknowledgment), overcoming TCP packet loss, 260
SANs (storage area networks), 70
SAS (serial attached SCSI), effect on application performance, 73
scenarios
 Acme Data Corporation, 292–302
 Almost Write Inc., 314–319
 C3 Technology LLC, 302–307
 Command Grill Corporation, 308–314
scheduling, 95, 101
SDRAM, 69
selecting
 advanced TCP implementations, 261–263
 BIC-TCP, 266
 High-Speed TCP, 264–265
 S-TCP, 265
 content acquisition method, 174–178
serialization delay, 41
server consolidation, 11
 compliance, 15–17
 data protection, 15–17
servers
 centralized management, 8
 components, 12–13
 infrastructure consolidation, 13–14
 managing in distributed environment, 6
service architecture of accelerators, 121
 nontransparent accelerators, 122
 transparent accelerators, 122, 125
services in distributed environments, 11
session layer (OSI model), 30
session-based compression, 284–286
share mode, 137
SLB (server load balancers), 120
sliding windows, 39, 244
 Large Initial Windows, 253–254
slow start (TCP), 247–248
slow-start threshold, 247

SNMP (Simple Network Management Protocol)

- centralized control, 192
- monitoring network devices, 59

sockets, 241**soft QoS, 98****software-based CDNs, 224–226**

- streaming media
 - executive demand*, 228–230
 - native protocol playback*, 226–228

SSL (Secure Sockets Layer), 145–146**stabilizing client/server network connections, 61****static message prediction, 116****S-TCP (Scalable TCP), 265****streaming media, 147–149**

- centralized control, 189
- executive demand*, 228–230
- live streaming, 199–200
- native protocol playback*, 226–228
- VoD content, 148, 200

streaming media storage, 233–234**strict content coherency, 132–133****Sun Microsystems operating systems, effect on application performance, 63–64****SYN/ACK messages, 35****synchronous replication, 15****syslog, monitoring network devices, 60****T****TCP, 240**

- accelerator proxy functionality, 267–269
- ACK messages, 35, 38
- advanced implementations, 261–263
 - BIC-TCP*, 266
 - High-Speed TCP*, 264–265
 - S-TCP*, 265
- bandwidth, fairness, 259
- blocking condition, 243
- checksum verification, 38
- congestion avoidance, 48, 249–250
- connection establishment, 35
- cwnd, 47
- effect on latency, 34–41
- effect on throughput, 46
- flow control, 39

functions

- bandwidth discovery*, 246–250
- connection-oriented service*, 241
- guaranteed delivery*, 242–244
- large initial windows, 253–254
- LFNs, 255
 - fill-the-pipe optimization*, 257–258
- link capacity limitations, overcoming, 270
 - compression*, 271
 - with data suppression*, 274–275, 281
 - with traditional data compression*, 274

mice connections, 251

optimizing, 107–109, 237

packet loss, overcoming, 259–261

PSH flag, 244

rate control, 271

retransmission management, 245–246

sliding window, 244

SYN messages, 35

throttling mechanism, 246

TCP Chimney Offload, effect on Microsoft Windows performance, 62**TCP proxy, 108****TCP slow start, 48, 247–248**

- performance limitations, overcoming, 252–253

testing applications, 4**TFTP, 199****throttling mechanism, TCP, 246****throughput, effect on application performance, 45, 48–50****TLS (Transport Layer Security), 146****TOE (TCP/IP Offload Engine), 79****traditional data compression, overcoming TCP link capacity limitations, 274****traffic policing, 99****traffic shaping, 104****transparent accelerators, 122, 125****transport layer (OSI model), 30, 238****transport protocols, 31, 238**

- effect on latency, 40
- effect on throughput, 46
- optimizing, 239
- TCP, 240. *See also* TCP
 - advanced implementations*, 261–263
 - bandwidth discovery*, 246–250
 - BIC-TCP*, 266
 - blocking condition*, 243
 - congestion avoidance*, 48

- connection-oriented service, 241*
- effect on latency, 34–41*
- fairness, 259*
- guaranteed delivery, 242–244*
- High-Speed TCP, 264–265*
- large initial windows, 253–254*
- LFNs, 255–258*
- mice connections, 251–253*
- PSH flag, 244*
- retransmission management, 245–246*
- sliding window, 244*
- S-TCP, 265*
- throttling mechanism, 246*
- UDP, 251

- window scale factor, 257**
- window scaling, 257**
- WRED (Weighted Random Early Detection), 104**
- write-back optimization, 141**
- write-behind, 115**

U-V

- UDP (User Datagram Protocol), 251
- unicast distribution, 218
- unidirectional compression library, 287
- UNIX HP-UX, effect on application performance, 65–66
- utopian environments versus real-world enterprise WANs, 22
- viewing network utilization
 - with NBAR, 90–92
 - with NetFlow, 84–87, 90
- VoD (video on demand), 200
 - content streaming, 148

W-X-Y-Z

- WAFS (Wide Area File Services), 116
- WAN optimization, 9, 107
 - compression, 111
 - data suppression, 110–112
 - TCP optimization, 107–109
- wandering log, 79
- WCCPv2 (Web Cache Control Protocol v2), 118
- web browsers as business-critical applications, 9
- web-based database applications, 149–154
- websites, Hewlett-Packard, 349
- WFQ (Weighted Fair Queuing), 103–104
- WHQL (Windows Hardware Quality Labs), 62