EtherChannel - 802.3ad

Pavel Jeníček, JEN022
Martin Milata, MIL051

Ostrava, 1.06.2005

Úvod:

EhterChanel a jedna z jeho konkrétních implementací protokol LACP (Link Aggregation Control Protocol) je protokol pro agregaci spojů na druhé vrstvě ISO/OSI modelu pro zvýšení šířky pásma.

Norma IEE802.3ad specifikuje principy agregace spojů, které dosud řešili výrobci síťových zařízení každý po svém (Cisco – PAgP). Agregace je vhodná pro sdružení více fyzických spojů, typu Ethernet nebo FastEthernet, případně GigabitEthernet, do jediného logického spoje. Výsledkem je nejen zvýšení šířky pásma, ale i zvýšení spolehlivosti pro případ poruch na některém spoji LAN.

Norma se týká nejen přepínačů, ale i síťových karet (NIC). Typický případ komunikace mezi přepínačem a serverem dnes mnohdy převyšuje šířku pásma několik set Mbit/s. Při použití více karet v serveru chrání před poruchami jak na jednotlivé kartě, tak na kabelu.

EtherChannel:

Protokol EtherChannel, jak již bylo řečeno výše, umožnuje agregaci několika fyzických spojů v jeden spoj logický. Například eth0, eth1 můžou být agregovány v jediný logický adaptér eth2. Tento adaptér pak dostane přiřazenu IP adresu a lze s ním pracovat jako s jakoukoliv jinou síťovou kartou. Všechny adaptéry v EtherChannelu, stejně jako výsledný logický adaptér, mají přiřazenu stejnou MAC adresu. Na vnější pohled se tváří jako jediné síťové rozhraní a pro vyšší vrstvy je tedy agregace zcela transparentní.

Největší výhodou technologie EtherChannel nebo také IEE802.3ad je, že virtuální kanál disponuje přenosovým pásmem, které je součtem pásem všech kanálů, které se na agregaci podílejí.

Podpora EtherChannel na přepínačích Cisco:

EtherChannel a IEE802.3ad agregace je povolena na následujících Ethernet adaptérech:

- 10/100 Mbps Ethernet PCI Adapter
- Universal 4-Port 10/100 Ethernet Adapter
- 10/100 Mbps Ethernet PCI Adapter II
- 10/100/1000 Base-T Ethernet PCI Adapter
- Gigabit Ethernet-SX PCI Adapter
- 10/100/1000 Base-TX Ethernet PCI-X Adapter
- Gigabit Ethernet-SX PCI-X Adapter
- 2-port 10/100/1000 Base-TX Ethernet PCI-X Adapter
- 2-port Gigabit Ethernet-SX PCI-X Adapter

Pouze základní funkcionalita (v módu "standard" nebo "round-robin") je podporována na těchto Ethernet adaptérech:

- PCI Ethernet BNC/RJ-45 Adapter
- PCI Ethernet AUI/RJ-45 Adapter

Adaptéry, které patří do EtherChannelu musí být připojeny k jednomu přepínači, který povoluje na svých portech spustit protokol LACP (PAgP). Jednotlivé porty musí být manuálně konfigurovány jako porty, které tvoří jeden EtherChannel.

Síťový provoz je rozdělován mezi porty na základě algoritmu round-robin (pakety jsou rovnoměrně posílány přes všechny porty), lze také použít load balancing a to podle zdrojové nebo cílové MAC adresy, na vyšších verzích přepínačů např. i podle typu služby nebo podle protokolu 4 vrstvy.

Konfigurace EtherChannelu:

Konfiguraci EtherChannelu je možno provést manuálně nebo použít protokol Port Aggregation Control Protocol (PAgP – proprietární řešení od Cisca) nebo, od Cisco IOS verze 12.1 EW, standardizovaný LACP (Link Aggregation Protocol). Tyto protokoly umožnují vytvořit EtherChannel z portů se stejnou konfigurací (rychlost, full-duplex...)

Módy portů pro EtherChannel:

Mód	***Popis***
auto	PAgP mód, který nastaví LAN port do pasivního stavu, ve kterém port odpovídá na PAgP pakety, které přijme, ale sám neiniciuje spojení
desirable	PAgP mód, který nastaví LAN port do aktivního stavu, ve kterém port iniciuje spojení s ostatními LAN porty posíláním PAgP paketů
passive	LACP mód, který nastaví LAN port do pasivního stavu, ve kterém port odpovídá na LACP pakety, které přijme, ale sám neiniciuje spojení
active	LACP mód, který nastaví LAN port do aktivního stavu, ve kterém port iniciuje spojení s ostatními LAN porty posíláním LACP paketů

PAgP umožnuje automatické vytvoření EtherChannelu výměnou PAgP paketů mezi LAN porty. PAgP pakety jsou posílány pouze mezi porty v **auto** a **desirable** módu. Podobně je to i s protokolem LACP.

LAN porty mohou vytvořit EtherChannel v těchto případech (platí obdobně i pro LACP):

- LAN port v desirable módu může vytvořit EtherChannel s jiným portem, který je také v desirable módu
- LAN port v desirable módu může vytvořit EtherChannel s jiným portem, který je v auto módu
- LAN port, který je v auto módu nemůže vytvořit EtherChannel s jiným portem, který je také v auto módu, protože žádný z portů neiniciuje spojení

LoadBalancing:

EtherChannel může provádět vyvažování zátěže algoritmy, které určují, který z portů se použije pro přenos daného paketu. Jako kritérium lze použít MAC adresy, IP adresy (u L3 switchů) a čísla portů.

Konfigurace přepínače CiscoCatalyst 2950:

Switch(config)# interface fastEthernet 0/1
Switch(config-if)# channel group 1 mode active
Switch(config-if)# switchport mode access
Switch(config-if)# no shutdown
Switch(config-if)# end

Tímto je nakonfigurován port 1 pro příjem i vysílání LACP paketů. Takto byl nakonfigurován i port 2 a to na obou směrovačích CiscoCatalyst 2950.

SW4#show running-config

interface FastEthernet0/1

```
 switchport mode access
 speed 100
 duplex full
 channel-group 1 mode active
!
interface FastEthernet0/2
 switchport mode access
 speed 100
 duplex full
 channel-group 1 mode active
!
interface FastEthernet0/3
!
…
monitor session 1 source interface Po1
monitor session 1 destination interface Fa0/11
```

Jak lze vidět s výpisu konfigurace, jako monitorovací port byl konfigurován port Fa0/11, kam směrovač zrcadlil veškerý provoz na EtherChannelu (zde označen jako Po1). Tento port byl použit pro měření rychlostí a kontroly funkčnosti na Etherealu. (Tato konfigurace není nezbytná pro správnou funkci EtherChannelu, použita pouze pro ladící účely). Pokud se jako zdroj pro monitoring zvolil fyzický port (Fa0/1 nebo Fa0/2) EtherChannel se rozpadl, je tedy nutno zvolit přímo rozhraní Po1.

Oba porty musí být pro správnou funkci nastaveny jako full-duplex, half-duplex není podporován.

Kontrola konfigurace portů:

SW4#show interfaces fastEthernet 0/1 etherchannel

```
Port state    = Up Mstr In-Bndl
Channel group = 1           Mode = Active          Gcchange = -
Port-channel  = Po1         GC   =   -             Pseudo port-channel = Po1
Port index    = 0           Load = 0x00            Protocol =   LACP

Flags:  S - Device is sending Slow LACPDUs   F - Device is sending fast LACPDUs.

        A - Device is in active mode.        P - Device is in passive mode.

Local information:
                            LACP port     Admin     Oper    Port     Port
Port      Flags   State     Priority      Key       Key     Number   State
Fa0/1     SA      bndl      32768         0x1       0x1     0x1      0x3D

Partner's information:

                  LACP port                         Oper    Port     Port
Port      Flags   Priority  Dev ID          Age     Key     Number   State
Fa0/1     SA      32768     0005.dcd2.2040  15s     0x1     0x1      0x3D
```

Age of the port in the current state: 0d:00h:07m:08s

Jak lze vidět, port FastEthernet 0/1 je součástí EtherChannelu. Obdobná konfigurace je i pro port FastEthernet 0/2. Na těchto portech beží protokol LACP a porty jsou v aktivním módu (příznaky S a A). Sumarizované informace o celém EtherChannelu pak vypadají takto:

```
SW4#show etherchannel 1 detail
Group state = L2
Ports: 2   Maxports = 16
Port-channels: 1 Max Port-channels = 16
Protocol:   LACP
               Ports in the group:
               -------------------
Port: Fa0/1
------------

Port state    = Up Mstr In-Bndl
Channel group = 1           Mode = Active          Gcchange = -
Port-channel  = Po1         GC   =   -             Pseudo port-channel = Po1
Port index    = 0           Load = 0x00            Protocol =   LACP

Flags:  S - Device is sending Slow LACPDUs   F - Device is sending fast LACPDUs.

        A - Device is in active mode.        P - Device is in passive mode.

Local information:
                            LACP port     Admin     Oper    Port     Port
Port      Flags   State     Priority      Key       Key     Number   State
Fa0/1     SA      bndl      32768         0x1       0x1     0x1      0x3D

Partner's information:
```

```
                  LACP port                          Oper     Port     Port
Port      Flags   Priority  Dev ID          Age      Key      Number   State
Fa0/1     SA      32768     0005.dcd2.2040   0s      0x1      0x1      0x3D

Age of the port in the current state: 0d:00h:08m:14s

Port: Fa0/2
------------

Port state    = Up Mstr In-Bndl
Channel group = 1           Mode = Active          Gcchange = -
Port-channel  = Po1         GC   =   -             Pseudo port-channel = Po1
Port index    = 0           Load = 0x00            Protocol =   LACP

Flags:  S - Device is sending Slow LACPDUs   F - Device is sending fast LACPDUs.

        A - Device is in active mode.        P - Device is in passive mode.

Local information:
                            LACP port     Admin     Oper    Port     Port
Port      Flags   State     Priority      Key       Key     Number   State
Fa0/2     SA      bndl      32768         0x1       0x1     0x2      0x3D

Partner's information:

                  LACP port                          Oper     Port     Port
Port      Flags   Priority  Dev ID          Age      Key      Number   State
Fa0/2     SA      32768     0005.dcd2.2040   1s      0x1      0x2      0x3D

Age of the port in the current state: 0d:00h:04m:12s

               Port-channels in the group:
               ---------------------------

Port-channel: Po1    (Primary Aggregator)

------------

Age of the Port-channel   = 0d:02h:17m:38s
Logical slot/port   = 1/0          Number of ports = 2
HotStandBy port = null
Port state          = Port-channel Ag-Inuse
Protocol            =   LACP

Ports in the Port-channel:

Index   Load   Port     EC state        No of bits
------+------+------+------------------+-----------
  0     00     Fa0/1    Active             0
  0     00     Fa0/2    Active             0

Time since last port bundled:    0d:00h:04m:13s    Fa0/2
Time since last port Un-bundled: 0d:00h:04m:38s    Fa0/2
```

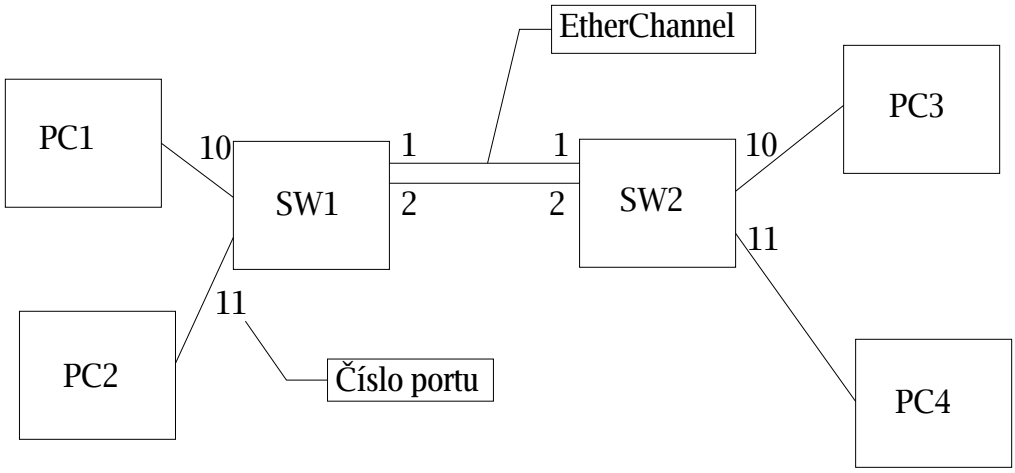
Sumarizovaný výpis popisuje konfiguraci obou portů, které se účastní EtherChannelu (viz. výše), pak také výpis konfigurace samotného virtuálního kanálu (Port Channel), kde je vidět jeho název (Po1), počet portů, které se EtherChannelu účastní, konkrétní čísla portů a mód, ve kterém se nacházejí.

Konfigurace LoadBalancingu:

Přepínač CiscoCatalyst 2950 umožnuje vyvažování zátěže pouze podle zdrojových a cílových MAC adres.

Switch(config)# port-channel load-balance src-mac (nebo dst-mac)
Switch(config)# end

Testovací zapojení pro ověření LoadBalancingu:



Následují výpisy z jednotlivých rozhraní, na kterých byla testována rychlost a LoadBalancing EtherChannelu. Jako zdroj datových toků byl použit flood ping. Nejdříve jsou uvedeny podmínky, při jakých byl test prováděn, poté výpis stavu portů (zvýrazněny jsou datové toky – rychlosti a počet paketů) a následuje zhodnocení daného měření.

Flood ping ze stanice PC1 na stanici PC3 (není zapnut LoadBalancing):

```
SW4#show interfaces fastEthernet 0/2
FastEthernet0/2 is up, line protocol is up (connected)
  Hardware is Fast Ethernet, address is 0005.dcd2.2302 (bia 0005.dcd2.2302)
  MTU 1500 bytes, BW 100000 Kbit, DLY 100 usec,
     reliability 255/255, txload 1/255, rxload 8/255
  Encapsulation ARPA, loopback not set
  Keepalive set (10 sec)
  Full-duplex, 100Mb/s, media type is 100BaseTX
  input flow-control is unsupported output flow-control is unsupported
  ARP type: ARPA, ARP Timeout 04:00:00
  Last input 00:00:22, output 00:00:00, output hang never
  Last clearing of "show interface" counters 00:01:10
  Input queue: 0/75/0/0 (size/max/drops/flushes); Total output drops: 0
  Queueing strategy: fifo
  Output queue: 0/40 (size/max)
  30 second input rate 3741000 bits/sec, 4583 packets/sec
  30 second output rate 1000 bits/sec, 2 packets/sec
     224571 packets input, 22906157 bytes, 0 no buffer
     Received 3 broadcasts (0 multicast)
     0 runts, 0 giants, 0 throttles
     0 input errors, 0 CRC, 0 frame, 0 overrun, 0 ignored
     0 watchdog, 3 multicast, 0 pause input
     0 input packets with dribble condition detected
     10 packets output, 1087 bytes, 0 underruns

SW4#show interfaces fastEthernet 0/1
FastEthernet0/1 is up, line protocol is up (connected)
  Hardware is Fast Ethernet, address is 0005.dcd2.2301 (bia 0005.dcd2.2301)
  MTU 1500 bytes, BW 100000 Kbit, DLY 100 usec,
     reliability 255/255, txload 10/255, rxload 1/255
  Encapsulation ARPA, loopback not set
  Keepalive set (10 sec)
  Full-duplex, 100Mb/s, media type is 100BaseTX
  input flow-control is unsupported output flow-control is unsupported
  ARP type: ARPA, ARP Timeout 04:00:00
  Last input 00:00:04, output 00:00:02, output hang never
  Last clearing of "show interface" counters 00:01:22
  Input queue: 0/75/0/0 (size/max/drops/flushes); Total output drops: 0
  Queueing strategy: fifo
  Output queue: 0/40 (size/max)
  30 second input rate 0 bits/sec, 0 packets/sec
  30 second output rate 4218000 bits/sec, 5168 packets/sec
     55 packets input, 4350 bytes, 0 no buffer
     Received 46 broadcasts (0 multicast)
     0 runts, 0 giants, 0 throttles
     0 input errors, 0 CRC, 0 frame, 0 overrun, 0 ignored
     0 watchdog, 46 multicast, 0 pause input
     0 input packets with dribble condition detected
     302775 packets output, 30883272 bytes, 0 underruns
```

Lze vidět, že požadavky na echo prochází jedním rozhraním a odpovědi rozhraním druhým

Flood ping ze stanice PC1 na stanici PC3 a současně flood ping ze stanice PC4 na PC2 (není zapnut LoadBalancing):

```
SW4#show interfaces fastEthernet 0/1
FastEthernet0/1 is up, line protocol is up (connected)
  Hardware is Fast Ethernet, address is 0005.dcd2.2301 (bia 0005.dcd2.2301)
  MTU 1500 bytes, BW 100000 Kbit, DLY 100 usec,
     reliability 255/255, txload 13/255, rxload 1/255
  Encapsulation ARPA, loopback not set
  Keepalive set (10 sec)
  Full-duplex, 100Mb/s, media type is 100BaseTX
  input flow-control is unsupported output flow-control is unsupported
  ARP type: ARPA, ARP Timeout 04:00:00
  Last input 00:00:05, output 00:00:06, output hang never
  Last clearing of "show interface" counters 00:01:02
  Input queue: 0/75/0/0 (size/max/drops/flushes); Total output drops: 0
  Queueing strategy: fifo
  Output queue: 0/40 (size/max)
  30 second input rate 0 bits/sec, 0 packets/sec
  30 second output rate 5114000 bits/sec, 6267 packets/sec
     42 packets input, 3199 bytes, 0 no buffer
     Received 36 broadcasts (0 multicast)
     0 runts, 0 giants, 0 throttles
     0 input errors, 0 CRC, 0 frame, 0 overrun, 0 ignored
     0 watchdog, 36 multicast, 0 pause input
     0 input packets with dribble condition detected
     407085 packets output, 41522737 bytes, 0 underruns

SW4#show interfaces fastEthernet 0/2
FastEthernet0/2 is up, line protocol is up (connected)
  Hardware is Fast Ethernet, address is 0005.dcd2.2302 (bia 0005.dcd2.2302)
  MTU 1500 bytes, BW 100000 Kbit, DLY 100 usec,
     reliability 255/255, txload 11/255, rxload 24/255
  Encapsulation ARPA, loopback not set
  Keepalive set (10 sec)
  Full-duplex, 100Mb/s, media type is 100BaseTX
  input flow-control is unsupported output flow-control is unsupported
  ARP type: ARPA, ARP Timeout 04:00:00
  Last input 00:00:14, output 00:00:03, output hang never
  Last clearing of "show interface" counters 00:01:09
  Input queue: 0/75/0/0 (size/max/drops/flushes); Total output drops: 0
  Queueing strategy: fifo
  Output queue: 0/40 (size/max)
  30 second input rate 9522000 bits/sec, 11668 packets/sec
  30 second output rate 4385000 bits/sec, 5372 packets/sec
     827565 packets input, 84411659 bytes, 0 no buffer
     Received 3 broadcasts (0 multicast)
     0 runts, 0 giants, 0 throttles
     0 input errors, 0 CRC, 0 frame, 0 overrun, 0 ignored
     0 watchdog, 3 multicast, 0 pause input
     0 input packets with dribble condition detected
     393702 packets output, 40157697 bytes, 0 underruns
```

Lze vidět,že nyní jsou datové toky přepínány vnitřím round-robin algoritmem, na venek se subjektivně přepínání jeví jako „nedeterministické – náhodné“.

Flood ping ze stanice PC1 na stanici PC3 a současně flood ping ze stanice PC4 na PC2 (zapnutý LoadBalancing):

```
SW4#show interfaces fastEthernet 0/1
FastEthernet0/1 is up, line protocol is up (connected)
  Hardware is Fast Ethernet, address is 0005.dcd2.2301 (bia 0005.dcd2.2301)
  MTU 1500 bytes, BW 100000 Kbit, DLY 100 usec,
     reliability 255/255, txload 1/255, rxload 12/255
  Encapsulation ARPA, loopback not set
  Keepalive set (10 sec)
  Full-duplex, 100Mb/s, media type is 100BaseTX
  input flow-control is unsupported output flow-control is unsupported
  ARP type: ARPA, ARP Timeout 04:00:00
  Last input 00:00:15, output 00:00:03, output hang never
  Last clearing of "show interface" counters 00:03:00
  Input queue: 0/75/0/0 (size/max/drops/flushes); Total output drops: 0
  Queueing strategy: fifo
  Output queue: 0/40 (size/max)
  30 second input rate 4852000 bits/sec, 5947 packets/sec
  30 second output rate 0 bits/sec, 0 packets/sec
     1182580 packets input, 120620093 bytes, 0 no buffer
     Received 100 broadcasts (0 multicast)
     0 runts, 0 giants, 0 throttles
     0 input errors, 0 CRC, 0 frame, 0 overrun, 0 ignored
     0 watchdog, 99 multicast, 0 pause input
     0 input packets with dribble condition detected
     28 packets output, 3197 bytes, 0 underruns
     0 output errors, 0 collisions, 0 interface resets
     0 babbles, 0 late collision, 0 deferred
     0 lost carrier, 0 no carrier, 0 PAUSE output
     0 output buffer failures, 0 output buffers swapped out

SW4#show interfaces fastEthernet 0/2
FastEthernet0/2 is up, line protocol is up (connected)
  Hardware is Fast Ethernet, address is 0005.dcd2.2302 (bia 0005.dcd2.2302)
  MTU 1500 bytes, BW 100000 Kbit, DLY 100 usec,
     reliability 255/255, txload 23/255, rxload 11/255
  Encapsulation ARPA, loopback not set
  Keepalive set (10 sec)
  Full-duplex, 100Mb/s, media type is 100BaseTX
  input flow-control is unsupported output flow-control is unsupported
  ARP type: ARPA, ARP Timeout 04:00:00
  Last input 00:00:20, output 00:00:08, output hang never
  Last clearing of "show interface" counters 00:03:05
  Input queue: 0/75/0/0 (size/max/drops/flushes); Total output drops: 0
  Queueing strategy: fifo
  Output queue: 0/40 (size/max)
  30 second input rate 4353000 bits/sec, 5335 packets/sec
  30 second output rate 9280000 bits/sec, 11373 packets/sec
     1045284 packets input, 106619283 bytes, 0 no buffer
     Received 9 broadcasts (0 multicast)
     0 runts, 0 giants, 0 throttles
     0 watchdog, 9 multicast, 0 pause input
     0 input packets with dribble condition detected
     2260168 packets output, 230537363 bytes, 0 underruns
     0 output errors, 0 collisions, 0 interface resets
     0 babbles, 0 late collision, 0 deferred
     0 lost carrier, 0 no carrier, 0 PAUSE output
     0 output buffer failures, 0 output buffers swapped out
```

V posledním měření jsme provedli LoadBalancing dle zdrojové a poté i cílové MAC adresy, v obou případech byly datové toky podobné, výpis výše ukazuje LoadBalancing podle zdrojové MAC adresy. I zde se, naproti očekávání, jeví vyvažování datových toků mezi jednotlivé kanály EtherChannelu subjektivně spíše jako náhodné.

Využití monitorovacího portu:

Pro odchytávání komunikace nelze použít HUB, jelikož umožňuje pouze half-duplex provoz, při kterém nelze provozovat EtherChannel. Při použití monitorovacího portu je pravděpodobné, že se na něj neposílají pakety, které mají za úkol sestavit EtherChannel, ale pouze servisní komunikace na již sestaveném EtherChannelu. Sestavení se tedy (standarní cestou) nedá odchytit.

Odchycené LACP z monitorovacího portu (Ethereal v. 0.10.10):

No.	Time	Source	Destination	Protocol Info
1	0.000000	Cisco_d2:20:41	Slow-Protocols *(protokoly, které nevysílají více než cca jeden paket za sekundu)*	LACP Link Aggregation Control ProtocolVersion 1. Actor Port = 1 Partner Port = 1
9	0.039693	Cisco_d2:23:02	Slow-Protocols	LACP Link Aggregation Control ProtocolVersion 1. Actor Port = 2 Partner Port = 2

Závěr:

Podle provedených měření se z hlediska LoadBalancingu „nejlépe choval“ provoz, který šel pouze jedním směrem. V tomto případě docházelo k rovnoměrnému rozložení zátěže mezi obě linky tvořící EtherChannel. Konfigurace LoadBalancingu nesplnila očekávání a zátěž linek se rozdělila (ze subjektivního pohledu) náhodně.

Při simulovaném výpadku jedné linky došlo okamžitě ke směrování veškerého provozu do redundantních spojů.

Nutno vyzdvihnou jednoduchost konfigurace EtherChannelu.

Výborně posloužil i monitorovací port, kde lze odchytnout pomocí síťového analyzátoru pakety LACP.

Použitá literatura:

Pužmanová,R.: TCP/IP v kostce. Kopp, České Budějovice 2004. ISBN 80-7232-236-2.

Cisco Systems: Understanding and Configuring EtherChannel [http://www.cisco.com]