Jihočeská univerzita v Českých Budějovicích

Přírodovědecká fakulta

Metody vyvažování zátěže na vícekanálových spojích typu PtP

Bakalářská práce

Jana Bakalová

Školitel: Ing. Rudolf Vohnout, Ph.D.

České Budějovice 2015

Bibliografické údaje

Bakalová J., 2015: Metody vyvažování zátěže na vícekanálových spojích typu PtP. [Methods of load balancing multichannel communications type PtP. Bc. Thesis, in Chech.] – 61 p., Faculty of Science, The University of South Bohemia, České Budějovice, Czech Republic.

Anotace

Tato bakalářská práce se zabývá výběrem vhodné metody rozložení zátěže na vícekanálovém hybridním spoji typu bod-bod (Point-to-Point, PtP). Dostupné metody rozložení zátěže jsou teoreticky analyzovány a prakticky ověřeny na směrovačích Mikrotik a přepínačích Cisco. Na základě rozboru jednotlivých metod jsou provedeny praktické testy pro charakteristické provozní situace, které mohou na spojích daného typu vzniknout. Podle porovnání kvalitativních parametrů celého spoje a zjištěného chování jednotlivých metod jsou vybrány nejvhodnější metody rozložení zátěže v hybridním spoji při různých provozních podmínkách.

Annotation

This bachelor thesis deals with a selection of convenient methods for load balancing over multichannel hybrid Point-to-Point (PtP) link. Available load balancing methods are theoretically analysed and verified using Mikrotik routers and Cisco switches. Based on the analysis of particular methods, the practical tests for characteristic operating situations that may occur in hybrid multichannel links are performed. The most convenient methods of load balancing at specific operating conditions for hybrid PtP links are chosen according to the comparison of the link quality parametres and discovered behaviour of particular methods

Prohlašuji, že svoji bakalářskou práci jsem vypracovala samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své bakalářské práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích, 12. 12. 2015

Podpis studenta

Obsah

1 Úvod

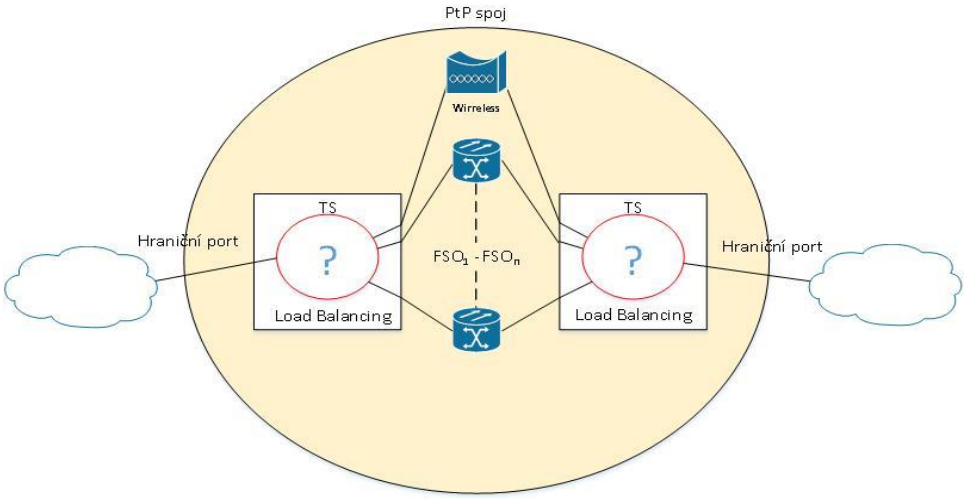
Základními prioritami v současném internetu je zajištění vysoké rychlosti přenosu a dostupnosti páteřních spojů mezi hlavními body přístupu k internetu POP (Point of Presence). Přitom je snaha o eliminaci souběhů tras pomocí tzv. zakruhování, kde páteřní spoje většinou nemají stejný fyzický průběh trasy. Při realizaci jednotlivých tras je ale vhodné z důvodu prostorového uspořádání a vlastností spojů využít vícekanálový paralelní přenos např. pomocí optických vláken, kde jednotlivé kanály mají stejné a v čase neproměnné vlastnosti (DWDM – Dense Wavelength Division Multiplexing), nebo jednotlivé paralelní kanály určité trasy mají rozdílné vlastnosti dané použitou technologií (FSO – Free Space Optic, Wi-Fi). V takovém případě prostorový souběh kanálů využívá přenos pomocí optického záření nebo pomocí rádiových vln volným prostorem a jedná se tedy o hybridní vícekanálový Point-to-Point (PtP) spoj. Při realizaci kanálů technologií FSO je nevýhodou velmi nestálé přenosové prostředí, kdy se útlum atmosféry může v poměrně krátkém čase změnit. Podobné nevýhody mají i kanály realizované pomocí technologie Wi-Fi, kde se navíc projevuje Radio-frequency (RF) rušení. Dochází tedy ke změnám propustnosti a dostupnosti dílčích spojů - kanálů, což má vliv na výběr vhodné metody vyvažování zátěže mezi nimi.

Hlavním cílem práce je navrhnout, implementovat a experimentálně ověřit různé metody vyvažování zátěže na vícekanálovém spoji typu PtP s důrazem na optimální přenosové charakteristiky pro využití v hybridních bezdrátových páteřních i přístupových sítích.

2 Teoretický základ

V hybridním spoji předpokládejme využití technologie optických spojů a rádiového spoje Wireless (Wi-Fi). Jedná se buď o samostatné zařízení zakončené metalickým rozhraním Ethernet, nebo je přímo součástí zařízení, které realizuje propojení s jeho dalšími rozhraními s možností vyvažování zátěže, přepínání a směrování (Routerboard).

Zvýšení spolehlivosti a propustnosti spoje je možné dosáhnout realizací hybridního spoje pomocí vícekanálového přenosu svazkem[1] (FSO_1 až FSO_n zálohovaného rádiovým spojem Wi-Fi). Spoj z hlediska datové struktury přenáší paketový provoz, který by měl být transparentní a neměl by vnášet do okolí žádnou další režii. Úlohu rozdělení provozu mezi paralelní linky plní dvojice Traffic Splitterů (TS) připojených hraničními porty do okolí. Protější hraniční porty mohou spojovat segmenty jedné logické sítě (jedna broadcast doména, více fyzických adres) nebo oddělovat PtP spoj od ostatních logických sítí pomocí směrovačů (pouze dva protější uzly se dvěma fyzickými adresami v tomto PtP spoji), což je typické pro páteřní a přístupové sítě PtP. Pro tuto topologii bude v praktické části navržena vhodná metoda vyvažování zátěže.



Obrázek 2.1: Hybridní vícekanálový PtP spoj

Pro testování vyvažování zátěže na PtP paralelním spoji je možné uvažovat dvě různé logické topologie, které mají na algoritmus vyvažování vliv:

- Oba konce PtP spoje patří do stejného segmentu logické sítě (jedné broadcast domény) s více fyzickými adresami. V takové síti by bylo možné využít pro vyvažování algoritmy založené na zdrojové a cílové MAC adrese, přitom předpokládejme provoz i na vyšších vrstvách ISO/OSI modelu.

[1] Svazek = alespoň dvě linky (např. Etherchannel, Bonding)

- PtP spoj tvoří samostatný logický segment se dvěma fyzickými adresami. V tomto případě nelze využít vyvažování reálného provozu založeného na L2 vrstvě ISO/OSI modelu.

2.1 Požadavky na vysokokapacitní přenosové sítě

Požadavky na vysokokapacitní přenosové sítě většinou stanovuje smlouva Service Level Agreement (SLA) mezi poskytovatelem připojení (ISP) a uživatelem a určuje, k čemu se poskytovatel zavazuje uživateli a v jaké kvalitě. K tomu se používají technické parametry poskytované služby. Tyto parametry popisuje standard RFC 2544[2][1] a v SLA smlouvách se většinou používají jen některé vybrané [2]:

- **Dostupnost spoje –** doba, kdy je spoj dostupný, vyjádřena v procentech
- **Latence** – maximální doba odezvy na PtP spoji
- **Ztrátovost paketů** (Packet loss) – poměrný počet nedoručených paketů vyjádřený v procentech
- **Propustnost** – udává objem přenesených dat za časovou jednotku

Všechny tyto parametry může ovlivnit kvalita vyvažování zátěže na paralelním PtP spoji. S tím souvisí požadavky na kvalitní vyvažování zátěže[3]:

- **Nízké přetížení** (Low Overhead) – rozdělení provozu pro všechny pakety, což přináší zvýšení režie. Algoritmus dělení provozu by měl být proto velmi jednoduchý s minimálním navýšením režie.
- **Vysoká efektivnost** (High Efficiency) – špatné rozdělení provozu má za následek nerovnoměrnou funkci linek vedoucí ke ztrátě šířky pásma. TS by se měl snažit distribuovat provoz co nejvíce podobný ideálnímu modelu.
- **Řazení paketů stejného toku přes stejnou linku** (Per-Flow Ordering) – porušení pořadí doručení paketů TCP může vytvářet úzká místa a způsobuje zbytečnou degradaci propustnosti. Je proto nezbytné, aby algoritmus dělení provozu udržoval uspořádání toku paketů Per-Flow (popsáno v následující kapitole).

2.2 Charakterizace metod vyvažování zátěže

Z obecného hlediska vyvažování zátěže je možné rozdělit vyvažování zátěže na základě charakteru provozu [4]:

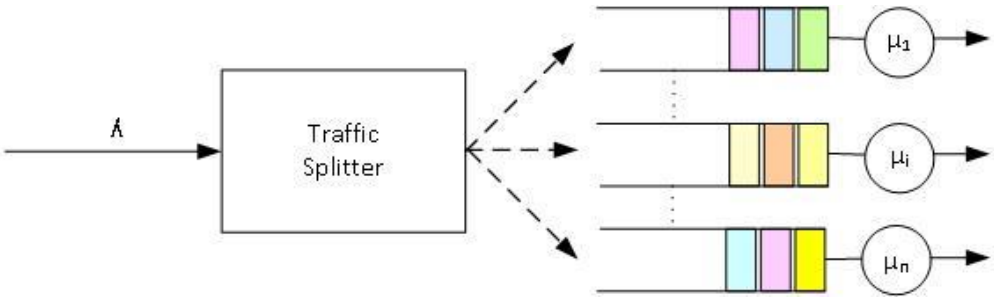
[2] Viz příloha A

- **Sub-Packet Load Balancing** – umožňuje rozdělení jednoho paketu rovnoměrně mezi více kanálů buď v jedné fyzické lince, nebo přes více fyzických linek, kde více kanálů běží na různých linkách (load balancing). Tento mechanismus musí podporovat oba protilehlé uzly.
- **Per-Packet Load Balancing** – pracuje na bázi jednotlivých paketů. Provoz je rovnoměrně rozdělen mezi linky ve svazku bonding. Pakety určené jednomu cíli jsou směrovány cyklicky jednotlivými rozhraními.
- **Per-Connection/Per-Flow Load Balancing** – pracuje na bázi datových toků. Všechny pakety, které jsou určené pro jeden cíl a odpovídající stejné charakteristice „flow“, jsou směrovány na stejná výstupní rozhraní. Tím je zamezeno problémům s pořadím paketů na cílovém zařízení a potřebě je znovu řadit, což by mohlo značně zpomalit přenos pomocí protokolu TCP.
- Ostatní metody vyvažování zátěže jsou založeny na využití směrovacích technik mezi sítěmi, na vytváření VPN tunelů s možností modifikací paketů a nastavení firewall rozhraní směrovačů, případně na vyvažování zatížení na vyšších vrstvách ISO/OSI modelu, např. vyvažování zátěže serverů pomocí sítí BitTorrent.

Charakter uvažovaného hybridního vícekanálového PtP spoje předpokládá paketový provoz, a tím se zužuje problematika vyrovnávání zátěže na Per-Packet a Per-Connection Load Balancing.

2.3 Obecný rozbor vyvažování zátěže na vícekanálovém PtP spoji

Uvažujme obecný model vyvažování zátěže:



Obrázek 2.2: Obecný model systému vyvažování zátěže pomocí Traffic Splitteru[3]

Přicházející pakety z hraniční linky s přenosovou kapacitou λ [bps] jsou předávány do jedné z nízko rychlostních výstupních linek$i \in \{1, N\}$ a přenosovou kapacitou μ_i [bps]. Kvalitní systém na vyvažování zátěže by měl rozdělovat provoz do více výstupních linek ve stejném nebo předdefinovaném poměru.

V matematickém modelu s ideálním plynulým provozem je provoz možné nekonečně dělit systémem vyvažování zátěže do N odchozích linek s kapacitou μ_i linky i. Množství přeneseného provozu linkou i během periody (τ, t) je $S_i(\tau, t)$. Systém s ideálním vyvážením by měl mít poměr kapacity linek ve stejném poměru s množstvím přenesených dat:

$$\frac{S_i(\tau, t)}{S_j(\tau, t)} = \frac{\mu_i}{\mu_j}$$

Vztah je možné upravit do tvaru:

$$\left| \frac{S_i(\tau, t)}{\mu_i} - \frac{S_j(\tau, t)}{\mu_j} \right| = 0$$

Pokud je provoz dokonale vyvážen, odchozí linky jsou ve stejnou dobu všechny buď zaneprázdněny, nebo nečinné. Tento stav vyjadřuje pojem „work-conserving“, ve kterém ani jedna odchozí linka není v klidu, pokud čekají další data na předání. Takové vyvažování zátěže je v reálné síti s paketovým provozem netypické. Základní jednotkou datového přenosu je minimálně jeden paket, provoz nelze nekonečně dělit a systém vyvažování zátěže paketového provozu již proto není „work-conserving“.

Pro vyjádření nejlepšího vyvažování zátěže uvažujme systém dvou linek stejné kapacity. Systém je ve výchozím stavu v klidu a obdrží jeden paket, který je předán na jednu ze dvou odchozích linek. Paket je obsluhován polovinou celkové možné šířky pásma a vysílání bude trvat dvakrát tak dlouho ve srovnání s ideálním systémem. Během této doby je jedna odchozí linka zaneprázdněna obsluhou paketu, zatímco druhá je v klidu. V reálném systému může TS dokonce poslat několik paketů do stejné odchozí linky, tím ještě navýší frontu na obsloužení a sníží propustnost. Uvažujme nejhorší případ v přenosu paketů, kdy všechny odchozí linky z TS jsou v klidu od času τ a dorazí paket o maximální velkosti P_{max}, který je poslán do linky i, přitom žádný další paket do času t nepřijde, a proto $S_j(\tau, t)$ je rovno nule. Během doby tohoto přenosu neplatí vztah 1, protože $\frac{S_i(\tau,t)}{\mu_i} - \frac{S_j(\tau,t)}{\mu_j} = \frac{C}{\mu_i} > 0$, kde C je část přenosu na lince i, která byla obsloužena během časového intervalu (τ, t). Paketový systém s nejlepším vyvažováním zátěže by měl rozdělovat nepřetržitý provoz do obou linek co nejplynuleji, přinejhorším by byl zahájen provoz na další lince ještě před skončením obsluhy nejpomalejší linky s maximální velikostí paketu. Měla by tedy být splněna následující podmínka:

$$\left|\frac{S_i(\tau,t)}{\mu_i}-\frac{S_j(\tau,t)}{\mu_j}\right|\leq\frac{P_{max}}{\min(\mu_i,\mu_j)}$$

v intervalu (τ,t), kde P_{max} je maximální velikost paketu.

To znamená, že rozdíl mezi okamžikem, kdy je linka i obsazenaa časem, kdy je obsazena linka j, by neměl být větší než čas pro odeslání největšího paketu přes nejpomalejší linku, neboli čím větší bude rozdíl časů potřebných na přenos jednotlivých paketů po různých linkách, tím hůř bude možné vyvažovat zátěž jednotlivých linek.

Při vyvažování zátěže je cílem nejlépe využít jednotlivé paralelní kanály, což vyžaduje vhodný algoritmus (hashovací funkci) pro rozdělení provozu v TS. Algoritmus rozdělení provozu by měl co nejlépe eliminovat situaci, při které dochází ke změně pořadí během doručování datových segmentů pomocí transportního protokolu (aby protokol nemusel čekat na segmenty doručené ve špatném pořadí a tím by došlo ke snížení propustnosti spoje). Proto bychom měli v této práci vybrat vhodný algoritmus rozdělování provozu do více linek při zachování datového toku Per-Packet. Existuje úzká vazba mezi rozdělováním do front a vyvažováním zátěže (fair queuing - load balancing).

V obecném modelu vyrovnávání zátěže, který zahrnuje TS a více odchozích linek, uvažujme vytvoření M zásobníků (front) přiřazených do N odchozích linek pomocí vhodné rozdělovací funkce s dynamickým přizpůsobením a nerovným rozložením zatížení.

Z výsledků simulace uvedené v práci [3] vyplývá, že přímé řazení paketů podle cílové IP adresy do dvou linek způsobuje značnou nerovnováhu zátěže. Kvalitní vyrovnávání zátěže poskytuje použití hashe, který je vytvořen z 16-ti bitového Cyklického redundantního součtu CRC (Cyclic redundancy check) kombinací pěti parametrů (zdrojová adresa, cílová adresa, zdrojový port, cílový port a protokol id). Tento hash udržuje rovnoměrné zatížení a délku fronty na obou linkách. Podobných vlastností při vyrovnávání zátěže může být dosaženo s použitím hashování založeného na tabulce (tablebased). Toto hashování vyžaduje nižší výpočetní výkon než metoda pomocí CRC. Je ale třeba sledovat zatížení linek a nastavení mapování z tabulky zásobníků do linek. Pro nalezení nejvhodnější metody vyvažování zátěže je zapotřebí definovat metriku výkonu pomocí následujících parametrů [3]:

Metrika výkonu

- **Rozložení zátěže** (Load Distribution) – časové rozdělení bajtů mezi více odchozích spojů. V ideálním systému je dopravní zatížení poměrně rozloženo podle kvality jednotlivých spojů.
- **Délka fronty** (Queue Length) – rozložení zatížení obvykle kolísá v čase, čemuž se zabraňuje pomocí vyrovnávací paměti a délka fronty odráží potřebu vyrovnávání zátěže. Vliv metriky délky fronty je menší v okamžicích menší zátěže spoje a uplatňuje se až při vyšším zatížení spoje. Vhodný rozdělovací algoritmus nemusí mít nutně dokonalé rozložení zátěže, ale měl by být schopen udržet fronty malé a vyvážené.
- **Doba, po kterou je linka v nečinnosti** (Non-Work-Conserving Idle Time) – je doba, kdy je alespoň jedna linka v klidovém stavu, zatímco jiné jsou plně zaneprázdněny.

2.4 Metody vyvažování zátěže podle hash algoritmu

2.4.1 Přímé hashování

Jedná se o jednoduchou metodu rozdělení provozu, TS použije transformační funkce na soubor polí z kombinace pěti parametrů (zdrojová adresa, cílová adresa, zdrojový port, cílový port a protokol id) a podle vypočtené hashovací hodnoty vybere odchozí linku. Implementace je jednoduchá, nevyžaduje zachování žádného zvláštního stavu. Mezi tento princip patří[3]:

- **Hashing of Destination Address** – nejjednodušší schéma, kde hash určuje cílová IP adresa modulo počtem odchozích linek N:

  $$H(\cdot) = DestIP mod N$$

  Pokud je možné vyjádřit počet linek jako k-tou mocninu binárního základu $N = 2^k$, nejnižší k bity z cílové adresy určují přímo index odchozího spojení (tuto funkci dříve používaly směrovače).
- **XOR Folding of Destination Address** – poskytuje dobrý výkon, může být vyjádřena jako:

  $$H(\cdot) = (D_1 \oplus D_2 \oplus D_3 \oplus D_4)\ mod\ N$$

  kde D_i je i-tý oktet cílové IP adresy. Tento přístup využívá více bitů cílové adresy při výběru linky.

- **XOR Folding of Source and Destination Addresses** – jedná se o jednoduchou modifikaci předchozí hashovací funkce, která zahrnuje do výpočtu jak cílovou, tak i zdrojovou adresu:

$$H(\cdot) = (S_1 \oplus S_2 \oplus S_3 \oplus S_4 \oplus D_1 \oplus D_2 \oplus D_3 \oplus D_4)\ mod\ N$$

  Kde S_i a D_i jsou i-té oktety ze zdrojové a cílové IP adresy.
- **Internet Checksum algorithm** – pro výpočet hashe se využívá 16-ti bitový kontrolní součet [5] z výše uvedené 5-tice modulo N:

$$H(\cdot) = CheckSum(5-tice)\ mod\ N$$

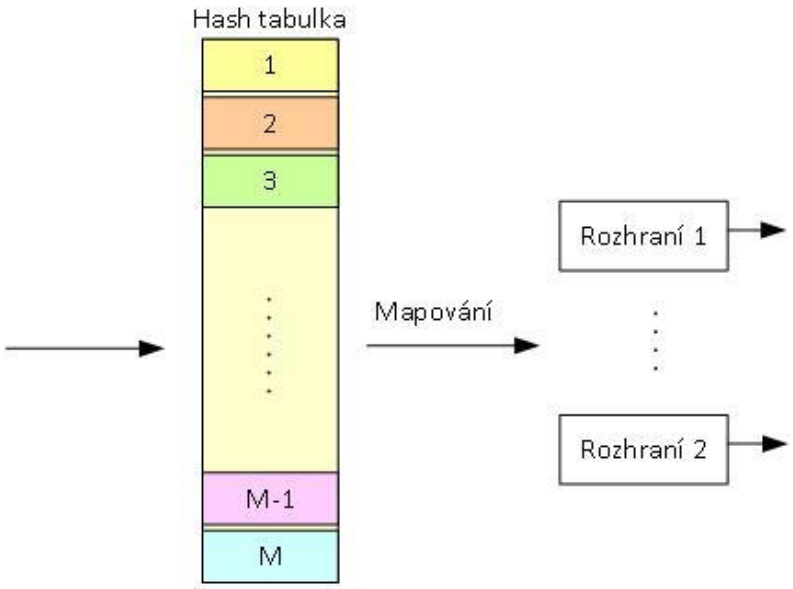
- **CyclicRedundant Checksum (CRC)** – CRC16 je algoritmus využívající 16-bit kontrolní součet CRC [6]. Je úspěšně využíván ve vysokorychlostních systémech. V principu TS aplikuje na výše uvedenou pětici CRC16 modulo N k získání odchozí linky. Hashovací funkce může být vyjádřena jako:

$$H(\cdot) = CRC16(5-tice)\ mod\ N$$

2.4.2 Hashování pomocí tabulky

Přímé hashování rozděluje provoz na stejné části do více odchozích cest. To ale není vždy žádoucí, např. uvažujme Wi-Fi spoj, který je dvakrát pomalejší než spoj FSO. Je proto potřeba rozdělit provoz v poměru 2:1. Použití metody s přímým hashováním je pro výše popsaný požadavek na rozdělení provozu v poměru 2:1 nevhodné.

Hashování založené na Table-Based řeší problém separátně – rozdělení provozu a přidělení zátěže. Toto schéma nejprve rozdělí provoz do M zásobníků, které jsou mapovány do N odchozích linek pomocí přiřazovací tabulky.



Obrázek 2.3: Rozložení zátěže založené na hash tabulce [3]

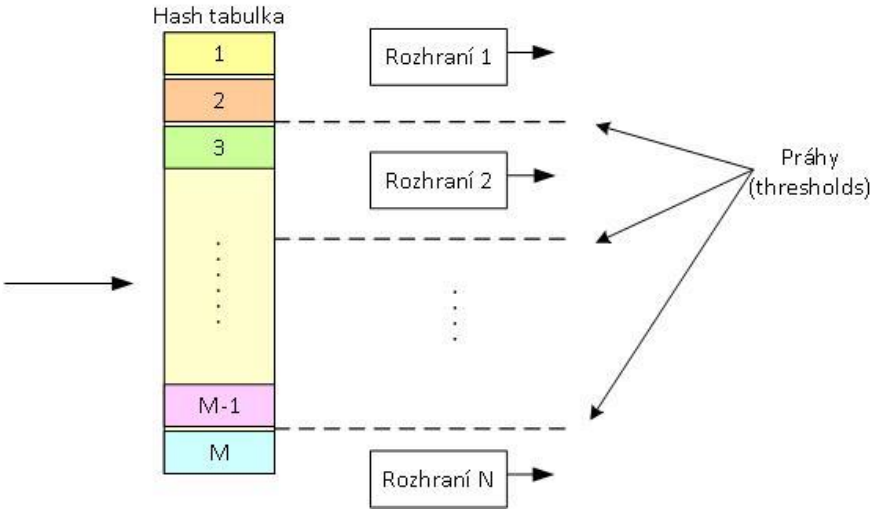
Různým přidělováním ze zásobníku na odchozí linky je možné distribuovat provoz v předem definovaném poměru, např. úpravou alokační tabulky (hash tabulky). Poměr mezi M a N určuje možnosti rozlišení. Typicky je M o jeden až dva řády větší než N a tím je možné rozdělit zatížení při poměrně jemném rozlišení. Při $M = N$ se jedná o hashování one-to-one. Při aplikaci metody Table-based existují dva základní přístupy [3]:

- **S využitím prahových hodnot**

  Při vyvažování zátěže tímto způsobem je vytvořen hash (např. z CRC16) z hlavičky paketu, který identifikuje daný tok. Odchozí linka je určena jedinečnému regionu v prostoru hashovací tabulky (mezi dvěma prahy). Traffic splitter určí podle hashe a regionu odchozí linku. Existuje několik hledisek pro výběr algoritmu:

  - Nároky na výpočetní výkon (Performance**)** – výběr regionu, získání klíče, určení odchozí linky porovnáním klíče a regionu
  - Četnost změn odchozích linek pro jeden tok (Disruption) – poměr celkového toku k toku se změněnou odchozí linkou
  - Volba hashovací funkce (Balancing)

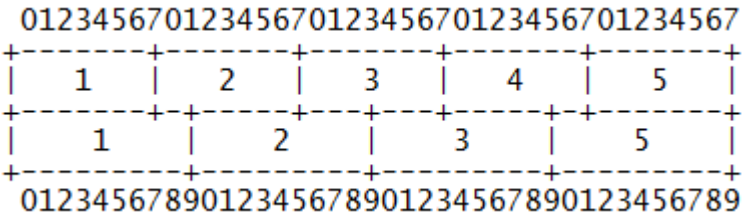
  Předpokládejme např., že chceme, rozdělit zatížení přes dvě odchozí spojení v poměru 2:1. Prahové hodnoty na odchozí spojení jsou tedy ve stejném poměru 2:1. Můžeme jednoduše stanovit prahovou hodnotu na $M/3$. Pro každý přicházející paket vypočítáme hodnotu hash a pak ji porovnáme s prahovou hodnotou. Pokud je hodnota větší než $M/3$, paket je odeslán na první linku (s větším zatížením), jinak se odešle na druhou linku. Metoda tedy umožňuje zatížit linky v různém poměru.



Obrázek 2.4: Rozdělení zátěže hashováním s využitím prahových hodnot [3]

Druhý příklad ukazuje vyvažování zátěže, kde jsou linky zatíženy ve stejném poměru a lze uplatnit podobné principy jako pro Equal-Cost Multi-Path (ECMP)[7].

Regiony jsou tedy stejné velikosti. Výstup z hashovací funkce je rovnoměrně rozdělen podle toků mezi odchozí linky. Pokud hranice regionu zůstávají beze změny, bude pro daný tok vybírána vždy stejná odchozí linka. V uvedeném případě ovlivní změnu hranic regionu pouze přidání nebo odstranění linky. V takovém případě dojde k přerozdělení hranic a velikostí regionů, což má za následek změnu linek pro jednotlivé toky.

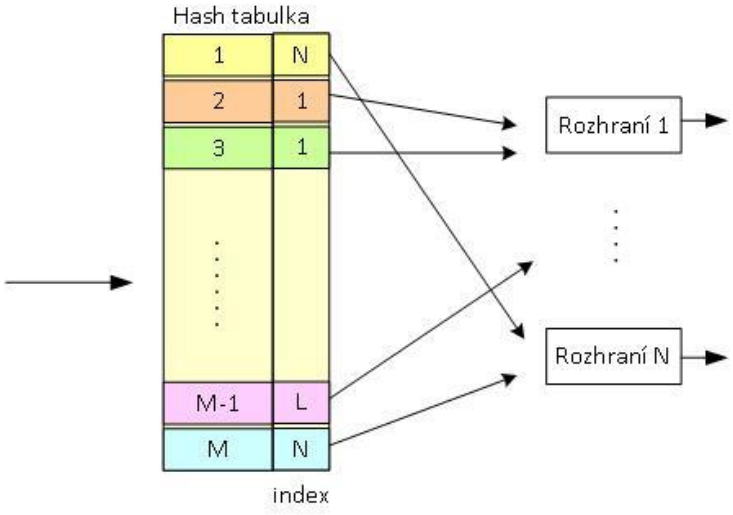


Obrázek 2.5: Změna prahu při odebrání 4-té linky [7]

Podle [7] je možné odvodit, že k nejmenšímu narušení (disruption) dojde, pokud dojde ke změně stavu linek odpovídajících regionů uprostřed hashovací tabulky (jak je patrné i z předchozího příkladu).

- **Podle indexu odchozí linky**

  Tento přístup založený na indexování vyžaduje více paměti než přístup na bázi prahových hodnot (M indexů oproti $N-1$ prahových hodnot). Na druhé straně mapování z hodnoty hash pro odchozí spojení je jednodušší s přístupem založeným na indexech. Tento způsob může být proveden s pomocí přímého vyhledání v tabulce, zatímco s přístupem na bázi prahů je hodnota hash srovnávána s $N-1$ prahovými hodnotami.



Obrázek 2.6: Rozdělení zátěže hashováním s využitím indexování [3]

Přístup na bázi indexu je pružnější, protože každý ze zásobníku M může být přiřazen k N odchozímu spojení nezávisle na sobě. Je možné ho použít k minimalizaci narušení existujícího provozu, pokud je nastaveno rozdělování zátěže nebo je přidána (odebrána) linka. Naopak přístup založený na prahové hodnotě může způsobit značné množství změn existujících toků na odchozí linky. Např. se domníváme, že je přidána nová linka ke dvěma vyvažovaným linkám. V případě, že zatížení musí být rovnoměrně rozloženo, je při použití metody založené na prahové hodnotě přesměrována na jinou odchozí linku $\frac{1}{2}$ provozu, zatímco pokud je použit index-based přístup, je ovlivněna $\frac{1}{3}$ toku. Přerozdělení odchozích linek může tedy způsobit změnu pořadí doručených paketů přes ovlivněné linky [3].

2.5 Monitorování stavu linek

Předchozí kapitoly se zabývaly možnostmi rozdělování provozu do jednotlivých paralelních linek za předpokladu, že tyto linky jsou funkční. V reálném provozu však může nastat situace, kdy některá linka není dostupná. Pokud by se funkce monitorování nepoužila, mohl by TS pokračovat v posílání paketů na linku, která selhala, a došlo by k degradaci celkové rychlosti spoje. Technologie vyvažování zátěže používají k monitorování dva základní způsoby ověřování stavu linky Address Resolution Protocol (ARP) a Media Independent Interface (MII) monitoring, přitom není možné používat oba způsoby najednou.

2.5.1 ARP monitoring

Posílá ARP dotazy s nastavenou frekvencí a používá odpovědi jako znamení toho, že spojení je funkční a provoz skutečně prochází přes linku. Je vhodné nastavit více ARP cílů, aby při výpadku jednoho z nich byla zajištěna spolehlivost ARP monitorování. Při využití hash metody vyvažování je dobré stanovit cíle tak, aby pokryly všechny slave linky (linky v jednom svazku)[8]. Způsob ověřování se dělí na:

- None – ověřování je vypnuté
- Active – je ověřována pouze aktivní linka
- Backup – jsou ověřovány pouze záložní linky
- All – jsou ověřovány všechny linky
- Filter – filtrování je rozděleno na všechny linky, bez ověřování
- Filter active – filtrování je aplikováno na všechny linky, ověřuje se pouze aktivní linka
- Filter backup - filtrování je aplikováno na všechny linky, ověřují se pouze záložní linky

Ověřování zjišťuje přicházející ARP žádosti a odpovědi pouze na rozhraních, která jsou zapnutá. Odpověď ARP na aktivní lince potvrzuje její vytvoření na ARP IP cíle. Vzhledem k tomu, že záložní linky běžně nepřijímají ARP odpovědi, je pro jejich ověření použita broadcast ARP žádost zaslaná přes aktivní linku. V případě konfigurace sítě, která vylučuje obdržení ARP žádosti na cíle pro záložní linky, by měl být tento způsob ověřování zakázán. TS využívá ověřování ARP žádostí na záložních linkách při určování výběru náhradní aktivní linky.

Filtrování používá při monitorování pouze příchozí ARP pakety pro zjištění dostupnosti. Filtruje (doručuje, ale ignoruje) ostatní pakety pro ověření dostupnosti linky. Při filtrování je zjišťován příjem ARP paketů (jakýkoliv bez ohledu na zdroj nebo cíl) k určení dostupnosti linky. Filtrování je užitečné, pokud by jinak generovaný ARP broadcast provoz způsoboval třetí straně problémy[9].

2.5.2 MIImonitoring

Způsob ověřování MII sleduje periodicky pouze přítomnost nosného signálu na lokálním síťovém rozhraní MII buď přímo na úrovni síťové karty, nebo pomocí operačního systému.

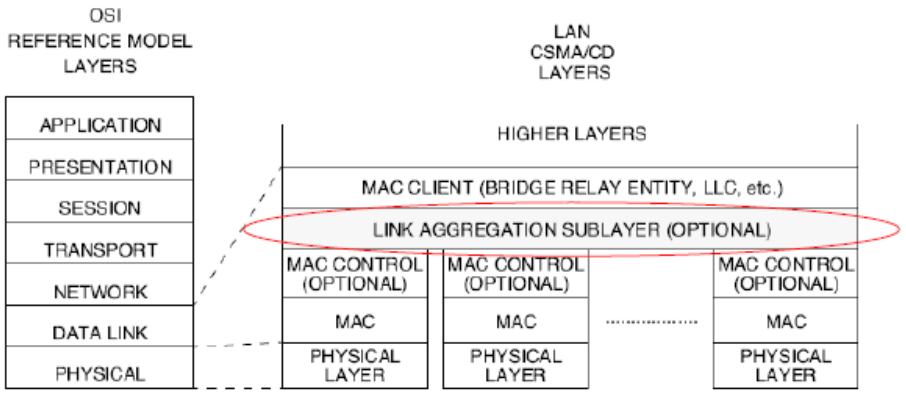
Typy monitorování MII:

- MII typ 1 – ovladač síťového rozhraní určuje, zda je na lince přítomen nosný signál
- MII typ 2 - metoda využívá operační systém[3]ke sdělení, zda je na lince přítomen nosný signál. Tato metoda je méně účinná, ale může být použita na všech zařízeních. Je vhodná spíš v případě, že MII typ 1 není podporován.

Volba MII monitorování je vhodná pro load balancing módy, které nepodporují ARP monitoring. U FSO PtP spoje při výpadku optického signálu je vhodné nastavit reakci vypnutí metalického rozhraní na optickém modulu (propojeného k slave lince ve svazku bonding) a tím zajistit MII monitoring na rozhraní příslušné slave linky. Parametrem updely je možné nastavit čas [ms] před zapnutím linky po detekci jejího obnovení[8].

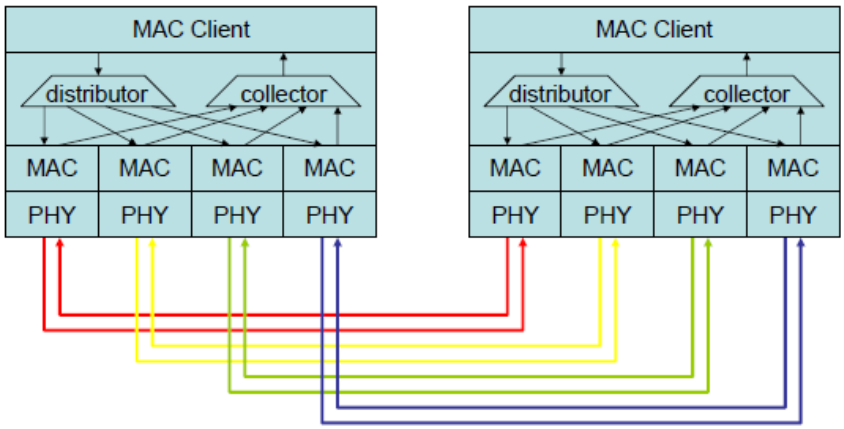
2.5.3 Standard 802.3ad

Nejedná se o přímé monitorování linek, ale o jejich automatické sestavení a monitorování pomocí protokolu Link Aggregation Control Protocol (LACP). Protokol LACP pracuje na Link Aggregation podrvrstvě (LAG) umístěné mezi podvrstvou MAC a LLC v ISO/OSI modelu. Pomocí nastavení rozhraní do aktivního módu (Actor) jsou rozesílány datagramy LACPDU na protější stranu. Tyto datagramy dynamicky vyjednají svázání více paralelních linek do jednoho logického spojení PtP (agregovaný port). Tím se na základě stavu a parametrů portu dynamicky vytvoří skupina podobně nakonfigurovaných portů (porty se stejnou rychlostí, full duplexním režimem, s nativní VLAN, s VLAN rozsahem a trunk statusem a typem). Pravidelným rozesíláním datagramů LACPDU s využitím stavu rozhraní MII je agregovaná linka monitorována [10].



Obrázek 2.7: Umístění LAG sublayer v ISO/OSI modelu [10]

[3]Volání registru daného zařízení nebo pomocí systémových nástrojů (např. v OS Linux – ethtool)



Obrázek 2.8: Architektura LACP protokolu – role distributoru a kolektoru [10]

Samotný standard 802.3ad nenařizuje žádný distribuční algoritmus, ale požaduje při příjmu rámce v kolektoru zajištění správného uspořádání rámců příslušejících do stejného spojení (důležité u protokolů connection oriented) a zamezení zdvojení rámců. Pro zajištění těchto podmínek se používá algoritmus, který zasílá všechny rámce tvořící jedno spojení do stejné linky ve stejném pořadí, v jakém jsou generovány klientem.

Vlastnosti standardu 802.3ad:

- Nemění formát rámců – nepřidává záhlaví ani pořadové číslo, nemění pole Type/Length
- Nevyžaduje přídavnou vyrovnávací paměť
- Nemění pořadí rámců
- Nepřidává významnou latenci
- Nezvyšuje šířku pásma pro jedno spojení
- Dosahuje vyvážení zátěže pouze při přenosu více souběžných spojení
- Výběr logiky agregace:
  Stable – active aggregator je zvolen podle nejvyšší agregované šířky pásma, změní se pouze, pokud jsou vypnuty všechny agregované linky
  Bandwidth - active aggregator je zvolen podle nejvyšší agregované šířky pásma, změna výběru nastane při přidání nebo odebrání linky, změny jejího stavu
  Count - active aggregator je zvolen podle nejvyššího počtu svázaných portů, změna výběru je stejná jako Bandwidth.
- Výměna LACPDU paketů nastává buď v režimu slow (každých 30 sekund – výchozí nastavení) nebo režimu fast (každou sekundu).
- Pro monitorování stavu linek nelze využít ARP monitoring.

Z uvedených vlastností standardu 802.3ad vyplývá, že je vhodný pro load balancing různorodého provozu se zachováním pořadí doručení, přičemž jeden typ provozu prochází vždy pouze jednou linkou [10].

2.6 Známé metody vyvažování zátěže

Při výběru vhodné metody pro vyvažování zátěže paralelního hybridního PtP spoje pro využití v páteřních a přístupových sítích uvažujme parametry uvedené v kapitole 2.1 Požadavky na vysokokapacitní přenosové sítě doplněné parametry módů pro vyvažování zátěže:

- Typ svázaných linek
- Metoda vyvažování zátěže
- Řazení paketů
- Monitorování stavu linek
- Vyvažování odpovědí na odchozí provoz

2.6.1 EtherChannel

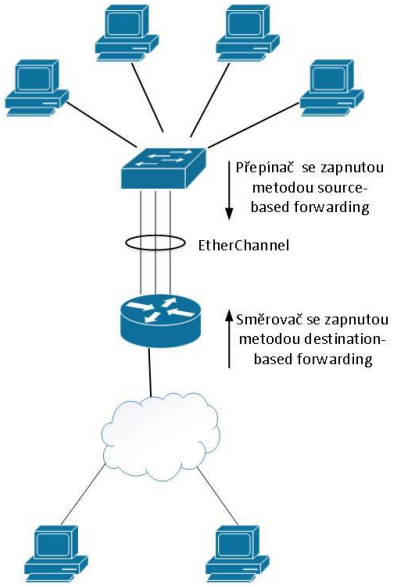
EtherChannel je technologie vyvinutá pro přepínače firmy CISCO, při které jsou využity paralelní redundantní spoje stejného typu ke zvýšení propustnosti provozu a zajištění větší spolehlivosti. Skládá se z jednotlivých Ethernetových linek stejných vlastností svázaných do jedné logické linky (port-channel logical interface).

EtherChannel je možné nakonfigurovat v Cisco proprietárním módu s protokolem Port Aggregation Protocol (PAgP) nebo standard 802.3ad s protokolem Link Aggregation Control Protocol (LACP), případně lze použít mód On. Oba konce EtherChannel se konfigurují do stejného módu.

- Při konfiguraci jednoho konce EtherChannel do jednoho PAgP nebo LACP módu systém vyjedná na druhé straně kanálu, který port má být aktivní. Nekompatibilní port se dostane do nezávislého stavu a pokračuje v přenosu jako normální samostatná linka. Konfigurace portu se nezmění, ale port se nezúčastní v EtherChannel.
- Při konfiguraci EtherChannel do módu On se vyjednávání neuplatní. Přepínač přinutí všechny kompatibilní porty k účasti v EtherChannel. Druhý konec kanálu musí být rovněž nastaven do On módu (jinak může dojít ke ztrátě paketů).

Pro vyrovnání zatížení provozu linek v kanálu EtherChannel používá výše popsané hashovací algoritmy XOR. Výběr konkrétního způsobu vyvažování zátěže by měl být

založen na umístění přepínače v síti a druhu provozu, který potřebuje vyvážit. Na obrázku je EtherChannel vyvažování zátěže metodou „Load Distribution and Forwarding“, která využívá hashovací algoritmus na bázi MAC adres. Přepínač, na kterém jsou koncentrována data ze čtyř pracovních stanic, komunikuje se směrovačem. Na přepínači je nastavená metoda source-based forwarding EtherChannel, která zajišťuje, že přepínač rozděluje provoz z pracovních stanic na všechny dostupné linky port-channel. Vzhledem k tomu, že směrovač má jednu MAC adresu, je nakonfigurován na metodu destination-based forwarding, která umožňuje rozdělení provozu podle cílových MAC adres pracovních stanic.



Obrázek 2.9: EtherChannel s metodou „Load Distribution and Forwarding [11]

Všechny porty EtherChannel jsou přiřazeny do stejné VLAN nebo nakonfigurovány jako trunk. Porty s různou nativní VLAN nemohou tvořit EtherChannel [11].

2.6.2 Bonding

Bonding je technologie navržená pro unixové operační systémy a je využita v technologii MikroTik. Umožňuje podobně jako Etherchannel svázání více síťových rozhraní do jednoho svazku (bonding) s využitím load balancing za účelem zvýšení propustnosti, zajištění větší spolehlivosti nebo vytvoření aktivní zálohy. Technologie bonding rozděluje provoz do jednotlivých linek na základě paketů síťové vrstvy nebo rámců linkové vrstvy ISO/OSI modelu. Přitom některé metody využívají k vyvažování výše uvedené hashovací algoritmy.

a) Balance-rr (Round robin policy)

Jedná se o vyvažování zátěže typu Per-Packet, kdy pakety stejného datového toku jsou pravidelně rozesílány na všechny odchozí linky. Výsledkem může být změna pořadí doručených paketů stejného toku do cíle, případně změna pořadí přijatých paketů z navazovaných spojení (TCP). Transportní protokoly nevyžadující navázání spojení (UDP) spoléhají na řešení tohoto problému pomocí vyšších vrstev ISO/OSI modelu.

- Na straně příjmu bonding linek je vyžadována konfigurace přepínače v módu balance-rr.
- Metoda balance-rr vyžaduje použití linek o stejné rychlosti a duplexního módu.
- V případě, že linky nemají stejnou rychlost, degraduje se rychlost ostatních linek na rychlost nejpomalejší linky.
- Při použití ARP monitorování jsou ARP dotazy distribuovány rovnoměrně mezi všechny aktivní linky. Jinak by odpovědi ARP přicházely po stejné lince, což by bylo vyhodnoceno jako porucha ostatních linek [9].

b) Active backup – aktivní záloha

Aktivní je pouze jedna linka ve skupině bonding. Jiná linka se stává aktivní pouze v případě, že současná aktivní linka selže. V případě výpadku aktivní linky je řešeno použití zdrojové MAC adresy obecně třemi způsoby podle použité technologie a vybraného módu:

1. **None** – Jako aktivní zálohy jsou nastaveny všechny slave linky na stejnou MAC adresu rozhraní bonding (výchozí nastavení v OS Linux).
2. **Active** – Žádná slave linka MAC adresu nemění. MAC adresa rozhraní bonding je MAC adresa vybrané slave linky. Tato metoda je použitelná pro zařízení (realizující bonding), které neumí změnit svoji MAC adresu, nebo odmítají příchozí broadcast s vlastní zdrojovou MAC adresou (která koliduje s ARP monitorem). Nevýhodou této politiky je, že každé (koncové) zařízení v takové síti musí aktualizovat ARP tabulku pomocí gratuitous ARP - gARP (oproti běžné aktualizaci přepínací tabulky…). Pokud není gratuitous ARP doručen, může dojít k narušení komunikace. Zařízení, která monitorují stav linky před aktuálním vysláním nebo

příjmem, jsou zvláště náchylná ke ztrátě gARP a je vhodné nastavit příslušné zpoždění updelay.

3. **Follow** – adresa rozhraní bonding je určena podle MAC adresy první slave linky, která je do bondingu přidána a je externě viditelná na aktivní lince. Ostatní linky mají roli backup a jejich MAC adresa není externě viditelná, aby nedošlo k duplikaci MAC adresy na více portech přepínače. Tento mód využívá technologie MikroTik, OS Linux od verze bonding 2.6.2 podle [9] při výpadku linky generuje bonding jeden nebo více gratuitous ARP na nové aktivní lince. Jeden gratuitous ARP je generován na bonding master rozhraní (tagovaně na každé rozhraní VLAN na něm konfigurované). Při změně aktivní linky se tedy přesune „promiskuitní nastavení“[4] na novou aktivní linku.

- Aktivní linka je určena nastavením primární linky (pokud je dostupná)
- Nevyžaduje žádné specifické nastavení na protilehlém přepínači.
- Je vhodný v případě, že není k dispozici jiný způsob vyvažování zátěže s monitorováním linek
- ARP monitoring nepracuje správně, pokud jsou k bondingu připojeny směrovače na obou stranách (v takovém případně je potřeba použít monitorování mii nebo musí být mezi směrovače vložen přepínač) [9].

c) Balance – xor

Je založen na jednoduché metodě hash, která využívá exkluzivní součet mezi zdrojovou a cílovou MAC adresou a packet type ID (source MAC address XOR destination MAC address XOR packet type ID) modulo počtem svázaných linek. Podle [12[5] a 9[6]] módy balance-xor, 802.3ad, tlb využívají tyto metody xmit_hash_policy:

Metoda výpočtu hash pomocí layer 2

Pomocí linkové vrstvy k vygenerování hashe využívá XOR z MAC adres modulo počtem linek N:

Vztah podle [12]:

$$H(\cdot) = (S_{MAC} \oplus D_{MAC}) \bmod N$$

[4] Nastavení mění síťovému rozhraní jeho adresu MAC, podporují jen vybrané síťové karty.
[5] Publikováno roku 2009
[6] Publikováno roku 2011

Vztah podle [9] využívá ve výpočtu navíc typ ID paketu:

$$hash = (S_{MAC} \oplus D_{MAC} \oplus packet\ type\ ID)\ mod\ N$$

Metoda je vhodná pro použití v LAN síti, kde jsou rozdílné MAC adresy, a kde je stejná adresa sítě. Nevhodná je pro propojení mezi dvěma směrovači (MAC adresy uzlů jsou konstantní). [9]

Metoda výpočtu hash pomocí layer 2 a 3

Ke generování hashe využívá kombinaci informací protokolu vrstvy 2 a 3 ISO/OSI modelu podle [9]:

$$hash1 = ((S_{MAC} \oplus D_{MAC} \oplus packet\ type\ ID) \oplus (S_{IP} \oplus D_{IP})$$

$$hash = hash1 \oplus (\text{hash RSHIFT } 16) \oplus (\text{hash RSHIFT } 8))\ mod\ N$$

Literatura [12] neuvádí způsob výpočtu pro tuto metodu.

Metoda poskytuje vyrovnanější rozdělení provozu než metoda layer2 zejména v prostředí, kde jsou použity směrovače do různých cílových sítí. Provoz neobsahující 3. a vyšší vrstvy používá metodu layer 2. [9]

Metoda výpočtu hash pomocí layer 3 a 4

Využívá informace protokolů síťové a transportní vrstvy ISO/OSI modelu. Jednotlivá spojení nejsou rozložena přes více linek. Pomocí síťové a transportní vrstvy (u nefragmentovaných TCP nebo UDP segmentů) využívá k vygenerování hashe XOR z čísla portů a IP adres modulo počtem linek N.

Vztah podle [12]:

$$H(\cdot) = \left((S_{port} \oplus D_{port}) \oplus (S_{IP} \oplus D_{IP})\right)\ mod\ N$$

Vztah podle [9]:

$$hash1 = \left((S_{port} \oplus D_{port}) \oplus (S_{IP} \oplus D_{IP})\right.$$

$$hash = hash1 \oplus (\text{hash RSHIFT } 16) \oplus (\text{hash RSHIFT } 8))\ mod\ N$$

Pro fragmentované TCP a UDP segmenty a ostatní IP provoz je zdrojový a cílový port vynechán. Pro síťový provoz, který nepoužívá protokol IP, je použitá stejná hashovací funkce jako u metody layer 2. Jedna TCP nebo UDP relace, která obsahuje fragmentované pakety prokládané přes více linek, může mít za následek

změnu pořadí doručených segmentů. Tato situace v praxi nenastává, poněvadž TCP a UDP provoz většinou nefragmentuje.

Při použití ARP monitorování jsou ARP dotazy distribuovány rovnoměrně mezi všechny aktivní linky. Jinak by odpovědi ARP přicházely po stejné lince, což by bylo vyhodnoceno jako porucha ostatních linek [9].

d) Broadcast

Veškerý provoz je vysílán přes všechny linky skupiny bonding. Používá se ve speciálních případech k zabezpečení vysoké dosažitelnosti v multiple switch topologiích. V důsledku tohoto chování je tento mód nevhodný pro rozkládání zátěže do paralelních linek a nebude v praktické části testován[9].

e) 802.3ad

Mód 802.3ad využívá výše uvedený standard. Do svazku bonding automaticky agreguje linky se stejnou rychlostí a nastavením duplexu a minimalizuje manuální konfiguraci přepínače. Umožňuje doručení rámců ve správném pořadí. Využívá stejné metody hashování jako již výše popsaný mód balance-xor a technologie Etherchannel, se kterými je kompatibilní a je vhodný pro vícekanálový PtP spoj[9].

f) Balance – tlb (adaptive transmit load balancing)

Adaptivní vyvažování zátěže, které nevyžaduje speciální podporu zařízení. Umožňuje dva režimy:

- Dynamický režim (výchozí nastavení) rozděluje odchozí provoz podle aktuálního zatížení (počítán vzhledem k rychlosti portů) na každé lince. Mícháním toků může dojít k přehození pořadí doručených paketů.
- Při vypnutém dynamickém režimu je provoz rozdělován pouze pomocí přímého hashe.

Aktivní linka je určena nastavením primární linky (pokud je dostupná). Při výpadku aktivní linky, případně jejím obnovením, se ke zvolení nové primární (aktivní) linky využívá těchto metod:

- Always – primární linka se stane aktivní vždy po obnovení jejího zapnutí.

- Better - primární linka se stane aktivní po obnovení jejího zapnutí, pokud má primární linka vhodnější parametry (rychlost, duplex) než současná aktivní linka.
- Failure - primární linka se stane aktivní, pokud současná aktivní linka selže a primární linka je zapnutá.

Při změně aktivní linky se přesune promiskuitní nastavení na novou aktivní linku. Nevyžaduje žádné specifické nastavení na protějším přepínači. Odpovědi na odchozí provoz jsou přijímány odpovídající aktivní linkou. Pokud tato linka selže, převezme její MAC adresu jiná linka.

Poněvadž se vyvažování zátěže provádí podle MAC adresy, v případě PtP propojení dvou směrovačů je posílán veškerý provoz po jedné lince. V lokální síti s různými MAC adresami balancuje tento mód odlišným způsobem od balance-xor nebo 802.3ad, aby nevznikaly matematicky „nešťastné“ hodnoty xor a tím byl provoz zasílán po jedné lince. V tomto módumohou být zapojeny do svazku linky různých rychlostí a není možné použít ARP monitoring.[9]

g) Balance – alb (adaptive load balancing)

Zahrnuje balance tlb a receive load balancing (rlb). Nevyžaduje žádné specifické nastavení na přepínači a je založen na vyjednávání pomocí protokolu ARP.

Bonding driver přepisuje zdrojovou MAC adresu provozu odcházejícího z bondingu na cíl s adresou slave linky bondingu tak, že různé protějšky spojení (cíle) používají rozdílnou MAC adresu původního zdroje. O změně zdrojové MAC adresy informuje bonding příslušný cíl vysláním gratuitous ARP reply. Protější přepínač si tuto MAC adresu zapíše do tabulky k příslušnému portu linky bondingu, a tím je vyvažován přijímaný provoz od protějších uzlů. Pokud ARP žádost protějšího uzlu využívá při vysílání cílovou MAC adresu svazku bonding, má použití gratuitous ARP vyjednávání při vyvažování zátěže problematický výsledek. Pro možnost změny MAC adresy rozhraní bonding při provozu je nutné použít NIC (Network Interface Card) s touto funkcí (např Intel Pro 100 nebo Pro 1000)[13], konkrétně Intel Pro 1000[14].

Přijímaný provoz je rovněž přerozdělen, když je přidána nová slave linka do svazku bonding, nebo když je neaktivní linka znovu aktivována. Během této doby je provoz rozdělován módem round robin mezi linky s nejvyšší rychlostí ve svazku bonding. Když je linka znovu připojena nebo se připojí nová, je proces

rozložení přijímaného provozu zahájen gARP odpovědí s vybranou MAC adresou každého klienta mezi všechny aktivní linky ve svazku bonding. Parametr updelay musí být nastaven na větší hodnotu, než je zpoždění v přepínači, aby gARP odpovědi nebyly blokovány přepínačem při zaslání k protějšímu uzlu.

Metoda vyvažuje vysílaný a přijímaný provoz na základě cílové IP adresy.

- Při změně aktivní linky se přesune promiskuitní nastavení na novou aktivní linku.
- Aktivní linka je určena nastavením primární linky (pokud je dostupná)
- Primární linka je použita pro řídící provoz specifický pro tento mód.
- pro každou slave linku ve svazku bonding posílá bonding driver 3 LLC rámce/1s, o velikosti 60B [15]
- rámce používají MAC adresu slave linky jako zdrojovou a cílovou adresu, ip pakety neaktualizují arp tabulku
- z pohledu uživatele jsou pakety posílány a přijímány pouze jednou linkou
- přijatý provoz je ovlivňován posíláním arp žádostí na hosta ve stejné VLAN
- režijní provoz generovaný mechanismem ARP monitor tagován interně uvnitř svazku linek (bonding zjišťuje VLAN ID z přicházejícího provozu a používá tyto ID k tagování rámců, které sám vytváří). [9]

2.7 Srovnání známých metod vyvažování

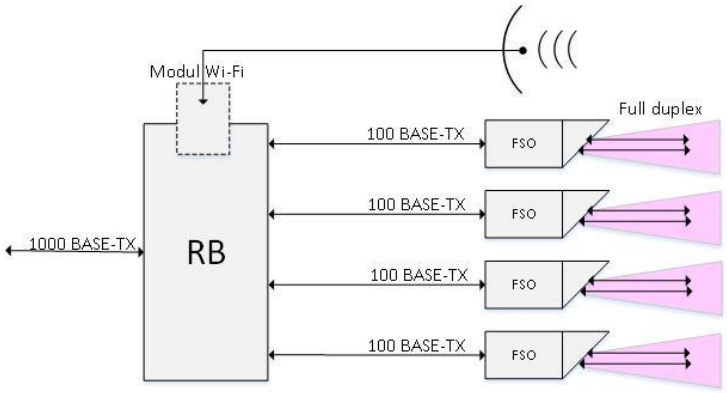
Pro výběr vhodné metody vyvažování je možné vyloučit metody, které nevyvažují provoz. Dále můžeme vyloučit metody rozdělující provoz na základě L2 ISO/OSI modelu (PtP spoj - dvě konstantní MAC adresy).

Tabulka 2.1: Srovnání známých metod vyvažování

technologie	mód	rozdělování provozu	citlivý provoz	vhodná metoda pro PtP
Etherchannel	LACP	Tx, Rx - hash L2, L3	L3	ano
Bonding	round robin	cyklicky	libovolný	ano
	active backup	nerozděluje	x	ne
	balance-XOR	Tx, Rx - hash L2, L2+L3, L3+L4	L3+L4	ano
	Broadcast	nerozděluje	x	ne
	802.3ad	Tx, Rx - hash L2, L2+L3, L3+L4	L3+L4	ano
	balance-tlb	Tx-table based, Rx nerozděluje	x	ne
	balance-alb	Tx-table based, Rx-gARP	libovolný	ano

3 Praktická část

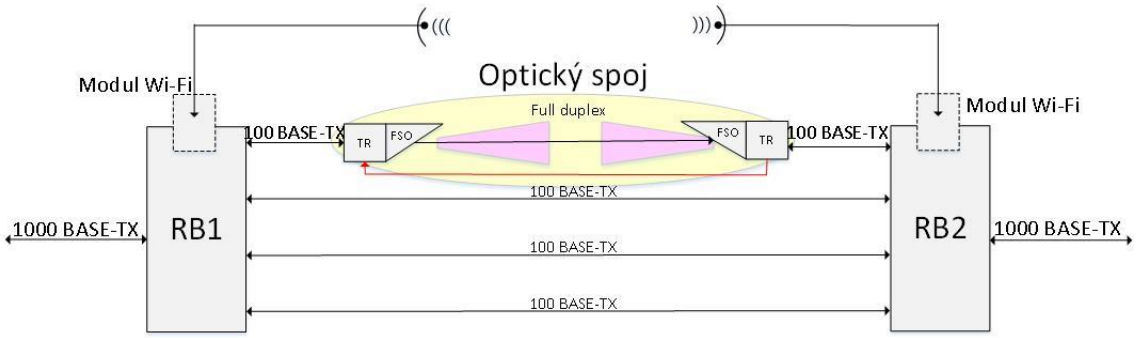
Praktická část této práce vychází z potřeb návrhu vyvažování vícekanálového paralelního spoje podle konstrukce navržené v diplomové práci „Realizace optického spoje volným prostorem pro kratší vzdálenosti“, 2013 [16]. Při návrhu testovací topologie pro vyvažování zátěže na hybridních PtP spojích byla tedy snaha navrhnout takovou topologii, která by umožnila budoucí nasazení modulů FSO. Na obrázku je blokové schéma jedné strany takového spoje se čtyřmi moduly FSO zálohovaného modulem Wi-Fi.



Obrázek 3.1: Blokové schéma jedné strany FSO spoje

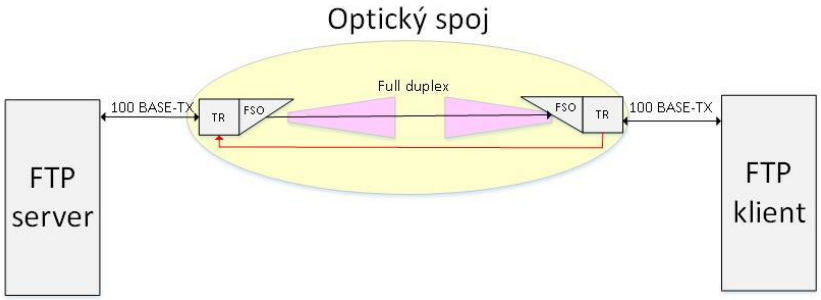
Příchozí provoz vstupuje přes hraniční port do TS (realizovaný Routerboardem - RB), který určitou metodou rozděluje provoz do jednotlivých paralelních spojů (FSO_1 až FSO_4 a záložního bezdrátového spoje) připojených metalickým rozhraním. Toto rozhraní je vybaveno funkcí LFP (Link Fault Pass-through), která umožňuje monitorovat stav optického spoje a vyvolat proces link down a následně vypnout metalickou linku na portu zařízení. Tento kontrolní mechanizmus existuje i pro rádiové spoje [16] a lze ho využít při monitorování MII. Na druhé straně FSO spoje je provoz opět sloučen na RB do jednoho datového proudu dostupného na hraničním portu.

Pro testování byl k dispozici pouze jeden optický spoj s jedním směrem přes FSO a opačným směrem přes optický kabel. Poskytnutý optický spoj byl tedy full duplexní, ale každý směr byl realizován jinou technologií. Byla tedy navržena topologie svazku jednoho optického spoje společně se třemi metalickými spoji a zálohováním rádiovým modulem Wi-Fi. Topologii znázorňuje následující obrázek:



Obrázek 3.2: Topologie hybridního paralelního PtP spoje

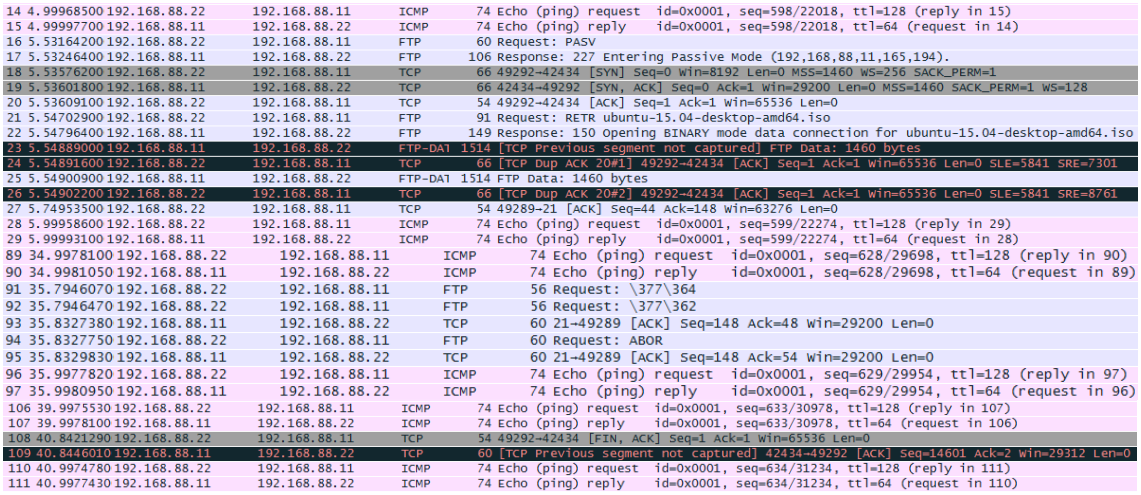
Při testování vykazoval samotný FSO modul nespolehlivost a proměnlivé chování, aniž by byly měněny podmínky testování. Pro ověření vlastností optického spoje bylo provedeno testování, při kterém byl optický spoj připojen na jedné straně k FTP serveru (Ubuntu 15.04) s IP adresou 192.168.88.11 a na druhé straně k FTP klientu (Windows 7) s IP adresou 192.168.88.22.



Obrázek 3.3: Topologie testovaného FSO modulu

Spoj byl otestován provozem ICMP (ping) společně s provozem TCP (FTP) tak, aby datový tok využíval FSO spoj. Z následujícího výpisu provozu je patrné, že linka přenáší pakety oběma směry. Provoz ICMP je sice přenášen beze ztrát, ale v provozu TCP dochází k duplikacím (TCP Dup ACK 20#1, 20#2)[7] a následnému ukončení TCP spojení. Na takovém spoji nebylo možné dále provádět testování metod vyvažování zátěže, poněvadž nebylo možné stanovit stejné a popsatelné podmínky v celém průběhu testování.

[7] Pokud příjemce příjme paket s TCP sekvencí vyšší než je očekávaná, znamená to ztrátu paketů a příjemce zašle TCP Dup ACK vysílací straně. Přijetím dvou TCP Dup ACK se aktivuje mechanismus pro zaslání TCP Fast Retransmission pro znovu zaslání paketu bez čekání na časovač TcpInitialRtt.



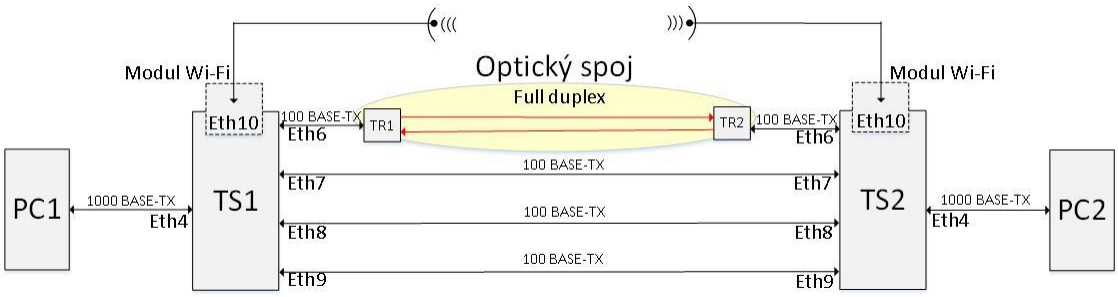
Obrázek 3.4: Duplikované odpovědi TCP dup ACK20 a následné ukončení

Bylo proto nutno testovací topologii upravit nahrazením FSO spoje optickými kabely zapojenými přes transceivery na metalická rozhraní. Eventuální výpadky optiky (nosného optického signálu) bylo možné simulovat vytažením optického kabelu. Pro využití MII monitorování byly transceivery nastaveny na funkci LFP, která vypne metalické rozhraní při ztrátě nosného optického signálu.

3.1 Topologie testovaného spoje

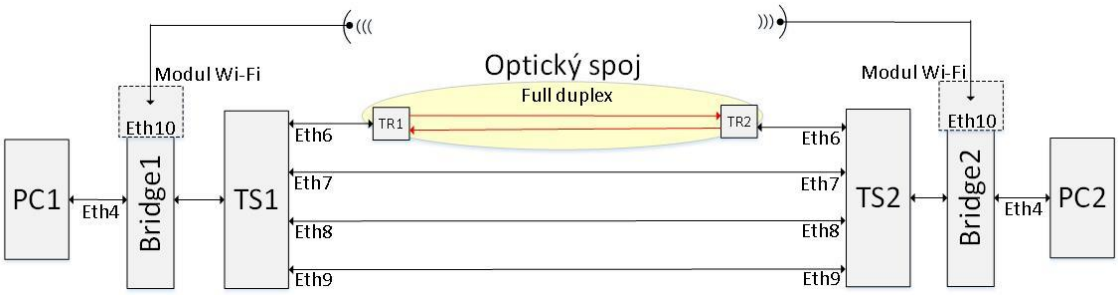
Pro testování metody Bonding byla použita v roli TS technologie MikroTik Routerboard s rozhraním Bonding a pro metodu Etherchannel byl použit přepínač Cisco Catalyst 2960 v propojení s Routerboard-bridge[8]. Rozhraní Eth7 až Eth9 (100BASE-TX) byla propojena metalickými kabely a rozhraní Eth6 (100BASE-TX) bylo metalicky připojeno k páru transceiverů na optické rozhraní nahrazující FSO spoj. Z principu zpracování optického signálu vyplývá, že nosný optický signál je buď přítomen rychlostí přenosu 100Mbps nebo přítomen není, což lze simulovat vytažením optického kabelu v transceiveru. Porucha optického spoje je v testování simulována výpadkem odchozího provozu na optickém spoji připojeném k metalickému rozhraní Eth6 RB1. Proto v testovacích krocích není uvažován provoz s rozdílnou rychlostí linek svazku Bonding.

[8] Viz příloha B



Obrázek 3.5: Fyzická testovací topologie

Logická topologie je tvořena konfigurací Routerboardů RB1 a RB2. Rozhraní TS je propojen s hraničním portem a náhradním spojem Wi-Fi pomocí přepínače (MikroTik-bridge) se zapnutým Spanning Tree protokolem realizujícího zálohu prioritního spoje TS přes Wi-Fi moduly. PtP spoj obvykle tvoří samostatný logický segment (používaný v páteřních sítích). Poněvadž takový PtP spoj využívá pouze dvě fyzické adresy (MAC) nebylo testováno rozdělování provozu podle L2 ISO/OSI modelu.



Obrázek 3.6: Logická testovací topologie

3.2 Způsob testování

3.2.1 Kritéria pro zvolení nejvhodnější metody vyvažování zátěže

Pro jednotlivé metody vyvažování zátěže popsané v teoretické části byly stanoveny následující hodnotící kritéria:

- Doba nedostupnosti - při výpadku linky (místo dostupnosti spoje) měřeno dobou ztráty jednotlivých provozů při změně stavu linek
- Latence - doba odezvy od protějšího uzlu spoje PtP měřená odezvou na provoz ping při zatíženém spoji udávaná v milisekundách
- Propustnost (bandwidth - BW) - množství přenesených dat za časovou jednotku pro daný provoz udávaných v Mbps
- Ztrátovost paketů - procentuální poměr odeslaných a přijatých paketů měřená během jednoho nastavení, nebo při vypínání a zapínání linek, kde celkový součet BW linek není nižší, než generovaný BW

- Efektivnost - rozložení provozu do jednotlivých linek
- Pakety mimo pořadí (out-of-order) – procentuální počet doručených paketů v nesprávném pořadí
- Jitter – nerovnoměrnost latence udávaná v milisekundách

Při měření výše uvedených hodnot bylo sledováno maximální vytížení procesorů nepřesahující 80%.

3.2.2 Generování testovacího provozu

Reálný provoz na páteřních sítích pochází z mnoha různorodých zdrojů navazovaného i nenavazovaného spojení. Při stanovení konkrétních metod testování bylo vycházeno ze standardu RFC 2544 [1]. K testování pomocí tohoto standardu se využívají speciální a nákladná zařízení, která průběh testu automatizují a umožňují jeho snadnou konfiguraci. Při snaze dodržet parametry uváděné ve standardu RFC 2544 bez speciálních zařízení s využitím aplikace Wireshark vzniká značné vytížení procesorů a nároků na paměť pro odchycený provoz (řádově GB). Testovací metody byly redukovány na nižší rychlosti a kratší časové intervaly.

Při měření (v topologii se čtyřmi svázanými linkami) je takový provoz simulován pomocí SW nástroje Iperf. Nástroj Iperf generuje provoz, pro který lze nastavit příslušné parametry. Pro testování byly využity koncové uzly v režimu server-klient a klient-server. Přiblížení k náhodnému reálnému provozu je zajištěno generováním provozu z deseti náhodných zdrojových portů známým mechanismem navazování spojení na L4 ISO/OSI modelu. Pomocí Iperfv2 byl generován provoz UDP, při kterém bylo možné naměřit hodnoty ztráty paketů, jitter a změny pořadí doručení paketů. Pomocí Iperfv3 byl generován provoz TCP, kterým bylo možné určit šířku pásma pro jednotlivá spojení, počet znovu zaslaných paketů a parametr cwnd[9] .Vytvořený provoz pro testování určují následující parametry příkazu iperf:

Pro server: iperf –u –s –p 5001

- -u: použitý UDP/bez parametru TCP protokol
- -s: označuje stranu serveru
- -p: určuje port 5001, na kterém server naslouchá

Pro klienta: iperf –u –c 192.168.88.11 –p 5001 –t 60 –b 10M –P 10 > udp-5001.txt

[9] CWND (congestion window) je počet paketů vyslaných během času round-trip time (RTT – čas přenosu na všech linkách mezi koncovými uzly). Z poměru CWND/RTT lze určit propustnost spoje.

- -u: použitý UDP/bez parametru TCP protokol
- -c: označuje stranu klienta
- IP adresa 192.168.88.11: k jakému serveru se připojuje
- -p: port 5001, na kterém server naslouchá
- -t: doba měření je 60s
- -b: bandwidth 10Mbps bez využití zálohy (Wi-Fi), 1Mbps s využitím zálohy (Wi-Fi)[10]
- -P: 10 navázaných spojení s náhodným zdrojovým portem
- > poslání výsledků do textového souboru udp-5001.txt

3.2.3 Testovací kroky

Nejprve byly ověřeny vybrané metody vyvažování zátěže popsané v teoretické části. Z těchto výsledků vyplynula potřeba testovat svazek Bonding minimálně jako celek s funkčními všemi linkami a stav, ve kterém je funkční pouze jedna linka, která je podrobněji testována.

Testování je provedeno pro nenavazované UDP spojení, při kterém jsou využity koncové uzly v režimu server-klient. Pro navazované TCP je sice provedeno měření ve stejném rozsahu, ale pro velký objem dat jsou do práce zahrnuty pouze výsledky pro kandidáty na nejvhodnější metodu vyvažování zátěže vybrané testováním UDP. Testovací kroky jsou navrženy tak, aby napodobovaly chování reálných hybridních spojů (výpadky, přechod na záložní linku) při různém způsobu monitorování linek ve svazku Bonding. Test trvá 60s a změny jsou prováděny po 20s. Změny stavu linek jsou provedeny pouze v jednom směru. Testovací kroky probíhají s monitorováním MII, pouze v krocích, ve kterých nelze určit stav linky (on/off), je použito monitorování ARP.

Jednotlivé testovací kroky:

a) Všechny linky ve svazku zapnuty
b) Vypnutí linky Eth6 RB1 po celou dobu měření
c) Vypnutí a opětné zapnutí metalického rozhraní Eth6 RB1 (Tx, Rx)
d) Vypojení a opětné zapojení linky Eth6 RB1 směr Tx na optickém spoji - vypnuté LFP a zapnuté monitorování MII
e) Vypojení a opětné zapojení linky Eth6 RB1 směr Tx na optickém spoji - zapnuté LFP a zapnuté monitorování MII

[10] Protože provoz je generován nepřetržitě, dojde ke ztrátě paketů z důvodů přetečení bufferů a nelze tím otestovat ztráty, kdy svazek bonding přechází na zálohu Wi-Fi.

f) Vypojení a opětné zapojení linky Eth6 RB1 směr Tx na optickém spoji - vypnuté LFP a zapnuté monitorování ARP

g) Vypnutí a opětné zapnutí rozhraní svazku, přechod na záložní linku Wi-Fi

h) Zapnutí a opětné vypnutí metalického rozhraní Eth6 RB1 (Tx, Rx), přechod ze záložní linky Wi-Fi

i) Zapojení a opětné vypojení linky Eth6 RB1 směr Tx na optickém spoji, přechod ze záložní linky Wi-Fi - vypnuté LFP a zapnuté monitorování MII

j) Zapojení a opětné vypojení linky Eth6 RB1 směr Tx na optickém spoji, přechod ze záložní linky Wi-Fi - vypnuté LFP a zapnuté monitorování ARP

3.3 Testován vybraných metod vyvažování provozem UDP

Pro každý krok testování je uvedena tabulka naměřených průměrných hodnot ztrátovosti, jitter, latence a paketů mimo pořadí pomocí nástrojů Iperf a ping.

Časové průběhy celkové propustnosti PtP spoje jsou uvedeny pouze při měření, ve kterých nastávají výrazné změny od normálního provozu. Rozlišení jednotlivých provozů popisuje následující legenda:

Obrázek 3.7: Legenda k časovým průběhům

3.3.1 Všechny linky ve svazku zapnuty

Měření proběhlo provozem generovaným z každé strany pomocí deseti spojení rychlostí 10 Mbps. Pro metodu Etherchannel byl provoz generován z každé strany na dvou cílových IP adresách po pěti spojení rychlostí 10 Mbps.

Tabulka 3.1: Naměřené průměrné hodnoty na 10-ti spojení pro jednotlivé metody

směr	metoda	ztrátovost [%]	jitter [ms]	latence [ms]	pkt mimo pořadí [%]
PC1 > PC2	round robin	0,017	0,54	0,293	0,02
	802.3ad	0,002	0,69	0,616	0,00
	balance-XOR	0,002	0,41	0,621	0,02
	Etherchannel	0,007	0,28	0,774	0,00
PC2 > PC1	round robin	0,057	0,20	1,071	0,74
	802.3ad	0,004	0,26	0,661	0,00
	balance-XOR	0,004	0,13	0,411	0,01
	Etherchannel	0,006	0,15	0,607	0,00

Při zapnutých všech linkách bez simulace poruch vykazuje metoda round robin nejvyšší ztrátovost a doručování paketů mimo pořadí, rozdělení provozu do linek je ale u této metody nejrovnoměrnější.

Name	Type	L2 MTU	Tx	Rx
ether6	Ethernet	1598	26.0 Mbps	25.9 Mbps
ether7	Ethernet	1598	26.0 Mbps	26.0 Mbps
ether8	Ethernet	1598	26.0 Mbps	26.0 Mbps
ether9	Ethernet	1598	26.0 Mbps	26.0 Mbps

Obrázek 3.8: Rozložení zátěže při metodě round robin

3.3.2 Vypnutí linky po celou dobu měření

Při vypnutí linky Eth6 RB1se rozloží provoz do ostatních linek podle jednotlivých metod stejným poměrem, podobně jako když jsou všechny linky zapnuty. Množství paketů mimo pořadí je podle očekávání nulové.

Tabulka 3.2: Naměřené průměrné hodnoty na 10-ti spojení pro jednotlivé metody

směr	metoda	ztrátovost [%]	jitter [ms]	latence [ms]	pkt mimo pořadí [%]
PC1 > PC2	round robin	0,004	0,41	0,675	0,00
	802.3ad	0,002	0,35	0,640	0,00
	balance-XOR	0,002	0,39	0,506	0,00
	Etherchannel	0,003	0,20	0,631	0,00
PC2 > PC1	round robin	0,006	0,1	0,576	0,00
	802.3ad	0,004	0,36	0,537	0,00
	balance-XOR	0,005	0,41	0,431	0,00
	Etherchannel	0,001	0,31	0,593	0,00

3.3.3 Vypnutí a opětné zapnutí metalického rozhraní

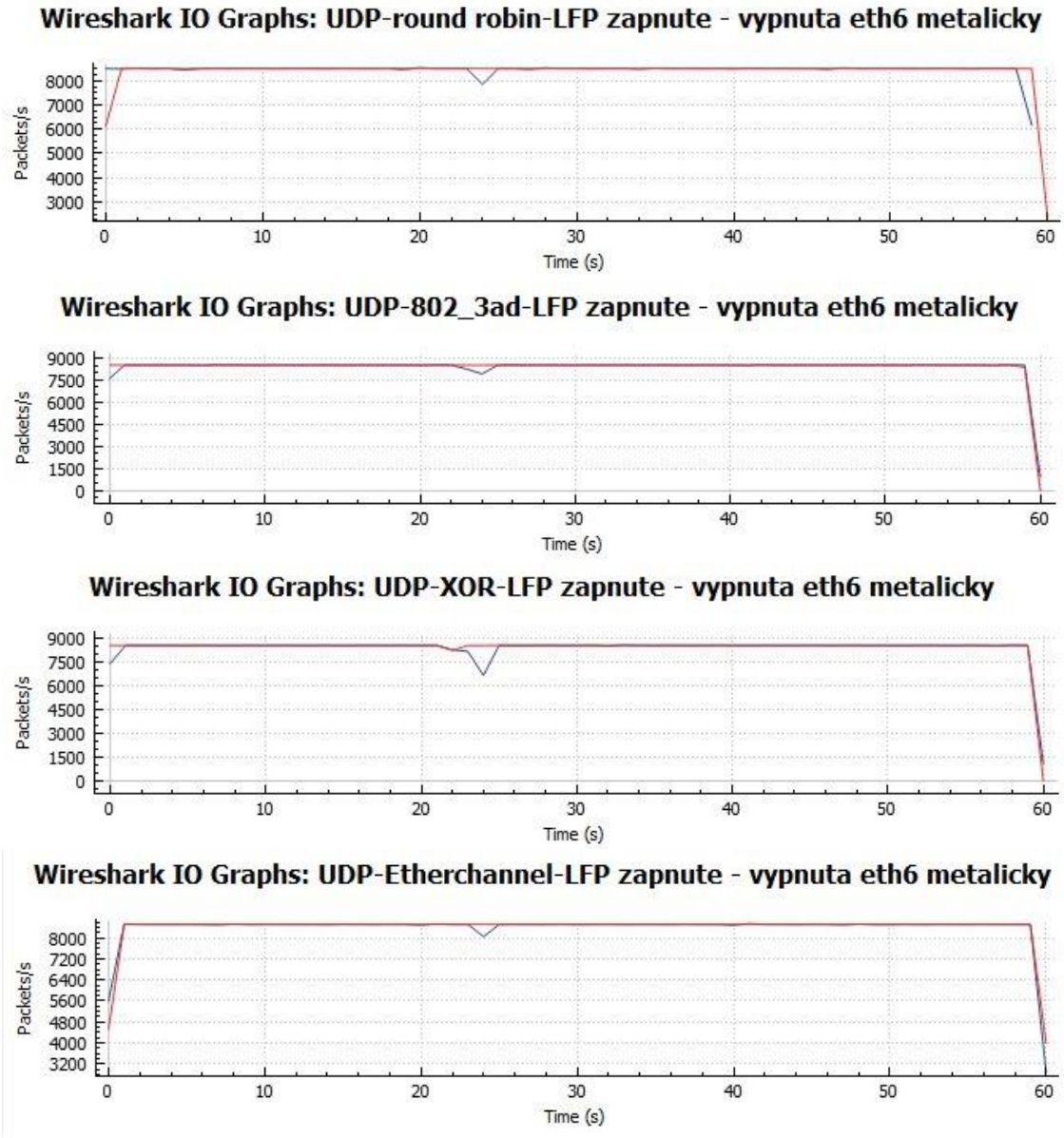
Při vypnutí a po 20s opětné zapnutí linky Eth6 RB1 (Tx, Rx) dochází v době přerozdělení linek k postupnému snižování propustnosti způsobené pravděpodobně vyprazdňováním vyrovnávacích pamětí jednotlivých rozhraní. Toto krátkodobé snížení propustnosti má nepatrný vliv na hodnotu ztrátovosti.

Tabulka 3.3: Naměřené průměrné hodnoty na 10-ti spojení pro jednotlivé metody

směr	metoda	ztrátovost [%]	jitter [ms]	latence [ms]	pkt mimo pořadí [%]
PC1 > PC2	round robin	0,011	0,54	0,392	0,02
	802.3ad	0,076	0,49	0,380	0,00
	balance-XOR	0,007	0,55	0,428	0,00
	Etherchannel	0,022	0,21	0,819	0,00
PC2 > PC1	round robin	0,140	0,36	0,425	0,40

	802.3ad	0,175	0,37	0,321	0,00
	balance-XOR	0,442	0,28	0,333	0,00
	Etherchannel	0,009	0,25	0,648	0,00

Podle grafu lze předpokládat, že po vypnutí linky nedojde k úplnému vyprázdnění vyrovnávacích pamětí rozhraní, kdy provoz neklesne až k nule. Při zapnutí linky ve 40s se linka připojí do svazku, aniž by došlo ke snížení propustnosti.



Obrázek 3.9: Grafy propustnosti při vypnutí linky Eth6 ve 20s a opětném zapnutí ve 40s

3.3.4 Vypojení a opětné zapojení linky optického spoje při vypnutém LFP

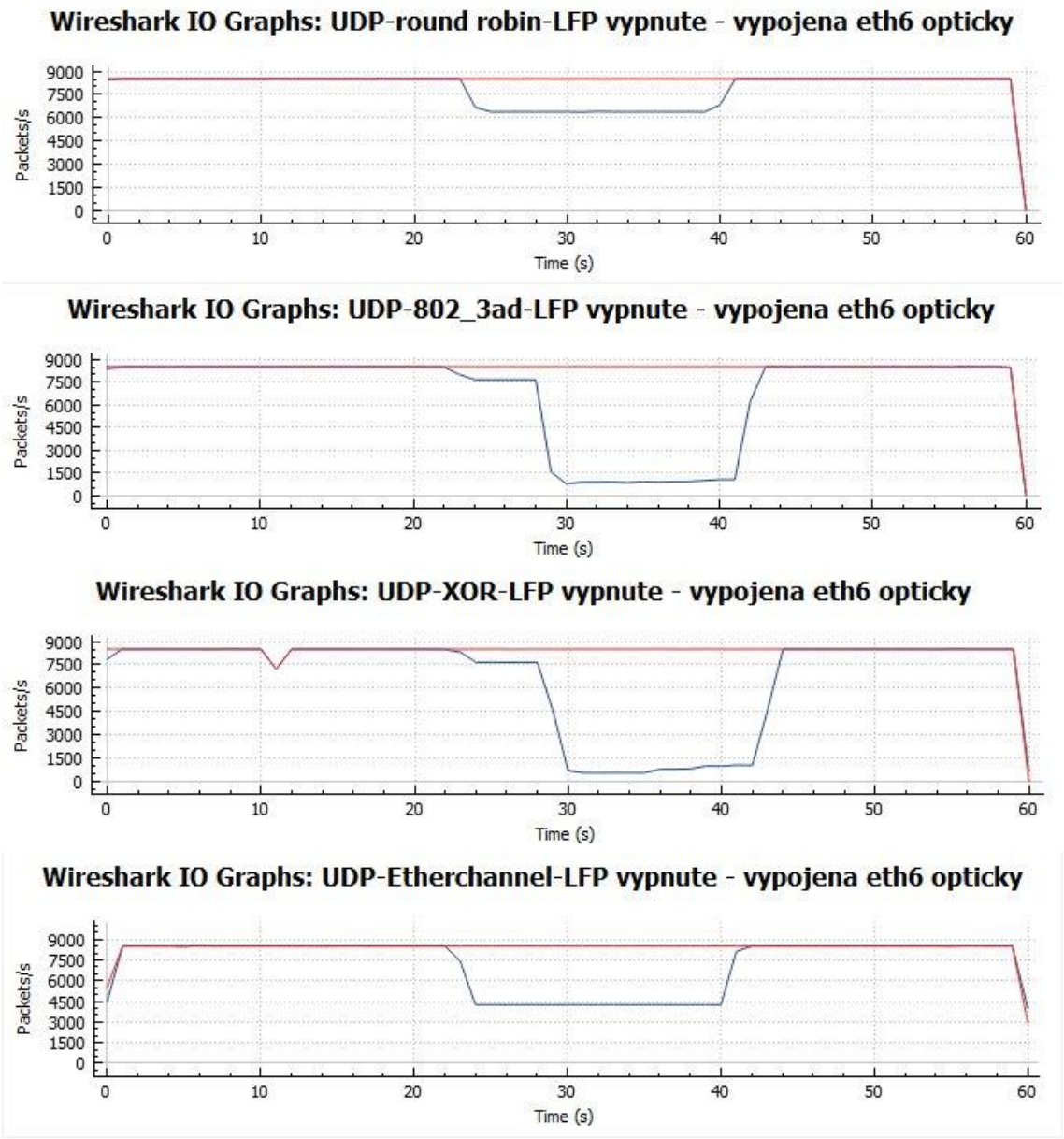
Při vypnuté funkci LFP a zapnutém monitorování MII na rozhraní Eth6 RB1 směr Tx je vypojena a opět zapojena optická linka. Poněvadž při monitorování MII je určována

funkčnost linky Eth6 pomocí stavu metalického rozhraní optického spoje, nedochází při vypnutém LFP k této detekci a odesílané pakety se ztrácí.

Tabulka 3.4: Naměřené průměrné hodnoty na 10-ti spojení pro jednotlivé metody

směr	metoda	ztrátovost [%]	jitter [ms]	latence [ms]	pkt mimo pořadí [%]
PC1 > PC2	round robin	6,950	0,86	36,793	0,02
	802.3ad	21,600	0,48	32,242	0,03
	balance-XOR	23,700	0,34	30,718	0,04
	Etherchannel	15,504	0,17	0,729	0,00
PC2 > PC1	round robin	7,000	0,34	37,429	0,48
	802.3ad	20,900	0,42	49,163	0,06
	balance-XOR	22,000	0,40	32,352	0,06
	Etherchannel	0,005	0,24	0,620	0,00

V případě přerušení optického spoje na lince Eth6 bez funkce LFP a při monitorování MII klesne u metod s přímým hashováním rychlost všech linek ve svazku až k nule. Mechanismus protokolu Spanning Tree použitý pro zálohování spoje tento stav detekuje a tím dojde k přechodu na náhradní spoj Wi-Fi. Metoda Etherchannel využívá monitorování pomocí protokolu LACP ke zjištění stavu linky Eth6 a při optickém vypojení po 60s přejde port Eth6 ze stavu bundle do stavu independent. Po tuto dobu dochází v rozpojeném směru ke ztrátám paketů. Poté se provoz z linky Eth6 rozloží s využitím hashe do ostatních linek. Po detekci funkčnosti optického spojení přejde po 15s port Eth6 opět do stavu bundle a obnoví na něm provoz. Z důvodu délky trvání celého procesu přechodu linky Eth6 do stavu independent (60s) a zpět (15s) je tento proces pouze popsán a časový graf je prezentován stejným způsobem jako ostatní (obnovení optického spoje ve 40s). Při měření linkou Eth6 procházel z důvodu algoritmu hash pouze jeden provoz ze dvou stejným směrem. Tomu odpovídají hodnoty ztrátovosti uvedené v tabulce (1/3 času a ½ provozu odpovídá ztrátovosti 16%). V opačném směru je optický spoj funkční a ke ztrátám nedochází. Ztráty u ostatních metod využívající hash dojde k vypnutí celého svazku a k přepnutí na záložní spoj Wi-fi, kde dochází ke značným ztrátám způsobených malou propustností tohoto spoje. Metoda round robin výpadek optického spoje nezaznamená, a proto ztrátovost odpovídá $\frac{1}{3}$ času a $\frac{1}{4}$ provozu ve 4 linkách.



Obrázek 3.10: Grafy propustnosti při přerušení linky Eth6 ve 20s na 20s

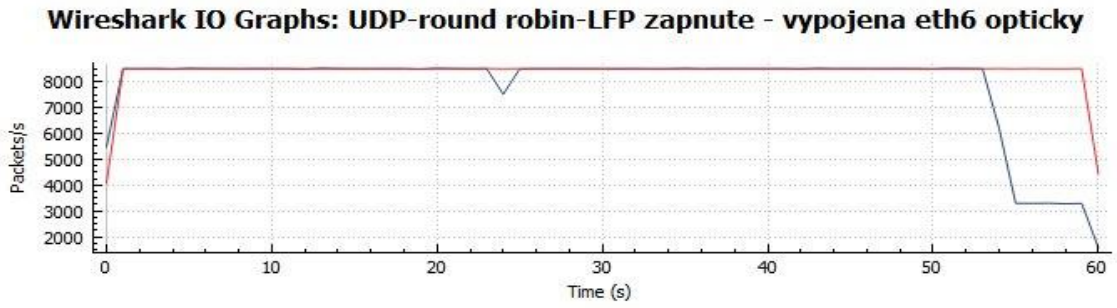
3.3.5 Vypojení a opětné zapojení linky optického spoje při zapnutém LFP

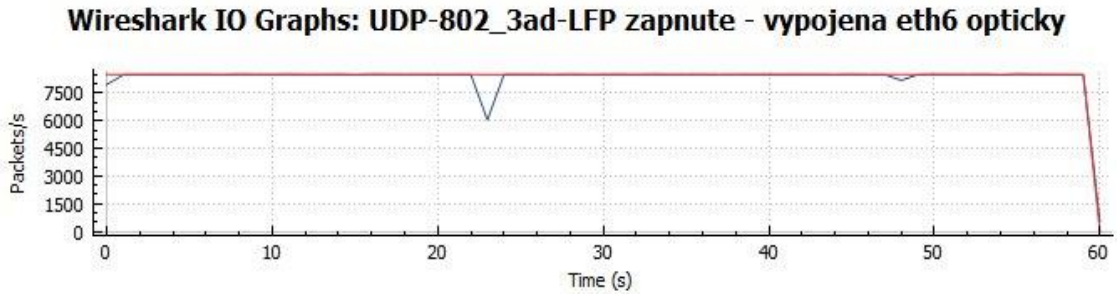
Při zapnuté funkci LFP a zapnutém monitorování MII na rozhraní Eth6 RB1 směr Tx je vypojena a opět zapojena optická linka. Zapnutá funkce LFP společně s monitorováním MII umožňuje spolehlivou detekci přerušení optického spojení. Metody vyvažování v takovém případě přepočítají rozdělení do ostatních linek a po tuto dobu může dojít k případným ztrátám. Hodnoty naměřené v tabulce ukazují nejvyšší ztrátovost a nejhorší hodnoty ostatních parametrů pro metodu round robin.

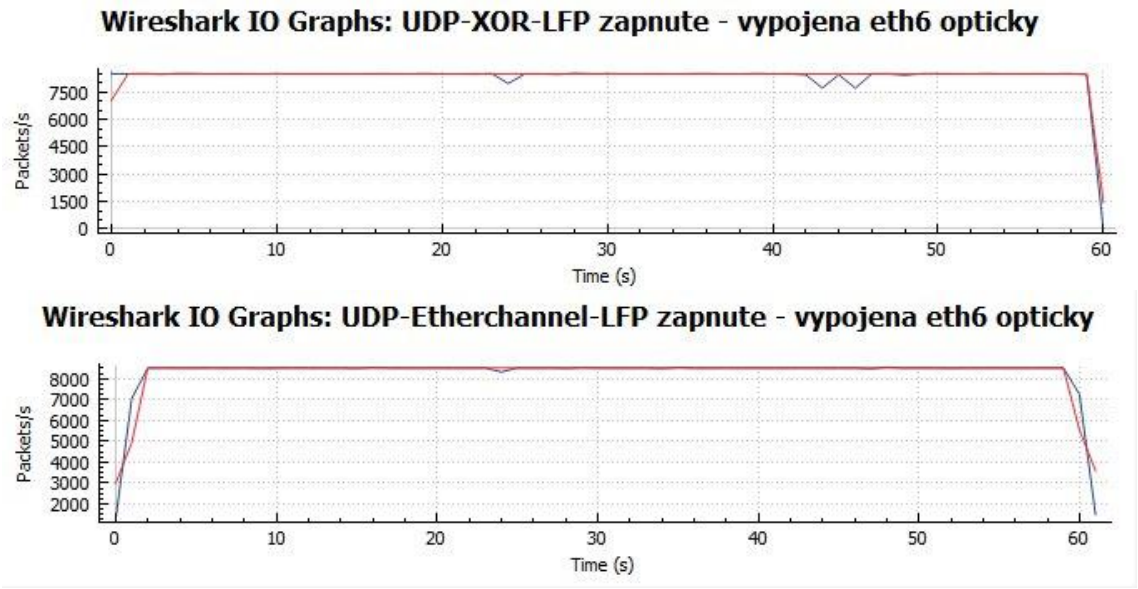
Tabulka 3.5: Naměřené průměrné hodnoty na 10-ti spojení pro jednotlivé metody

směr	metoda	ztrátovost [%]	jitter [ms]	latence [ms]	pkt mimo pořadí [%]
PC1 > PC2	round robin	2,040	4,89	1,715	0,71
	802.3ad	0,513	0,82	0,623	0,00
	balance-XOR	0,455	0,43	0,751	0,00
	Etherchannel	0,035	0,18	0,626	0,00
PC2 > PC1	round robin	6,060	11,9	2,468	1,07
	802.3ad	0,563	0,50	0,518	0,00
	balance-XOR	0,445	0,33	0,686	0,00
	Etherchannel	0,002	0,15	0,723	0,00

Z následujících časových grafů je možné stanovit dobu poklesu propustnosti, která je srovnatelná a má hodnotu 2s. Při opětném zapojení optického spoje ve 40s přejde plynule provoz na všechny linky bez poklesu propustnosti.



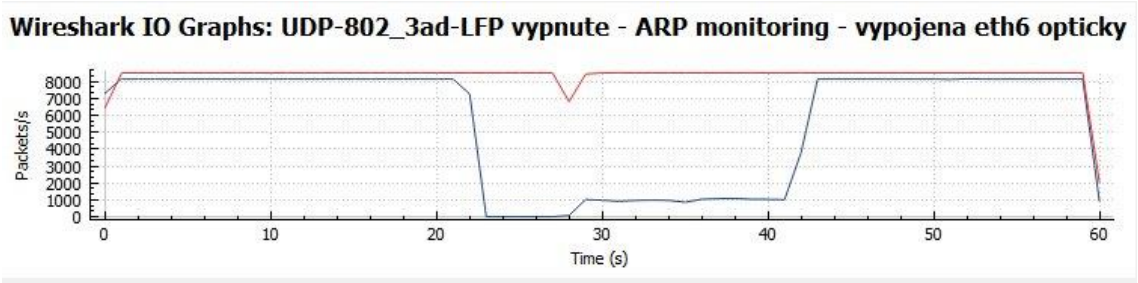




Obrázek 3.11: Grafy propustnosti při vypojení a opětném zapnutí optického spoje na lince Eth6 po 20s

3.3.6 Vypojení a opětné zapojení linky optického spoje s monitorováním ARP

Při vypnuté funkci LFP a zapnutém monitorování ARP na rozhraní Eth6 RB1 směr Tx je vypojena a opět zapojena optická linka. ARP monitorování nepodporují metody Etherchannel a 802.3ad, protože používají pro monitorování protokol LACP. Při volbě kombinace 802.3ad s ARP monitorováním dojde k vypnutí svazku Bonding a přechodu na záložní linku, což ukazuje následující graf:



Obrázek 3.12: Graf propustnosti při ARP monitorování pro metodu 802.ad

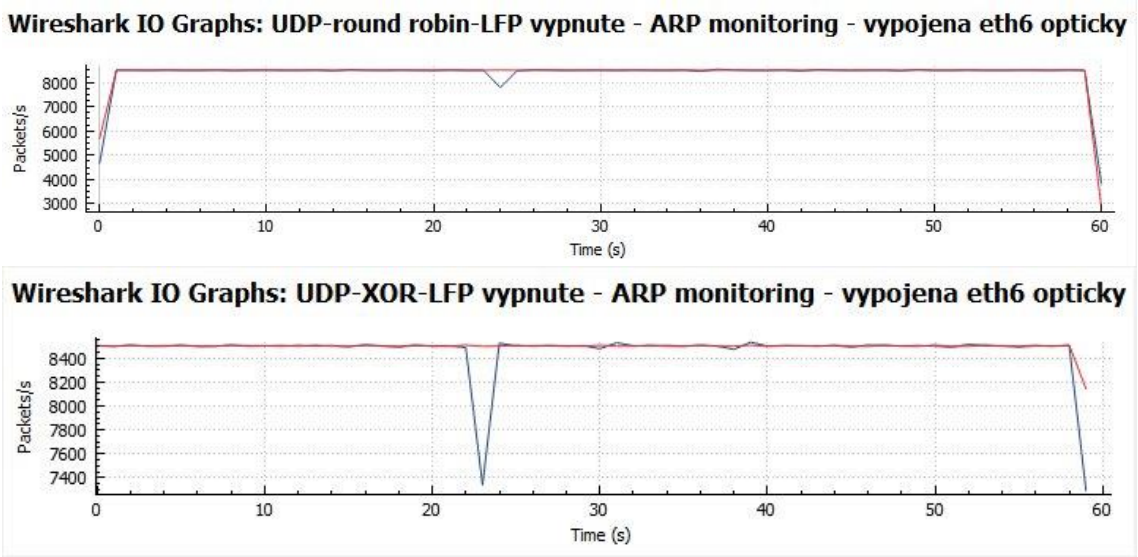
Ostatní metody při využití ARP monitorování vykazují optimální hodnoty parametrů.

Tabulka 3.6: Naměřené průměrné hodnoty na 10-ti spojení pro jednotlivé metody

směr	metoda	ztrátovost [%]	jitter [ms]	latence [ms]	pkt mimo pořadí [%]
PC1 > PC2	round robin	0,113	0,5	0,974	0,02
	802.3ad	-	-	-	-
	balance-XOR	0,093	0,49	0,560	0,00
	Etherchannel	-	-	-	-
PC2 > PC1	round robin	0,151	0,39	0,639	0,50
	802.3ad	-	-	-	-

	balance-XOR	0,235	0,27	0,589	0,00
	Etherchannel	-	-	-	-

Při monitorování ARP byly pro ověřování dostupnosti cílů zadány adresy koncových uzlů spoje PtP. Odečtená doba snížené propustnosti při ARP monitorování z časového grafu je 2s.



Obrázek 3.13: Grafy propustnosti při vypojení a opětném zapnutí optického spoje na lince Eth6 po 20s

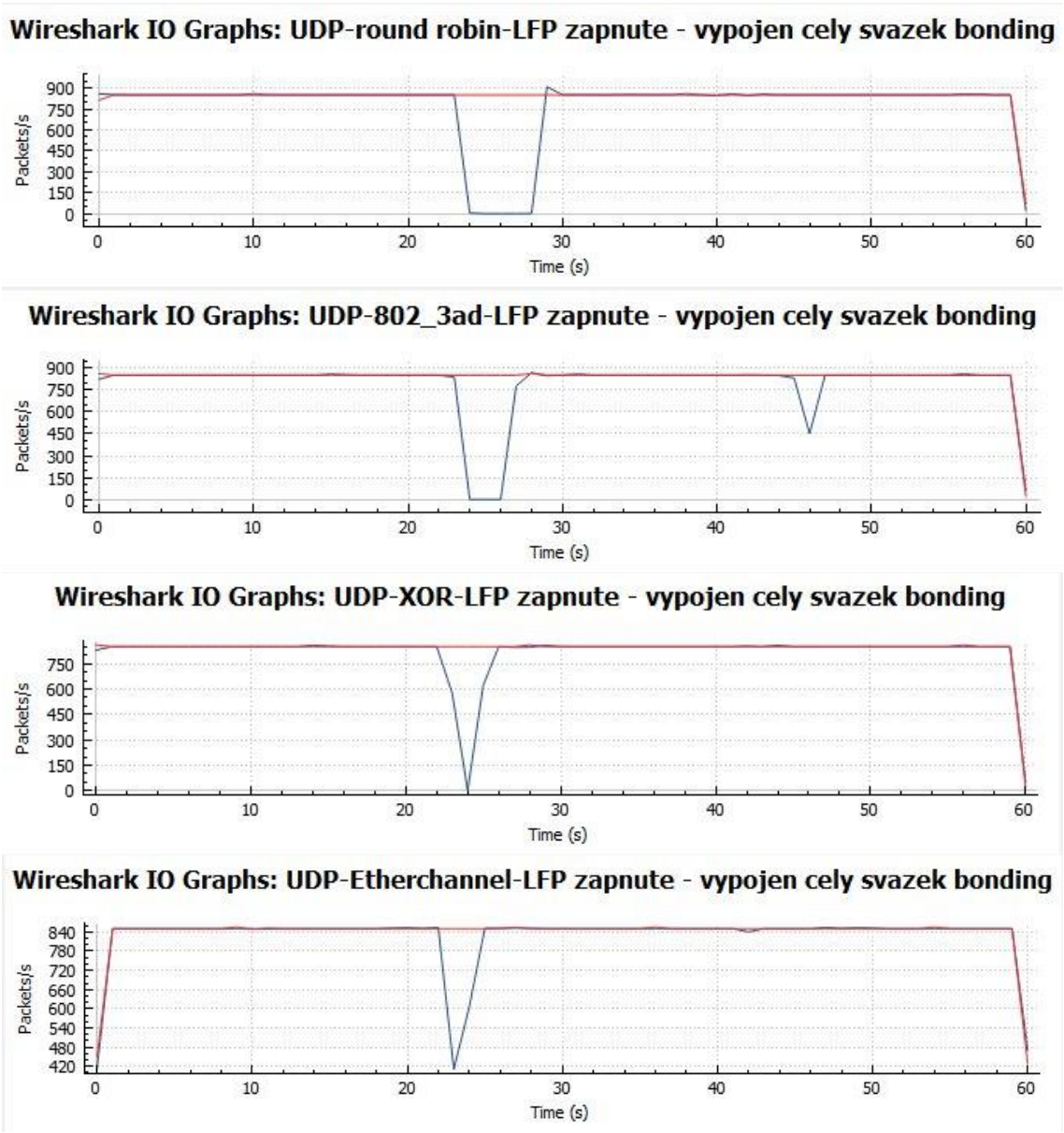
3.3.7 Přechod na záložní linku vypnutím rozhraní svazku

Při vypnutí svazku na rozhraní Bonding (Portchannel) zajistí mechanismus spanning-tree protokolu přechod na záložní linku Wi-fi. K této situaci by mohlo dojít, pokud by všechny linky ve svazku Bonding selhaly. Velká ztrátovost u metody round robin je způsobená úplným výpadkem provozu. Metoda Etherchannel vykazuje v tomto kroku přechodu na záložní linku nejlepší výsledky.

Tabulka 3.7: Naměřené průměrné hodnoty na 10-ti spojení pro jednotlivé metody

směr	metoda	ztrátovost [%]	jitter [ms]	latence [ms]	pkt mimo pořadí [%]
PC1 > PC2	round robin	8,240	0,48	3,576	0,01
	802.3ad	6,636	0,37	0,924	0,13
	balance-XOR	2,620	0,29	2,370	0,01
	Etherchannel	0,828	0,19	1,470	0,00
PC2 > PC1	round robin	8,230	0,59	7,964	0,10
	802.3ad	6,000	0,10	42,861	0,02
	balance-XOR	2,680	0,13	2,460	0,00
	Etherchannel	0,816	0,20	1,321	0,00

Z časového průběhu je patrná odlišná doba potřebná k přechodu na záložní linku pro různé metody při stejném mechanismu spanning-tree protokolu v rozmezí 2,5s až 5s. Během návratu ze zálohy vykazuje metoda 802.3ad jako jediná pokles propustnosti pravděpodobně způsobená vyjednáváním LACP a výpočtem hash. U metod round robin a 802.3ad dojde až k úplnému výpadku provozu.



Obrázek 3.14: Grafy propustnosti při přechodu na záložní spoj Wi-fi a zpět

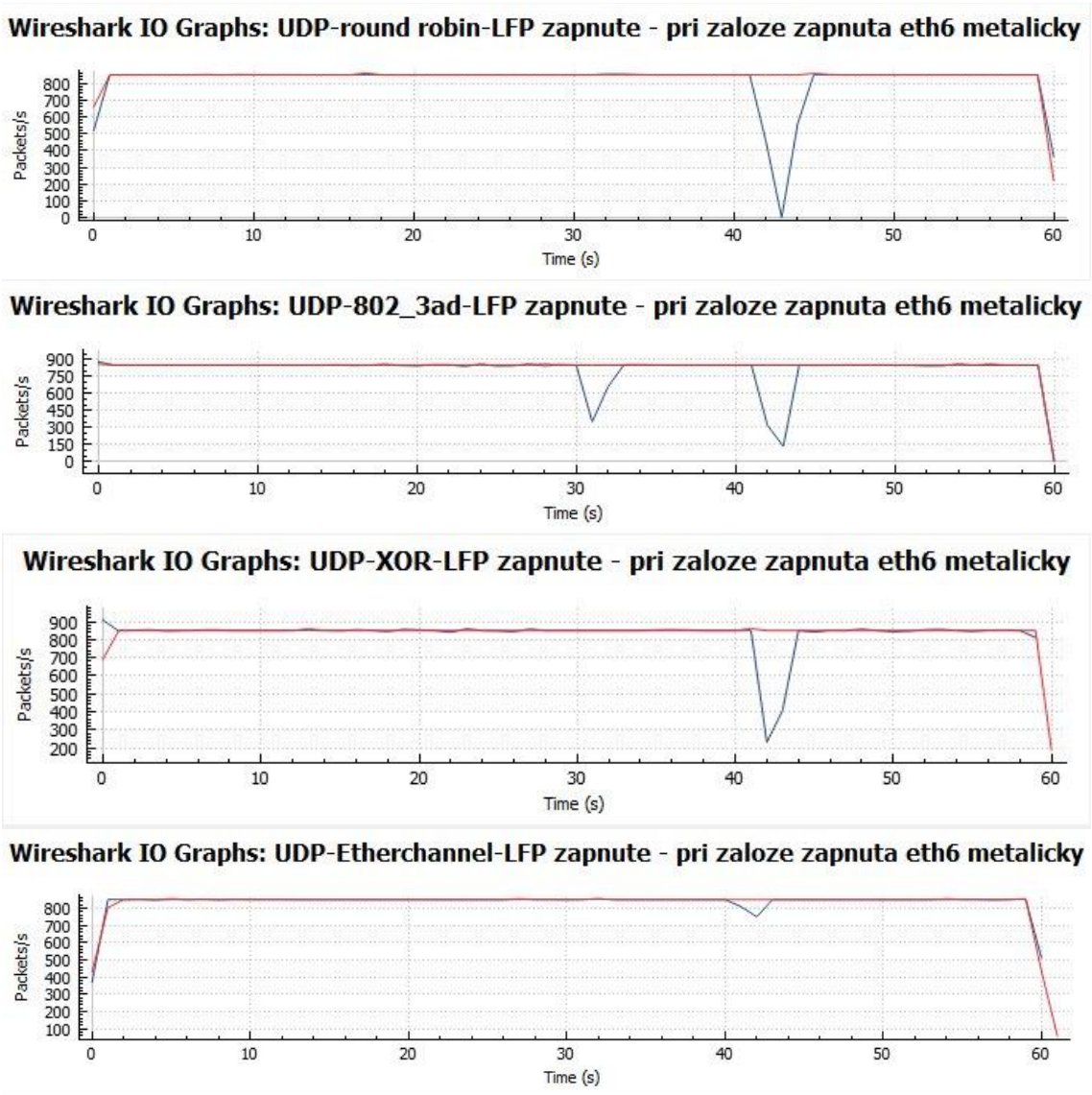
3.3.8 Přechod ze záložní linky zapnutím jedné linky ve svazku

Při přechodu ze záložní linky na linku Eth6 RB1 (Tx, Rx) ve svazku Bonding (zapnutí funkce Bonding) dochází k omezení propustnosti pouze u metody 802.3ad, což se projevuje největší ztrátovostí.

Tabulka 3.8: Naměřené průměrné hodnoty na 10-ti spojení pro jednotlivé metody

směr	metoda	ztrátovost [%]	jitter [ms]	latence [ms]	pkt mimo pořadí [%]
PC1 > PC2	round robin	3,010	0,99	5,273	0,00
	802.3ad	3,820	2,43	4,385	0,00
	balance-XOR	2,100	1,81	6,374	0,00
	Etherchannel	0,113	2,26	3,451	0,00
PC2 > PC1	round robin	3,010	1,2	4,379	0,00
	802.3ad	3,800	4,39	5,839	0,00
	balance-XOR	2,100	3,18	6,325	0,00
	Etherchannel	0,198	0,68	2,684	0,00

Přechod ze záložní linky je plynulý, pouze metoda 802.3ad vykazuje čas přechodu 2,5s. Návrat na záložní linku je podobný jako v předcházejícím kroku.



Obrázek 3.15: Grafy propustnosti při přechodu ze záložního spoje Wi-fi na linku Eth6 a zpět

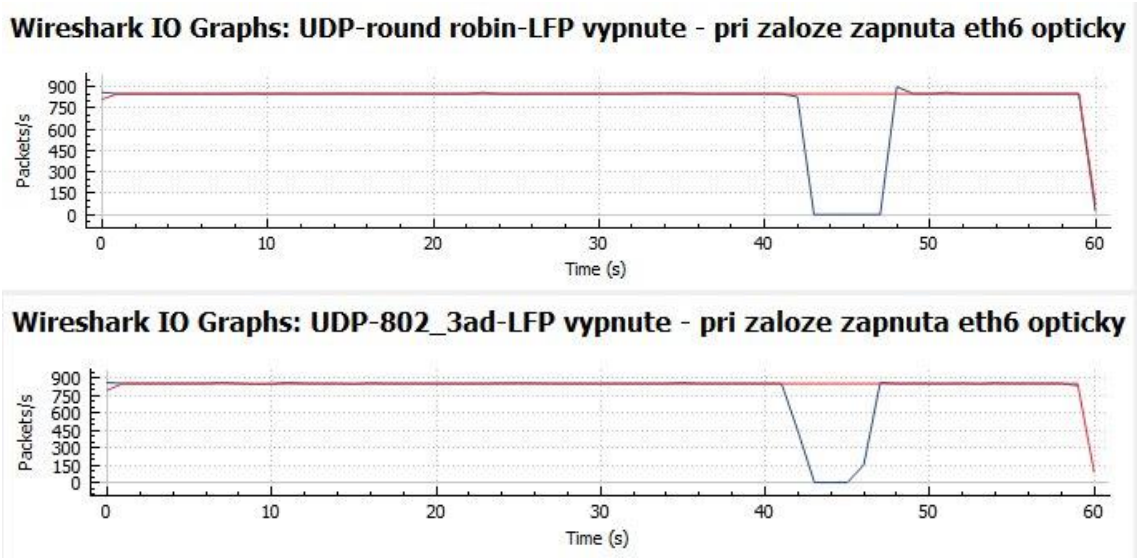
3.3.9 Přechod ze záložní linky zapojením jedné linky svazku při vypnutém LFP

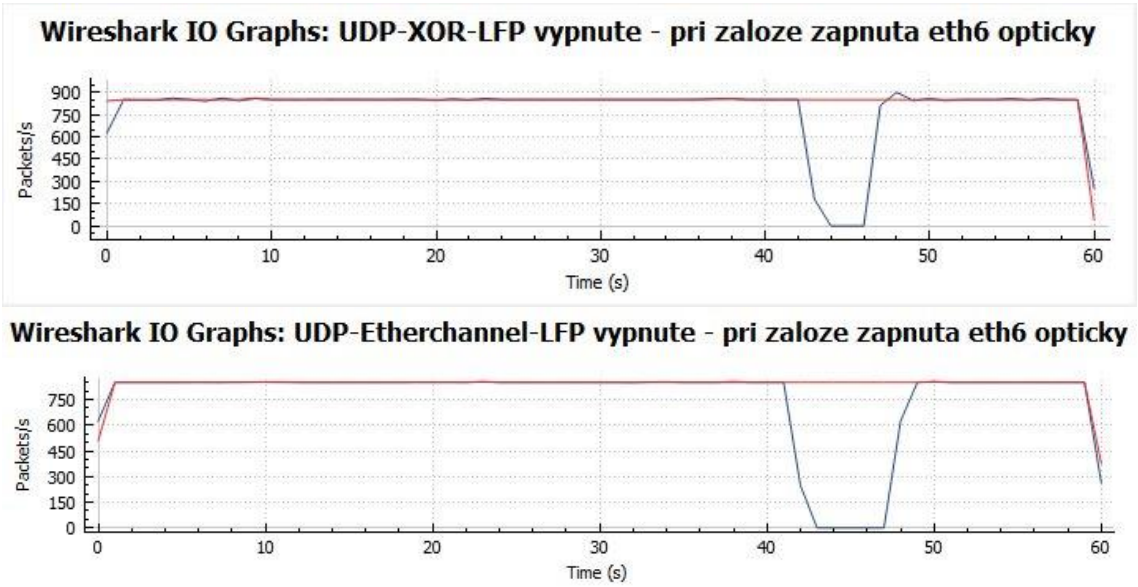
V testování je zapojena a opětně vypojena linka Eth6 RB1 směr Tx na optickém spoji při vypnuté funkci LFP a zapnutém monitorování MII. Při přechodu ze záložní linky na linku Eth6 jsou naměřené parametry metod srovnatelné a souvisí s podobným průběhem časových grafů propustnosti.

Tabulka 3.9: Naměřené průměrné hodnoty na 10-ti spojení pro jednotlivé metody

směr	metoda	ztrátovost [%]	jitter [ms]	latence [ms]	pkt mimo pořadí [%]
PC1 > PC2	round robin	8,310	1,68	8,184	0,00
	802.3ad	7,320	1,49	3,435	0,00
	balance-XOR	6,330	1,64	5,353	0,00
	Etherchannel	7,040	1,60	7,324	0,00
PC2 > PC1	round robin	8,300	1,10	8,361	0,00
	802.3ad	7,250	1,49	3,407	0,00
	balance-XOR	6,300	2,49	5,981	0,06
	Etherchannel	7,160	1,07	7,221	0,00

V časových grafech je patrný hladký průběh přechodu ze záložní linky na již metalicky zapnutou linku Eth6 zajištěný pravděpodobně vyjednáváním spanning-tree protokolu. Při zpětném přechodu na záložní linku je doba vyjednávání srovnatelná se zapnutím celého svazku Bonding a je v rozmezí 5s až 8s.





Obrázek 3.16: Grafy propustnosti při přechodu ze záložního spoje Wi-fi na linku Eth6 a zpět

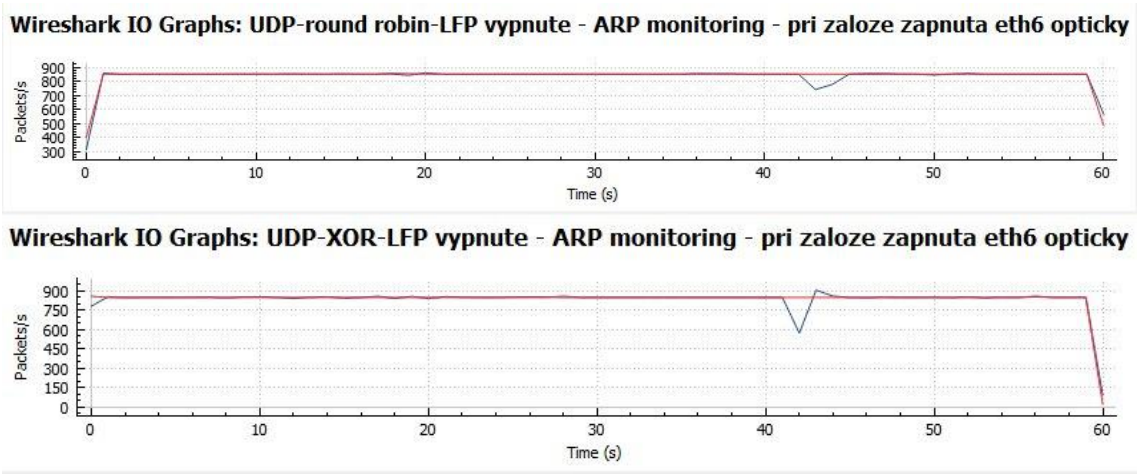
3.3.10 Přechod ze záložní linky zapojením jedné linky svazku při vypnutém LFP a monitorování ARP

V testování je zapojena a opětně vypojena linka Eth6 RB1 směr Tx na optickém spoji při vypnuté funkci LFP a zapnutém monitorování ARP. Poněvadž metody 802.3ad a Etherchannel ARP nepodporují, bylo provedeno testování pouze pro metody round robin a balance-XOR. Testované parametry u obou metod nabývají optimálních hodnot.

Tabulka 3.10: Naměřené průměrné hodnoty na 10-ti spojení pro jednotlivé metody

směr	metoda	ztrátovost [%]	jitter [ms]	latence [ms]	pkt mimo pořadí [%]
PC1 > PC2	round robin	0,472	1,29	5,891	0,00
	802.3ad	-	-	-	-
	balance-XOR	0,480	1,34	6,035	0,00
	Etherchannel	-	-	-	-
PC2 > PC1	round robin	0,354	1,08	3,167	0,00
	802.3ad	-	-	-	-
	balance-XOR	0,428	2,29	5,447	0,00
	Etherchannel	-	-	-	-

Časové průběhy propustnosti při přechodu ze zálohy zapnutím optického spoje na lince Eth6 nevykazují pokles propustnosti. Při zpětném návratu na zálohu dochází k nepatrnému zakolísání.



Obrázek 3.17: Grafy propustnosti při přechodu ze záložního spoje Wi-fi na linku Eth6 a zpět

3.4 Vyhodnocení vybraných metod vyvažování

Pro hodnocení metod je z testování patrné, že první dva testovací kroky (všechny linky zapnuté, vypnutá linka Eth6) není nutné uvažovat. Provoz v těchto krocích probíhá ve statickém režimu, který se chová podobně při plném nebo jiném počtu aktivních linek ve svazku a mění se pouze celá propustnost svazku. Z charakteru dalších testovacích kroků (použití LFP, monitorování MII a ARP) vyplývá, že není nutné hodnotit některé metody vyvažování v krocích, které jsou jednoznačně nevhodné (v následující tabulce označené *). Jedná se o kombinaci vypnuté LFP a současné monitorování MII, monitorování ARP v metodách 802.3ad a Etherchannel. Proto testovací kroky pro takové metody se ve srovnávání neuvažují.

3.4.1 Výběr vhodných metod pro různé provozy

Srovnání chování testovaných metod v jednotlivých testovacích krocích znázorňuje následující tabulka. Hodnota plus znamená vhodnost metody v testovacím kroku, hodnota nula označuje vhodnou alternativu a hodnota mínus označuje nevhodnost metody pro daný testovací krok. Při stanovení těchto hodnot byla hlavním kritériem ztrátovost, doručování paketů mimo pořadí a doba výpadku.

Tabulka 3.11: Výběr vhodných metod vyvažování zátěže

	metody vyvažování zátěže			
kroky měření	**round robin**	**802.3ad**	**balance-XOR**	**Etherchannel**
Vypnutí a opětné zapnutí jedné linky na metalické části spoje	-	-	0	+
Při zapnutém LFP a při monitorování MII vypojení a opětné zapojení jedné linky na optickém kabelu	-	-	0	+
Při vypnutém LFP a při monitorování ARP vypojení a opětné zapojení jedné linky na optickém kabelu	+	-*	+	-*
Vypnutí a opětné zapnutí celého svazku při monitorování MII – využití záložní linky	-	-	0	+
Zapnutí a opětné vypnutí jedné linky při monitorování MII na metalické časti spoje - využití záložní linky	-	-	0	+
Při vypnutém LFP a při monitorování ARP zapojení a opětné vypojení jedné linky na optickém kabelu - využití záložní linky	+	-*	+	-*

Z tabulky vyplývá, že pro různý typ PtP hybridních spojů jsou vhodné různé metody vyvažování:

a) Pro PtP hybridní spoje umožňující monitorování MII ve spolupráci s LFP je nevhodnější metoda Etherchannel.
b) Pro PtP hybridní spoje bez možnosti funkce LFP je nutné využít monitorování ARP a nejvhodnější metoda je balance-XOR.
c) V případě provozu bez možnosti využití hashování, je nutné využít metodu round robin s využitím monitorování ARP a přitom hybridní PtP spoj nepotřebuje využívat funkci LFP.

3.4.2 Ověření vybraných nejlepších metod pomocí provozu TCP

a) Metoda Etherchannel při zapnutém LFP a monitorováním MII

Je vybrán testovací krok pro ověření metody Etherchannel při vypojení a opětném zapojení optického spoje pro odchozí provoz na lince Eth6 RB1. Souhrný graf TCP provozu je uveden v příloze C.

- Průměrná měřená latence pomocí funkce ping při provozu TCP je 5,607.
- Hodnota celkových chyb v provozu TCP je vyrovnaná při prováděných změnách na optické lince.
- Z celkových chyb v provozu TCP převažují duplikované odpovědi a ve stejném poměru k hodnotě celkových chyb jsou opětně přeposílány pakety (retransmission).

a) Metoda balance-XOR při vypnutém LFP a při monitorování ARP

Je vybrán testovací krok pro ověření metody balance-XOR při vypojení po 20s a opětném zapojení po 20s optického spoje pro odchozí provoz na lince Eth6 RB1. Souhrný graf TCP provozu je uveden v příloze D.

- Průměrná měřená latence pomocí funkce ping při provozu TCP je 8,150.
- Hodnota celkových chyb v provozu TCP nepatrně narůstá při vypojení optického spoje pravděpodobně v důsledku omezení celkové kapacity svazku
- Z celkových chyb v provozu TCP rovněž převažují duplikované odpovědi a dochází k přeposílání paketů ve stejném poměru k hodnotě celkových chyb.

b) Metoda round robin při vypnutém LFP a při monitorování ARP

Je vybrán testovací krok pro ověření metody round robin při vypojení po 20s a opětném zapojení po 20s optického spoje pro odchozí provoz na lince Eth6 RB1. Souhrný graf TCP provozu je uveden v příloze E.

- Průměrná měřená latence pomocí funkce ping při provozu TCP je 0,921.
- Množství celkových chyb v provozu TCP je značný s porovnáním s výsledky v bodě a), b). Toto měření ukazuje velké množství celkových chyb v provozu TCP, když jsou použity v hybridním svazku různé typy linek (metalika, optika – 1 a 3 část doby testování) na rozdíl od druhé části doby testování, kdy optický spoj vypadne ze svazku a provoz prochází pouze přes metalické linky.

- Z celkových chyb v provozu TCP narůstá především počet duplikovaných odpovědí, přičemž změny počtu přeposíláných paketů a paketů mimo pořadí jsou průběhu testování nevýrazné.

4 Závěr

Při zpracování této práce byly v teoretické části obecně rozebrány principy známých metod vyvažování zátěže s ohledem na charakteristické vlastnosti jejich chování a možnosti nasazení na vícekanálových PtP hybridních spojích. Při studii dostupné literatury byly jednotlivé metody popisovány v konkrétních nasazeních, např. při vyvažování zátěže linuxových serverů s více síťovými rozhraními. Poněvadž nebyly nalezeny jiné konkrétní příklady nasazení metod vyvažování, především pro vyvažování na vícekanálových hybridních PtP spojích, stala se tato literatura základem pro popsání chování jednotlivých metod. Rozbor chování jednotlivých metod pro využití na PtP spojích, výběru hodnotících kritérii, stanovení postupu testování a vyhodnocení výsledků testů považuji za hlavní přínos v této práci.

V práci byla vybrána jako nejvhodnější metoda pro vícekanálové PtP hybridní spoje metoda balance-XOR a v případě provozu bez možnosti rozlišení na vrstvách L3 a L4 ISO/OSI modelu metoda round robin. V tomto případě jsou uvažovány PtP hybridní spoje, které využívají linky bez možnosti detekce přerušení spoje vypínáním a zapínáním přípojného metalického rozhraní (funkce LFP). V případě možnosti využití funkce LFP na PtP spoji je vybrána jako nejvhodnější proprietární metoda Etherchannel – Cisco. Metoda 802.3ad již podle názvu předpokládá využití protokolu LACP podobně jako metoda Etherchannel, ale testováním se prokázalo, že při nedodržení podmínek stejných linek ve svazku metoda 802.3ad celý svazek vypíná, kdežto metoda Etherchannel linku vyloučí ze svazku a pokračuje v provozu.

Diskuse a otevřené otázky:

V průběhu zpracování této práce byla postupně objevována složitost původně stanoveného cíle, především nejasností principu a fungování metody balance-alb, kterou proto nebylo možné zahrnout do výběru metod v praktické části. Tato metoda vyžaduje použití nesymetrické topologie (na jedné straně balance-alb, na druhé straně přepínač v základním režimu). Dále předpokládá použití speciálních síťových karet (s možností změn adres MAC za běhu provozu) z důvodu posílání rámců protokolu gratuitous ARP pro určení linky zpětného provozu. V důsledku celkového rozsahu a objemu práce tato metoda nebyla testována a mohla by se stát vhodným tématem pro zpracování v případné navazující práci. Rovněž nebylo možné otestovat funkci modulu Free Space Optic (FSO) především z důvodu jeho nekompletnosti (pouze jeden směr jednoho spoje) a nespolehlivosti projevující se změnou chování v průběhu testu. V případné navazující

práci by bylo vhodné se také zaměřit na testování hybridního vícekanálového PtP spoje s využitím více stabilních FSO modulů.

5 Seznam literatury

[1] GIGUÉRE, Bruno. RFC 2544: HOW IT HELPS QUALIFY A CARRIER ETHERNET NETWORK. In: www.EXFO.com [online]. 3. 3. 2008, Quebec [cit. 2015-08-20]. Dostupné z:

http://www.3-edge.de/export/sites/default/.content/3Edge_Datasheets-pdf/RFC2544-Howithelps-qualify-a-carrierEthernet-Network.pdf

[2] Macourek, Tomáš. Co je SLA a jaké jsou její doby (příklady). In: https://helpdesk.microware.cz [online]. 28. 6. 2014 [cit. 2015-11-15]. Dostupné z: https://helpdesk.microware.cz/index.php?/Knowledgebase/Article/View/87/0/co-je-sla-a-jake-jsou-jeji-doby-piklady

[3] CAO, Zhiruo, ZEGURA, Ellen (College of Computing, Georgia Institute of Technology, Atlanta, GA 30332-0280), WANG, Zheng (Bell Labs, Lucent Technologies, Holmdel, NJ 07733). Performance of Hashing-Based Schemes for Internet Load Balancing [online]. 7. 11. 2007 [cit. 2015-06-10]. Dostupné z:

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.41.3646&rep=rep1&type=pdf

[4] MikroTik. MikroTik RouterOS Workshop Load Balancing Best Practice. In: http://www.slideshare.net [online]. 8.10.2013[cit. 2015-10-18]. Dostupné z: http://www.slideshare.net/kkeker/mikrotik-load-balansing?related=1

[5] Braden, Borman. Computing the Internet Checksum. In:https://tools.ietf.org [online]. Září 1988, aktualizováno leden 1990 [cit. 2015-10-03]. Dostupné z:

https://tools.ietf.org/html/rfc1071

[6] Cyclic redundancy check. In: Wikipedia: the free encyklopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2015-10-6]. Dostupné z: https://en.wikipedia.org/wiki/Cyclic_redundancy_check

[7] Hopps, C. Analysis of an Equal-Cost Multi-Path Algorithm. In: https://tools.ietf.org [online]. Listopad 2000 [cit. 2015-9-29]. Dostupné z:

https://tools.ietf.org/html/rfc2992

[8] Manual:Interface/Bonding. In: wiki.mikrotik.com [online]. 2. 2. 2015 [cit. 2015-05-20]. Dostupné z:

http://wiki.mikrotik.com/wiki/Manual:Interface/Bonding

[9] Davis, Thomas. Linux Ethernet Bonding Driver HOWTO. In: https://www.kernel.org [online]. 15. 10. 2000, aktualizováno 27. 4. 2011 [cit. 2015-09-30]. Dostupné z:

https://www.kernel.org/doc/Documentation/networking/bonding.txt

[10] FRAZIER, Howard (Broadcom), VAN DOORN, Schelto, HAYS, Robert (Intel), MULLER, Shimon (Sun Microsystems), TOLLEY, Bruce (Solarflare Communications), KOLESAR, Paul (CommScope), GEOFF Thompson (Nortel), TURNER, Brad (Juniper Networks). IEEE 802.3ad Link Aggregation (LAG). ieee802.org [online]. 17. 4. 2007 Ottawa [cit. 2015-08-20] Dostupné z:

http://www.ieee802.org/3/hssg/public/apr07/frazier_01_0407.pdf

[11] Cisco Network Academy II

http://www.cisco.com/c/en/us/td/docs/switches/lan/catalyst3550/software/release/12-1_13_ea1/configuration/guide/3550scg/swethchl.html

[12] Linux Foundation. Bonding. In: linuxfoundation.org [online]. 19. 11. 2009 [cit. 2015-08-20]. Dostupné z:

http://www.linuxfoundation.org/collaborate/workgroups/networking/bonding

[13] Schroder, Carla. Tips and Tuning for Ethernet Bonding with Linux. In: http://www.enterprisenetworkingplanet.com [online]. 4. 9 2007 [cit. 2015-11-1]. Dostupné z:

http://www.enterprisenetworkingplanet.com/netsysm/article.php/3697756/Tips-and-Tuning-for-Ethernet-Bonding-With-Linux.htm

[14] Network Connectivity. In: http://support.intel.co.jp [online]. Santa Clara (CA):Intel Corporation. 27. 2. 2014 aktualizováno: 2. 11. 2015 [cit. 2015-11-08]. Dostupné z:

http://support.intel.co.jp/support/network/sb/cs-009747.htm

[15] Linux ethernet bonding configuration. In: http://hacktracking.blogspot.cz [online]. Květen 2013 [cit. 2015-10-14]. Dostupné z:

http://hacktracking.blogspot.cz/2013/05/linux-ethernet-bonding-configuration.html

[16] BAKALA, Břetislav. Realizace optického spoje volným prostorem pro kratší vzdálenosti. Praha, 2013. Diplomová práce. ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE. Fakulta elektrotechnická.

[17] RB2011L. In: http://routerboard.com [online]. 1996- [cit. 2015-08-23]. Dostupné z: http://routerboard.com/RB2011L

[18] RouterBOARD 2011L series. In: http://www.mikrotik.com, Riga (Litva) [online]. 22. 6. 2012 [cit. 2015-08-23]. Dostupné z: http://i.mt.lv/routerboard/files/rb2011L-qg.pdf

[20] Introduction. In:http://ostinato.org [online]. 2015 [cit. 2015-07-08]. Dostupné z: http://ostinato.org

[21] Manual:Winbox. In: wiki.mikrotik.com [online]. 2. 2. 2015 [cit. 2015-10-12]. Dostupné z: http://wiki.mikrotik.com/wiki/Manual:Winbox

Seznam obrázků

Přílohy

A. Standard RFC 2544

Popisuje zkoušky potřebné k měření a prokázání výkonnosti provozu Ethernetových sítí. Tato metodika definuje velikost rámce, dobu trvání zkoušky a počet zkušebních iterací. Testovací sada podporuje sedm předdefinovaných velikosti rámců (64, 128, 256, 512, 1.024, 1280 a 1518 bajtů), aby se zajistilo, že síť Ethernet je schopná podporovat různé služby (jako je například VoIP, video, atd). Při stejném objemu dat zvýší malé rozměry rámce jejich přenášený počet, tím se musí přenést velké množství rámců, které zvýší síťové zatížení. Standard definuje následující testy:

Test propustnosti

Propustnost je definována jako maximální počet rámců za sekundu, které jsou přeneseny bez chyby. Metodika začíná na maximální frekvenci rámců při dané ztrátě rámců (frame loss). V případě, že je překročena daná ztráta frame loss, zvolí se poloviční přenosová rychlost a test se znovu opakuje. Pokud nedochází k frame loss, přenosová rychlost se zvýší o polovinu rozdílu z předchozího procesu. Tato metodika se opakuje, dokud není nalezena rychlost odpovídající danému frame loss. Test je proveden pro každou velikost rámce a dobu 60s. Výsledky jsou vyhodnoceny v tabulce v jednotkách frame/s nebo bps.

Back-to-Back Test

Je známý rovněž jako burst test. Hodnotí schopnost vyrovnávací paměti zařízení tím, že měří maximální počet rámců přijatých při plné rychlosti linky před první ztrátou. Jedná se o přenos s minimální mezirámcovou mezerou (float). Zkušební délka musíbýt minimálnědvěsekundy a měření seopakuje nejméně50 krátsezaznamenáním průměrných hodnot pro každou velikost rámce.

Frame Loss Test

Měří reakci sítě na podmínky přetížení. Jedná se o kritický ukazatel schopnosti sítě podporovat aplikace v reálném čase, při kterých velké množství ztrát rámců vede k degradaci kvality služeb. Testovací zařízení vysílá provoz na maximální rychlosti linky a měří frame loss. Pokud k němu dochází, test se opakuje s rychlostí o 10% nižší, až ke ztrátě nedochází. Výsledky jsou vyneseny do grafu pro různé rychlosti.

Latency Test

Test měří čas potřebný k přenosu rámce od zdroje k cíli (end-to-end testování) nebo od zdroje k cíli a zpět. Velká nebo proměnná latence způsobuje problémy v realtime službách. Test začíná měřením a porovnáním propustnosti pro každou velikost rámce (ověření bezztrátovosti rámců). Měření latence probíhá při naplnění všech vyrovnávacích pamětí. Ve druhém kroku, který trvá 120s, jsou rámce v polovině cesty označkovány (time-stamp) a při zpětném příjmu měřícím zařízením je měřena latence. Měření se opakuje 20x pro každou velikost rámce a vyhodnocuje se jako průměr.

Jitter

Test měří variabilitu času doručení po sobě jdoucích paketů při maximální rychlosti rámců.

B. Použitý Hardware

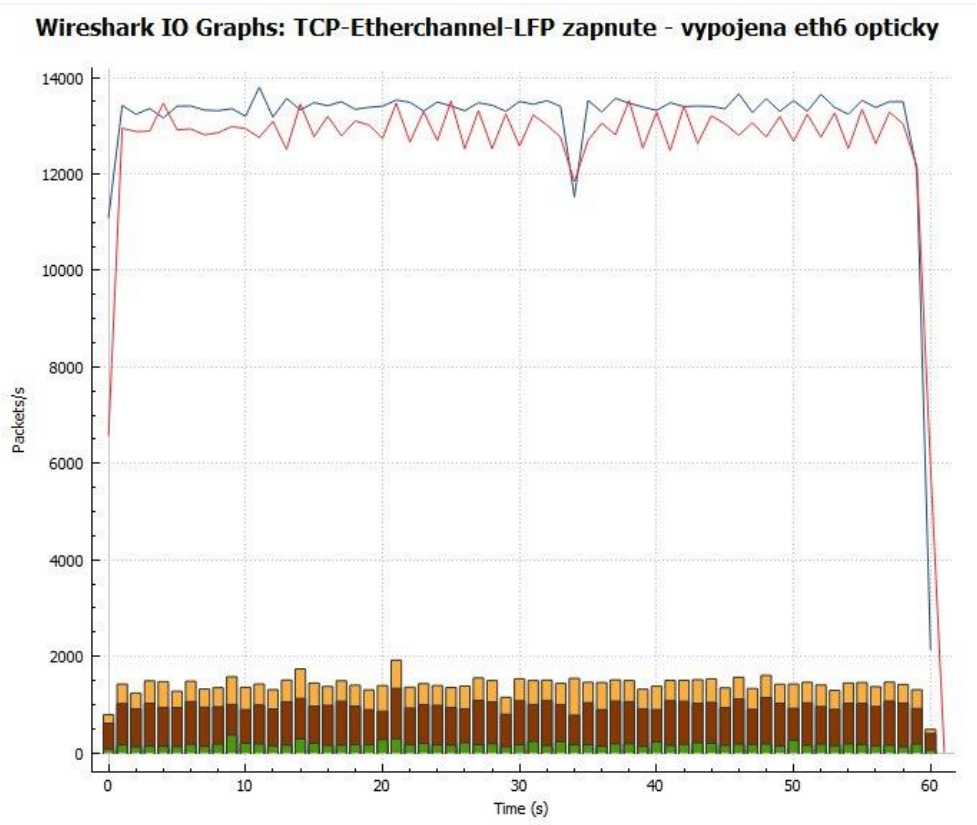
Je využita testovací infrastruktura v napojení dvou koncových bodů (PC1-notebook Toshiba InteCore i3, 4GB RAM a PC2-IntelCore 2 Duo 2,2 Ghz, 2GB RAM s OS Linux Ubuntu 15.04 přes 1Gbps síťovou kartu) na dvě zařízení plnící funkci traffic splitteru (Routerboard MikroTik, Cisco) mezi sebou propojených třemi 100Mbps metalickými linkami a jedním párem Fiber/T Transceiver. 10/100Base-Tx To 100Base-Fx Media Converter. Testování probíhalo na technologii MikroTik RouterBOARD 2011Lseries, firmware: 3.22, RouterOS: 6.29.1 [17]a CISCO Catalyst 2960, Boot Loader (C2960-HBOOT-M) Version 12.2(25r)SEE6, IOS: 12.2(35)SE5.

C. Testování TCP Etherchannel

Name	Display filter	Color
PC2>PC1	tcp.port==5201	
PC1>PC2	tcp.port==5202	
retransmission	tcp.analysis.retransmission	
Out-of-order	tcp.analysis.out_of_order	
Duplikate ACK	tcp.analysis.duplicate_ack	
TCP errors	tcp.analysis.flags	

Obrázek 18: Legenda pro souhrnné grafy testování provozu TCP

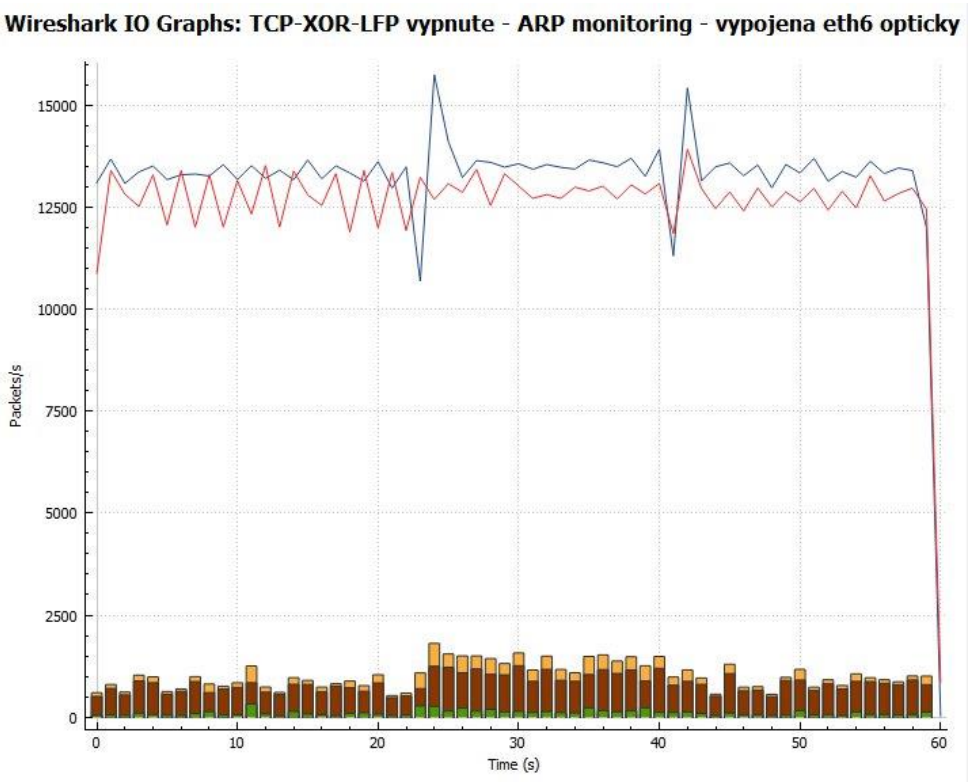
a) Etherchannel Při zapnutém LFP a při monitorování MII vypojení a opětné zapojení jedné linky na optickém kabelu



Obrázek 19: Souhrný graf TCP provozu pro metodu Etherchannel

D. Testování TCP balance-XOR

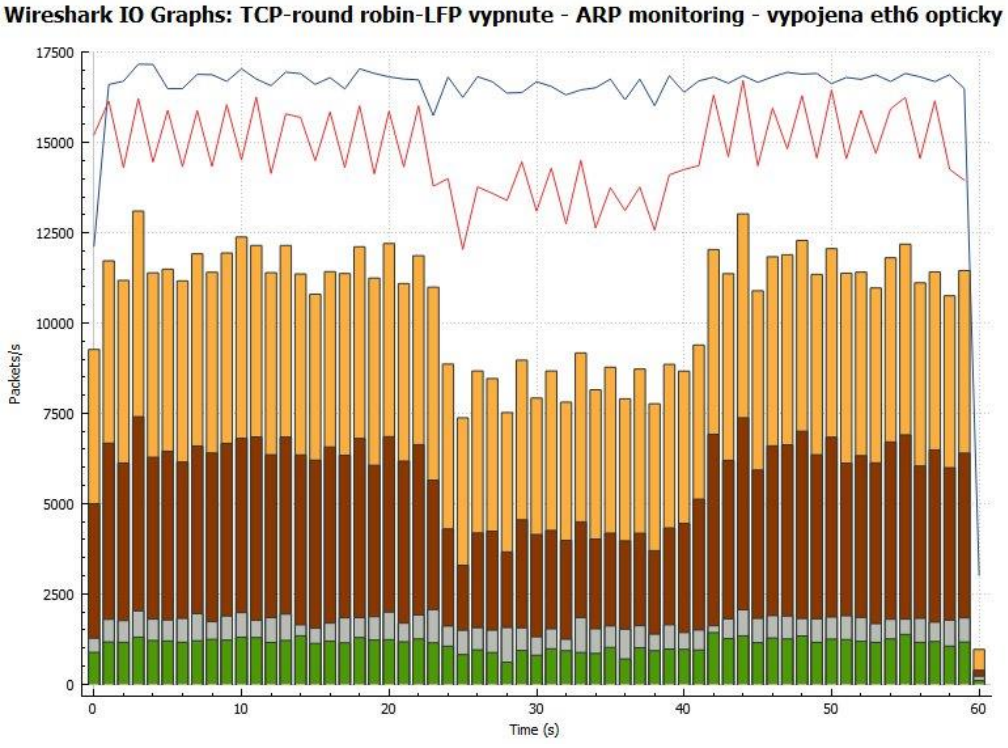
a) Balance-XOR při vypnutém LFP a při monitorování ARP vypojení a opětné zapojení jedné linky na optickém kabelu



Obrázek 20: Souhrný graf TCP provozu pro metodu balance-XOR

E. Testování TCP round robin

a) Round robin při vypnutém LFP a při monitorování ARP vypojení a opětné zapojení jedné linky na optickém kabelu



Obr. Souhrný graf TCP provozu pro metodu round robin