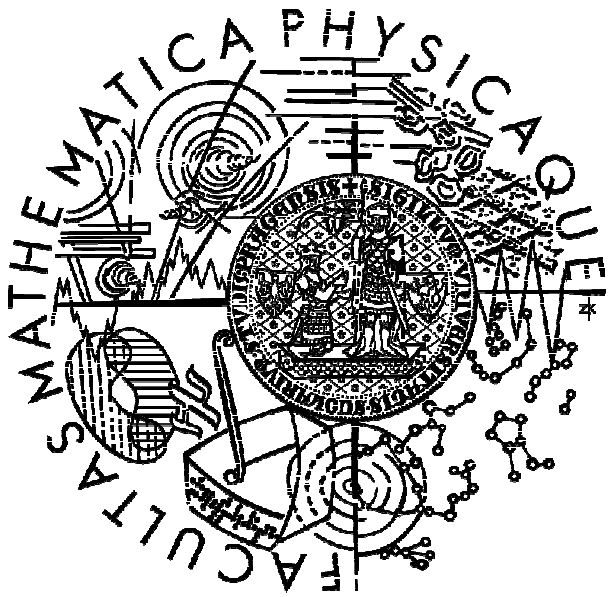
Univerzita Karlova v Praze

Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Miroslav Hnilička

Zajištění kvality služby v bezdrátových sítích

Katedra softwarového inženýrství

Vedoucí diplomové práce: Doc. Ing. Jan Janeček, CSc.

Studijní program: Informatika, Softwarové systémy

Děkuji vedoucímu Doc. Ing. Janu Janečkovi, CSc za nabídnutí zajímavého tématu diplomové práce a za připomínky a rady při jejím zpracování. Dále děkuji své ženě Kateřině za všestrannou podporu.

Prohlašuji, že jsem svou diplomovou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce.

V Praze dne 7.prosince 2005

Miroslav Hnilička

……

OBSAH

Název práce: Zajištění kvality služby v bezdrátových sítích
Autor: Miroslav Hnilička
Katedra: Katedra softwarového inženýrství
Vedoucí diplomové práce: Doc. Ing. Jan Janeček, CSc.
E-mail vedoucího: janecek@cs.felk.cvut.cz
Abstrakt:

Diplomová práce se věnuje mechanismům zajištění kvality služeb v bezdrátových paketových sítích. Pozornost se soustředila na sítě s více bezdrátovými přeskoky (multihop sítě), kde je problém obecnější, než u přístupových sítí. Teoretická část práce podává rozbor problému i ucelený přehled přístupů k jeho řešení.

Praktická část práce se zaměřuje na užší problém. Studuje vlivy použití konkrétních mechanismů linkové a síťové vrstvy na dosažitelné kvalitativní charakteristiky end-to-end přenosů v mesh sítích založených na technologii IEEE 802.11b (WiFi). Na bázi diskrétního simulačního systému OMNeT++ je navržen simulační model MeshQoS. V rámci tohoto modelu jsou implementovány dva směrovací protokoly – AODV podle IETF standardu a nově navržený hybridní oportunistický protokol OMR (Opportunistic Mesh Routing).

Výsledky experimentů s modelem MeshQoS ve formě grafů dávají představu o úrovni kvality služeb, kterou lze při konkrétních konfiguracích mesh sítí očekávat. Jednoznačně se přitom prokazují pozitivní efekty použití oportunistického přístupu ke směrování. Ve velké většině simulací dosahuje OMR významně nižšího přenosového zpoždění, vyšší úspěšnosti doručení paketů i vyšší propustnosti než AODV.

Klíčová slova: zajištění kvality služeb (QoS), bezdrátové mesh sítě, IEEE 802.11b (WiFi), oportunistický směrovací protokol OMR, diskrétní simulační systém OMNeT++

Title: Quality of Service Provision in Wireless Networks
Author: Miroslav Hnilička
Department: Department of Software Engineering
Supervisor: Doc. Ing. Jan Janeček, CSc.
Supervisor's e-mail address: janecek@cs.felk.cvut.cz
Abstract:

This thesis aims at quality of service provisioning mechanisms in wireless packet networks. The main target of the work is multiple hops wireless networks area in which the problem is more general than in access networks. Theoretical part of the thesis analyses the problem and shows detailed insight of possible solutions.

Practical part of the thesis focuses on more specific problem. Impacts of using different particular mechanisms on link and network layers are studied here, considering wireless mesh networks based on IEEE 802.11b technology (WiFi). Simulation model MeshQoS is designed on the basis of discrete simulation system OMNeT++. Inside this model, there are implemented two routing protocols – AODV defined by IETF and newly proposed hybrid opportunistic protocol OMR (Opportunistic Mesh Routing).

Results of experiments with this model helps to get the notion of the level of quality of services achievable in particular mesh networks configurations. They also show the positive effects of using opportunistic routing principle. In most scenarios OMR provide much lower latency, higher delivery probability and higher throughput.

Keywords: Quality of Service (QoS) Provision, Wireless Mesh Networks, IEEE 802.11b (WiFi), Opportunistic Mesh Routing OMR, Discrete Event Simulation System OMNeT++

1. Úvod

Předmětem této studie jsou technické aspekty zajištění kvality služeb v bezdrátových datových (paketových) sítích. V úvodní kapitole je rozebrána podstata problému, tedy co to "kvalita služeb" je, jak se hodnotí a o co při jejím zajišťování jde. Dále je stručně přiblížen a vymezen kontext, na který se tato práce zaměřuje – bezdrátové sítě. Poté jsou uvedeny základní problémy týkající se kvality služeb v bezdrátových sítích.

Druhá, teoretická kapitola se věnuje obecným i konkrétním principům a přístupům k problému. Rozebrány jsou jednotlivé konkurenční síťové modely (tj. "celkové strategie" řešení problému) a jejich komponenty. Dál jsou uvedeny některé QoS signalizační protokoly a na celý problém je také nahlédnuto z vrstevnatého pohledu (pozornost se zde věnuje především linkové a síťové vrstvě).

Zbylá část práce je praktická. Podrobněji se přitom zaměřuje na užší problém – kvalitu služeb v sítích s bezdrátovou infrastrukturou (tzv. mesh sítích) založených na technologii WiFi. Pro tento účel byl navržen diskrétní simulační model MeshQoS. Pomocí něj lze studovat vlivy různých topologií a konkrétních použitých mechanismů linkové a síťové úrovně na dosažitelné kvalitativní parametry datových přenosů. V rámci modelu MeshQoS byl navržen, implementován a odladěn nový směrovací protokol OMR (Opportunistic Mesh Routing) a standardizovaný směrovací protokol AODV (Ad-hoc On-demand Distance Vector). Na konkrétních scénářích bylo poté provedeno několik experimentů. Jejich hlavním cílem přitom bylo studovat možné přínosy oportunistického směrování na kvalitu služeb. Výsledky experimentů jsou prezentovány a analyzovány na konci práce.

1.1. Kvalita služeb počítačových sítí, QoS metriky

V souvislosti s počítačovými sítěmi je "kvalita služby" (Quality of Service, QoS) v poslední době hojně používaný termín. Žádná široce akceptovaná exaktní definice tohoto pojmu ale neexistuje a stále se debatuje o tom, co by QoS měla znamenat. Přívlastkem "QoS" bývají označovány mnohá proprietární řešení a protokoly uvažující nejrůznější specifika služeb či charakteristiky sítí. Doporučení ITU-T E.800 [1] Mezinárodní telekomunikační unie (International Telecommunication Union, ITU) definuje kvalitu služby jako "celkový efekt provozních charakteristik, který určuje stupeň satisfakce uživatele služby". Tato nekonkrétní a do jisté míry subjektivní definice se ujala zřejmě právě proto, že je obecná pro mnoho kategorií služeb, nezmiňuje žádné jejich konkrétní charakteristiky či sledované parametry (pro přenosové služby sítí např. zpoždění nebo propustnost) a zároveň ani žádné aspekty nebo prostředky jejího zajišťování (např. Service Level Agreement, řízení přístupu či signalizační protokoly).

Konkrétněji lze na pojem "kvalita služby" nahlížet z různých perspektiv [2]:

- **Vnitřní kvalita služby (Intrinsic QoS)**
  Vnitřní kvalita služby popisuje úroveň přenosových služeb, kterou jednotlivým aplikacím poskytuje, nebo je schopna poskytnout síť či její část. Protože úkolem sítě je umožnit implementaci a provoz nejrůznějších aplikací, vnitřní kvalita služby se posuzuje především podle očekávaných provozních hodnot jednotlivých výkonnostních charakteristik sítě. Ty se měří v místech, kde je služba poskytována, tj. na jejích "okrajích". Vnitřní kvalita služby je určena jak technickým návrhem sítě, tedy síťovou architekturou, tak konfigurací konkrétní sítě – ta určuje, zda jsou prostředky adekvátní

očekávaným potřebám. Používá se také termín síťová kvalita služby či přesnější kvalita síťových služeb (network-level QoS).

- **Vnímaná kvalita služby (Perceived QoS)**
  Jiný pohled hodnotí kvalitu služby konkrétní aplikace, jak ji vnímá její koncový uživatel (někdy též Quality of Experience, QoE). Ta vyplývá z konkrétního použití aplikace a závisí na tom, jaký efekt má vnitřní kvalita služby na komunikační aktivity aplikace, ale i na parametrech koncového systému (hardware, operační systém), specifících dané služby, organizačních záležitostech a také na subjektivním očekávání uživatele. Vnímaná kvalita služby se označuje také jako aplikační kvalita služby / kvalita aplikačních služeb (application-level QoS).

Dále bude používána především terminologie síťová/aplikační QoS, která je pro prostředí počítačových sítí výstižnější. Zatímco síťová QoS odpovídá pohledu poskytovatele síťových služeb, aplikační QoS odpovídá pohledu uživatele konkrétní aplikace a proto nebývá čistě technickou, přesně "měřitelnou" záležitostí. Formalizace vztahů mezi aplikační a pro ni potřebnou síťovou QoS je úkolem systémových analytiků, potažmo aplikačních programátorů. Ti navíc musí obvykle hledat kompromis mezi satisfakcí uživatele a hospodárným využitím omezených síťových prostředků. Předmětem této diplomové práce jsou technické aspekty zajišťování síťové kvality služeb, proto bude nadále obecným pojmem "kvalita služby" (nebude-li uvedeno jinak) označována síťová QoS.

Tradiční model síťových přenosových služeb založený na principu maximální snahy (Best Effort, BE) je pro moderní, intenzivně komunikující "real-time" a "mission-critical" aplikace nevyhovující. Je totiž obtížné zajistit, aby síť splňovala požadavky všech používaných aplikací, aniž by tyto požadavky znala. Síť se stává kriticky sdíleným prostředkem. Proto v takovém modelu dochází k tomu, že jistá část provozu v síti nepříznivě ovlivňuje jinou část provozu používající stejné prostředky. Prioritní přenosy pak mohou trpět na úkor nedůležitých. Tento důvod je hlavní motivací pro vznik a používání QoS mechanismů. Ty umožňují sítím díky vyššímu stupni organizace a efektivnějšímu využití síťových prostředků nabízet a garantovat služby definované úrovně. Cílem podpory QoS je tedy přizpůsobení typu služeb nabízených sítí poptávce aplikací. Aplikace přitom mají na přenosové služby sítí obecně velice různorodé požadavky a kvůli omezeným prostředkům typicky není možné je všechny bezezbytku splnit. Proto je důležitou součástí podpory QoS schopnost sítě poskytovat různým aplikacím (resp. různým přenosům) služby různé úrovně a případně také aplikační požadavky zamítat. Velmi zhruba lze rozlišovat datové a multimediální přenosy:

- **Datové přenosy**
  Neprioritní datové přenosy, jako např. transport souborů a elektronické pošty nebo zálohy databází mají výrazně dávkový charakter a požadují "absolutní" spolehlivost (co do ztrátovosti). Jsou ale tzv. elastické, tedy vysoce tolerantní k přenosovému zpoždění. Naproti tomu transakční a interaktivní datové přenosy (např. webových aplikací) mají určité požadavky i na zpoždění.

- **Multimediální přenosy**
  Obecně mají multimediální přenosy proudový charakter. Audio a hlavně video přenosy mívají vysoké nároky na propustnost spojení, v závislosti na použitém kódování ale mohou být poměrně odolné např. vůči chybovosti a ztrátovosti. Streamingové přenosy (multimedia-on-demand) nejsou příliš citlivé ke zpoždění a díky možnosti předzásobení daty na straně klienta jsou do určité míry tolerantní i k nestabilitě zpoždění a k nestabilitě propustnosti. Naproti tomu interaktivní multimediální přenosy vyžadují stabilní a velice malé zpoždění. Při konferenčním přenosu rostou již tak vysoké požadavky na celkovou propustnost kvadraticky s počtem účastníků.

Z uvedeného je zřejmá vícedimenzionální povaha problému. Jsou-li parametry ovlivňující kvalitu služby nezávislé, nelze vyvážit nedostatečnou úroveň jednoho parametru výraznou převahou jiného. Např. ani technologie gigabitového Ethernetu není dostatečně "rychlá" pro implementaci distribuované sdílené paměti – a to nikoli z důvodu nedostatečné propustnosti, ale kvůli vysokému zpoždění. Kvalita služby se tedy nevyjadřuje jediným měřítkem, ale pomocí více charakteristik. Ke kvantitativnímu vyjádření úrovně jednotlivých charakteristik slouží tzv. QoS metriky. Nejdůležitější jsou:

- **Přenosové zpoždění**
  Přenosové zpoždění (latence) je doba od zahájení odesílání dat zdrojem do dokončení jejich úspěšného příjmu adresátem. Mimo zpoždění při samotném předávání zahrnuje možné zpoždění při zpracování v mezilehlých uzlech sítě. Obvykle se zkoumají dva aspekty tohoto zpoždění – jeho střední či maximální hodnota a rozptyl (tzv. jitter). Interaktivní real-time aplikace, jako např. hlasová komunikace nebo videokonference jsou na přenosové zpoždění vysoce citlivé. Ačkoli latence nemá žádný vliv na kvalitu přenesených dat, vysoké zpoždění v příjmu (a tedy i v odpovědi) negativně ovlivňuje vnímanou kvalitu služby, protože snižuje interaktivitu komunikace. U účastníků to vyvolává pocit poloduplexního spojení. ITU-T G.114 [3] Mezinárodní telekomunikační unie definuje některá doporučení pro interaktivní komunikaci mezi lidmi. End-to-end zpoždění v jednom směru by mělo být maximálně 150 ms (např. v případě přenosu hlasu to zahrnuje i zpoždění při zpracování zvukových vzorků příslušnými kodeky na obou koncích komunikace, tj. tzv. "od úst k uchu"). Obousměrné zpoždění nad 200 ms už většina lidí registruje a může obtěžovat. Nad 400 ms komunikace vázne, účastníci si často skáčou do řeči. Dalším problémem je nestabilita přenosového zpoždění. Tu způsobuje proměnné vytížení přepojovacích uzlů. Další příčinou může být přístup, kdy je každý paket přenášen nezávisle – tedy ne nutně po stejné přenosové cestě. V real-time aplikacích tato nestabilita způsobuje "trhání" a další nedobré efekty. Pomocí tzv. jitter bufferu ji lze na straně příjemce do jisté míry kompenzovat. Jitter buffer nicméně přináší dodatečné zpoždění a proto se obvykle omezuje na vyrovnávání pouze malých odchylek. Ideálně je jitter v jednom směru nižší než 30 ms.

- **Propustnost**
  Propustnost vyjadřuje dosažitelnou přenosovou rychlost, tj. množství dat, které lze přenést za jednotku času. Požadavky na propustnost se liší jak v závislosti na typu aplikace (resp. přenosu), tak na použitém kódování. Např. přenos hlasu technologií VoIP (Voice over IP) v závislosti na vzorkovací frekvenci a kodeku využije 20 až 320 Kbps.

- **Spolehlivost**
  Spolehlivost se obvykle definuje kvalitativně, tj. jako záruka doručení všech paketů, a to nepoškozených, bez duplicit a v původním pořadí. Obecněji se dá vyjádřit kvantitativně mírou ztrát, duplicit, resp. porušení pořadí paketů. Příčiny nespolehlivosti i její možná řešení mohou být v různých vrstvách síťového modelu. Např. ztrátovost bývá v drátových sítích způsobena především zahlcením, zatímco v bezdrátových mnohem častěji špatnou kvalitou signálu a kolizemi. Požadavky na spolehlivost jsou jednou z rozlišujících charakteristik datových / multimediálních přenosů (zcela striktní / docela benevolentní požadavky). V závislosti na způsobu kódování jsou multimediální přenosy odolné vůči výpadku až cca čtvrtiny vyslaných paketů. Videopřenosy bývají obecně odolnější než audiopřenosy. U multimediálních přenosů se obecně preferují nespolehlivé síťové služby, která obvykle mají lepší ostatní QoS parametry (zpoždění, jitter). Vysoká nespolehlivost však může degradovat vnímanou kvalitu multimediálních přenosů různými "přeskoky" a dalšími nepříjemnými efekty.

Uvažovat lze i mnohé další měřitelné charakteristiky kvality služby. Klíčová je například její dostupnost, resp. spolehlivost (zde ve smyslu pravděpodobnosti výpadku celého systému). Nároky na spotřebu energie v koncových uzlech a pokrytí (tedy rozsah lokalit, v nichž je síť schopna poskytovat své služby) jsou další metriky, důležité především v prostředí mobilních sítí. Podstatná může být také rychlost a úspěšnost navázání spojení. Existují také snahy vnímat bezpečnost jako QoS parametr a měřit ji (resp. její konkrétní aspekty jako je důvěrnost, autenticita a integrita dat či nepopiratelnost) QoS metrikami (např. [4]).

Obecně se QoS metriky mohou týkat stavů (např. úroveň konektivity, bezpečnosti), událostí (rychlost navázání spojení, míra ztrátovosti atd.) i aktivit (přenosové zpoždění, spotřeba energie apod.) a vztahovat se k různým entitám (uzel sítě, přímý spoj, cesta sítí, resp. spojení, celá síť). Lze jimi vyjadřovat jak okamžitou úroveň daných parametrů, tak aplikační požadavky a úrovně garancí (střední hodnota a stabilita, mezní hodnoty apod.).

Aplikace mohou klást požadavky na jeden či více QoS parametrů, které nemusí být nutně nezávislé. Závislosti QoS parametrů nejsou neobvyklé. Např. doručování paketů mimo pořadí (při paralelním používání více přenosových cest) jde obvykle ruku v ruce s vysokým rozptylem zpoždění. Konkrétní použité techniky mohou vytvářet další vztahy – např. protokoly různých vrstev kompenzující ztráty opakováním přenosu. Opakování přenosu má totiž vliv na latenci a jitter.

Většinu QoS metrik vztahujících se k cestám v síti lze rozdělit do tří základních skupin. Nechť M je metrika a $P=(N_1, N_2, \ldots, N_{k-1}, N_k)$ libovolná cesta sítí. Podle povahy závislosti $M(P)$ na metrikách jednotlivých úseků cesty P se rozlišují tyto typy metrik [5]:

- **Aditivní metriky**
  Metrika M se nazývá aditivní, pokud platí $M(P) = M(N_1,N_2) + M(N_2,N_3) + \ldots + M(N_{k-1}, N_k)$. Zpoždění a jeho rozptyl, vzdálenost, spotřeba energie a další parametry typu "cena", které se po cestě akumulují, jsou aditivní metriky.

- **Multiplikativní metriky**
  Platí-li rovnost $M(P) = M(N_1,N_2) * M(N_2,N_3) * \ldots * M(N_{k-1}, N_k)$, metrika M je multiplikativní. Například vyjadřuje-li metrika M pravděpodobnost ztráty paketu, platí $1-M(P) = (1-M(N_1,N_2)) * (1-M(N_2,N_3)) * \ldots * (1-M(N_{k-1}, N_k))$. Ztrátovost (přesněji řečeno metrika doplňkového jevu) je proto multiplikativní.

- **Konkávní metriky**
  Konkávní se nazývají metriky, pro které platí $M(P) = min(M(N_1,N_2), M(N_2,N_3), \ldots, M(N_{k-1}, N_k))$. Například propustnost je konkávní metrika, neboť její určitá úroveň je vyžadována na každém spoji podél cesty. Obdobně bezpečnost lze také považovat za konkávní metriku, protože bezpečnost celku odpovídá (vágně řečeno) bezpečnosti nejslabšího článku.

1.2. Bezdrátové sítě

1.2.1. Kategorie bezdrátových sítí

Fixní počítačové sítě využívají jako přenosové médium metalické nebo optické vedení. Bezdrátové sítě chápeme jako sítě, které namísto takovéto fixní infrastruktury využívají šíření elektromagnetických vln volným prostorem. Podle použitého frekvenčního pásma lze rozlišovat rádiové a optické spoje. Rádiové spoje obecně mají do značné míry schopnost "projít" překážkou nebo ji "obejít". Naproti tomu optické spoje (infračervené i ve viditelné části spektra) tuto schopnost nemají a proto se většinou používají jen jako point-to-point spoje. Předností bezdrátových sítí je samozřejmě jejich pružnost, mobilita (resp. možnost mobility) a také rychlost a snadnost jejich instalace/deinstalace. Nevýhodou naopak může být horší možnost zabezpečení a případně nutnost získat licenci pro provoz takové sítě.

Hlavní problémy bezdrátového prostředí, které mají vztah k zajištění kvality služeb, jsou pojmenovány v následující podkapitole. Pro vymezení předmětu zájmu této práce je zde uvedeno jedno (z mnoha možných) rozlišení kategorií bezdrátových sítí.

Osobní bezdrátové sítě, bezšňůrové sítě

Osobní bezdrátové sítě WPAN (Wireless Personal Area Networks) podle specifikace IEEE 802.15 [6] jsou sítě určené pro propojení spolupracujících zařízení (typicky počítačových periferií, PDA, mobilní telefonů apod.) na malé vzdálenosti (v řádu jednotek metrů). Do této kategorie patří např. technologie Bluetooth a IrDA. Sítě WPAN slouží typicky potřebám jednotlivce, případně velmi omezené skupině uživatelů. Jejich trvání je často omezené na dobu plnění konkrétního úkolu. Podpora kvality služeb v těchto sítích vychází z požadavků na přenos hlasu. Např. technologie Bluetooth pro hlasové přenosy nabízí vyhrazené synchronní spoje SCO s garantovanou propustností 64 kbps.

Bezdrátové přístupové sítě

Přístupové sítě slouží k připojení koncových terminálů do již existujících fixních sítí (internetu, podnikových a komunitních sítí), tj. pouze k překonání "poslední míle", resp. "posledního metru". Tyto sítě využívají celulárního principu a pracují v režimu point-to-multipoint, tj. jeden přístupový bod/základnová stanice obsluhuje obecně více koncových stanic (klient/server model). Jde např. o lokální bezdrátové sítě WLAN (Wireless Local Area Networks) podle IEEE 802.11 [7] nebo sítě pro bezdrátový přístup (Fixed Wireless Access), resp. metropolitní bezdrátové sítě WMAN (Wireless Metropolitan Area Networks) podle IEEE 802.16 [8]. Patří sem technologie WiFi, HIPERLAN/2, WiMAX a další. Do přístupových sítí lze ale zařadit i satelitní sítě (symetrické i asymetrické, tj. používající pro zpětný kanál jinou technologii).

Požadavky na kvalitu služeb v bezdrátových přístupových sítích se většinou řeší centrálním, determininistickým přidělováním média. Technologie využívající jako přístupovou metodu časový multiplex (Time Division Multiple Access, TDMA) alokují určitou část časových slotů pro konkrétní přenos. Přístupovou metodu CSMA/CA (Carrier Sense Multiple Access/Collision Avoidance), kterou používají sítě WLAN rozšiřuje standard IEEE 802.11e [9] o funkci HCF (Hybrid Coordination Function), která řídí médium centrálně – pomocí mechanismu výzev v pravidelných intervalech (tzv. Contention-Free Periods) přiděluje stanicím médium deterministicky.

Mobilní ad-hoc sítě

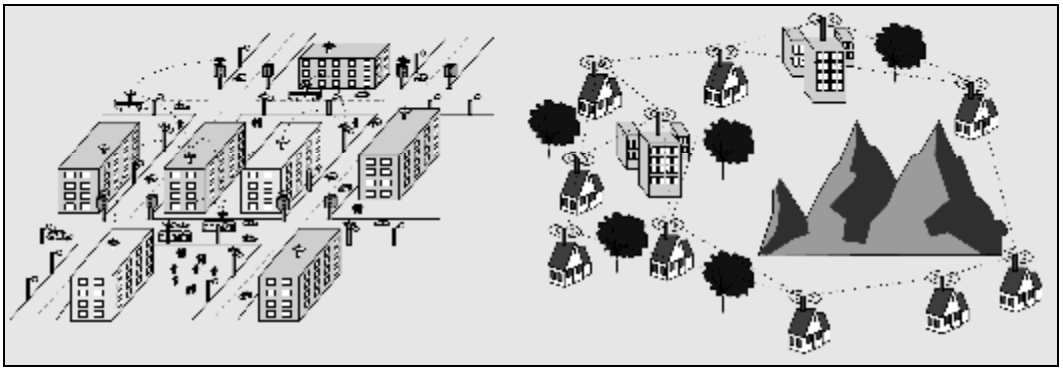
Mobilní ad-hoc sítě (Mobile Ad-hoc NETworks, MANETs) [10] tvoří skupina rovnocenných uzlů propojených bezdrátovými multihop cestami. Jejich organizace a správa je distribuována mezi všechny uzly. Sítě MANET nemají žádnou předdefinovanou nebo preferovanou topologii, žádné centrální řízení ani jakoukoli jinou fixní podporu. Jejich uzly představují terminály a současně směrovače provozu pro ostatní uzly. Na rozdíl od fixních sítí a sítí, kde je bezdrátový pouze poslední přeskok, nejsou koncové uzly pouze na "okrajích", ale také "uvnitř" sítě. Některé uzly MANETů přitom mohou sloužit jako brány a poskytovat rozhraní do fixní sítě. Peer-to-peer koncept sítí MANET byl navržen jako komunikační základna např. pro vojenské nebo záchranné operace, kde je cílem okamžité zajištění internetworkingu v prostředích bez existující infrastruktury. Z toho vyplývají i požadavky na vysokou škálovatelnost a schopnost zajistit vysokou úroveň konektivity mobilním zařízením. Časté změny topologie způsobené malým dosahem a vysokou mobilitou uzlů jsou pro sítě MANET charakteristické, nicméně některé jejich nasazení může předpokládat i zcela fixní pozice uzlů (senzorové sítě monitorující prostředí, např. síť inteligentních detektorů ohně rozmístěná v určitém objektu či oblasti). Kvalita služeb v sítích MANET je již delší dobu živě diskutovaný problém (např. [11,12]), nicméně standardy na bázi IETF či IEEE ani komerční řešení v produkční kvalitě se zatím neobjevují.

Mesh sítě

Sítě s bezdrátovou infrastrukturou (neboli mesh sítě) se skládají z mobilních i fixních uzlů propojených bezdrátovými multihop cestami. Mobilní uzly zde fungují stejně jako v MANETech, mohou se tedy sítí libovolně pohybovat, dynamicky se připojovat a odpojovat. Pracují současně jako terminály i jako směrovače provozu pro ostatní uzly sítě, čímž rozšiřují její pokrytí (a případně i kapacitu). Bezdrátovou infrastrukturu tvoří fixní uzly. Ty pracují výhradně jako bezdrátové směrovače (tj. nevzniká ani nekončí v nich žádný provoz). Některé bezdrátové směrovače mohou navíc poskytovat rozhraní do fixní sítě (typicky do lokální sítě a internetu) – fungují tedy stejně jako přístupové body a také se tak označují.

Mesh sítě tedy vycházejí ze sítí MANET, nicméně zaměřují se především na civilní aplikace (viz obr. 1). Do značné míry slevují na požadavcích kladených na MANETy (plná mobilita sítě, vysoká škálovatelnost a rychlé nasazení), protože tyto vlastnosti nejsou pro typické aplikace klíčové. Mesh sítě nemají sloužit jako izolované, samo-konfigurující se sítě pro speciální účely, ale spíše jako pružné a víceúčelové rozšíření drátové infrastruktury. Předpokládá se tedy, že velká část provozu v mesh síti směřuje právě z/do fixní sítě. Úkolem mesh sítí je zajišťovat ty přenosy, jejichž zdrojem nebo cílem je některý bezdrátový terminál, tedy nikoli propojovat sítě a poskytovat tak "tranzitní" prostor pro vnější provoz.

Mesh sítím je v současnosti věnována velká pozornost (např. [13, 14]) a některá řešení mesh sítí již v praxi fungují (např. [15, 16]). Vývoj je nicméně zatím ve fázi, ve které byly lokální bezdrátové sítě v 90. letech, tj. jde o proprietární (a často drahá) řešení. Nedávno však pod IEEE vznikla pracovní skupina 802.11s, která se zabývá návrhem a standardizací fyzické a linkové vrstvy pro mesh sítě (Mesh Wireless Local Area Networks, MWLANs) [17]. Podpora pro "mesh" rozšíření se objevila také ve standardu IEEE 802.16(a).



Obrázek č. 1. Možné aplikace mesh sítí
a) inteligentní dopravní systémy zpracovávající real-time informace
b) širokopásmový přístup v odlehlých nebo hůře dostupných oblastech.

Předmětem zájmu obecné části této práce jsou pouze sítě s více bezdrátovými přeskoky, tj. MANETy a mesh sítě. Praktická část práce se pak úžeji zaměřuje na QoS v mesh sítích založených na standardu IEEE 802.1b (tj. především na přístupové metodě CSMA/CA).

1.2.2. Specifika bezdrátových QoS přenosů

Techniky zajišťující kvalitu služeb bezdrátových sítí musí řešit některé úkoly a problémy, které v tradičních drátových sítích nenastávají. Část z nich je společná pro všechny kategorie bezdrátových sítí, jiné jsou specifické pro mobilní, multihop nebo ad-hoc sítě. Dále jsou pojmenovány ty nedůležitější.

Relativně nízká propustnost a nestabilní vlastnosti bezdrátových spojů

Přesné chování bezdrátového média je, především u mobilních sítí, velmi obtížně předvídatelné. Jeho použití totiž komplikují různé deformace signálu (např. vícecestné interference), silný útlum, šum a případně kolize. Kapacita a chybovost bezdrátových kanálů se může s časem a místem rychle a významně měnit. Bezdrátové spoje nebývají příliš stabilní a dynamicky se mění jejich provozní parametry, jako např. chybovost, propustnost a přenosové zpoždění. Tyto efekty jsou nejvážnější v bezdrátových multihop sítích (přenos je ovlivněn problémy na více přeskocích). Nespolehlivou povahu bezdrátového spoje lze částečně maskovat na linkové úrovni, např. technikami "dopředné" ochrany dat (FEC, Forward Error Correction) nebo automatickým opakováním přenosu (ARQ, Automatic Repeat Request). V multihop sítích si ale těchto problémů musí být vědomy i směrovací protokoly, protože ty jej mohou řešit obecněji (např. obcházením problematické oblasti). Kapacita sdíleného bezdrátového média je v daném pásmu pevně omezená a bezdrátové technologie mají obecně i za příznivých podmínek řádově nižší propustnost než drátové technologie.

Problém skryté a předsunuté stanice

Použití "soutěžních" MAC protokolů třídy CSMA má v bezdrátovém prostředí určitá omezení. Mimo fakt, že vysílající uzel není schopen spolehlivě detekovat současné vysílání ostatních uzlů (a detekovat tak kolize), jsou hlavním důvodem kolizní mechanismy označované jako problém skryté a předsunuté stanice (Hidden/Exposed Terminal Problem). Pokud uzly α a β navzájem komunikují, α vysílá, β přijímá (viz obr. 2a) a uzel γ rovněž chce vysílat pro β, musí si v souladu s CSMA nejdřív "příposlechem" ověřit, že je médium volné. Protože ale γ není bezprostředním sousedem α a jeho vysílání "neslyší", médium pokládá za volné. Může tedy začít vysílat. Tím ale vyvolá na uzlu β kolizi. Tato situace se označuje jako problém skryté stanice. V jiné situaci dochází k problému předsunuté stanice. Nechť uzel α intenzívně komunikuje s β (α vysílá, β přijímá, viz obr. 2b). Má-li uzel γ zájem začít vysílat pro δ, příposlechem zjišťuje, že je médium prakticky trvale obsazeno (slyší vysílat α). V souladu s CSMA musí γ své vysílání odložit, přestože toto vysílání nemůže s vysíláním α na uzlu β kolidovat.

Obrázek č. 2. a) problém skryté stanice b) problém předsunuté stanice.

Problémy skryté a předsunuté stanice jsou navzájem duální. První problém snižuje kapacitu sítě kvůli zvýšenému počtu kolizí, zatímco druhý zbytečným odkládáním přenosu. Bezdrátové sítě neposkytují "vysoký stupeň konektivity", tj. přímá konektivita mezi jejich uzly není tranzitivní relace. Příčinou obou problémů je tedy fakt, že protokoly CSMA před vysíláním kontrolují stav média na vysílači a ne na přijímači, kde dochází ke kolizím.

Sítě WLAN zmírňují následky problému skryté stanice mechanismem RTS/CTS (Request To Send/Clear To Send) [7]. Problém předsunuté stanice ale samotná přístupová metoda CSMA/CA neřeší. V multihop scénářích přitom může způsobovat problémy s propustností, rozptylem zpoždění a neférovým přístupem jednotlivých přenosů k médiu (i přenosů se stejným počtem bezdrátových přeskoků) [18,19]. To představuje určité omezení pro její použití v multihop prostředí. Oba zmíněné problémy kompletně řeší např. přístupová metoda DBTMA (Dual Busy Tone Multiple Access) [20], která však mimo RTS paketů používá i dvou "obsazovacích tónů" mimo hlavní přenosové pásmo.

Decentralizované prostředí ad-hoc sítí

Ad-hoc sítě se nemohou opírat o existenci pevné infrastruktury nebo nějaké centrální autority, která by mohla zajišťovat např. směrování a bezpečnost. Všechny jejich funkce musí být navrženy tak, aby efektivně fungovaly v decentralizovaných podmínkách. Použité algoritmy tedy musí být plně distribuované a neměly by intenzivněji používat globálních výpočtů.

Omezené výpočetní a komunikační schopnosti mobilních uzlů

Uzly mobilních sítí představují většinou přenosná zařízení napájená z baterií. Omezená kapacita baterií limituje maximální vysílací výkon a energeticky náročné činnosti a komponenty těchto zařízení. Jde především o CPU (jak jeho výkon, tak vytížení), paměť (velikost i použití) a zpracování signálu [21]. Nepřetržitý provoz bývá v současnosti omezen na několik hodin. Algoritmy a techniky zajišťující síťové funkce těchto zařízení (tj. v podstatě celý protocol stack pod aplikací) by měly mimo komunikačního výkonu uvažovat i spotřebu energie (mj. ji šetřit pro samotnou aplikaci). Speciálně alokace prostředků při poskytování QoS by měla brát v úvahu stav baterií ve vztahu k energetické spotřebě alokovaných prostředků. Směrovací protokoly v bezdrátových multihop sítích by neměly být kvůli omezeným zdrojům, resp. schopnostem uzlů (např. kešování) příliš komplexní a paměťově náročné.

Dynamická topologie mobilních multihop sítí

V infrastrukturních sítích představuje mobilita klientských uzlů jasně vymezený problém předávání (handoverů) mezi základnovými stanicemi. V mobilních ad-hoc sítích je situace složitější. Uzly se připojují k síti a odpojují od ní kdykoli a kdekoli, při jejich vzájemném pohybu se dynamicky formují nové spoje a zanikají staré. Topologie mesh sítí a sítí MANET tedy může být (v závislosti na mobilitě a dosahu jejich uzlů) velice dynamická. Zavedené směrovací cesty mohou být přerušeny kdykoli během přenosu, proto je nutná jejich efektivní údržba a rychlá rekonstrukce. Mimo proměnných vlastností bezdrátových spojů komplikuje údržbu přesných informací o stavu sítě právě její dynamická topologie. Směrovací protokoly tak musí pracovat s informacemi, které jsou ze své podstaty nepřesné a rychle zastarávají.

Bezpečnost

Jak už bylo uvedeno, i na bezpečnost lze pohlížet jako na QoS atribut. Především v ad-hoc bezdrátových sítích představuje zajištění bezpečnosti obtížný úkol. Bezdrátové médium je ze své podstaty nezabezpečené a všesměrová povaha vysílání zvyšuje možná rizika (odposlech, padělání identity, zneužití konektivity apod.). Na médium se spoléhat nelze, jediným řešením je tedy kryptografie. Z jejího pohledu přitom nejde o "nové" problémy. Požadavky týkající se autenticity, důvěrnosti, integrity a nepopiratelnosti jsou stejné jako v ostatních veřejných sítích. Specifickým problémem ad-hoc sítí ale je vytváření zabezpečených relací mezi entitami bez pomoci certifikace důvěryhodnou "třetí stranou". Ad-hoc sítě totiž vznikají spontánně, tj. bez záruk, že každý uzel vlastní veřejné klíče ostatních uzlů, nebo že může ověřovat jejich certifikáty. Existující přístupy používají např. "delegování" důvěry mezi jednotlivými uzly [22], nebo rozložení autority a funkcionality autentifikačního serveru mezi všechny uzly sítě tak, že libovolná dostatečně velká skupina uzlů může sama poskytovat jeho služby [23]. Bezpečnost je velmi široká oblast a navíc vysoce specifická "kvalita" služby a proto nebude v této práci dále rozebírána.

2. Zajišťování kvality služeb v bezdrátových sítích

2.1. Obecné přístupy a techniky zajištění QoS

Kvalitu služeb lze zajišťovat jak technikami na aplikačními úrovni (application-based QoS, AQoS), tak i na síťové úrovni (network-based QoS, NQoS). AQoS techniky zahrnují všechny nástroje obsažené v samotné aplikaci, které mají za cíl zajistit a udržovat kvalitu jejího zamýšleného použití, tj. vnímanou kvalitu služby. U multimediálních aplikací jde např. o samotné získávání audiovizuálních dat, o jejich kompresi, dekompresi, jitter buffering, přehrávání a o adaptabilitu těchto procesů. Naproti tomu NQoS techniky zasahují tedy pouze do vrstev pod aplikační vrstvou. Dále jsou popsány hlavní NQoS mechanismy, ještě před tím ale budou ale uvedeny obecné přístupy k řešení problému.

Obvykle zajišťuje kvalitu služby pro koncové uživatele poskytovatel služby na základě určitého kontraktu, tzv. SLA (Service Level Agreement). SLA měřitelnými parametry specifikuje úroveň služeb, kterou se poskytovatel zavazuje zajistit. Mimo výkonnostních charakteristik může také obsahovat např. závazky týkající se dostupnosti dané služby, uživatelské podpory apod. Ačkoli forma takového kontraktu může být různá, např. bezdrátovým ad-hoc sítím koncept SLA nevyhovuje. V nich totiž žádná centrální autorita "poskytovatele" neexistuje (rozdíl mezi poskytovatelem a uživatelem služby zde zaniká). Konkrétně ad-hoc sítě tedy potřebují trochu jiný koncept rozdělení rolí a zodpovědnosti při zajištění QoS. SLA (ve své právní rovině) a další organizační či ekonomické aspekty zajišťování kvality služby nicméně vybočují z technického zaměření této práce a proto dále nebudou samostatně rozebírány.

Až na výjimky, jako je např. bezpečnost nebo spotřeba energie, popisují parametry QoS výkonnostní charakteristiky sítí. Nasnadě je tedy otázka, zda je nutná explicitní podpora QoS, když lze většinu problémů řešit zvyšováním přenosové kapacity sítě a naddimenzováním výkonu jejich uzlů (tzv. overprovisioning). Jak bude zdůvodněno, tento přístup řeší problém pouze částečně a v bezdrátových sítích je použitelný jen v omezené míře. Navíc pokud by měl overprovisioning sloužit výhradně k zajištění QoS, nebyl by (ekonomicky) příliš efektivní. V drátových sítích je nicméně tento přístup zatím nejrozšířenější, protože je nejméně komplikovaný a lze jej aplikovat postupně.

Datové přenosy běžných síťových aplikací často vykazují dávkový charakter, tj. mají silně nárazové požadavky. V klasických "best-effort" sítích, bez ohledu na jejich kapacitu, tedy občas dochází k zahlcení. Dalším problémem je, že směrovací protokoly bez explicitní podpory QoS mají jen kusé informace (pokud vůbec nějaké) o vytížení jednotlivých uzlů, kanálů a cest a proto většinou neobsahují dostatečně účinný mechanismus vyhýbání se úzkým hrdlům. Zatímco některé cesty jsou zahlcené, na jiných může zůstávat nadbytek nevyužitých prostředků. Samotný dostatek přenosové, případně i výpočetní kapacity navíc nezajišťuje predikovatelné chování co se týká zpoždění při přenosu a férovosti přístupu k jednotlivým tokům (třebaže úkolem zajištění QoS není "všem měřit stejně"). Stále tak hrozí váznutí real-time přenosů kvůli méně důležitým přenosům na pozadí. Zmíněný přístup "hrubou silou" tedy nepřináší žádný druh garancí – problém neřeší cíleně, pouze snižuje pravděpodobnost jeho výskytu a jeho intenzitu.

V bezdrátových sítích je overprovisioning často realizovatelný jen v omezené míře. Na rozdíl od drátových sítí zde existuje pouze jediné, všemi sdílené přenosové médium. Přes pokroky v kódování a modulaci signálu je jeho kapacita v daném pásmu pevně omezená. Je sice možné ho v různých prostorových lokacích bez výrazných interferencí využívat násobně, nelze ho ale např. trvale vyhradit pro konkrétní spoj. Bezdrátových technologií se spíše než zpochybňovaný Gilderův zákon (zdvojnásobení propustnosti každých 9 měsíců) týká méně progresivní Cooperův zákon a limitem samozřejmě zůstává Shannonův zákon [24]. Další omezující faktory se týkají výpočetního výkonu, paměťové kapacity a vysílacího výkonu uzlů mobilních sítí (kvůli spotřebě energie).

Overprovisioning tedy není pro bezdrátové sítě schůdnou cestou ("maximální snaha" nestačí). Dále popisované přístupy (společně s dalšími označované jako "network traffic engineering") se týkají systémových řešení kvality služby, jakožto samostatného problému.

Přestože přenosové požadavky aplikací bývají velmi různorodé, mechanismy implementace QoS v různých sítích bývají podobné. Při zajišťování kvality služeb je klíčové rozlišování provozu a poskytování různých úrovní služeb jeho různým druhům. Ze všeho nejdříve je tedy potřeba provoz nějak klasifikovat. Na základě této klasifikace se rozhoduje, zda vůbec bude vpuštěn do sítě (o to se starají mechanismy řízení přístupu). Pokud ano, jeho klasifikace poté ovlivňuje mechanismy plánování. Ty rozhodují o způsobu, jakým je daný provoz zpracováván na přenosové cestě. Pro QoS v bezdrátových sítích je zásadní rovněž mechanismus přístupu k médiu.

Mimo zmíněné mechanismy je nutný také nějaký management síťových prostředků. Navíc se používají další techniky, jako např. tvarování provozu a buffer management. Kombinací uvedených přístupů pak lze zajistit podporu kvality služeb.

Klasifikace provozu

Identifikace jednotlivých druhů provozu může mít různou granularitu – od několika předem definovaných tříd (per-class granularita) až po jednotlivé přenosy (per-flow granularita). Klasifikace se mohou účastnit různé vrstvy OSI modelu. Např. na linkové vrstvě se pro klasifikaci používá 3-bitové pole určující prioritu Ethernetového rámce, na síťové vrstvě pak TOS (Type Of Service) v hlavičce IP paketu pro označení typu služby požadované pro daný paket. V transportní vrstvě lze pomocí pětice (zdrojová adresa, cílová adresa, zdrojový port, cílový port, protokol) identifikovat konkrétní TCP přenos. Klasifikace se provádí přímo na terminálu, příp. na prvním (hraničním) směrovači či jiném síťovém zařízení. "Ideální" rozlišování jednotlivých přenosů je samozřejmě možné pouze u služeb se spojovaným charakterem. Jeho nevýhodou ovšem může být vysoká režie a omezená škálovatelnost odpovídajících mechanismů.

Řízení přístupu

Při nedostatku prostředků nelze všechny přenosové požadavky beze zbytku splnit a některé je třeba krátit nebo zamítat. Obvykle se k zamítnutí požadavku přistupuje až v okamžiku, kdy mu nelze vyhovět. Obecněji ale jakékoli filtrování požadavků šetří prostředky a tím umožňuje poskytnout vyšší kvalitu služby "ostatním". Cílem filtrování požadavků přitom nemusí být zajištění vysoké kvality pro "vyvolenou" část provozu, ale třeba naopak zajištění férového přístupu jednotlivých uživatelů nebo částí síťového provozu ke "kvalitní službě". Filtrování

provozu z tohoto důvodu obvykle definuje poskytovatel v tzv. FUP, Fair Usage Policy. Podstata filtrování požadavků může být podobná tomu, když např. železniční společnost přeznačí část vagónů 2. třídy na 1.třídu. Aniž by došlo např. k výměně sedaček za pohodlnější, samotné snížení poptávky o jízdu těmito vagóny (zde kvůli vyšší ceně) způsobí, že budou méně přecpané – tj. zvýšení kvality služby.

O přijetí či zamítnutí konkrétních aplikačních požadavků rozhodují mechanismy řízení přístupu (Admission Control, AC). Mimo poskytnutí garance požadované QoS pro nový přenos je cílem tohoto rozhodnutí také zajistit, že již existující QoS přenosy nebudou degradovány. Mechanismy řízení přístupu jsou obvykle distribuované. Např. při požadavku na QoS spojení se do rozhodování zapojují mezilehlé uzly, případně i cílový uzel. K finálnímu rozhodnutí přitom může dojít v jiném uzlu, než kde požadavek vznikl. Přijetím aplikačního QoS požadavku vzniká pro síť závazek tento požadavek splnit. Tento závazek se chápe jako striktní, třebaže samotné QoS požadavky striktní být nemusí. Při per-flow granularitě musí QoS požadavek definovat mimo vyžadované úrovně jednotlivých parametrů (zpoždění, ztrátovost apod.) také profil příslušného toku (např. špičkovou/průměrnou rychlost a velikost dávky). Závazek se potom týká pouze provozu v rámci tohoto profilu.

Plánování

Plánování se týká způsobu manipulace s frontami paketů. Nejjednodušší frontový algoritmus FIFO (First-In-First-Out) zachází se všemi pakety stejně, tj. nepodporuje QoS. Vylepšením může být schéma se striktními prioritami, kde do fronty s nižší prioritou se přistupuje až v okamžiku, kdy jsou všechny fronty s vyššími prioritami prázdné (obsloužené). Striktně prioritní fronty ovšem trpí problémem "vyhladovění", proto se častěji používá algoritmus WFQ (Weighted Fair Queue). Ten přiřazuje každé frontě poměrnou váhu, která určuje podíl vybírání z dané fronty. Zobecněním WFQ je algoritmus CBQ (Class-Based Queuing), který navíc umožňuje hierarchické členění vážených prioritních front do tříd a podtříd. Obecně lze kombinací různých plánovacích algoritmů na různých úrovních dosáhnout komplexního QoS řešení.

Přístup k bezdrátovému médiu

Striktní garance přenosového zpoždění a jitteru nelze poskytovat, pokud není pro všechny uzly sítě zajištěn spolehlivý přístup k médiu, ideálně v pravidelných (a krátkých) intervalech. Proto je zde podstatné rozlišení protokolů řízení přístupu k médiu na soutěžní a nesoutěžní, resp. nedeterministické a deterministické. Např. mechanismy virtuálního příposlechu nosné frekvence a exponenciálního prodlužování prodlevy při nepotvrzení odeslaného paketu, které tvoří základ soutěžního protokolu CSMA/CA, vůbec negarantují přístup k médium v konečném čase. Naproti tomu nesoutěžní protokoly pravidelný přístup k médiu zajistit mohou, např. mechanismem výzev (polling) nebo pomocí rezervací části časových slotů při časovém multiplexu (TDMA). Nesoutěžní protokoly nicméně fungují centralizovaně. To u přístupových sítí nevadí, ale brání to jejich nasazení v multihop sítích. Toto téma bude ještě rozebráno dále.

Management síťových prostředků

Při zajišťování QoS jsou obvyklé dva generické přístupy ke správě síťových prostředků: rezervace a prioritizace. Rezervace (typicky spojené s per-flow granularitou) představují explicitní vyhrazení určitého množství síťových prostředků (přenosová, výpočetní a paměťová kapacita) pro konkrétní přenos v uzlech, které se tohoto přenosu účastní. Jsou-li

rezervace absolutní povahy, přinášejí pro konkrétního "konzumenta" absolutní garance kvality služby (v souladu s principem přepojování okruhů). Na druhé straně pevně vyhrazené prostředky, které ale daný přenos nevyužívá, typicky nelze přenechat ostatním přenosům. To je z celkového pohledu neefektivní a snižuje to "agregovanou" kvalitu služeb sítě.

Prioritizace poskytuje garance přednostního zpracování konkrétního toku či paketu na úkor ostatních, méně prioritních. Jde tedy pouze o relativní garance – sám o sobě tento přístup neimplikuje dosažitelné QoS parametry. Prioritizace se proto používá v kombinaci s per-class granularitou a dalšími mechanismy (především řízení přístupu). QoS architektura založená na principu prioritizací typicky definuje několik tříd provozu (tj. několik úrovní priorit), pro které jsou implicitně vyhrazeny určité prostředky v přepojovacích uzlech sítě. Distribuované mechanismy řízení přístupu pak mají za úkol "nevpustit" do sítě příliš velký objem provozu daných tříd, aby nedocházelo k výrazné degradaci kvality služby.

Tvarování provozu

Pokud je zajištěno, že generovaný provoz nevyhovuje vyjednanému profilu, je možné jej do sítě nevpustit. Lepším řešením nicméně je provoz před vstupem do sítě podle příslušného profilu "dotvarovat". Obvykle se používá kombinace dvou technik – algoritmu děravého vědra (Leaky Bucket) a algoritmu Token Bucket. Leaky bucket omezuje maximální rychlost provozu vpouštěného do sítě na rychlost, kterou vytéká vědro (tj. vědro omezuje dávkový charakter přenosu). Token bucket naproti tomu dávkový charakter přenosu zachovává (až do velikosti vědra), ale omezuje průměrnou rychlost provozu vpouštěného do sítě na rychlost, kterou se vědro plní tokeny.

Buffer management

Při zahlcení sítě dramaticky klesá její propustnost a je obtížné garantovat QoS. Mechanismus RED (Random Early Detection) částečně předchází zahlcení tak, že ještě před úplným zaplněním front zahazuje některé pakety. Tím lze "ještě včas" omezit rychlost zdroje TCP spojení. Pakety se začínají zahazovat, pokud dosáhne zaplnění front určité úrovně. Pravděpodobnost zahození paketu pak se zaplněním front roste. Weighted RED (WRED) [25] je rozšíření RED, které navíc uvažuje klasifikaci paketů a tím zachovává diferenciaci služeb.

Tradiční pojetí kvality služeb vycházející z fixních sítí předpokládá, že po akceptování požadavku na QoS spojení musí síť kvalitu tohoto spojení garantovat po celou dobu jeho trvání (až na velmi výjimečné případy). Tento druh záruk se označuje jako hard-QoS. Bezdrátové sítě ale služby s hard-QoS garancemi mohou poskytovat pouze omezeně, protože zdroje, o které se opírají, nejsou stabilní. Částečným řešením (jak to dnes dělají např. mobilní telefonní sítě) může být omezit se na takovou fixní úroveň QoS, pro kterou ještě lze poskytnout rozumně pevné garance i v nestabilním prostředí. Tento přístup ale zcela opomíjí heterogenní požadavky různých aplikací. Mobilní ad-hoc sítě mají v tomto směru ještě mnohem těžší úkol. Ztráta konektivity např. kvůli vyčerpání energie některého jiného uzlu,

vysoká chybovost, nebo třeba rozpad sítě do více nesouvislých segmentů jsou v MANETech zcela běžné "provozní" záležitosti. Přitom jde o tak vážné problémy, že samotné snahy o zajišťování QoS v tomto prostředí někdy bývají zpochybňovány. Mnoho multimediálních aplikací nicméně hard-QoS nepožaduje – jsou totiž schopny poměrně dobře pracovat, i když není po jistou dobu jejich požadavkům plně vyhověno. Přistupuje se proto ke kompromisním řešením a benevolentnějším (resp. obecnějším) definicím kvality služby – např. soft-QoS nebo dynamic-QoS.

Soft-QoS

Soft-QoS [26] namísto striktních garancí obsahují pouze závazek vyhovět požadavkům v určité míře. Připouští se situace, kdy po jistou dobu není žádaná úroveň služeb dodržena. Míra splnění QoS požadavků (tzv. "target satisfaction", *TS*) se měří např. jako podíl doby, kdy bylo (či má být) QoS parametrům vyhověno k celkové době trvání spojení. Jiným měřítkem *TS* může být podíl paketů přenesených v souladu s QoS požadavky. Mimo samotných QoS parametrů tedy aplikace určuje požadovanou (minimální) hodnotu *TS*. Při *TS*=1 jde aplikaci o hard-QoS garance, zatímco při *TS*=0 o čistě best-effort službu. Někdy se pro jiné vyjádření téhož "statistického" uvolnění hard-QoS používá tzv. "soft-index". Ten je pouze jednotkovým komplementem *TS* – vyjadřuje tedy tolerovanou pravděpodobnost výpadku QoS.

Dynamic-QoS

Obvykle vyjadřuje aplikace své požadavky na kvalitu přenosu *K* hodnotami QoS parametrů dle příslušných metrik. Toto vyjádření odpovídá bodu v *K*-dimenzionálním prostoru se souřadnicemi definujícími minimální úrovně daných charakteristik (zpoždění, propustnost apod.). Dynamická QoS (dynamic-QoS) [27] umožňuje aplikacím vyjádřit své požadavky určením jak minimální úrovně kvality služby, kterou jsou ochotny akceptovat, tak i maximální úrovně, kterou jsou schopny využít. Místo jediného bodu tedy požadavek určuje *K*-rozměrný interval. Závazek splňovat dynamickou QoS znamená, že síť garantuje kvalitu služby odpovídající určitému bodu uvnitř tohoto intervalu. Aktuální úroveň služby, tedy o který bod intervalu přesně jde, určuje sama síť a určitým signalizačním mechanismem o tom informuje aplikaci. Ta se musí být schopna v daném rozsahu adaptovat. Dynamická QoS tedy část zodpovědnosti přenáší na samotné aplikace. Tím poskytují určitou flexibilitu pro jejich dobré fungování v nestabilním prostředí.

Následující kapitoly popisují některé konkrétní QoS modely, protokoly a další komponenty či mechanismy. Nejdříve jsou ale popsány vztahy mezi těmito stavebními prvky síťového QoS řešení.

QoS model představuje celkový "strategický plán" fungování sítě, která má podporovat kvalitu služeb. S ohledem na síťové prostředí, pro které je určen, musí QoS model definovat svůj cíl (tj. povahu poskytovaných služeb) a příslušný aparát (tj. konkrétní vrstvy či komponenty jednotlivých typů uzlů, jejich úkoly, závislosti a interakce). QoS model nespecifikuje, jakým způsobem mají jeho jednotlivé komponenty plnit své úkoly (až na výjimky, kdy je detailní definice chování kritická pro fungování celku). Externě přístupné rozhraní služeb daného QoS modelu je tedy odděleno od jeho konkrétní implementace (QoS architektury). Pro ověření správnosti návrhu QoS modelu bývá nicméně potřeba referenční implementace.

Síťová QoS architektura definuje konkrétní algoritmy, mechanismy a protokoly. Jde např. o QoS signalizaci a rezervaci prostředků nebo o MAC a směrovací QoS protokoly. Jejich úkolem je řešit zadání příslušného QoS modelu. Pouze ve vztahu ke konkrétnímu modelu také lze hodnotit, jak jsou tyto mechanismy důležité, účinné či vhodné. Např. uvažuje-li QoS model různé priority pouze několika tříd provozu, je signalizace pro každý tok nadbytečná. V bezdrátových sítích navíc dochází mezi jednotlivými komponentami QoS architektury k netriviálním interakcím (často např. mezi linkovou vrstvou a vyššími vrstvami, jak bude doloženo dále). Proto ne vždy lze kombinací "kvalitních" komponent (kvalitních podle izolovaného posuzování) dosáhnout kvalitních vlastností celku. Pro "kompletní řešení" kvality služeb v bezdrátových sítích se proto preferuje pohled shora, tj. od QoS modelu.

2.2. QoS modely

Dále jsou přiblíženy některé konkrétní QoS modely, resp. odpovídající architektury. Pro úplnost to jsou nejdříve dnes už "historické" internetové QoS standardy IntServ a DiffServ. Poté je popsán QoS model FQMM a speciální iMAQ. U prvních dvou modelů jsou zvlášť diskutovány omezení jejich použití v bezdrátovém prostředí. Druhé dva modely – jak iMAQ, tak i FQMM jsou již navrženy přímo pro bezdrátové sítě MANET. Popisují nicméně pouze lokální provoz v rámci ad-hoc domény, tj. přímo nepodporují propojení s dalšími QoS architekturami (především internetu). QoS interakce ad-hoc a přístupových sítí různých architektur jsou tedy dalším aktuálním tématem. Návrh možného řešení ukazuje např. systém PYLON [28].

IntServ

Integrated Services (IntServ) [29] je první komplexní QoS architektura navržená pro IP sítě. Jejím cílem jsou absolutní QoS garance pro individuální spojované datové přenosy. Těchto garancí má být dosaženo explicitní rezervací přenosové a výpočetní kapacity na směrovačích podél aktivních cest. IntServ obsahuje čtyři hlavní komponenty – signalizační protokol, rutinu řízení přístupu, klasifikátor paketů a plánovač. Jde přitom o rozšíření původní architektury internetu o tyto nové komponenty. Základní IP-služba je tedy (ve snaze usnadnit nasazování nové architektury) zachována. Mimo původní best-effort službu přináší IntServ dva nové druhy služeb: garantovanou (Guaranteed Service) a službu řízené zátěže (Controlled Load Service).

Guaranteed Service (GS) [30] je vhodná pro přenosy požadující pevné end-to-end garance propustnosti i přenosového zpoždění. Tyto garance jsou přitom založeny na předpokladu nejhoršího možného vlivu chování ostatních přenosů. Odpovídá-li daný tok vyjednanému profilu a nedojde-li k výpadku nebo změně topologie sítě (!), GS zajistí včasný přenos paketů beze ztrát způsobených případným zahlcením sítě. GS tedy emuluje "vyhrazený drát". Žádné snahy minimalizovat střední hodnoty zpoždění nebo jitteru ale nevyvíjí.

Přestože se Controlled Load Service (CLS) [31] nesnaží dodržel limit pro přenosové zpoždění u každého jednotlivého paketu, poměrně spolehlivě docílí jeho nízké střední hodnoty (uvažuje očekávaný vliv ostatních přenosů). Zhruba tato služba odpovídá chování best-effort služby při nízkém zatížení sítě, navíc s garancí požadované propustnosti. Nabízí tedy nulové nebo nízké ztráty způsobené zahlcením a pro většinu přenesených paketů také malé zpoždění. S provozem nad rámec parametrů vyjednaných pro CLS je zacházeno jako s best-effort provozem. CLS je vhodná speciálně pro real-time aplikace tolerantní ke zpoždění,

které při malém zatížení pracují dobře i v best-effort sítích, ale s vyšším zatížením se rychle dostávají do problémů.

Před navázáním GS nebo CLS spojení musí být nejdříve nalezena vhodná cesta a v uzlech podél této cesty zarezervovány prostředky odpovídající požadovaným parametrům spojení. To zajišťuje signalizační protokol (IntServ používá RSVP – Resource reSerVation Protocol [32], viz dále) ve spolupráci se směrovacím protokolem a rutinou řízení přístupu. Tato rutina provádí na každém uzlu podél cesty lokální rozhodnutí, zda lze požadavku vyhovět. Signalizační protokol pak zpětně informuje aplikaci o přijetí nebo zamítnutí požadavku.

Během vlastního QoS přenosu provádí směrovače pro každý přijatý paket klasifikaci podle zdrojových a cílových IP adres a portů, TOS (Type Of Service) a ID protokolu z IP hlavičky. Podle výsledku klasifikace je paket zařazen do konkrétní fronty. Plánovač (používá se CBQ) poté zajistí přenos tohoto paketu v souladu s vyjednanými QoS parametry.

Směrovače podporující IntServ nutně musí udržovat per-flow informace o rezervacích prostředků. Objem těchto informací je přitom úměrný počtu aktivních spojení. IntServ je proto vhodný pouze pro malé sítě s nízkým počtem toků, případně pro omezený počet specifických přenosů ve větších sítích (velké a páteřní sítě vyžadují škálovatelnější mechanismy). Služba GS vyžaduje implementaci IntServ ve všech uzlech sítě. Službu CLS je možné nasazovat inkrementálně. [33]

Nasazení IntServ v bezdrátových sítích je problematické. Přestože je tento model navržen pro obecně různé přenosové technologie linkové vrstvy, předpokládá některé jejich vlastnosti, které jsou typické pro drátové sítě (nízkou chybovost a vysokou kapacitu), ale prakticky nedosažitelné u bezdrátových spojů. Pro bezdrátové přístupové sítě by mohla být relevantní alespoň služba CLS. Bezdrátové sítě nicméně mají mnohem bohatší sadu parametrů ovlivňujících charakteristiky přenosů než drátové sítě a ani CLS neposkytuje dostatečné možnosti pro jejich nastavení. Řešením by mohlo být službu CLS rozšířit např. o akceptovatelnou míru chybovosti (resp. ztrát), nebo očekávané maximální zpoždění (např. pro stanovení počtu pokusů o opakování neúspěšného přenosu) [34]. Pro moderní vrstevnatá kódovací schémata by přitom byla ideální možnost určit citlivost vůči chybám a výpadkům individuálně pro každý paket.

Pro sítě MANET je architektura IntServ zcela nevhodná. Udržování per-flow informací ve směrovačích (tj. všech uzlech ad-hoc sítě) může znamenat příliš vysokou režii (vzhledem k jejich omezeným možnostem). Protokol RSVP také spotřebovává velkou část přenosové kapacity a neobsahuje mechanismy, které by dokázaly flexibilně reagovat na změny topologie. Podle IntServ by musel každý mobilní uzel mimo směrovacího a signalizačního protokolu implementovat také řízení přístupu, klasifikátor paketů a plánovač. To vyžaduje množství dalších prostředků, které v sítích MANET nemusí být dostupné [35]. Zásadním problémem je ale především nutnost pevného vyhrazení části přenosové kapacity v přepojovacích uzlech, což je v ad-hoc sítích prakticky nemožné. Nutno dodat, že IntServ je velmi "těžkotonážní" řešení, které se neujalo ani ve fixních sítích.

DiffServ

Model DiffServ (Differentiated Service) [36] vznikl jako možná alternativa či doplněk modelu IntServ, který je poměrně náročný na implementaci a reálné nasazení. DiffServ definuje omezený počet tříd provozu, kterým jsou na principu prioritizace poskytovány různé úrovně služeb. Nepoužívá žádné explicitní rezervace síťových prostředků. Poskytované QoS garance mají pouze relativní charakter, tj. nadměrné množství provozu v dané třídě tak může způsobit zahlcení a degradaci služby.

Podle modelu DiffServ zajišťuje poskytovatel své služby na základě kontraktu SLA s uživatelem. SLA mj. specifikuje podporované třídy provozu a povolený objem provozu pro jednotlivé třídy (což představuje určitou formu implicitní rezervace prostředků). Tzv. statický SLA se vyjednává a obnovuje v pravidelných intervalech. Dynamický SLA se uzavírá prostřednictvím signalizačního protokolu při každém konkrétním požadavku na přenos. Provoz vstupující do sítě je na hraničním (ingress) směrovači nejdříve podroben zásadám odvozeným z příslušného SLA (klasifikace a značení jednotlivých paketů, tvarování podle příslušného profilu). Týká se to jak provozu generovaného koncovou aplikací, tak provozu vstupujícího z jedné DiffServ domény do druhé (v takovém případě se uvažuje SLA mezi doménami). Označení paketu implikuje konkrétní způsob, jakým bude paket zpracován ve všech vnitřních (interior) směrovačích podél své cesty. To v důsledku určuje míru jeho upřednostnění před ostatním provozem. Vnitřní směrovače se řídí výhradně označením paketu a neudržují žádné per-flow informace – proto mohou být velice jednoduché a rychlé. DiffServ tedy nevyžaduje end-to-end signalizaci.

Architektura DiffServ definuje nový význam položky TOS v hlavičce IP paketu [37]. K označování paketů se používá tzv. DSCP (DiffServ Codepoint) obsažený v šesti bitech TOS. Dále DiffServ definuje základní sadu tzv. PHB (Per-Hop-Behavior) [39]. PHB detailně popisuje způsob nakládání s přenášenými pakety (ve smyslu jejich plánování) ve všech směrovačích tak, aby bylo dosaženo diferenciace jednotlivých tříd provozu. DSCP v hlavičce paketu tedy určuje PHB směrovačů přenášejících tento paket. DiffServ definuje několik povinných a doporučených DSCP a jejich mapování na konkrétní standardizované PHB. Obecně je možné mapovat více DSCP na stejné PHB.

S pomocí klasifikace paketů, tvarování provozu a konkrétních plánovacích mechanismů lze pod DiffServ zajistit širokou škálu různých typů služeb. Jejich konkrétní podoba závisí především na poskytovateli služeb, DiffServ definuje pouze některé DSCP a PHB. Např. od služby Premium Service [39] mohou aplikace s pevnou špičkovou rychlostí generování provozu očekávat nízké ztráty, zpoždění i jitter. Assured Service [39] představuje spolehlivou službu, která i v případě zahlcení sítě dosahuje alespoň očekávané propustnosti. Tuto službu lze navíc zavádět inkrementálně [33].

Sofistikovanější mechanismy, které DiffServ parametrizuje v SLA (klasifikace, tvarování provozu), stačí implementovat na "okrajích" sítě, tj. v ingress směrovačích či koncových uzlech. Vnitřní směrovače jsou naopak "odlehčené" – implementují pouze jednoduché PHB. Díky tomu lze model DiffServ poměrně snadno zavádět a ve srovnání s IntServ je mnohem lépe škálovatelný.

DiffServ používá k zajištění QoS prioritizaci namísto neflexibilních rezervací prostředků a je tedy pro použití v bezdrátových sítích principiálně vhodnější než IntServ. Nicméně i DiffServ byl navržen pro fixní a poměrně rychlé sítě a neuvažuje vysokou ztrátovost, a možnou mobilitu uzlů. Mechanismus implicitních rezervací, který DiffServ používá, nemusí být dostatečný při silně "nevyvážených" topologiích (v mobilních sítích zcela běžných). Může tedy vzniknout potřeba pomocí end-to-end signalizace zajišťovat prostředky dynamicky (např. změnit PHB pro konkrétní třídu provozu v uzlu, který tvoří úzké hrdlo). Takový signalizační protokol by přitom měl být kvůli omezené propustnosti bezdrátových spojů velmi jednoduchý. Vysoká míra ztrát u bezdrátových přenosů si dále může vynutit nasazení nějakých kompenzačních mechanismů (např. vyhrazení určitého podílu přenosové kapacity pro speciální kompenzační třídu).

Nasazení DiffServ do prostředí mobilních ad-hoc sítí má i další aspekty. Každý uzel sítě MANET funguje současně jako zdrojový uzel i jako směrovač pro "cizí" přenosy. To zvyšuje nároky na něj kladené, protože musí implementovat funkcionalitu jak ingress, tak interior směrovačů. Jak už bylo řečeno, koncept SLA není v ad-hoc sítích přímočaře použitelný. Pro plnou implementaci DiffServ je přitom nějaká forma SLA nutná.

FQMM

První QoS model určený pro mobilní ad-hoc sítě – FQMM (Flexible Quality-of-service Model for Manets) [35] navrhli Xiao a kol. Jde o hybridní model, který se snaží kombinovat přednosti IntServ a DiffServ, přičemž uvažuje specifika a omezení sítí MANET. Přenosy nejvyšší priority jsou rozlišovány per-flow, ostatní per-class. FQMM je primárně určen pro menší sítě (cca desítky uzlů) s plochou, nehierarchickou topologií.

Stejně jako DiffServ rozlišuje i FQMM tři typy uzlů – ingress, interior (resp. core) a egress. Ve FQMM nicméně nemá typ uzlu přímou souvislost s jeho fyzickou polohou v síti (neboť ta se dynamicky mění). Ingress jsou uzly, které vysílají data, interior uzly přenáší data ostatních uzlů a egress jsou cílové uzly přijímající data. Pro různé přenosy (a vzhledem k mobilitě) tedy uzly hrají různé role. Podobně jako v DiffServ používají ingress uzly traffic conditioner, který klasifikuje pakety, označuje je a podle daného profilu tvaruje provoz. Interior uzly pak přeposílají pakety v souladu s PHB určeným DSCP. Celá síť reprezentuje jedinou DiffServ doménu, v níž veškerý provoz vzniká v aplikacích na ingress uzlech a končí v egress uzlech. Chování FQMM tedy vychází z DiffServ – s tím, že pro jistou třídu provozu je poskytována per-flow granularita.

Cílem FQMM je maximální zužitkování konektivity při udržení pevné relativní diferenciace jednotlivých toků a tříd provozu. Za nejkritičtější považují autoři FQMM kapacitu bezdrátového spoje a proto zvolili za parametr diferenciace služeb přidělenou šířku pásma. Používaný profil je tedy definován procentuelním podílem daného toku nebo třídy na aktuální efektivní kapacitě spoje. Předpokládá se také adaptabilita samotného traffic conditioneru – např. při použití algoritmu děravého vědra by se měla s efektivní kapacitou spoje měnit i velikost "vědra". Pro směrování lze použít protokoly, které již při vyhledávání cest uvažují QoS omezení na dostupnou přenosovou kapacitu. FQMM to nicméně nevyžaduje – např. při použití DSR (Dynamic Source Routing) [40] se teprve po vyhledání cest provádí jejich dodatečná kontrola. Tato kontrola ověřuje, zda má každý uzel podél cesty v dané třídě provozu dostatek volné přenosové kapacity. Vyhovuje-li více cest, lze z nich vybrat nejlepší nebo třeba náhodnou. Implicitní formu rezervace prostředků modelu DiffServ prostřednictvím SLA tedy FQMM nahrazuje řízením přístupu prostřednictvím signalizace.

Model FQMM je založen na předpokladu, že pouze malá část přenosů požaduje vysoké QoS záruky (tj. per-flow zacházení), zatímco většině stačí agregované diferencování služeb per-class. Díky tomu je množství udržovaných per-flow informací nižší a problém nízké škálovatelnosti typický pro IntServ se zmírňuje.

iMAQ

Intergrated Mobile Ad-hoc QoS framework (iMAQ) [41] je vícevrstvá architektura navržená pro zajištění kvality služeb v sítích MANET. Jako specifický QoS parametr explicitně zajišťuje vysokou úspěšnost přístupu ke sdíleným datům pro aplikace třídy "multimedia-on-demand". Předpokládá se (zatím velmi vzácná) schopnost mobilních uzlů průběžně určovat vlastní polohu.

Architektura iMAQ popisuje vztahy aplikační vrstvy, vrstvy middleware a síťové vrstvy. Pro dosažení QoS garancí přitom middleware a síťová vrstva vzájemně aktivně spolupracují a sdílí systémové informace (tzv. profily). Primárně jsou uvažovány aplikace, které produkují multimediální data a následně je sdílí v rámci nějaké skupiny uživatelů sítě. Middleware takovým aplikacím asistuje při vyhledávání těchto dat a při přístupu k nim. V odůvodněných případech provádí také jejich replikaci a tím zajišťuje jejich dostupnost i v případech, kdy kvůli mobilitě ztratí skupina vzájemnou konektivitu. Síťová vrstva zajišťuje vlastní QoS směrování, tj. výpočet a údržbu cest splňujících sadu QoS požadavků (typicky zpoždění a propustnost).

Uzly sítí MANET se mohou relativně rychle pohybovat. To způsobuje rychlé "stárnutí" aktualizačních informací a směrovacím protokolům (i těm dynamičtějším) velké problémy s údržbou cest. Pokud neexistuje záložní cesta, oprava nebo rekonstrukce porušené cesty může znamenat významnou prodlevu. iMAQ se snaží tento problém řešit implementací QoS směrovacího protokolu s predikcí polohy ostatních uzlů. V rámci tohoto protokolu každý uzel pravidelně (a mimořádně i při náhlých změnách směru nebo rychlosti pohybu) rozesílá zprávy obsahující jeho aktuální polohu, dosah vysílání, zbývající energii a vytížení (tzv. profil uzlu). To mj. umožňuje zjišťovat charakteristiky pohybu uzlů, z nich predikovat jejich budoucí polohu a tedy i konektivitu v síti. S použitím těchto informací uzel provádí výpočty a údržbu QoS cest. Při směrování se nejdříve provede odhad času potřebného pro doručení paketu cílovému uzlu (tj. end-to-end zpoždění). Na základě tohoto odhadu a dostupných informací o pohybu cílového uzlu se určí jeho očekávaná poloha v momentu předpokládaného doručení paketu. Rozhodnutí o směru dalšího přenosu paketu se opírá mj. právě o tuto očekávanou polohu cílového uzlu. Popsaná procedura se provádí iterativně, dokud není dosaženo cíle.

V průběhu trvání spojení může síťová vrstva predikovat přerušení cesty (resp. snížení její kvality pod požadovanou úroveň) a v předstihu ji modifikovat, příp. zajistit novou vyhovující cestu. Pokud se to nezdaří, je upozorněna vrstva Middleware. Té jsou současně poskytnuty charakteristiky možných náhradních cest. Middleware pak s aplikací znovu projedná kvalitu služby – aplikace se tak včas může adaptovat na nové (horší) podmínky.

Vrstva middleware zajišťuje služby spolehlivého sdílení dat jejich inzerováním, lokalizací a replikacemi. Periodicky rozesílá zprávy (tzv. datové profily) s informacemi o lokální dostupnosti dat v daném uzlu ostatním kooperujícím uzlům. Mimo identifikátoru sdílených dat obsahují datové profily prioritu přiřazenou těmto datům a úroveň QoS potřebnou pro jejich přenos (např. propustnost a tolerovanou ztrátovost). Specifické aplikační požadavky lze řešit rozšířením seznamu parametrů datových profilů. Middleware udržuje v každém uzlu seznam všech dostupných dat a jejich zdrojů. Na základě informací o poloze a pohybu ostatních uzlů se pak snaží predikovat možné rozdělení sítě do více nesouvislých segmentů. Před tím, než se tak potenciálně stane, jsou v těchto segmentech vybrány "spolehlivé" uzly, mezi které se potřebná data zreplikují. Díky sdílení profilů uzlů se síťovou vrstvou může middleware při těchto replikacích i při zajišťování normálních přenosů zužitkovat heterogenitu uzlů (zbývající energie, vytížení apod.).

2.3. Signalizace a rezervace prostředků

QoS signalizace hraje při zajišťování kvality služeb roli řídícího centra. Signalizační protokol vychází z QoS modelu a koordinuje chování jeho jednotlivých komponent – řízení přístupu, směrování a MAC protokolu. Úkolem signalizace je zprostředkovávat vytváření, adaptace a rušení QoS přenosů a případné rezervace a uvolňování síťových prostředků. Signalizace je obvykle nejsložitější komponentou QoS modelu, protože představuje řízení složitých funkcí a kladou se u ní vysoké nároky na spolehlivost a výkonnost. Pro správné fungování signalizace je nezbytný spolehlivý přenos signálů (signalizačních informací) a jejich korektní interpretace (vyvolání patřičných mechanismů) ve všech uzlech. V svých důsledcích je to signalizační protokol, který předurčuje "pevnost" garancí QoS a na druhé straně i efektivitu využití sítě [42].

Podle způsobu přenosu signálů se rozlišují dva druhy signalizačních protokolů – v pásmu (in-band), tj. "přibalením" k běžným datovým paketům, a mimo pásmo (out-of-band), tj. prostřednictvím speciálních řídících paketů. Out-of-band signalizační protokoly představují svým způsobem čistější řešení. Při přenosu signalizačních informací totiž nespoléhají na přenos datových paketů a oproti in-band signalizaci, která musí vystačit s velmi omezeným prostorem, jsou snadněji rozšiřitelné a mnohem pružnější (mohou podporovat složitější funkce). Řídící pakety lze navíc přenášet obecně po jiných cestách než datové pakety. Některé funkce dokonce nelze implementovat pomocí čistě in-band signalizace – např. v případě, že po sestavení spojení je všechen datový provoz pouze jednosměrný. Out-of-band signalizace nicméně spotřebovává podstatně více přenosové kapacity než in-band. Její řídící pakety musí soupeřit o médium s datovými pakety. Přenos řídících informací by navíc měl mít přednost před přenosem běžných datových paketů. V mobilních ad-hoc sítích se většinou preferuje jednodušší a "levnější" in-band signalizace. Out-of-band signalizace totiž pomaleji reaguje na změny topologie – musí např. explicitně žádat dotčené uzly o rezervace, příp. uvolnění prostředků.

Rezervační protokol RSVP představující de-facto standard pro IntServ je příkladem out-of-band signalizace. Protokol RSVP a některé jeho rozšíření budou přiblíženy dále. In-band signalizaci používá např. protokol INSIGNIA, kombinaci in-band a out-of-band signalizace používá protokol ASAP. Protokoly INSIGNIA a ASAP navržené pro mobilní ad-hoc sítě jsou rovněž popsány dále.

Rezervace prostředků (vyžaduje-li ji QoS model) probíhá podél cest, po kterých směrovací protokol přenáší příslušné signály. Obecně není nutné, aby sám směrovací protokol uvažoval QoS charakteristiky cest, nicméně použitím QoS směrování se snižuje riziko, že zvolená cesta nebude obsahovat dostatek prostředků (tj. že rezervace selže).

Při zajišťování QoS v drátových sítích se lze na rezervované prostředky spolehnout. Rezervace prostředků při sestavování spojení a jejich uvolnění po skončení přenosu tedy mohou představovat dvě nezávislé, explicitní akce. V bezdrátovém prostředí tento předpoklad neplatí a proto se používá flexibilnější a robustnější správa prostředků prostřednictvím soft-state rezervací. Soft-state rezervace se musí periodicky obnovovat, jinak je uzel automaticky zruší. V případě out-of-band signalizace se rezervace "občerstvují" pomocí explicitních zpráv. In-band signalizace předpokládá průběžné potvrzování rezervací samotnými přenášenými daty. Mechanismus soft-state rezervací tedy předchází neplatným rezervacím, přičemž nevyžaduje jejich explicitní rušení. Interval pro obnovování rezervací představuje kompromis mezi režií a adaptabilitou rezervací na změny topologie a přenosů. Ideálně by tento interval měl mít nějaký vztah k dynamice sítě.

Rezervaci prostředků může iniciovat buď zdrojový uzel (např. INSIGNIA) nebo cílový uzel (např. RSVP). Rezervace cílovým uzlem s sebou přináší větší režii, nicméně může zohlednit i požadavky cílového uzlu. Výsledně je tedy obecnější – např. u multicast přenosů tak lze podporovat heterogenní QoS (tj. různí příjemci multicast přenosů mohou požadovat různou úroveň QoS). Na druhou stranu rezervace zdrojovým uzlem obecně nevyžaduje spolupráci cílového uzlu (a ten ani nemusí signalizační protokol implementovat).

RSVP

Resource reSerVation Protocol (RSVP) [43] byl navržen jako signalizační protokol pro architekturu IntServ. Jde nicméně o obecný out-of-band rezervační protokol pro IP sítě. RSVP nedefinuje ani interní formát svých zpráv, ani přesnou sémantiku rezervací. Tyto záležitosti považuje za "neprůhledné" – zabývá se výhradně mechanismem správy rezervací prostředků pro jednotlivé jednosměrné datové toky. Vlastní rezervace prostředků v RSVP jsou vždy soft-state a iniciuje je cílový uzel. RSVP podporuje heterogenní QoS. Pro jednoduchost zde bude ilustrován pouze případ unicast přenosu.

Při sestavování spojení zašle zdrojový uzel zamýšlenému příjemci zprávu PATH. Tato zpráva dohodnutým způsobem specifikuje charakteristiky uvažovaného přenosu. Po obdržení zprávy cílový uzel na základě těchto charakteristik a vlastních specifických potřeb rozhodne o přesné podobě rezervací. Specifikaci požadovaných rezervacích zahrne cílový uzel do zprávy RESV, kterou zašle zpět ke zdroji. Během přenosu zprávy RESV jednotlivé uzly rezervují pro daný přenos požadované prostředky. Není-li možné rezervaci v některém uzlu zajistit, informuje tento uzel příjemce chybovou zprávou a sestavování spojení končí neúspěchem.

Statické rezervace, které RSVP zajišťuje, obecně příliš neodpovídají povaze bezdrátových přenosů. Rozšíření dRSVP (dynamic RSVP) [27] nahrazuje toto schéma dynamickou QoS se zpětnou vazbou pro adaptivní aplikace.

Protokol RSVP dále nebere v úvahu mobilitu uzlů. Pro prostředí sítí WLAN proto Talukdar a kol. navrhli rozšíření MRSVP (Mobile RSVP) [44], které mobilním uzlům zajišťuje dodatečné rezervace prostředků v sousedních buňkách. Kvalitu služby je pak možné udržet i při přechodu uzlu k jiné základnové stanici. Podobný princip používá také HMRSVP (Hierarchical Mobile RSVP) [45] – díky integraci RSVP s mechanismem registrace mobilních uzlů v Mobile IP[46] ale dosahuje vyšší efektivity (méně blokovaných nevyužitých prostředků). LRSVP (Localized RSVP) [47] zase při přechodu mobilního uzlu k jiné základnové stanici umožňuje lokální opravy dotčené části cesty v přístupové síti.

Obecně se má zato, že proaktivní přístup protokolu RSVP s "předpřipravenými" rezervacemi nedokáže dostatečně pružně reagovat na časté změny topologie mobilních ad-hoc sítí. Pro ně je principiálně vhodnější in-band signalizace, která navíc přináší nižší komunikační režii.

INSIGNIA

Signalizační protokol INSIGNIA [48] byl navržen speciálně pro mobilní ad-hoc sítě. Prostřednictvím in-band signálů zajišťuje rychlé rezervace prostředků při sestavování přenosů, lokální opravy těchto rezervací při změnách topologie a průběžné hlášení kvality spojení zdrojovému uzlu. Mimo best-effort uvažuje INSIGNIA adaptivní real-time přenosy. Pro ty zajišťuje garanci propustnosti prostřednictvím soft-state rezervací přenosové kapacity mezilehlých uzlů.

INSIGNIA vkládá do volitelné části IP hlavičky každého odesílaného paketu signalizační informace ve formátu <*ReservationMode, ServiceType, PayloadIndicator, BandwidthIndicator, BandwidthReqMin, BandwidthReqMax*> (čtyři bitové příznaky a dvě osmibitové hodnoty). Příznak *ReservationMode* určuje, zda zdroj daného přenosu žádá o rezervaci prostředků (*REQ*), nebo jen chce využít již rezervovaných prostředků (*RES*). V případě *REQ* se INSIGNIA pokusí pro daný přenos rezervovat požadované prostředky (tj. přenosovou kapacitu *BandwidthReqMax* nebo alespoň *BandwidthReqMin*). Zdaří-li se, rezervace je potvrzena nastavením typu služby (*ServiceType*) na real-time (*RT*). V opačném případě je přenos degradován na best-effort (*BE)*. V obou případech je paket předán dalšímu uzlu na cestě k cíli. Pro real-time přenosy příznak *BandwidthIndicator* průběžně určuje, zda všechny předchozí uzly dokázaly rezervovat *BandwidthReqMax*, nebo zda podél cesty existuje "úzké hrdlo", kde se podařilo rezervovat pouze *BandwidthReqMin.*

Cílový uzel aktivně monitoruje kvalitu spojení (propustnost, ztrátovost) a periodicky o ní, v intervalech určených aplikací, informuje zdroj. Pomocí této zpětné vazby se zdrojový uzel dozví o dokončení rezervací (poté odesílá pakety, které už o rezervaci prostředků nežádají) a o tom, zda má k dispozici pásmo *BandwidthReqMax* nebo pouze *BandwidthReqMin*. Aplikace se musí průběžně adaptovat na aktuální podmínky. Příznak *PayloadIndicator* určuje, zda je daný paket v rámci "základního profilu" (do *BandwidthReqMin*). Provoz nad tento základní profil je v případě nižší úrovně rezervací (tj. v situacích, kdy se aplikace zatím neadaptovala na nižší propustnost spojení) degradován na best-effort.

Požadavky aplikací na přenosy v prostředí sítí MANET s nízkou propustností často nelze uspokojit. INSIGNIA rozšiřuje rezervační model "všechno nebo nic" protokolu RSVP na dvě diskrétní úrovně – *BandwidthReqMin* a *BandwidthReqMax.* Tím může v mnoha situacích zajistit pro real-time aplikace alespoň nějaké garance. Kvůli jednoduchosti INSIGNIA nezasílá explicitní chybové zprávy ani v případě, že nelze vyhovět požadavkům na minimální propustnost (místo toho používá pouze zmíněné periodické QoS hlášení).

ASAP

Protokol ASAP (Adaptive reServation And Pre-allocation Protocol) [49] je pokročilý QoS signalizační protokol pro sítě MANET. Použitím dvoufázového rezervačního mechanismu se snaží poskytnou vyšší flexibilitu a omezit nadbytečné rezervace, ke kterým může vést INSIGNIA. Sledovaným QoS parametrem je propustnost end-to-end spojení. ASAP zajišťuje adaptivní dynamické QoS (tj. aplikace může průběžně měnit rozsah svých požadavků a síť průběžně informuje aplikaci o aktuálních poměrech).

ASAP definuje dva druhy rezervací – hard a soft (obojí fungují jako soft-state). Prostředky se soft rezervací lze použít pro QoS přenosy i best-effort přenosy. Nelze je však opětovně znovu rezervovat (hard ani soft). Soft rezervace vyjadřuje předběžný zájem o vyhrazení určitého množství prostředků. V druhé fázi rezervace přechází část soft-rezervovaných prostředků do režimu hard rezervace (tím je napevno vyhrazena pro konkrétní přenos), zbylá část soft rezervace se ruší. Každý uzel udržuje množství hard a soft rezervovaných prostředků aktivních přenosů.

Vlastní sestavení QoS cesty, její adaptace a lokální opravy zajišťují dva typy signalizačních zpráv – *SR* (Soft Reservation) a *HR* (Hard Reservation) ve formátu: *SR*=<*MinBW, MaxBW, SoftBW, HardBW*> a *HR*=<*SetBW, SoftBW, HardBW*>. Zprávy SR periodicky zasílá zdrojový uzel spojení cílovému uzlu v hlavičce datových paketů (in-band). Slouží k soft rezervacím prostředků při sestavování přenosů (nebo jejich opravách) a k monitorování kvality spojení. Out-of-band signálem HR (tj. speciálním paketem) potvrzuje

cílový uzel předběžné soft rezervace nebo modifikuje hard rezervace a informuje zdrojový uzel o změnách kvality spojení.

Při sestavování spojení vyšle zdrojový uzel zprávu *SR* s požadavkem na propustnost spojení v rozsahu [*MinBW*, *MaxBW*]. Jednotlivé uzly podél cesty k cílovému uzlu zajistí soft rezervaci co největšího dostupného množství přenosové kapacity (v daném rozsahu). *SoftBW* udržuje průběžné minimum množství soft-rezervovaných prostředků pro dané spojení v předchozích uzlech. Cílový uzel se tak v *SoftBW* dozví dosažitelnou propustnost spojení. Prostřednictvím zprávy *HR* pak zajistí hard rezervaci příslušného množství prostředků (nastaví *SetBW*) ve všech uzlech podél cesty. Při tom jsou současně uvolněny další prostředky, předběžně rezervované nad rámec *SetBW*.

Poté, co je spojení sestaveno, zasílá zdroj pravidelné in-band signály *SR*. Ty při svém přenosu obnovují rezervace a shromažďují informace o kvalitě spojení (analogicky k *SoftBW* udržuje *HardBW* průběžné minimum hard-rezervovaných prostředků). Současně se ASAP prostřednictvím SR signálů snaží průběžně upravovat propustnost spojení (přiblížit se požadovanému maximu). Cílový uzel sleduje aktuální QoS situaci a v případě změny informuje zdrojovou aplikaci vysláním signálu *HR*. Ukončení spojení nevyžaduje explicitní akce (soft-state rezervace), nicméně v případě potřeby může cílový uzel vyvolat okamžité uvolnění prostředků zprávou *HR* se *SetBW*=0.

V mobilních ad-hoc sítích běžně dochází k přerušení cest. Pro zajištění QoS je pak kritická rychlost zajištění nových rezervací od místa přerušení do cílového uzlu. ASAP proto obsahuje mechanismus rychlých lokálních oprav přerušených cest. Uzel, který obdrží zprávu *SR* týkající se spojení, které ještě "nezná", ale obsahující přitom nenulové *HardBW*, usoudí, že došlo k přesměrování (tj. jedná se o opravu přerušené cesty). Dále pak postupuje stejně jako při sestavování spojení – až na to, že pro část prostředků odpovídající *HardBW* provede rovnou hard rezervaci (je-li to možné). Cílový uzel ze zprávy *SR* případně zjistí, že rekonstruovaná cesta vykazuje horší parametry než původní cesta, a může o tom standardním způsobem informovat zdroj.

Kombinací hard a soft rezervací ASAP brání nadbytečným rezervacím prostředků, ke kterým dochází při použití protokolu INSIGNIA v uzlech před "úzkým hrdlem". Další významnou výhodou pro aplikace s adaptivními kodeky je to, že ASAP dokáže zajistit obecně jakoukoli (dostupnou) šířku pásma z požadovaného intervalu, nejen minimální nebo maximální – přenosovou kapacitu tedy využívá efektivněji. Nevýhodou protokolu ASAP zůstává, že mobilní uzly musí udržovat per-flow informace.

2.4. Kvalita služeb z vrstevnatého pohledu

Následujících oddíl se věnuje vrstevnatému pohledu na QoS. Cílem přitom není zde podat vyčerpávající přehled nebo kategorizaci, ale uvedením několika konkrétních technik ilustrovat možné přístupy. Především jde o protokoly přístupu k médiu a směrování, zmíněny jsou ale také některé konsekvence ve fyzické a transportní vrstvě.

Vyššími vrstvami se tato práce nezabývá. Možná je ale vhodné se zde okrajově zmínit, že existují i snahy o zajištění vnímané kvality služby v best-effort sítích založené čistě na aplikaci, tj. bez spolupráce či vědomí sítě. Příkladem může být poměrně kuriózní programovací model [50] umožňující aplikacím určeným pro přenos řeči adaptovat se na změny kvalitativních parametrů bezdrátových multihop sítí. V tomto modelu audio klient nepřetržitě monitoruje QoS parametry spojení, jako např. propustnost, ztrátovost paketů či rozptyl zpoždění a předává získané informace zpět audio serveru. Server ve velkém rozsahu přizpůsobuje charakteristiku audioproudu aktuálním podmínkám v síti. Za určitých podmínek je aktivována jistá "minimální" vrstva, která používá techniky rozpoznávání a syntetizace řeči (tj. nepřenášejí se zvukové vzorky, ale datově mnohem méně objemný text). Tento adaptivní model umožňuje přijatelně smysluplnou komunikaci i za velmi nepříznivých podmínek v síti. Za obvyklých podmínek přitom může zlepšovat charakteristiky přeneseného hlasu a zvyšovat srozumitelnost řeči [50].

2.4.1. Fyzická vrstva

Zvýšit efektivitu využití proměnné kapacity bezdrátových kanálů umožňují adaptivní modulačních techniky. Ty se přizpůsobují okamžitému stavu kanálu (např. odstupu signál/šum) tak ,že hledají vhodný poměr mezi efektivitou modulace (v bps/Hz) a odolností modulovaného signálu vůči chybám při přenosu. Cílem je samozřejmě maximalizace přenosového výkonu. Schopnost přesně odhadovat aktuální vlastnosti kanálu přitom závisí na účasti přijímače i vysílače a spolehlivé zpětné vazbě. Zpoždění této zpětné vazby může ve vysoce dynamickém prostředí omezovat přínosy adaptivní modulace. Zlepšení pak lze dosáhnout např. lineární predikcí vývoje chování bezdrátového kanálu [51].

Jiné schéma adaptivní modulace je založeno na prioritě přenášených dat. Vysoce prioritní data jsou modulována tak, aby "prošly" kanálem i za velmi nepříznivých podmínek. Méně důležitým datům (ať už přenášeným současně nebo zvlášť) je přiznán nižší stupeň ochrany. Tento přístup (tzv. Unequal Error Protection) se přitom může týkat i kódovacích schémat nebo např. maximálního počtu opakování přenosu rámce na linkové úrovni.

Mimo adaptivní modulace existují i další techniky, které na fyzické úrovni pomáhají efektivněji využívat přenosové pásmo. Např. dynamickou volbou kmitočtu (Dynamic Frequency Selection) se lze vyhýbat zarušeným nebo intenzivně využívaným kanálům. Automatická regulace vysílacího výkonu (Transmitter Power Control) omezuje vysílací výkon na minimum nezbytné pro spolehlivý přenos – omezuje tak rušení a navíc se snižují nároky na napájení mobilních zařízení.

Na přenosový výkon má vliv i použité kódování na fyzické úrovni. Techniky s delšími kódovými slovy obecně zvyšují odolnost signálu proti zkreslení při přenosu – za cenu snížení propustnosti lépe chrání přenášená data. Opačný vliv na chybovost má ale komprese zdrojových dat na aplikační úrovni. Většina kódovacích schémat na fyzické úrovni je navržena pro specifické vlastnosti kanálu a cílovou chybovost (Bit Error Rate) a nebere ohled na charakteristiku přenášených dat. Analogie platí i na aplikační úrovni – současné multimediální kodéry jsou většinou optimalizované pro dosažení co nejvyšší (vnímané)

kvality služby pro danou bitovou rychlost a předpokládají, že všechny zakódované bity budou přeneseny bezchybně. Především kvůli jednoduchosti se obvykle volí určitý pevný kompromis mezi oběma přístupy, tj. fixní úroveň komprese dat na aplikační úrovni a fixní ochrana těchto dat kódováním na fyzické úrovni.

Za jistých podmínek skutečně lze podle Shannonova teorému provádět kódování (kompresi) zdrojových dat a fyzické kódování kanálu odděleně (např. sekvenčně) bez jakýchkoli ztrát na celkovém výkonu. Tyto podmínky (mj. stabilní vlastnosti kanálu) nicméně v reálných podmínkách bezdrátových sítí neplatí a Shannonův teorém nelze použít ani jako dobrou aproximaci [52]. Optimálního přenosového výkonu lze dosáhnout pouze koordinovanou volbou míry komprese dat a jejich ochrany fyzickým kódováním v závislosti na aktuálních vlastnostech bezdrátového kanálu [52]. Proto Qian a kol. navrhli adaptivní parametrický model pro bezdrátové videopřenosy odolné vůči chybám [52], který bere v úvahu jak charakteristiky zdroje, tak i aktuální chybovost kanálu. Zmíněný přístup se tedy netýká pouze fyzické vrstvy, ale spíše její interakce a těsnější spolupráce s vyššími vrstvami.

2.4.2. QoS protokoly linkové vrstvy

Protokoly linkové vrstvy musí zajišťovat minimálně přístup k médiu, formování a správu bezdrátových spojů a spolehlivé unicast přenosy. Komponenty QoS architektury vyšších vrstev (např. QoS směrování a signalizace) nadto většinou požadují podporu rezervací prostředků nebo prioritizace a případně také nějaké zpětnovazební mechanismy dovolující reagovat na změny vlastností bezdrátového kanálu.

Hlavní pozornost je zde věnována soutěžním metodám, které lze implementovat distribuovaně, a tedy použít i v multihop prostředí. Pro srovnání je ale nejdříve popsán centralizovaný deterministický mechanismus, kterým standard 802.16 zajišťuje kvalitu služeb v metropolitních bezdrátových sítích.

IEEE 802.16

Technologie standardu IEEE 802.16 [8] jsou určeny pro širokopásmový bezdrátový přístup na střední vzdálenosti (v řádu kilometrů). Tento standard podporuje spojově orientované služby. Je založen na deterministickém přístupu k bezdrátovému kanálu prostřednictvím časového multiplexu (TDMA). Používá se centralizovaný přístup, kdy základnová stanice (Base Station, BS) řídí přístup a alokaci prostředků pro koncové stanice (Subscriber Stations, SS). Podle směru přenosu dat se rozlišují dva druhy spojů – downlink (od BS k SS) a uplink (od SS k BS). Struktura linkového rámce je u těchto dvou druhů spojů různá. Uplink rámce se dělí do tří částí – části podpory (Maintenance Period), soutěžní části (Contention Request Period) a část plánovaných datových přenosů (Scheduled Data Grant Period). První dvě části slouží koncovým stanicím k inicializaci spojení a k vyjádření požadavků na přidělení přenosového pásma. Koncové stanice v těchto dvou částech soutěží o časové sloty. Downlink rámce obsahují mimo dat pro koncové stanice také řídící informace pro downlink i uplink. Z hlediska QoS je zde podstatná mapa UL-MAP, která indikuje přidělení časových slotů v uplinku a DL-map, která indikuje přidělení časových slotů v downlinku. Alokace přenosového pásma se tedy provádí prostřednictvím UL-MAP a DL-MAP a rozhoduje o ní základnová stanice.

Podpora QoS v 802.16 je založena na konceptu servisních toků (Service Flows). Servisní tok je jednosměrný tok paketů, kterému je zajištěna určitá úroveň QoS. Může přitom existovat, aniž by byl aktivován pro přenos paketů. Existují čtyři typy servisních toků: Unsolicited Grant Service (UGS), Real-Time Polling Service (rtPS), Non-Real-Time Polling Service (nrtPS) a Best Effort (BE). UGS se používá jako emulace okruhů pro real-time přenosy s konstantní bitovou rychlostí (CBR). USG vyhradí pro daný přenos fixní přenosové pásmo, přičemž zde není možnost žádat o jeho rozšíření. Servisní toky rtPS jsou určeny pro real-time přenosy s variabilní bitovou rychlostí (VBR), jako např. streaming videa nebo audia. Pásmo se pak přiděluje dynamicky, výzvami základnové stanice. Aby se zamezilo nepředvídatelnému zpoždění, rtPS toky nemohou získat dodatečné pásmo soutěžním způsobem. Pro aplikace, které by rády něco více než best-effort, ale nevyžadují nutně real-time služby (např. vysokorychlostní FTP), jsou navrženy servisní toky typu nrtPS. Ty mohou přenášet data jak na základě výzev základnové stanice, tak na soutěžním principu. BE může pracovat pouze na soutěžním principu. Plánovač na základnové stanici zajišťuje zmíněným typům servisních toků tyto QoS parametry: minimální garantovanou rychlost (Minimum Reserved Rate, MRR), maximální zpoždění, maximální jitter a prioritu (při soutěžení). Navíc se definuje maximální rychlost nad hranicí MRR, které se ještě plánovač snaží vyhovět na principu best effort.

Registrací do sítě získá síťová aplikace unikátní identifikátor servisního toku (Service Flow Identifier). Kvalita služby pro konkrétní spojení se pak zajistí asociací identifikátoru tohoto spojení (Connection Identifier, CID) s příslušným servisním tokem. Standard přitom definuje dva druhy alokací – Grant Per Connection (GPC) a Grant Per Subscriber Station (GPSS). GPC znamená vyhrazení pásma konkrétnímu spojení, tj. o plánování se stará základnová stanice. Naproti tomu při GPSS je pásmo vyhrazeno pro jednotlivé koncové stanice – ty tak mají možnost poskytovat QoS "svým" přenosům lokálním plánováním.

IEEE 802.11

Lokální bezdrátové sítě WLAN podle standardu IEEE 802.11 používají přístupovou metodu CSMA/CA, konkrétně distribuovanou verzi DCF (Distributed Coordination Function), případně centralizovanou PCF (Point Coordination Function) [53]. DCF nedokáže diferencovat různé druhy provozu a tedy nenabízí žádnou podporu QoS. Varianta PCF sice nabízí deterministický přístup k médiu, ale v praxi se neujala. Nedokáže totiž diferencovat služby pro různé zdroje provozu. Navíc když chce základnová stanice podle PCF začít přidělovat médium deterministicky, musí nejdřív soutěžním způsobem "získat" médium. Efektivní intervaly, ve kterých je médium skutečně přidělováno deterministicky, se tak mohou měnit.

Standard 802.11e [9] doplňuje jak distribuovaný, tak centralizovaný přístup k médiu o podporu QoS (resp. v případě PCF tuto podporu vylepšuje). Rozšíření EDCF (Enhanced Distribution Coordination Function) zobecňuje definici mezirámcových mezer. Zavádí AIFS (Arbitration Inter-Frame Space), která relativně (budiž zdůrazněno, že absolutní garance EDCF neposkytuje) diferencuje osm kategorií provozu – s rostoucí prioritou od přenosů na pozadí až po služby řízení sítě. Přenos s vyšší prioritou má přiřazenu nižší hodnotu AIFS, než přenos s nižší prioritou. Má tedy při soutěži o médium vyšší pravděpodobnost výhry. K předcházení kolizím při více vysílání v rámci stejné kategorie provozu se používá (stejně jako u DCF, kde je ale jediná kategorie provozu) dodatečná náhodná čekací doba v rámci soutěžního okénka (Contention Window). EDCF používá pro různé kategorie provozu různé spodní hranice tohoto okénka. Dalším rozšířením, které přináší EDCF je packet bursting. Stanice, která se dostane k médiu, může vyslat více rámců za sebou, aniž by musela o médium

opakovaně soutěžit. Celkový čas vysílání přitom nesmí překročit určitou mez. Tím se zvyšuje utilizace média.

Druhým rozšířením původního standardu 802.11 je HCF (Hybrid Coordination Function), kdy je přístup k médium řízen výzvami koordinátora (Hybrid Coordinator). Koordinátorem je obvykle přístupový bod. Stejně jako původní PCF (Point Coordination Function) rozděluje i HCF čas do superrámců, které mají soutěžní (Contention-Based) a nesoutěžní (Contention-Free) část. V průběhu nesoutěžní části superrámce mohou stanice vysílat výhradně na výzvu koordinátora. HCF rozšiřuje mechanismus výzev PCF tak, že každá stanice může být vyzývána s různou frekvencí. Výzva přitom obsahuje povolený začátek a maximální délku vysílání (problémem PCF bylo, že doba vysílání vyzvané stanice nebyla předem známa). Kvalitu služby pro jednotlivé přenosy lze tedy s HCF definovat kvantitativně. QoS signalizace od stanice ke koordinátorovi, která se k tomu používá, definuje dva druhy signálů. QSI (Queue State Indicator) informuje koordinátora o stavu front jednotlivých datových toků. TS (Traffic Specification) obsahuje požadavek na rezervaci části pásma. Během soutěžní části superrámce se k médiu přistupuje podle EDCF i na základě výzev koordinátora. Samotnou alokační politiku, která na základě přenosových požadavků stanovuje rezervace částí kanálu, ani mechanismy pro řešení případných variací kvality kanálu standard nespecifikuje.

Ani nový standard neřeší problém předsunuté stanice. Použití přístupové metody CSMA/CA v multihop prostředí proto s sebou nadále přináší problémy s propustností, možnou nestabilitou TCP spojení, férovostí přístupu k médiu jednotlivých přenosů a možným vysokým rozptylem zpoždění [18,19]. Dvě zajímavé QoS rozšíření CSMA/CA pro multihop sítě jsou přiblíženy dále.

MACA/PR

MACA/PR (Multiple Access Collision Avoidance with Piggyback Reservation) [54] je MAC protokol pro multihop bezdrátové sítě kombinující asynchronní povahu CSMA/CA s podporou QoS podobnou alokaci média v TDMA sítích. Pro real-time provoz zajišťuje rychlý a spolehlivý přenos a navíc může prostřednictvím rezervací garantovat určitou úroveň propustnosti. MACA/PR je protokol přístupu k médiu a jako takový je schopen vytvářet real-time spojení pouze přes jeden přeskok. Je ale navržen tak, aby ve spolupráci se směrovacím protokolem a mechanismem rychlého navázání spojení poskytoval end-to-end garance.

K běžnému provozu MACA/PR přistupuje obdobně jako CSMA/CA. Významným způsobem se při tom ale opírá o tabulku rezervací slotů tzv. "cyklu". Do ní se zaznamenávají rezervovaná vysílací i přijímací časová okénka všech uzlů v dosahu. Uzel, který chce vysílat rámec, musí respektovat rezervace v tabulce a čekat na volné okénko. Dále musí čekat po určitou náhodnou dobu. Během ní příposlechem zjišťuje, zda je médium volné. Pokud je volné, může začít klasický RTS/CTS dialog a v případě dohody pak vyslat rámec a čekat na potvrzení (dialog DATA/ACK). V případě, že je médium obsazeno, musí čekat na jeho uvolnění a stejným způsobem zkoušet znovu. Neprioritní provoz tedy klasickým soutěžním způsobem vyplňuje volná okénka rezervační tabulky.

K real-time provozu se MACA/PR chová odlišným způsobem. Pro první přenášený rámec real-time přenosu se na každém přeskoku použije dialog RTS/CTS a následně DATA/ACK. To zajistí jeho spolehlivý přenos až k cíli. Zároveň se při přenášení prvního rámce podél cesty vytvářejí soft-state rezervace. Pro následující rámce real-time přenosu se již používá pouze dialog DATA/ACK. V hlavičkách rámců DATA a ACK je vyhrazen prostor pro informace o rezervaci časového okénka pro daný real-time přenos. Odesílatel do hlavičky každého datového rámce DATA zaznamená rezervaci pro přenos následujícího rámce. Všechny uzly, které tento rámec následně uslyší, se tak dozví o čase dalšího vysílání a (vyjma uzlu, kterému

je rámec určen) zápisem do rezervační tabulky označí příslušné okénko jako nevhodné pro vlastní příjem. Uzel, kterému je rámec určen, rezervaci vloží do své rezervační tabulky jako okénko nevhodné pro vysílání. Navíc tuto informaci vloží i do rámce ACK, prostřednictvím kterého rezervaci potvrzuje. Uzly, které uslyší ACK, rezervaci zapíší do tabulky (také jako okénko nevhodné pro vysílání). Ani v případě, že odesílatel dat neobdrží potvrzení, se přenos real-time rámců (vyjma prvního) neopakuje. Odesílatel místo toho, neobdrží-li žádné potvrzení pro několik po sobě následujících rámců, informuje vyšší vrstvu (spojení již nevyhovuje QoS požadavkům na propustnost). Směrovací protokol pak může spojení restartovat, tím vyvolat mechanismus RTS/CTS a pokusit se o nové rezervace. Potvrzovací rámce ACK v tomto schématu tedy slouží pro potvrzování rezervací a nikoli pro zotavení z chyb při přenosu. Rezervace, které nejsou po několik následujících cyklů obnoveny, se automaticky ruší.

Real-time provoz ve schématu MACA/PR je před problémem skrytého uzlu chráněn namísto obvyklého dialogu RTS/CTS pomocí mechanismu údržby rezervací a jejich propagací mezi sousedními uzly. K občasných kolizím a ztrátám může docházet i u real-time provozu kvůli možné nedostatečné rychlosti aktualizace rezervačních tabulek (způsobené např. vysokou mobilitou uzlů).

Black-Burst

Black Burst Contention Scheme [55] je další technika, pomocí níž lze vylepšit chování MAC protokolů typu CSMA v multihop prostředí. U real-time provozu účinně minimalizuje zpoždění přístupu k médiu, brání kolizím a dokáže zajistit vysokou utilizaci bezdrátového kanálu [56]. Od fyzické vrstvy Black Burst očekává schopnost zkoušet přistupovat k médiu v pravidelných intervalech a schopnost bezdrátové médium po definovanou dobu nějakým způsobem "zarušit".

Uzel, jenž má zájem vysílat, si v souladu s CSMA/CA příposlechem ověří, že médium bylo alespoň po dobu mezirámcové prodlevy volné. Poté začíná soutěž o médium. Real-time provoz používá kratší mezirámcovou prodlevu a je tak upřednostněn před best-effort provozem. V okamžiku, kdy přijde na řadu vyslání real-time rámce (potenciálně se tak stane ve více uzlech současně), se vysílací stanice pokusí rezervovat médium tím, že vyšle krátkou dávku, tzv. Black burst (BB). BB na určitý čas zaruší médium. Podstatné přitom je, že tyto dávky mají proměnnou délku. Každý BB zabírá celočíselný počet časových slotů. Délka BB je rostoucí funkcí doby, která uplynula od prvního pokusu o přístup k médiu do doby začátku vysílání BB. Bezprostředně po odvysílání BB stanice po jistou krátkou dobu (méně než jeden slot) příposlechem zjišťuje, zda některý konkurent nadále neruší médium vysíláním delší BB. Stane-li se tak, musí stanice čekat na uvolnění média a znovu se je pokoušet získat stejným způsobem (už ale s delším BB). Schéma Black Burst zaručuje, že každá stanice používá BB dávky různých délek, tj. že vždy bude pouze jediný vítěz. Ten může odeslat svůj real-time rámec bez rizika kolize.

Po úspěšném odvysílání real-time rámce stanice naplánuje vysílání dalšího rámce daného real-time přenosu tak, aby byl mezi začátky jednotlivých vysílání pevný časový interval. Tento interval je shodný pro všechny stanice. To má za následek, že se vysílání jednotlivých real-time toků průběžně synchronizuje. Neporuší-li se synchronizace, real-time přenosy využívají médium způsobem podobným TDMA, bez kolizí a s velmi krátkými BB. K občasnému porušení synchronizace může dojít kvůli provozu s nižší prioritou.

2.4.3. QoS směrování

Best-effort směrování zajišťuje pouhou konektivitu. Úkolem QoS směrování je vyhledávat a používat takové cesty, které (nejlépe) vyhovují QoS požadavkům na end-to end spojení (a případně také ve spolupráci s QoS signalizací podél těchto cest zprostředkovat rezervace prostředků). Tento úkol implicitně zahrnuje i předcházení zahlcení (congestion control) a snahu o maximálně efektivní využití síťových prostředků. Je vhodné zde podotknout, že velká část QoS architektur nevyžaduje použití směrovacího protokolu s explicitní podporou QoS a že QoS směrování obvykle zajišťuje i best-effort službu.

Při hledání QoS cest lze uvažovat různé parametry a jejich kombinace. Sledovaná QoS charakteristika přitom může podle stanovených vah kombinovat víc parametrů (např. maximální latenci a jitter). Obtížný úkol nastává, pokud jsou explicitně kladeny požadavky na více nezávislých parametrů. Jak tvrdí Wang a Crowcroft [5], již pro dva nezávislé QoS požadavky s aditivní či multiplikativní metrikou je hledání příslušných cest NP-úplný problém (např. hledání nejlevnější cesty s minimálním zpožděním). Proto se používají aproximační algoritmy. Z hlediska vlastností výsledných cest přitom mohou být různé QoS požadavky v příkrém rozporu. Např. kratší cesty (co do počtu přeskoků) zpravidla nabízejí nižší přenosové zpoždění. Jde-li ale o maximální energetickou úspornost (tzv. Power-Aware Routing) a dovolují-li mobilní uzly řízení vysílacího výkonu, jsou za určitých předpokladů cesty s více kratšími (fyzicky) přeskoky výhodnější.

Předností může být, pokud si směrovací protokol dokáže předpřipravit jednu či více záložních cest – v případě výpadku nebo snížení kvality hlavní cesty se minimalizuje prodleva opravy. Některé protokoly (např. [57]) dokonce používají více takovýchto konkurenčních cest paralelně a uvažují při tom jejich kolektivní QoS. Tento přístup (Multipath Routing) může být přínosem v situacích, kdy žádná cesta sama o sobě QoS požadavky nesplňuje (jde zde především o propustnost). Na druhou stranu může multipath routing přinášet vysoký rozptyl zpoždění a časté doručování paketů mimo pořadí.

Směrovací protokoly fixních sítí jsou založené na předpokladu, že je síť většinu doby v konvergovaném stavu a že přechod z jednoho konvergovaného stavu do druhého je v podstatě výjimečnou záležitostí. Případná mobilita uzlů na "okrajích" sítě se přitom např. v Mobile-IP [46] řeší přesměrováním provozu. V sítích MANET je situace přesně opačná – mobilita je zde normální a musí se řešit přímo směrováním. Směrovací protokol přitom zpravidla konverguje k optimálním cestám pouze za stabilních podmínek.

Aktualizace směrovacích informací mohou znamenat kvůli častým změnám topologie sítí MANET u proaktivních protokolů nepřiměřenou režii. Proaktivní protokoly totiž udržují cesty do všech uzlů, včetně těch, do kterých aktuálně nesměřuje žádný provoz. Reagují tedy i na takové změny topologie, které nemohou ovlivnit žádný přenos. Aby mohly poskytovat platné cesty, musí periodicky přenášet aktualizační informace. Frekvence výměny aktualizačních informací mezi uzly přitom musí odpovídat dynamice sítě (tj. v sítích, kde se předpokládá vyšší mobilita, je i režie vyšší). Ze zmíněných důvodů se obecně vzato v MANETech preferují reaktivní (neboli On-Demand) protokoly – ty hledají cesty až v okamžiku, kdy jsou potřeba. Potenciální nevýhodou toho, že pak uzly kontinuálně nepředpřipravují nové cesty, je možné větší zdržení při navazování a opravách cest.

QoS směrovacích protokolů existuje široká škála. Pro ilustraci zde budou uvedeny pouze dva příklady QoS směrování v sítích MANET. První příklad ilustruje možný způsob integrace podpory QoS do směrovacích protokolů (zde konkrétně do AODV), které vyhledávají cesty pomocí záplavového šíření a ve svém návrhu s podporou kvality služeb původně nepočítaly. Druhým příkladem je protokol CEDAR, který ilustruje koncept dynamické virtuální infrastruktury (zde konkrétně jde o tzv. jádro sítě), která robustním a efektivním způsobem podporuje propagaci směrovacích informací sítí.

QoS rozšíření AODV

AODV (Ad-hoc On-demand Distance Vector) [58] je reaktivní protokol typu distance vector navržený pro mobilní uzly v ad-hoc sítích. Použitou metrikou je počet přeskoků (Hop Count) – protokol vždy volí jednu z nejkratších cest. AODV používá sekvenční čísla stejným způsobem jako protokol DSDV (Destination Sequenced Distance Vector) [59], ze kterého vychází. Zajišťuje tak, že se cesty stanovují na základě nejčerstvějších dostupných informací a že žádná cesta neobsahuje cyklus.

Pokud chce uzel poslat zprávu uzlu, který není jeho bezprostředním sousedem a k němuž nezná platnou cestu, vyvolá mechanismus Path Discovery. Broadcastuje zprávu RREQ (Route Request), kterou sousedé iterativně stejným způsobem předávají dál – až do cílového uzlu nebo do uzlu, který zná dostatečně aktuální cestu. Zprávy RREQ obsahují identifikaci zdroje i cíle, životnost požadavku, minimální sekvenční číslo cíle a jednoznačný identifikátor zprávy (vztažený ke zdroji). Každý uzel kešuje identifikátory zpráv RREQ, které již zpracoval. Opakované přijetí stejného RREQ se ignoruje (předpokládá se, že taková zpráva přišla "oklikou", po zbytečně dlouhé cestě). Zprávy RREQ obsahují seznam uzlů od zdroje do aktuálního uzlu – zde AODV vychází z DSR (Dynamic Source Routing) [40]. Každý uzel si při zpracování RREQ aktualizuje ve své směrovací tabulce nejkratší cesty do všech uzlů, jimiž daný RREQ prošel. Uzel, který obdrží RREQ a zná cestu do cíle (např. sám cílový uzel), pošle zpět zdroji zprávu RREP (Route Reply). Pro RREP se už používá unicast přenos (cesta do zdroje RREQ je už nutně známa). RREP obsahuje seznam uzlů podél celé cesty. Uzly, které RREP přenášejí, si na základě tohoto seznamu aktualizují nejkratší cesty do všech uzlů směrem k cíli podél uvedené cesty. To nakonec udělá i uzel, který jako první vyslal RREQ – dozví se tedy požadovanou cestu a Path Discovery končí. AODV nepoužívá source routing. I v případě, že má uzel k dispozici informace o celé cestě, si do směrovací tabulky poznamenává pouze délku cesty, její stáří a první přeskok.

Přerušení cest se oznamuje prostřednictvím zpráv RERR (Route Error) – obsahují seznam cílů, do kterých uzel ztratil konektivitu. V případě, že některý uzel přijme datový paket, který má být směrován dál, ale cestu do cíle nezná, vyšle RERR. Skutečným problémem zde přitom není to, že cestu nezná, ale fakt, že je některým jiným uzlem považován za vhodného prostředníka. Uzly, které přijmou RERR, případně zneplatní příslušné položky směrovací tabulky. Pokud tím sami přijdou o konektivity do některého uzlu, vysláním nového RERR to oznámí dál. RERR se vysílá i v případě, že uzel detekuje přerušení spoje k některému ze svých sousedů.

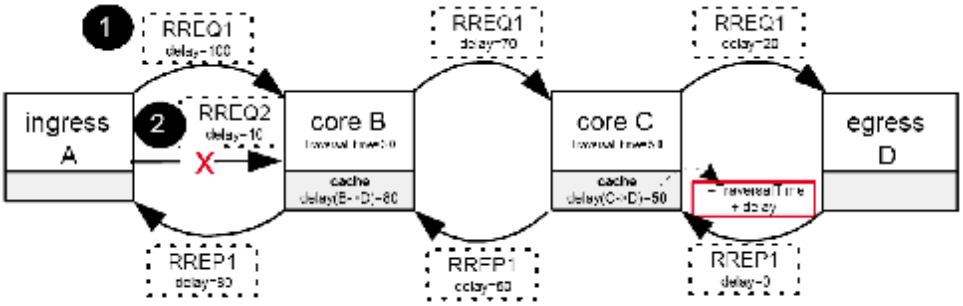
Hlavní myšlenka doplnění AODV o podporu QoS spočívá v rozšíření zpráv RREQ a RREP. Toto rozšíření specifikuje podmínky, za kterých uzly účastnící se Path Discovery předávají zprávy RREQ a RREP dál. QoS rozšíření přidává do zpráv RREQ a RREP dvě nové položky – *Delay* a *Bandwidth*, tj. požadavky na zpoždění a propustnost. Uzel, který obdrží některou ze zpráv s QoS rozšířením, musí být schopen tyto požadavky splnit. Není-li toho schopen, nepředává zprávu dál (tj. nebroadcastuje RREQ, resp. nepřeposílá RREP ke zdroji).

Pokud po sestavení QoS cesty její libovolný uzel zjistí, že nadále není schopen požadované QoS parametry plnit, vyšle dotčeným uzlům zprávu ICMP QOS_LOST.

Směrovací tabulka základního AODV obsahuje pro každý cílový uzel následující položky: *Destination Sequence Number*, *Interface*, *Hop Count*, *Next Hop* a *List of Precursors*. Popisované QoS rozšíření vyžaduje doplnění o další čtyři položky pro každou cestu: *Maximum Delay*, *Minimum Available Bandwidth*, *List of Sources Requesting Delay Guarantees* a *List of Sources Requesting Bandwidth Guarantees*.

Hodnota *Delay* v RREQ specifikuje maximální přípustné přenosové zpoždění do cílového uzlu (od zdroje nebo od konkrétního mezilehlého uzlu na cestě do cílového uzlu, který zrovna RREQ zpracovává). Při příjmu RREQ každý uzel odečte z hodnoty *Delay* čas potřebný pro zpracování paketu tímto uzlem (hodnotu *Traversal_time*). Jestliže zůstane hodnota *Delay* kladná (tj. zbývá ještě nějaký čas), RREQ je broadcastován dál. Cílový uzel na RREQ odpoví zprávou RREP, ve které nastaví *Delay* na 0. K hodnotě *Delay* v RREP se v každém uzlu přičte *Traversal_time* a výsledek (tj. hodnota zpoždění z daného uzlu do cíle) se kešuje v příslušné položce *Maximum Delay* směrovací tabulky. Při zpracování dalších zpráv RREQ se nejdříve "nahlédne" do této keše. RREP se následně unicastově přepošle dalšímu uzlu směrem ke zdroji, který se výsledně dozví celou cestu i očekávané zpoždění.

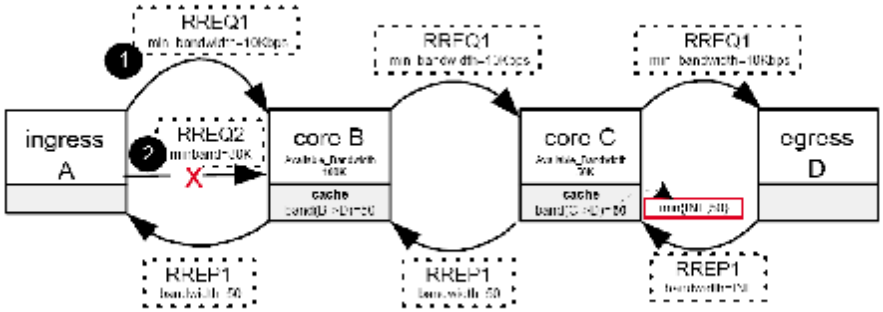
Jednoduchý příklad, jakým způsobem probíhá hledání cest splňujících omezení na maximální přenosové zpoždění, je uveden na obr. 3. První požadavek (RREQ1) zdrojového uzlu A uvádí maximální hodnotu zpoždění 100 ms. Mezilehlé uzly B a C jej přenesou až k cílovému uzlu D. Ten odpoví vysláním RREP1 zpět. Maximálnímu zpoždění 10 ms v druhém požadavku (RREQ2) nedokáže uzel B vyhovět (zpoždění z B do D je 80 ms), proto jej ignoruje. V případě, že se např. zvýší zatížení uzlu C (potažmo i příslušná hodnota Traversal_time), vyšle C zprávu ICMP QOS_LOST všem potenciálně dotčeným uzlům (tj. A a B). Seznam těchto uzlů se zaznamenává do *List of Sources Requesting Delay Guarantees*.



Obrázek č. 3. Mechanismus RREQ/RREP - hledání cest splňujících omezení na maximální přenosové zpoždění.

Zcela analogicky k hodnotě *Delay* se v RREQ a RREP používá i hodnota *Bandwidth*. Na mezilehlých uzlech se ale *Bandwidth* neupravuje – pouze se ověří, že má daný uzel k dispozici dostatečné množství volného přenosového pásma (tj. že je hodnota *Available_bandwith* větší než *Bandwidth* v RREQ). RREP, kterým odpovídá cílový uzel, definuje výchozí hodnotu *Bandwith* jako "nekonečno". Při přenosu RREP se nová hodnota *Bandwith* definuje jako minimum z původní hodnoty a *Available_bandwith*. Zdrojový uzel se tak nakonec dozví očekávanou propustnost cesty. Při snížení kapacity spoje se informují uzly

z *List of Sources Requesting Bandwidth Guarantees*. Na obr. 4 je jednoduchý příklad – požadavku RREQ1 na propustnost cesty 10 Kbps je vyhověno, požadavku RREQ2 na 80Kbps už ne.



Obrázek č. 4. Mechanismus RREQ/RREP - hledání cest splňujících omezení na minimální propustnost.

CEDAR

Sinha a kol. navrhli robustní a vysoce adaptivní protokol CEDAR (Core-Extraction Distributed Ad-hoc Routing) [60] pro QoS směrování v mobilních ad-hoc sítích. CEDAR předpokládá linkovou vrstvu schopnou odhadnout volnou kapacitu spoje a protokol přístupu k médiu typu CSMA/CA (používání dialogu RTS/CTS). Sledovaným parametrem kvality služby je propustnost end-to-end spojení.

CEDAR vytváří z dynamicky vybíraných uzlů tzv. jádro sítě (Core of The Network). V rámci něho inkrementálně propaguje link-state informace o všech stabilních spojích s vysokou propustností. Jádro sítě slouží jako směrovací infrastruktura, v níž na základě žádosti uzlů probíhají výpočty cest (s použitím pouze lokálních informací). Propagace směrovacích informací se omezuje pouze na informace o stabilních rychlých spojích a pouze na uzly jádra. Brání se tak nadměrné režii, kterou mohou v dynamickém prostředí sítí MANET trpět klasické link-state protokoly. Přestože v žádném uzlu neexistují detailní informace o stavu celé sítě, za stabilních podmínek dosahuje CEDAR přibližně stejného výkonu jako link-state směrování.

Protokol CEDAR se skládá ze tří hlavních komponent, jejichž význam a fungování jsou popsány dále. Stejně jako ostatní protokoly typu link-state je CEDAR vhodný pro malé a středně velké sítě (desítky až několik stovek uzlů). Pro velké ad-hoc sítě je vhodnější hierarchické klastrování, při kterém může CEDAR posloužit jako směrovací mechanismus na jednotlivých úrovních. Existuje rovněž rozšíření MCEDAR [62] obsahující podporu pro multicast vysílání.

Vytvoření a údržba jádra sítě

V ideálním případě tvoří jádro minimální vrcholové pokrytí sítě (MDS, Minimum Dominating Set). Cílem totiž je, aby každý uzel sítě buď přímo ležel v jádru, nebo měl na některý uzel jádra přímý spoj a aby jádro sítě by co nejmenší. Hledání MDS je obtížně aproximovatelný NP-těžký problém. Známé efektivní (tj. polynomiální) distribuované aproximační algoritmy (např. hladový algoritmus [63]) vyžadují globální výpočty. Proto

CEDAR používá velmi jednoduchý a robustní algoritmus, který se opírá pouze o lokální informace a přitom v průměrném případě generuje poměrně dobrou aproximaci [60].

Každý uzel sítě ležící mimo jádro si vybere jednoho svého souseda, který je součástí jádra (vždy existuje alespoň jeden takový soused) a označí ho za svůj dominátor. Uzly jádra sítě označí za svůj dominátor sebe sama. Při žádosti o výpočet cesty se uzly vždy obrací na svůj dominátor. V případě ztráty spojení mezi uzlem a jeho dominátorem postižený uzel buď vyhledá nový dominátor mezi ostatními sousedy, nebo požádá některého ze svých sousedů, aby se připojil k jádru sítě a stal se novým dominátorem. Třetí možností je, že se připojí k jádru sám [60].

Propagace link-state informací

Záplavové šíření může v sítích MANET vyvolávat broadcastové "bouře" a kvůli problémům skrytého a vystaveného uzlu bývá nespolehlivé a vysoce ztrátové [60, 61]. Link-state informace jsou navíc potřeba pouze v uzlech jádra sítě, protože pouze tam probíhají výpočty cest. CEDAR proto definuje mechanismus Core Broadcast (CB). Tento mechanismus propaguje informace jádrem sítě prostřednictvím spolehlivých virtuálních kanálů mezi sousedními uzly jádra (za sousední se považují každé dva uzly jádra, jejichž vzdálenost je nejvýše 3). Těsnou spoluprací s MAC vrstvou (pro pakety distribuované pomocí CB se kešují RTS a CTS rámce) CB výrazně zefektivňuje šíření link-state informací. Mechanismus CB se používá i při výpočtu cest.

Základním principem protokolu CEDAR je to, že informace o stabilních spojích s vysokou propustností se propagují do celé sítě (resp. celého jádra), zatímco informace o pomalých či méně stabilních spojích se udržují lokálně (nejsou totiž při vytváření QoS cest tak užitečné). Každý uzel, který o existenci spoje ví, může tento spoj potenciálně použít. O pomalé spoje tedy soutěží pouze omezený počet uzlů v jeho okolí, zatímco stabilní rychlé spoje jsou dány k dispozici co nejširšímu okolí. Tohoto stavu se dosahuje prostřednictvím tzv. rostoucích a klesajících vln (Increasing/Decreasing Waves).

Všechny uzly sítě monitorují volnou kapacitu svých linkových spojů. Kdykoli je u spoje *L* mezi uzly *U* a *V* překročena jistá prahová hodnota, uzly *U* a *V* požádají své dominátory, aby vyvolali core broadcast pro vzrůstající vlnu spoje *L*. Tato vlna jádrem sítě "roznese" informaci o tomto kvalitním spoji. Obdobně pokud kvalita spoje *L* poklesne pod jistou minimální hodnotu (příp. spoj úplně zanikne), *U* a *V* prostřednictvím svých dominátorů vyvolají klesající vlnu.

Vzrůstající vlny se šíří jádrem sítě výrazně pomaleji než klesající vlny. Klesající vlna navíc ruší účinek a další šíření předešlé vzrůstající vlny pro stejný spoj. Výsledně se informace o stabilních spojích rozšíří do vzdálenějších oblastí. Vzrůstající i klesající vlny se šíří pomocí mechanismu core broadcast. Šíření klesající vln se zastavuje v uzlech, do nichž ještě nedorazila žádná odpovídající vzrůstající vlna, nebo v nichž již předešlá klesající vlna odstranila link-state daného spoje. Nekvalitní spoje tedy žádným způsobem neovlivňují vzdálené oblasti sítě.

Výpočet QoS cest

CEDAR používá source-routing pro datové i řídící pakety. Zdrojový uzel *S*, který požaduje vytvoření spojení do cílového uzlu *D* o minimální propustnosti *B*, osloví svůj dominátor *Dom(S)* zprávou <*S, D, B*>. Pokud je toho *Dom(S)* na základě lokálních znalostí topologie schopen, sám stanoví cestu a vrátí ji uzlu *S*. V opačném případě *Dom(S)* vyvoláním mechanismu Core Broadcast (který je ve své podstatě prohledáváním do šířky) najde "přibližně nejkratší" cestu jádrem k uzlu *Dom(D)*. Tato cesta sice nemusí splňovat požadavky na propustnost, ale poskytuje dobré vodítko pro další postup. *Dom(S)* následně vybere cestu splňující QoS požadavky do nejvzdálenějšího možného uzlu na cestě jádrem sítě od *Dom(S)* k

Dom(D). Tento nejvzdálenější uzel pak osloví požadavkem, aby se při splnění QoS požadavků stejným způsobem pokusil iterativně "prodloužit" cestu až k cílovému uzlu *D*. Nebude-li toho oslovený uzel schopen, požadavek na vytvoření QoS cesty je zamítnut. V opačném případě vznikne výsledná cesta konkatenací jednotlivých úseků. Jako QoS parametr uvažuje CEDAR pouze propustnost. Ta zůstává konkatenací jednotlivých vyhovujících úseků garantována i pro celé end-to-end spojení.

Pro řešení výpadků spojů způsobených mobilitou uzlů nabízí CEDAR dva mechanismy. Výpadek spoje mezi uzly *K* a *L* oznámí zdrojovému uzlu *S* uzel *K* (nechť je to ten, který je k němu blíže). Jde-li o výpadek v blízkosti cílového uzlu *D*, uzel *K* vyhledá novou cestu do *D* a poté začne odpovídajícím způsobem přepisovat source-route informace ve všech paketech směřujících k *D*. Protože došlo k výpadku poblíž *D*, lze očekávat, že bude doba vyhledání nové cesty z *K* do *D* krátká. Pakety, které již jsou na cestě, tedy budou přesměrovány bez výraznějšího přerušení. Výpadek poblíž zdroje řeší sám uzel *S*. Při jeho detekci dočasně pozastaví vysílání a vyhledá novou cestu. I v tomto případě lze očekávat nízký počet nedoručených paketů. Výpadek v blízkosti zdroje je totiž detekován poměrně brzy a tedy počet paketů, které *S* vyšle od okamžiku výpadku do okamžiku jeho detekce, bude nízký.

2.4.4. Transportní vrstva

Na transportní vrstvě se real-time aplikace mohou opírat např. o protokol RTP (Real-time Transfer Protocol) [64]. RTP definuje standardizovaný formát paketů pro přenos audia a videa. Sám o sobě RTP nezprostředkovává rezervace prostředků ani žádné garance QoS, ale zajišťuje identifikaci typu přenášených dat (tj.použitého kodeku), sekvenční číslování a časová razítka (tj. podporu pro jitter buffering) a ve spolupráci s protokolem RTCP (RTP Control Protocol) [65] QoS feedback. RTCP poskytuje out-of-band řídící informace o hlavním RTP přenosu, samotného přenosu real-time dat se ale neúčastní. Jeho smyslem je zajišťovat monitoring kvality služby aktivního RTP přenosu. Shromažďuje údaje např. o počtu vyslaných bytů a paketů, počtu ztracených paketů, rozptylu zpoždění, době obrátky (Round-Trip Time) apod. Aplikace pak tyto informace typicky používají pro omezení nebo naopak rozšíření datového toku. RTP i RTCP jsou založeny na UDP a podporují unicast i multicast přenosy.

Je potřeba zdůraznit, že samotné protokoly TCP a UDP vůbec nevychází vstříc potřebám na zajištění QoS. Přestože multimediální přenosy a další real-time aplikace téměř bez výjimky místo protokolu TCP používají UDP, některé jeho nadstavby nebo jiné protokoly, na TCP je založeno množství existujících aplikací. Kvalita těchto služeb v bezdrátovém prostředí nás také může zajímat (např. propustnost FTP přenosů). Proto je vhodné zde uvést problémy, které má TCP v bezdrátových sítích a možné přístupy, jak lze v tomto prostředí QoS u TCP spojení zvýšit.

Protokol TCP nebyl navržen s ohledem na bezdrátové sítě. Předpokládá totiž, že ztráty paketů jsou způsobeny zahlcením sítě. To v prostředí bezdrátových sítí ale obecně neplatí – nejvíce paketů se zde ztrácí kvůli nespolehlivému přenosu (kolize/interference na bezdrátovém kanálu, úplné přerušení spoje atd.). Takové ztráty TCP vyhodnotí mylně a reaguje neadekvátním způsobem – exponenciálním zkracováním vysílacího okénka. Zcela bezdůvodně se tak snižuje propustnost spojení. Tento efekt se nejvíc projevuje u spojení s více bezdrátovými přeskoky (výpadky jsou pravděpodobnější), což ještě podporuje neférovou výhodu krátkých TCP spojení nad dlouhými [66].

Techniky, které umí těmto problémům čelit, lze rozlišovat na lokální a globální – podle toho, které síťové komponenty musí změnit svoje chování. Příkladem globální techniky může být Multiple Acknowledgements [67], používající dva typy potvrzení k rozlišení ztrát v drátové a bezdrátové části sítě. Mimo klasické potvrzení z TCP zavádí tzv. částečné potvrzení (partial acknowledgement). Sekvenční číslo S obsažené v částečném potvrzení informuje odesílatele o tom, že segment paket s číslem S-1 byl doručen do základnové stanice. Odesílatel může na různá potvrzení reagovat různě. Další globální technikou je Control Connection [68] – vytváří se dodatečné spojení, které používá stejnou cestu, ale končí na základnové stanici (zdroj spojení se zde předpokládá v drátové části sítě). Tímto spojením se periodicky posílají zprávy a detekují se případné ztráty (tj. zahlcení v drátové části). To pomáhá odhadnout důvod ztráty paketu v hlavním spojení. Zmíněné techniky ale vyžadují změny i mimo bezdrátovou část sítě, což představuje problém. Ačkoli takováto globální řešení lze teoreticky zavádět postupně, na to, že přinášejí výhodu jen pro bezdrátová spojení, představují neúměrnou režii. Vhodnější jsou tedy lokální techniky, které si vystačí s modifikacemi v bezdrátových komponentách sítě a přitom mohou komunikovat s existujícími verzemi TCP v drátové části sítě.

Cílem takových technik může být buď maskovat některé ztráty paketů před vyššími vrstvami tak, aby protokol TCP nevyžadoval změny, nebo detekovat a hlásit protokolu TCP důvody ztráty paketu. Protokoly založené čistě na linkové úrovni odpovídají filozofii maskování ztrát. Jejich hlavní myšlenkou je snaha problém řešit lokálně (neboť se zdá být lokálním). Např. použitím automatických retransmisí (ARQ) lze zajistit spolehlivost na linkové úrovni. Předpokad, že se tím skryje ztrátový charakter bezdrátového spojení před vyššími vrstvami ale nemusí být správný. Ve skutečnosti může opakovaný přenos nevhodnou interakcí s mechanismem Fast Retransmit [69] protokol TCP zmást [70]. Této nevhodné interakci zabraňuje např. algoritmus DDA (Delayed Duplicate Acknowledgements) [71] na cílovém uzlu. Pozdržením třetího a následných duplicitních potvrzení po dobu potřebnou pro opakovaný přenos rámce se předchází předčasnému vyvolání mechanismu Fast Retransmit, tj. opakování end-to-end přenosu.

Některá rozšíření TCP se snaží zužitkovat schopnosti základnových stanic a jsou tak vhodná jen pro single-hop přístupové sítě (např. zmíněné Multiple Acknowledgements a Control Connection). Sítě MANET bere v úvahu např. ATCP (Ad-hoc TCP) [72]. ATCP představuje novou tenkou vrstvu mezi IP a TCP, která se spoléhá na explicitní zpětnou vazbu ze sítě – na zprávy ECN (Explicit Congestion Notification) [73] a na ICMP zprávy "destination unreachable". Díky nim rozlišuje příčiny vzniklých problémů. Přijetí zprávy ECN informující o nastupujícím zahlcení v síti způsobí okamžité vyvolání ochrany proti zahlcení (bez zbytečného čekání na vypršení časovačů). Ztráty způsobené chybovostí bezdrátového kanálu ATCP naopak řeší opakovaným přenosem, přičemž se nezmenšuje vysílacího okénko. Úplný výpadek spojení "zmrazí" TCP odesílatele tak, že dokud není známa nová cesta k adresátovi, neodesílají se do sítě žádné pakety. Při znovunavázání spojení podél nové cesty se přitom přepočítá velikosti vysílacího okénka tak, aby se předešlo možnému zahlcení po obnovení přenosu. ATCP bere v úvahu i situace, kdy je mnoho paketů doručováno mimo pořadí (snaží se je před TCP vrstvou maskovat vlastním přeuspořádáváním).

3. Simulační model MeshQoS

Následující části práce jsou zaměřené prakticky. Jejich cílem je studium vlivů mechanismů linkové a síťové vrstvy na dosažitelné kvalitativní vlastnosti end-to-end spojení v mesh sítích. Problém je úžeji vymezen na sítě s bezdrátovou infrastrukturou založené na technologii WiFi (resp. konkrétním rozšíření přístupové metody CSMA/CA a standardu IEEE802.11b) a použití standardizovaného směrovacího protokolu AODV a nově navrženého oportunistického směrování OMR. Pro tento účel byl navržen diskrétní simulační model MeshQoS. Ten je založen na bázi simulačního systému OMNeT++ a potřebné možnosti nabízí. V rámci MeshQoS bylo na konkrétních scénářích provedeno několik experimentů, jejichž výsledky jsou prezentovány a analyzovány v další kapitole.

3.1. Systém pro diskrétní simulace OMNeT++

Pro simulace bylo zvoleno prostředí OMNeT++ (Objective Modular Network Testbed in C++) [74]. OMNeT++ je výkonný modulární, objektově orientovaný simulátor diskrétních událostí. Byl navržen především pro simulace komunikačních protokolů a validaci návrhu a vyhodnocování výkonnosti multiprocesových distribuovaných systémů.

Simulační model se v systému OMNeT++ popisuje hierarchickým skládáním jednoduchých modulů. Úroveň vnoření přitom není omezena. To usnadňuje implementaci logické struktury složitých systémů, které obvykle obsahují mnoho úrovní abstrakce. Moduly nejnižší úrovně (jednoduché moduly, Simple Modules) zapouzdřují konkrétní funkcionalitu. Tu implementuje programátor v metodách příslušné C++ třídy (za použití simulační knihovny, kterou OMNeT++ obsahuje). Z jednoduchých modulů se sestavují složené moduly (Compoud Modules). Nejvyšší úroveň hierarchie (systémový modul Network) obsahuje celý simulační model. Přesná topologie této hierarchie (tj. skládání a propojování modulů) se definuje prostřednictvím jazyka NED (Network Description). OMNeT++ pro tento účel nabízí i vizuální editor GNED (Graphical Network Description).

Moduly mohou definovat sadu parametrů, která upravuje jejich chování a parametrizuje topologii celého modelu. Komunikují mezi sebou pomocí předávání zpráv. Zprávy mohou obsahovat libovolně složité datové struktury. Modul přitom může zaslat zprávu jinému modulu buď přímo, nebo skrze předdefinované cesty. Tyto cesty se definují propojováním bran jednotlivých modulů prostřednictvím kanálů.

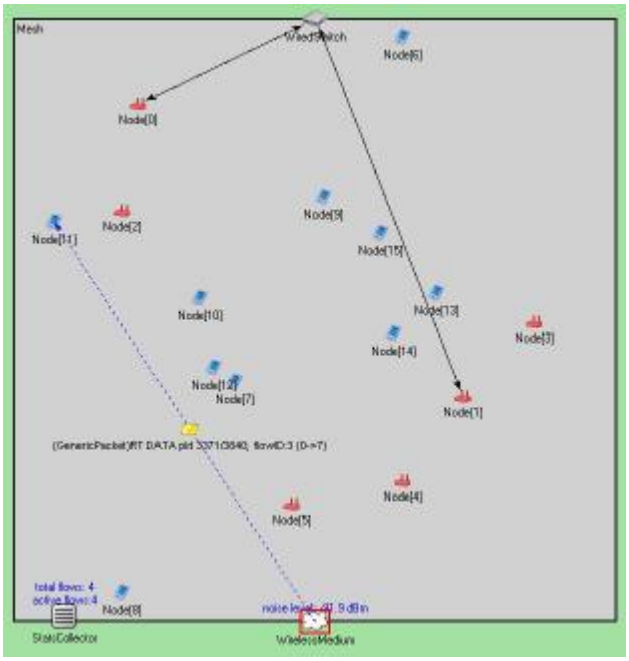
Pro různé způsoby použití výsledných simulačních programů nabízí OMNeT++ dvě uživatelská rozhraní. Grafické rozhraní TkEnv je určeno pro ladění a prezentace simulací. Uživatel může získat detailní vizuální přehled o tom, jak model pracuje a co se v simulaci zrovna děje, řídit průběh simulace a ovlivňovat jej změnou parametrů uvnitř modelu. Druhé rozhraní (CmdEnv) je pak určeno pro dávkové zpracování více simulačních "běhů". Výsledkem běhu simulačního programu je sada skalárních hodnot a vektorů (tj. průběhů hodnot sledovaných parametrů v simulačním čase).

OMNeT++ je open-source systém implementovaný v C++. Jeho simulační jádro i obě uživatelská rozhraní jsou přenositelná (platformově mezi Windows a některými klony Unixových systémů i mezi různými překladači C++). OMNeT++ poskytuje nástroje pro zpracování výsledků simulací a mj. také podporuje paralelní distribuované simulace a real-time simulace. Má dobře definované a dokumentované rozhraní simulační knihovny i jazyka NED. Vývoj OMNeT++ stále aktivně probíhá.

3.2. Popis modelu MeshQoS

Pro prostředí OMNeT++ již existuje několik volně dostupných modelů určených pro simulace ad-hoc sítí. Jsou však pro účel této práce buď příliš zjednodušující (např. adHocSim [74]), nebo naopak příliš komplexní a obtížně modifikovatelné (např. Mobility FrameWork [74]). Proto byl navržen, implementován a odladěn simulační systém MeshQoS. MeshQoS lze používat přímo "tak jak je", mimo simulačního jádra a knihovny OMNeT++ nevyžaduje žádné další komponenty. Implementací nové třídy jej přitom lze snadno rozšířit o další směrovací protokol.

Na obr. 5 je znázorněn příklad, jak může vypadat síť Mesh (nejvyšší úroveň simulačního modelu). Mesh se skládá z modulů několika typů. Jsou to jednak samotné uzly sítě v poli Node (modře jsou znázorněny mobilní uzly, červeně fixní uzly – bezdrátové směrovače a přístupové body). Přístupové body jsou drátově připojeny na centrální přepínač WiredSwitch. Dále MeshQoS obsahuje modul WirelessMedium, který zprostředkovává vlastní bezdrátové vysílání a modul StatsCollector, který shromažďuje většinu statistik o průběhu simulace. Chování MeshQoS lze přesně definovat velkým množstvím parametrů. Cílem této kapitoly přitom není podat vyčerpávající dokumentaci simulačního modelu, ale pouze přiblížit jeho možnosti a uvést základní principy, na kterých je založen.

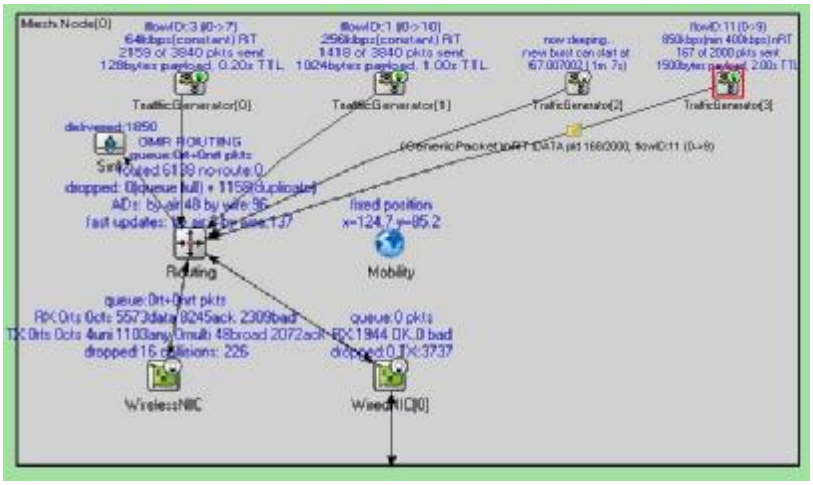


Obrázek č. 5. Simulační model MeshQoS v grafickém rozhraní TkEnv.

Počet uzlů jednotlivých typů se definuje prostřednictvím parametrů AccessPoints, WirelessRouters a NomadicUsers. Plocha, po které se mohou mobilní uzly pohybovat, se vymezí pomocí parametrů SizeX a SizeY. Úroveň mobility všech uzlů lze globálně upravovat parametrem SpeedFactor.

Vnitřní struktura uzlů sítě se skládá z několika komponent, které odpovídají komunikační logice. Na obr. 6. je znázorněn příklad pro nejobecnější typ uzlu – přístupový bod. Veškeré přenosy vznikají v generátorech provozu. Parametr TrafficGenerators určuje jejich počet v daném uzlu. Jednotlivé generátory v rámci uzlu přitom lze konfigurovat nezávisle. Centrální komponentou je modul směrování. Použitý směrovací protokol lze definovat parametrem

Routing. MeshQoS přitom umožňuje, aby různé uzly používaly různé protokoly. Sink je modul, v němž terminuje provoz, tj. společně s konkrétním jedním nebo více moduly TrafficGenerator reprezentuje aplikaci. Každý uzel obsahuje modul WirelessNIC, který představuje bezdrátové rozhraní WiFi sítě a modul WiredNIC představující rozhraní do Fast Ethernetové sítě. Poněkud "mimo" zmíněných modulů leží modul Mobility, který definuje polohu a (případný) pohyb uzlu.



Obrázek č. 6. Vnitřní struktura přístupového bodu v modelu MeshQoS.

Základní struktura ostatních typů uzlů je implementačně shodná. Rozdílného chování tří základních typů uzlů je dosaženo pouze prostřednictvím konfigurace jejich parametrů. Např. na bezdrátových směrovačích nevzniká žádný provoz, proto se počet generátorů provozu (parametr TrafficGenerators) nastaví na 0. Bezdrátové směrovače ani mobilní uzly nemají drátové rozhraní, proto vůbec neobsahují modul WiredNIC. O mobilitě či fixní pozici uzlu rozhoduje konfigurace modulu Mobility. Simulační model MeshQoS se tedy neomezuje pouze na typické mesh sítě. Lze pomocí něj simulovat i ad-hoc sítě (jakožto speciální případy mesh sítí) nebo např. definovat mobilní přístupové body, kde drátové rozhraní simuluje "rychlé" připojení jinou technologií než WiFi (např. satelitní připojení).

MeshQoS rozlišuje real-time (RT) a non-real-time (NRT) přenosy. Na linkové úrovni v modulu WirelessNIC umožňuje definovat priority těchto dvou druhů provozu. Pakety směrovacího protokolu jsou označeny jako RT a mají přednost před všemi ostatními (i RT) pakety. Každý paket mimo příznaku RT/NRT obsahuje časovou značku, kdy byl vygenerován a hodnotu TTL (Time To Live), která určuje jeho životnost. Na linkové úrovni se kontroluje stáří paketů. Pakety starší než TTL se automaticky zahazují. Používá se zde tedy původní význam "TTL", tj. ve smyslu času a ne dnes obvyklejšího maximálního počtu přeskoků (reálná implementace přitom nevyžaduje časovou synchronizaci uzlů).

Následuje popis fungování a možností konfigurace nejdůležitějších modulů.

Mobilita uzlů

Modul Mobility definuje okamžitou polohu uzlu a charakter jeho pohybu. Implementovány jsou dvě schémata pohybu uzlů – náhodná procházka (Random Waypoint) a sledování jiného uzlu (Pursuit). V modelu náhodné procházky se uzel začíná pohybovat ve směru InitialDirection rychlostí DefaultSpeed a klasicky se "odráží" od hranic simulované oblasti.

Navíc lze použít dva "variátory" směru pohybu. Ty definují, s jakou pravděpodobností dojde ke změně směru pohybu a jak významná změna to může být. V případě sledování jiného uzlu (kterého konkrétně lze definovat parametrem PursuitNode) se uzel pohybuje rychlostí DefaultSpeed směrem ke sledovanému uzlu. Aby se mohlo zabránit nepřirozeně hustému soustředění uzlů (když je např. pomalejší uzel sledován rychlejším), parametrem PursuitDistanceLimit lze určit limitní vzdálenost ke sledovanému uzlu. Pokud je sledující uzel blíže sledovanému, než je tato vzdálenost, jeho rychlost se násobí faktorem PursuitSlowDownFactor pokaždé, když sledující urazí 1 metr.

Bezdrátové médium

Šíření signálů bezdrátovým médiem simuluje modul WirelessMedium. Předpokládá se, že všechny uzly mají všesměrové antény (resp. antény, které za všech okolností směrově pokrývají všechny ostatní uzly) a pracují v pásmu daném parametrem Frequency. Základní konstantní úroveň šumu v tomto pásmu definuje parametr BackgroundNoiseLevel. Navíc lze prostřednictvím parametru AdditionalNoise definovat dodatečnou úroveň šumu, která se k základní úrovni přičítá. Parametr AdditionalNoiseDuration definuje dobu, po kterou si AdditionalNoise udržuje konstantní hodnotu (předpokládá se použití náhodných veličin).

Modul WirelessNIC simuluje všesměrové vysílání uzlů. Používá model šíření rádiových signálů popsaný v [75]. Úroveň signálu na vysílači je dána jeho vysílacím výkonem (TxPower). WirelessNIC určuje úroveň signálu na přijímači (RxPower) a odstup signál/šum (SNR) podle vztahů

$$\text{Attenuation} = -27.56 + 20*\log(\text{Frequency}) + 10*\text{DPLE}*\log(\text{Distance})$$

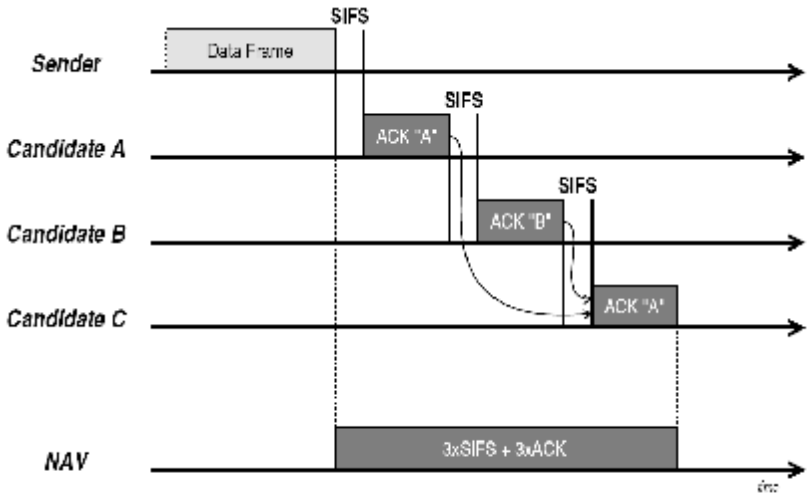
$$\text{RxPower} = \text{TxPower} - \text{Attenuation}$$

$$\text{SNR} = \text{RxPower} - (\text{BackgroundNoiseLevel} + \text{AdditionalNoise})$$

kde Attenuation vyjadřuje útlum při přenosu na vzdálenost Distance a DPLE (Distance Power Law Exponent) popisuje, jak rychle klesá úroveň signálu s vzdáleností od vysílače. Hodnota DPLE by měla reflektovat průměrnou "hustotu" prostředí, tj. množství překážek. Např. ve zcela volném prostoru je DPLE 2, pro prostředí hustě zastavěných centrech měst se doporučuje hodnota 3 až 4.

Rozhraní bezdrátové sítě

Modul WirelessNIC reprezentuje rozhraní do bezdrátové WiFi sítě. Simuluje chování fyzické a linkové vrstvy definované standardem IEEE 802.11b. Přístupová metoda je CSMA/CA EDCF s rozšířením o podporu anycastu a multicastu. Parametrem RTSTreshold lze určit minimální velikost paketů, pro které je použit mechanismus RTS/CTS. Parametry ShortRetryLimit a LongRetryLimit definují maximální počet opakování vysílání krátkých rámců (resp. RTS pro dlouhé rámce) a dlouhých rámců, tak jak je specifikováno v IEEE 802.11.



Obrázek č. 7. Rozšíření CSMA/CA o podporu anycastu a multicastu.

Podpora anycastu v MeshQoS vychází z návrhu prezentovaného v definici ExOR (Extremely Opportunistic Routing) [76]. Struktura datového rámce je rozšířena tak, že místo MAC adresy jediného příjemce je uveden seznam kandidátů (tj. jeho délka a příslušný počet MAC adres) sestupně podle priorit. Dialog "slotovaného" potvrzování anycast paketů na jednoduchém příkladu tří kandidátů A, B a C je na obr. 7. Jednotliví příjemci postupně podle pořadí adres v seznamu kandidátů vyšlou ve správný čas ACK. Všichni kandidáti, kteří se anycastu účastní, přitom "poslouchají" i ACK rámce ostatních kandidátů. Struktura rámce ACK je rozšířena o položku, která určuje číslo (pořadí) kandidáta s nejvyšší prioritou, jehož potvrzení daný uzel slyšel. To vytváří velmi robustní schéma, kdy se s velkou pravděpodobností (mj. proto, že ACK rámce jsou kratší než datové rámce, tj. více odolné vůči chybám) všichni kandidáti i samotný zdroj anycastu dozví, který "nejprioritnější" kandidát rámec přijal. Tento kandidát je vítězem anycastu, tj. skutečným příjemcem rámce. Ostatní kandidáti rámec zahodí.

MeshQoS navíc rozšiřuje popsaný mechanismus tak, že rámec ACK namísto čísla kandidáta s nejvyšší prioritou obsahuje bitovou mapu pro všechny kandidáty. Aniž by vzrostla režie pak tento mechanismus vylepšuje robustnost i multicast vysílání. U multicastu se postupuje stejně, pouze s tím, že priority kandidátů nejsou podstatné – každý kandidát, který rámce uslyší, "zvítězí" automaticky. Při anycastu ani multicastu se nepoužívá RTS/CTS, a to bez ohledu na velikost rámce. Retransmise se ale používají – u anycastů se vysílání opakuje, dokud nepotvrdí příjem alespoň jeden kandidát. U multicastu musí příjem potvrdit všichni kandidáti (při opakování vysílání se v tomto případě vynechají kandidáti, kteří již v předchozím pokusu vysílání potvrdili). Limitem počtu opakování opět zůstávají hodnoty ShortRetryLimit a LongRetryLimit. Režie zde popsaného mechanismu je v průměru cca 2% na každého dalšího kandidáta (pro pakety běžných délek). Je vhodné poznamenat, že popsaná rozšíření lze pouhou změnou firmwaru implementovat i ve stávajícím WiFi hardwaru.

Co se týká fyzické úrovně a zpracování signálu, prostřednictvím parametrů TxPower a RxSensitivity lze definovat vysílací výkon a citlivost při příjmu. Je-li úroveň signálu přijímaného rámce nižší než citlivost, rámec není rozpoznán (potenciálně ale vyvolá kolizi s jiným přijímaným rámcem, resp. po dobu vysílání znemožní příjem dalších rámců). Je-li úroveň signálu vyšší než citlivost, použije se tabulka hodnot závislosti mezi SNR a bitovou chybovostí (BER) empiricky zjištěná pro konkrétní chip 802.11b firmy Intersil zjištěná v rámci [77]. Lineární interpolací jejich hodnot se z odstupu signál/šum (SNR) přijímaného rámce zjistí BER pro jednotlivé modulace (BPSK, QPSK, CCK11), které 802.11b používá pro PLCP preambuli, PLCP hlavičku a vlastní MAC rámec. Jednoduchým vztahem

$$FER = 1 - (1 - BER)^{8*FrameLength}$$

se zjistí pravděpodobnost chyby při příjmu celého rámce (FER), resp. každé z jeho zmíněných tří částí. FrameLength je délka rámce, resp. jeho části v bytech.

Pakety čekající na odvysílání WirelessNIC bufferuje ve dvou frontách – zvlášť RT a NRT pakety. K těmto frontám se přistupuje s frekvencí úměrnou parametrům RTPriority a NRTPriority (odpovídá algoritmu WFQ). Každá z těchto front je navíc prioritní frontou – vždy se z ní vybírá ten paket, který je v daném okamžiku relativně nejstarší (vztaženo k jeho TTL). Výjimkou jsou pakety směrovacího protokolu, které mají přednost bez ohledu na stáří. Parametrem BufferSize lze určit souhrnnou velikost obou front. Pokud je jejich kapacita vyčerpána a vznikne potřeba zahodit některý paket, nejdříve je to nejstarší NRT paket, poté nejstarší z běžných RT paketů. Pakety směrovacího protokolu se zahazují až jako poslední.

WirelessNIC poskytuje modulu směrování zpětnou vazbu o úspěšnosti odeslání (resp. potvrzení) rámců, přičemž současně informuje o tom, kolikrát se vysílání opakovalo a (v případě anycastů a multicastů) o tom, kteří příjemci potvrdili přijetí. Modul směrování je navíc informován o tom, že WirelessNIC právě odvysílal poslední rámec a zcela tak vyprázdnil fronty. Podporován je i promiskuitní režim monitoringu média (RF_MON), kdy je modul směrování prostřednictvím speciálního kanálu informován o všech úspěšně přijatých rámcích (bez ohledu na zamýšleného příjemce).

Modul WirelessNIC (a vlastně celý simulační model MeshQoS) simuluje pouze základní, zde popsanou funkcionalitu – plánování, soutěž o médium a vlastní datové přenosy. Asociaci a autentizaci uzlů, přidělování a správu adres a mnohé další záležitosti, které nejsou z hlediska studia kvality služeb významné, neuvažuje.

Drátová část infrastruktury

Rozhraní drátové sítě (modul WiredNIC) obsahuje výstupní FIFO frontu o velikosti BufferSize kB. Měl by simulovat switchované připojení do spolehlivé vysokorychlostní sítě . Drátové médium má ve výchozí konfiguraci propustnost 100 Mbps, zpoždění 1 μs a bitovou chybovost 1e-8. Modul WiredSwitch implementuje chování Fast Ethernetového switche s výstupními FIFO frontami pro jednotlivé porty. Drátová část infrastruktury podporuje unicast a broadcast.

Generování provozu

TrafficGenerator produkuje pakety a v pravidelných intervalech je zasílá modulu směrování. Délka všech paketů daného generátoru je stejná, definovaná parametrem PacketPayloadLength. Frekvence generování paketů je taková, že produkovaný datový tok má konstantní rychlost TxRate. Parametr RealTime určuje typ paketů, které generátor produkuje a parametr TTL jejich životnost.

Pakety se generují v dávkách. Počet paketů v dávce určuje parametr BurstLength, příjemce dávky parametr BurstDestination a dobu mezi dávkami parametr BurstIATime. V okamžiku, kdy má začít nová dávka, ověřuje TrafficGenerator topologie sítě. Pokud je taková, že není možné zajistit ani minimální úroveň konektivity k cílovému uzlu, přenos vůbec nezačne (dávka je zrušena). Admission Control tedy provádí samotný TrafficGenerator – použitý směrovací protokol nemá vliv na to, jaké přenosy během simulace vznikají (srovnávání různých směrovacích protokolů je pak i na menším vzorku dat přesnější). Topologie se v daném okamžiku považuje za "příznivou" pro konkrétní přenos, pokud v ní existuje alespoň jedna cesta mezi zdrojovým a cílovým uzlem složená pouze z bezdrátových přeskoků kratších než ConnectivityRadius metrů a případně drátového přeskoku.

Nastavením příznaku Reciprocal lze vytvořit generátory, které generují provoz až na žádost některého z generátorů v jiném uzlu (takového, který má nastaven příznak InitiateReciprocal). Tímto způsobem lze simulovat přenosy, které vznikají v protisměrných párech (např. VoIP).

Parametr FlowControl určuje, zda daný generátor používá řízení toku či nikoli. Pokud ano, TrafficGenerator průběžně upravuje rychlost vysílání (resp. frekvenci generování paketů). V pravidelných intervalech (50ms) generátor ověřuje, zda od poslední kontroly nedošlo ke ztrátě některého z jím vygenerovaných paketů. Důvody případných ztrát se přitom nerozlišují. Nedošlo-li k žádné ztrátě, TrafficGenerator lineárně zrychluje (vždy o pevnou hodnotu 50 kbps). V opačném případě exponenciálně zpomaluje (vždy na polovinu). Simulované zpětné kontrolní spojení, které řízení toku v reálu zprostředkovává, produkuje pevně danou rychlostí 5 kbps pakety s délkou payloadu 32 bytů. Pro kontrolní spojení se používají stejné hodnoty TTL a RealTime jako u hlavního přenosu.

Popsaný mechanismus "zhruba" simuluje přenosy s řízením toku. Zpětné kontrolní spojení ve skutečnosti nepřenáší žádné informace o doručení nebo nedoručení paketů. O ztrátách se TrafficGenerator dozví na základě globální znalosti statistik. Cílem přitom nebylo přesně vystihnout chování TCP nebo velmi přesně zkoumat propustnost, ale mimo přenosů s konstantní rychlostí simulovat i přenosy, které vykazují charakteristický "pilovitý" průběh a dokážou "nasytit" síť.

Směrování

Modul Routing implementuje směrovací protokol. Součástí simulačního modelu je implementace protokolů AODV a OMR. Protokol AODV je implementován podle standardu definovaného v RFC 3561 [58] a revidovaného v IETF draftu <draft-perkins-manet-rfc3561bis-01.txt> [78]. Druhý implementovaný protokol – OMR je popsán v následující kapitole. Kvůli jednoduchosti jsou oba protokoly implementovány tak, že vlastní rozhodnutí o směru dalšího přenosu "nic netrvá". Implementované protokoly AODV a OMR přitom nejsou pravé "QoS" směrovací protokoly, jak je definuje a popisuje kapitola 2.4.3. Přestože je možné AODV o podporu QoS rozšířit, reálná implementace v prostředí založeném na výhradně soutěžním způsobu přístupu k médiu je velice obtížná. Až na prioritní fronty, které OMR "přesunuje" z linkové do síťové vrstvy ani OMR explicitně neuvažuje QoS charakteristiky používaných cest. Podpora QoS je v simulačním modelu MeshQoS soustředěna do linkové (a fyzické) vrstvy bezdrátového rozhraní WirelessNIC.

Sběr statistických dat

Simulační model umožňuje sledování velkého množství detailních statistik o průběhu simulace. Statistiky týkající se jednotlivých modulů, především směrování, drátového a bezdrátového rozhraní a bezdrátového média shromažďují přímo tyto moduly. Modul StatsCollector shromažďuje statistické údaje o všech probíhajících i dokončených datových přenosech. Přímo během simulace lze v uživatelském rozhraní TkEnv zobrazovat statistická data ve formě histogramů, časových průběhů i skalárních hodnot. Větší část získaných dat je na konci simulace uložena prostřednictvím modulu StatsCollector do výstupního skalárního souboru.

3.3. Směrovací protokol OMR

Nově navržený oportunistický směrovací protokol OMR (Opportunistic Mesh Routing) je založen na podobných ideách jako již zmíněný protokol ExOR (Extremely Opportunistic Routing) [76,79]. OMR však tyto myšlenky (především použití anycastu a metriky ETX) rozšiřuje a zaměřuje se speciálně na prostředí mesh sítí. OMR je hybridní protokol (tj. s proaktivními i reaktivními prvky) typu Distance Vector. Detailní popis protokolu OMR vybočuje z rámce i rozsahu této práce, proto se následující popis omezuje jen na základní uplatněné principy.

Klíčovou myšlenkou OMR je oportunistický přístup ke směrování. Většina unicastových směrovacích protokolů (např. AODV) podle nějaké metriky ohodnotí jednotlivé spoje (ať drátové či bezdrátové) a na základě toho zvolí "optimální" cestu (nejkratší, nejlevnější, příp. i s ohledem na QoS parametry). Podél zvolené cesty se poté přenášejí data od zdroje k cíli. Tento standardní přístup ale lze dobře zdůvodnit pouze ve fixních sítích, kde každý pár uzlů buď je nebo není propojen, tj. pouze přímo propojené uzly mohou komunikovat. Mesh sítě (resp. bezdrátové multihop sítě) se ale od fixních sítí odlišují ve třech podstatných věcech. Předně je principiálně možná přímá komunikace mezi libovolnými dvěma uzly sítě (třebaže potenciálně s vysokou chybovostí). Druhým rozdílem je to, že uzly nemusí předem určit konkrétního příjemce bezdrátového vysílání (na rozdíl od drátových směrovačů, které musí zvolit příslušný port) – na fyzické úrovni jsou totiž stejně všechna vysílání všesměrová (předpoklad celého modelu MeshQoS). Důležitý je také třetí rozdíl – bezdrátová komunikace je mnohem méně spolehlivá než drátová.

Unicastové směrovací protokoly se snaží (např. ve spolupráci s opakováním přenosu na linkové vrstvě) zmíněné charakteristické vlastnosti bezdrátových multihop sítí maskovat. OMR se naopak snaží jich využít. Místo toho, aby předem zvolil cestu do cíle, určuje OMR cestu na základě toho, které uzly úspěšně přijaly konkrétní vysílání (což závisí na přesných aktuálních vlastnostech bezdrátového kanálu). Při rozhodování o směru dalšího přenosu paketu tedy OMR nezvolí konkrétní uzel, ale celou skupinu vhodných kandidátů (s různými prioritami, tj. různě vhodné). Pokroku na cestě k cíli je přitom dosaženo pokud vysílání úspěšně rozpozná kterýkoli z kandidátů. Již dříve popsaný mechanismus anycastů s velkou pravděpodobností zaručí, že z těch uzlů, které paket přijaly jej přenáší dál pouze ten "nejvhodnější" (ten, který je nejblíže cíli).

Klíčovým problémem, který zásadně ovlivňuje chování protokolu, je samozřejmě rozhodování, do jaké míry je konkrétní sousední uzel vhodným kandidátem pro přenos paketu do cílového uzlu. OMR pro tento účel definuje a používá metriku ETXC.

Metrika ETXC

ETXC (Expected Trasmissions Cost) je "anycastovým" zobecněním metriky ETX (Expected Tranmissions Count) [80]. ETX je jednoduchá aditivní "radio-aware" metrika, která popisuje očekávaný počet vysílání nutných pro úspěšný přenos prostřednictvím konkrétního bezdrátového spoje, resp. cesty (tj. včetně případných retransmisí). ETX konkrétního spoje je definována vztahem

$$ETX = 1 / (FDR * RDR)$$

kde FDR (Forward Delivery Ratio) je změřená pravděpodobnost, že příjemce úspěšně rozpozná data a RDR (Reverse Delivery Ratio) pravděpodobnost, že v pořádku dorazí potvrzení (ACK). Způsob, jakým se FDR a RDR v protokolu OMR měří, je popsán dále. ETX cesty je součtem EXT jednotlivých přeskoků. Hlavní přínos ETX spočívá v tom, že penalizuje cesty, které používají méně spolehlivé spoje (na rozdíl např. od Min-Hop-Count metriky, která uvažuje pouze binární spolehlivost/nespolehlivost spojů).

Nová metrika ETXC (očekávaná "cena" přenosu) zobecňuje popsaný princip ve dvou směrech. Jednak reflektuje použití anycastu, ale také rozlišuje různé typy spojů. Lze definovat různou cenu (jednoho) bezdrátového a drátového vysílání. ETXC přitom nedefinuje cenu spoje nebo cesty, ale cenu uzlu (ve vztahu ke konkrétnímu uvažovanému cíli přenosu). ETXC(U,D) (hodnota metriky ETXC uzlu U pro cílový uzel D) je definována pomocí vztahů

$$ETXC(U,U) = 0$$

$$ETXC(U,D) = \min(ETXC_{wire}(U,D), ETXC_{air}(U,D))$$

$$ETXC_{wire}(U,D) = \min\{C_{wire} + ETXC(N,D) \mid N \in WiredNodes\}$$

$$ETXC_{air}(U,D) = \min\{ETXC_{aircand}(U,D,N_1,N_2,\ldots,N_k) \mid N_i \in Nodes, i \in \{1..k\}, k<=MaxCandidates\}$$

$$\begin{aligned} ETXC_{aircand}(U,D,N_1,N_2,\ldots,N_k) &= [FDR(U,N_1)*(C_{air}+ETXC(N_1,D))] \\ &+ [(1-FDR(U,N_1))*FDR(U,N_2)*(C_{air}+ETXC(N_2,D))] + \ldots \\ &\ldots + [(1-FDR(U,N_1))*(1-FDR(U,N_2))*\ldots \\ &\ldots*(1-FDR(U,N_{k-1}))*FDR(U,N_k)*(C_{air}+ETXC(N_k,D))] \end{aligned}$$

kde C_{wire} je cena drátového vysílání, WiredNodes množina všech přístupových bodů, C_{air} cena bezdrátového vysílání, Nodes množina všech uzlů sítě, MaxCandidates maximální počet kandidátů na další přenos paketu směrem k cílovému uzlu při anycastu a FDR (X,Y) pravděpodobnost, že Y úspěšně příjme vysílání uzlu X (tj. spolehlivost spoje $X \rightarrow Y$).

ETXC samotného cílového uzlu je tedy vždy 0. ETXC ostatních uzlů se stanovuje na základě jeho FDR k jednotlivým sousedům a ETXC těchto sousedů. Uzly, která mají přístup i k drátové infrastruktuře nejprve rozhodnou, prostřednictvím kterého média se budou pakety přenášet. Předpokládá se vysoká pravděpodobnost toho, že zdroj anycastového vysílání obdrží alespoň jedno potvrzení. ETXC proto na rozdíl od ETX neuvažuje RDR.

Uzel vybírá ty kandidáty na další přenos paketu, na základě nichž stanovil své vlastní ETXC. Jejich pořadí (priorita) odpovídá klesajícím hodnotám ETXC – kandidáti blíže cíle se preferují. Z uvedeného vztahu pro $ETXC_{aircand}(U,D,N_1,N_2,\ldots,N_k)$, (tj. hodnotu ETXC v případě použití uzlů $N_1,N_2,\ldots,N_k$ jako kandidátů) vyplývá, že metrika předpokládá nezávislost ztrát na $N_1,N_2,\ldots,N_k$. V reálných situacích ale často určitá korelace mezi těmito ztrátami existuje (mj. proto, že se většinou kandidátské uzly fyzicky nacházejí ve vzájemné blízkosti – možné rušení způsobené vysíláním třetích uzlů pravděpodobně ovlivní všechny kandidáty). Tuto korelaci tedy metrika ETXC nevystihuje. Nicméně i v případě vysoké korelace ztrát přináší anycastové vysílání ETXC výhodu – v průměru totiž vysílání "dosáhne" dál.

Proměřování průchodnosti bezdrátových spojů

OMR odhaduje FDR jednotlivých spojů pomocí tří mechanismů. Předpokládají se přitom i silně asymetrické spoje (tj. s výrazně rozdílným FDR v každém směru), které nejsou v reálném prostředí neobvyklé. Je potřeba zdůraznit, že FDR může pro různě velké pakety nabývat velice rozdílných hodnot. OMR nabízí určitou aproximaci tohoto problému.

Prvním mechanismem odhadu FDR je pravidelné vysílání proměřovacích (tzv. link-probing) paketů. Každý uzel v pravidelných intervalech broadcastuje pakety fixní velikosti (tj. bez potvrzování a retransmisí). Aby se zabránilo náhodné synchronizaci, pravidelný interval se "rozostří" použitím dodatečného jitteru. Tyto pakety obsahují samotné směrovací informace. Každý uzel si přitom vede historii, kolik těchto paketů od svých sousedů v poslední době úspěšně přijal. Na základě této historie stanoví RDR svých sousedů. Při výpočtu RDR mají starší údaje exponenciálně klesající váhu. Do pravidelně vysílaných

směrovacích informací každý uzel zahrne i RDR jednotlivých sousedů. Ti se tedy dozví své FDR (RDR(X,Y) = FDR(Y,X)).

Zmíněný mechanismus zajistí, že se kontinuálně proměřují FDR i u těch uzlů, které momentálně nepřenáší žádný datový přenos. Jakmile ale uzel začne intenzivněji vysílat běžné datové pakety, přichází na řadu druhý mechanismus. Ten poskytuje mnohem přesnější údaje. Každý uzel v pravidelně vysílaných proměřovacích paketech uvádí i počet vysílání datových rámců (všech druhů, tj. i broadcastů a retransmisí), které realizoval během několika posledních "proměřovacích" intervalů. Všechny uzly používají promiskuitní režim RF_MON. Díky němu mohou jednotlivě monitorovat vysílání všech ostatních uzlů a zaznamenávat počet úspěšně přijatých rámců. Podíl takto zjištěné hodnoty a počtu vysílání, který sousední uzel pravidelně hlásí, slouží ke zpřesnění výpočtu RDR (takto zjištěná hodnota tedy reflektuje průměrnou velikost vysílaných rámců).

Třetí mechanismus využívá toho, že bezdrátové rozhraní je schopno poskytovat zpětnou vazbu o tom, kteří kandidáti anycastové vysílání potvrdili a kteří ne (a na kolikátý pokus). Každý uzel vede historie těchto údajů a na základě ní měří své FDR ke konkrétním sousedům sám. Zohledňuje se zde přitom velikost vysílaných paketů.

Hodnoty průchodnosti jednotlivých spojů získané popsanými třemi mechanismy se pro různé účely kombinují podle různých vah. Např. pro stanovení vlastního ETXC se dá "konzervativně" větší váha hodnotě získané pravidelným proměřováním. Při rozhodování o dalším směru přenosu konkrétního paketu se naopak zohlední především hodnota získaná třetím postupem, protože ta nejlépe vystihuje průchodnost spoje pro konkrétní velikost paketů (zde se tedy záměrně porušuje výše uvedené pravidlo, že kandidáty na další přenos jsou tytéž uzly, na základě nichž se prováděl výpočet ETXC).

Protokol OMR používá některé další optimalizace. Např. co nejdéle odkládá finální rozhodnutí o směru dalšího přenosu paketu (bufferování paketů v prioritních frontách je přesunuto z linkové vrstvy do síťové). Lze tak zužitkovat nejaktuálnější informace dostupné v okamžiku, kdy je bezdrátové rozhraní připraveno uskutečnit nový přenos. Problém counting-to-infinity, kterým trpí jednodušší protokoly typu distance-vector, řeší OMR speciální variantou tzv. "Diffusing Computations" [81].

OMR v sobě obsahuje určitou formu implicitního load-balancingu. Pakety totiž obcházejí "zarušené" oblasti sítě. Další přínos oportunistického přístupu spočívá v tom, že jsou využity i potenciálně vysoce nespolehlivé spoje na velké vzdálenosti. Vylepšení, která přináší OMR, jsou obecného rázu, zaměřená na celkovou vyšší efektivitu využití bezdrátového média. Jak ukazují výsledky experimentů uvedené v následující kapitole, jejich přínos se jednoznačně projevuje na dosažených kvalitativních parametrech end-to-end spojení, např. propustnosti, ztrátovosti a přenosovém zpoždění.

4. Výsledky simulací a jejich analýza

V této kapitole jsou uvedeny a analyzovány hlavní výsledky experimentů se simulačním modelem MeshQoS. Všechny konfigurační soubory potřebné pro jejich ověření jsou uloženy na CD, které je součástí práce. CD obsahuje i samotný simulační program včetně zdrojových kódů, výstupní soubory všech experimentů a distribuci systému OMNeT++.

Provedené experimenty jsou dále rozebrány jednotlivě. Nejdřív jsou vždy popsány výchozí předpoklady (topologie sítě, mobilita jejich uzlů, druh a intenzita generovaného provozu atd.). Na takto definovaném konkrétním scénáři je pak studován vliv jednotlivých parametrů na výsledky relevantní pro QoS. Zjištěné závislosti jsou prezentovány ve formě grafů. Vždy jsou znázorněny výsledky pro oba implementované směrovací protokoly – AODV i OMR.

Ve všech experimentech jsou použity generátory provozu některých z následujících tří typů. Bude-li dále uvedeno "RT přenosy", má se na mysli souhrnně VoIP a video-streamingové přenosy. Bude-li uvedeno "NRT přenosy", mají se na mysli FTP přenosy.

- **VoIP**
  Real-time (RT) přenosy audia konstantní rychlostí 64 kbps (v jednom směru). Délka (payload) každého paketu je 128 bytů, tj. generátor vysílá 64 paketů za sekundu. Délka audiopřenosu je 1 minuta, během které generátor vyšle 3840 paketů (tj. 480 kB). Životnost každého paketu (TTL) je 200ms. VoIP přenosy vznikají vždy v protisměrném páru (duplex).
- **Video-streaming**
  Real-time streamingové audio/video přenosy konstantní rychlostí 256 kbps. Předpokládá se zde nižší citlivost na přenosové zpoždění, proto je TTL každého paketu 1 s. Pakety jsou dlouhé 1024 bytů, tj. generátor jich každou sekundu vyšle 32. Během jednosměrného přenosu, který trvá 2 minuty se vygeneruje 3840 paketů (tj. 3840 kB).
- **FTP**
  Non-real-time (NRT) datové přenosy nenáročné na zpoždění, ale požadující vysokou propustnost. Simuluje se řízení toku. Minimální rychlost je omezena na 400 kbps (celý přenos, během kterého se vyšlou 3 MiB tak trvá nejvýše 1 minutu).

Není-li u konkrétního experimentu uvedeno jinak, používají se výchozí hodnoty parametrů uvedené v tab. 1. Např. co se týká bezdrátových přenosů a média, frekvence 2437 MHz odpovídá 6. kanálu WiFi. Střední úroveň šumu je v součtu -90dB. DPLE 2.5 odpovídá prostředí např. okrajových částí měst. Vysílací výkon uzlů se pohybuje v rozmezí 17 a 18 dBm, citlivost při příjmu v rozmezí -81 až -83 dBm (obojí podle rovnoměrného rozdělení).

Simulovaná doba je u všech experimentů omezena na 1000 sekund, bez ohledu na případné nedokončené přenosy.

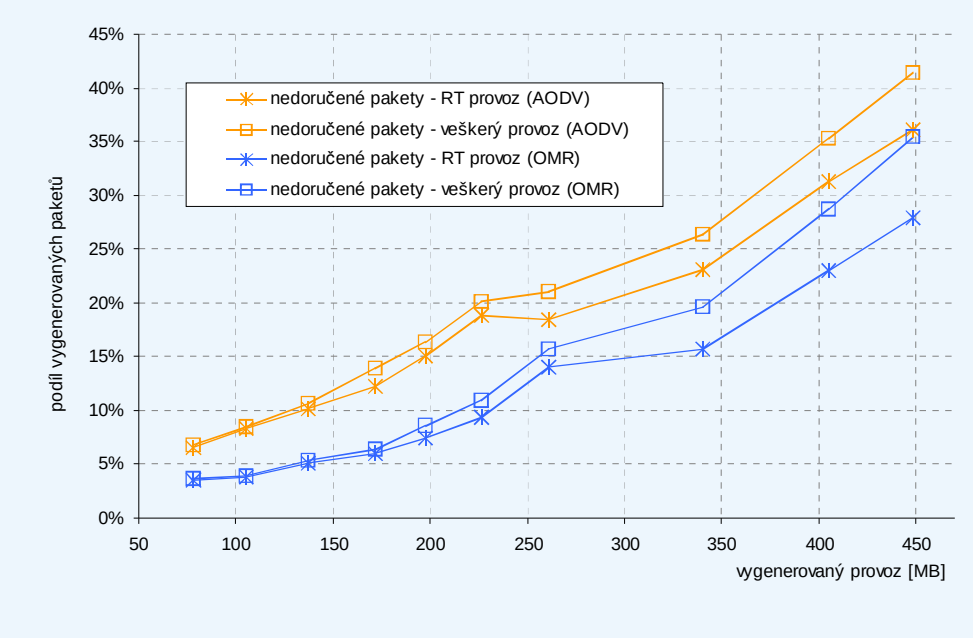
Mobilita uzlů		Generátor provozu "VoIP"	
DefaultSpeed	1.5 (směr.odch. 0.5) m/s	TxRate	64 kbps
Pursuit	false	PacketPayloadLength	128 B
Topologie A		BurstLength	3840
SizeX x SizeY	600 x 600 m	RealTime	true
AccessPoints	2	TTL	0.2 s
WirelessRouters	4	InitiateReciprocal	true
NomadicUsers	10	FlowControl	false
Topologie B0		**Generátor provozu "video streaming"**	
SizeX x SizeY	1000 x 1000 m	TxRate	256 kbps
AccessPoints	0	PacketPayloadLength	1024 B
WirelessRouters	17	BurstLength	3840
NomadicUsers	15	RealTime	true
Topologie B1		TTL	1 s
SizeX x SizeY	1000 x 1000 m	InitiateReciprocal	false
AccessPoints	4	FlowControl	false
WirelessRouters	13	**Generátor provozu "FTP"**	
NomadicUsers	15	TxRate	min. 400 kbps
Fyzická vrstva		PacketPayloadLength	1500 B
BackgroundNoiseLevel	-95 dBm	BurstLength	2000
AdditionalNoise	5 (směr.odch. 2) dBm	RealTime	false
AdditionalNoiseDuration	1 (směr.odch. 5) s	TTL	2 s
Frequency	2437 MHz	InitiateReciprocal	false
DPLE	2.5	FlowControl	true
Linková vrstva		**Hardwarové parametry**	
RTSTreshold	5000 B	BufferSize	128 kB
ShortRetryLimit	7	TxPower	17 až 18 dBm
LongRetryLimit	4	RxSensitivity	-81 až -83 dBm
RTPriority	85	**Další parametry**	
NRTPriority	15	sim-time-limit	1000 s

Tabulka č. 1. Výchozí hodnoty parametrů modelu MeshQoS použité v simulacích.

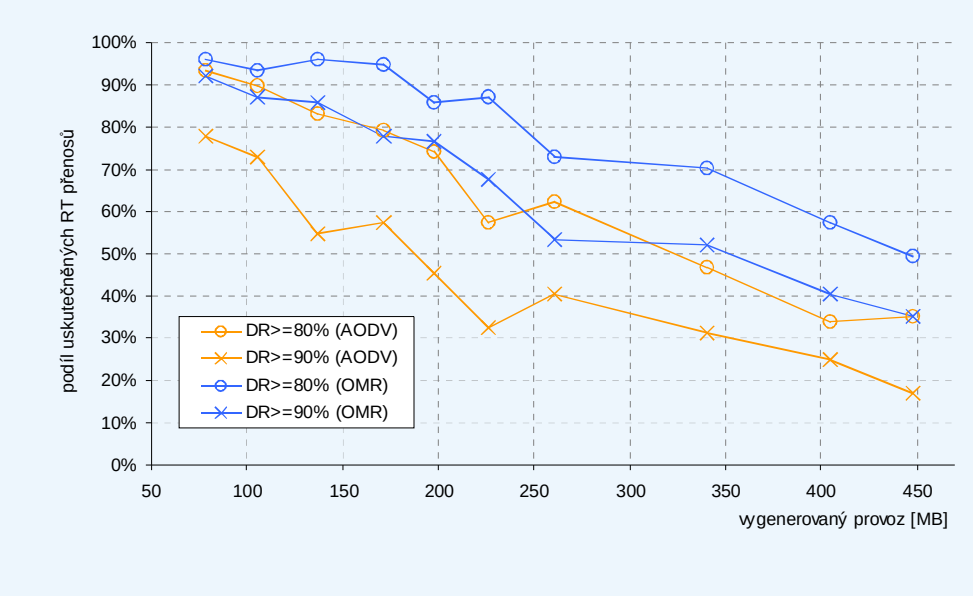
4.1. Vlivy zátěže

První experiment (konfigurační soubory s "A.01" v názvu) předpokládá topologii A (viz obr.5 a tab.1). Síť se tedy skládá ze dvou přístupových bodů připojených ke drátové infrastruktuře, dalších čtyř bezdrátových směrovačů rozšiřujících pokrytí a deseti mobilních "pěších" uzlů (střední rychlost jejich pohybu po vymezeném území 600x600 m je 1.5 m/s, což odpovídá rychlejší chůzi).

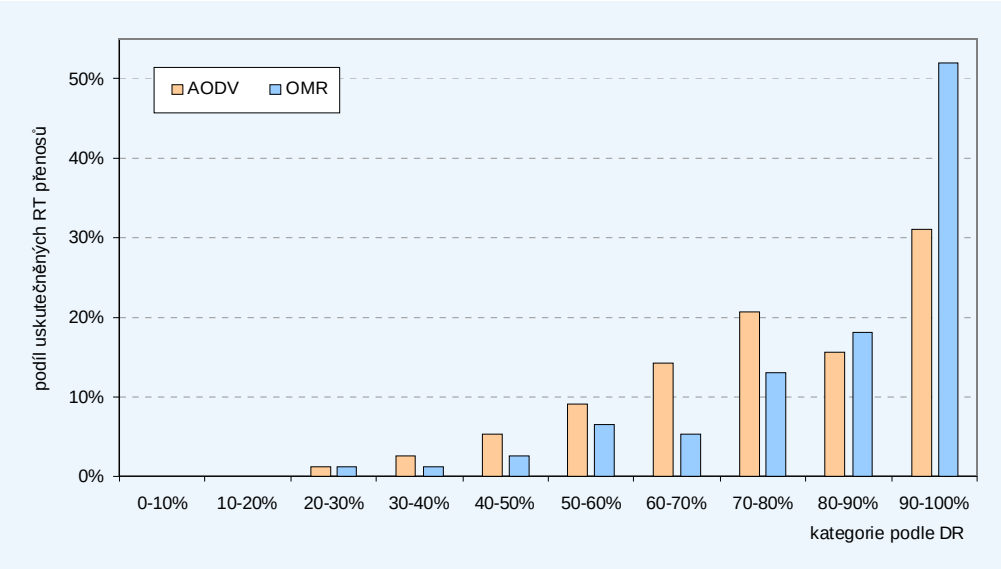
V každé simulaci tohoto experimentu vznikne stejných 66 VoIP přenosů (33 párů) a 11 video-streamingových přenosů. Celkem je v nich vygenerováno cca 71 MB real-time dat. Zkoumá se vliv intenzity provozu v síti jednak na zmíněné real-time přenosy, ale také na dodatečné FTP přenosy, kterými se intenzita provozu reguluje. Celková zátěž odpovídá vygenerování 77 až 448 MB dat (3 až 131 FTP přenosů). Větší část přenosů (RT i NRT) má přitom původce nebo cíl v drátové části sítě, což odpovídá předpokládanému použití mesh sítí jako bezdrátových multihop přístupových sítí. Průměrný počet přeskoků doručených paketů se u jednotlivých simulačních běhů pohybuje většinou v rozmezí 2.0 až 2.2.



Graf č. 1. Vztah ztrátovosti paketů (u RT přenosů a celkově) a zátěže.



Graf č. 2. Vliv zátěže na podíl RT přenosů s vysokou úspěšností doručení paketů.

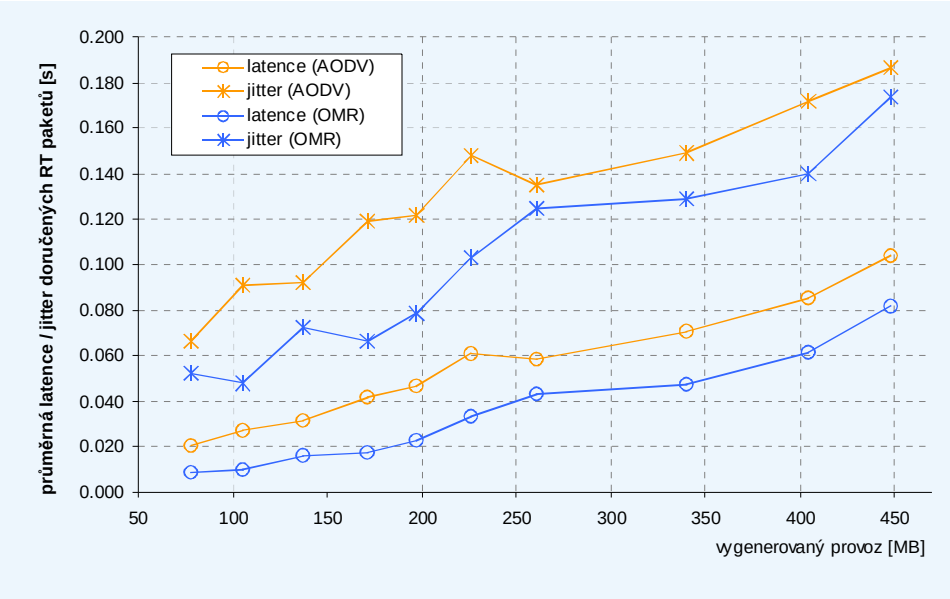


Graf č. 3. Histogram RT přenosů podle DR (8. simulační běh)

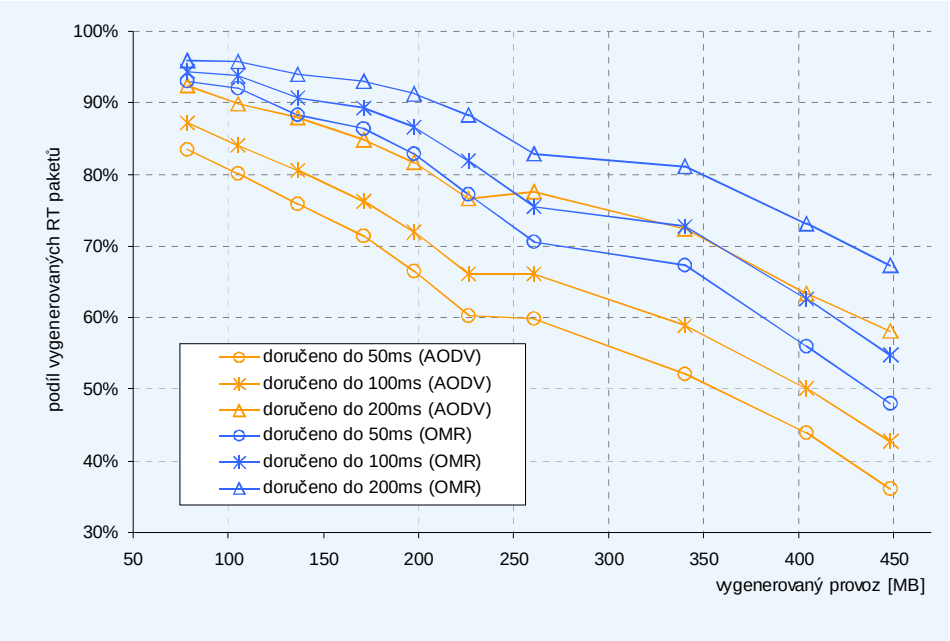
Grafy č. 1 a 2 ilustrují dva pohledy na věc týkající se úspěšnosti doručení paketu. První graf ukazuje, jakým způsobem ovlivňuje intenzita provozu ztráty paketů (ať už jsou způsobené kolizemi, zahlcením nebo vypršením TTL) u RT přenosu a celkově. Při nižší zátěži (cca do 200 MB) má OMR zhruba poloviční ztráty než AODV.

RT přenosy s podílem doručených paketů (Delivery Ratio, DR) minimálně 80 až 90% znamenají při použití vhodného kódování vysokou kvalitu služby pro uživatele. Druhý pohled zkoumá podíl takto "úspěšných" přenosů z celkového počtu 77 RT přenosů (graf č. 2). Např. ještě při zátěži 340 MB lze při použití OMR úspěšně realizovat cca 70% RT přenosů.

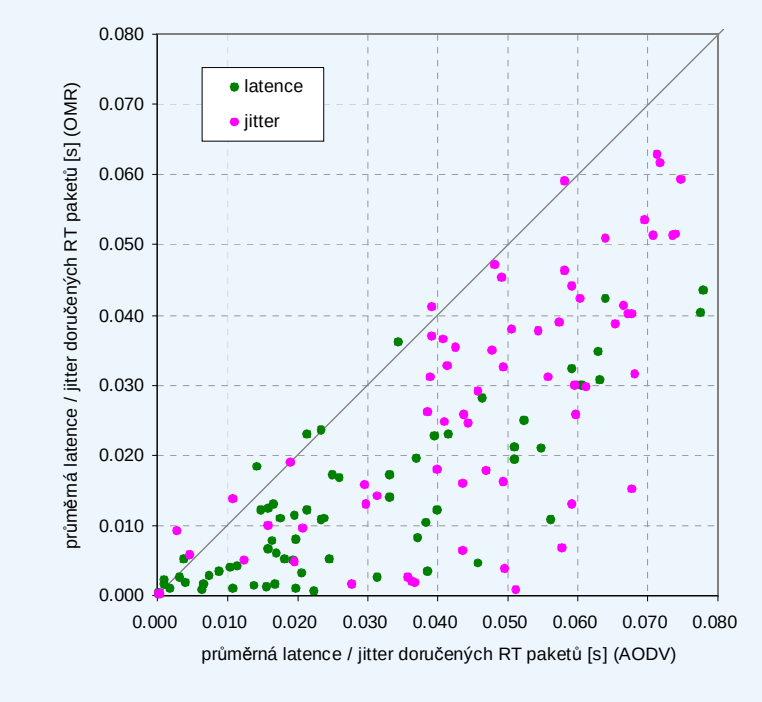
V obou uvedených pohledech vykazuje OMR lepší výsledky než AODV, a to bez ohledu na zátěž. V grafu č. 3 je ještě pro ilustraci detailněji uvedeno, jakých DR v 8. simulačním běhu (vygenerováno 340 MB) dosahuje jaká část RT přenosů.



Graf č. 4. Závislost průměrné latence a jitteru doručených RT paketů na zátěži.



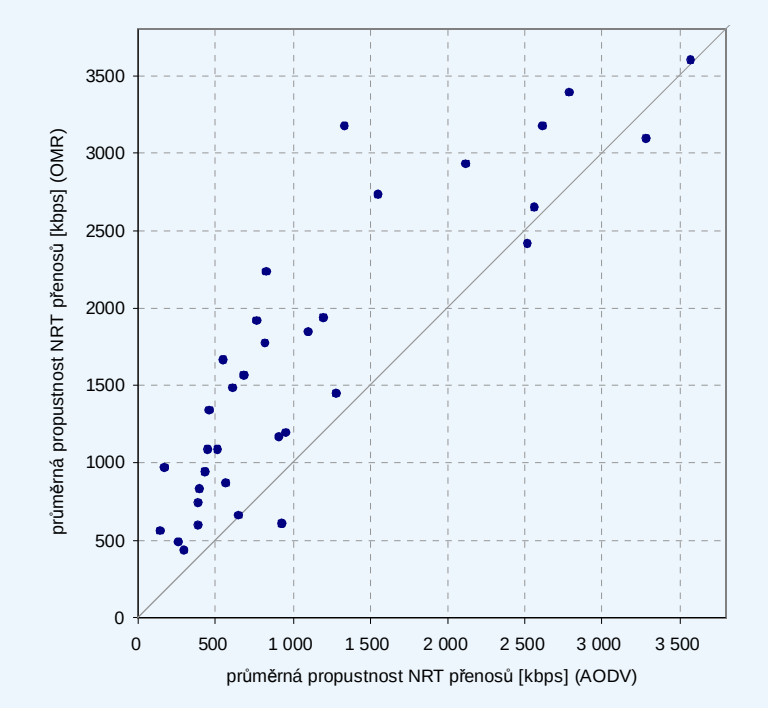
Graf č. 5. Závislost podílu RT paketů doručených do 50, 100 a 200 ms na zátěži.



Graf č. 6. Průměrná latence a jitter doručených paketů u jednotlivých RT přenosů při použití AODV a OMR (5. simulační běh).

Z časového pohledu hodnotí výsledky grafy č. 4, 5 a 6. Jak vyplývá z grafu č. 4, OMR dosahuje významně nižších latencí doručených RT paketů než AODV (při zátěžích do cca 200 MB zhruba polovičních) a při všech zátěžích i nižšího jitteru. Podíly paketů doručených v časových limitech 50, 100 a 200 ms od odeslání zobrazuje graf č. 5. Ještě při zátěži 226 MB doručí OMR v 50 ms více paketů než AODV ve 200 ms. Distribuci zpoždění a jitteru mezi jednotlivými RT přenosy pro 5. simulační běh (vygenerováno 198 MB) ukazuje graf č. 6. Drtivá většina RT přenosů má nižší latenci i jitter při použití OMR než při použití AODV.

Graf č. 7. Vliv zátěže na průměrnou propustnost NRT přenosů.

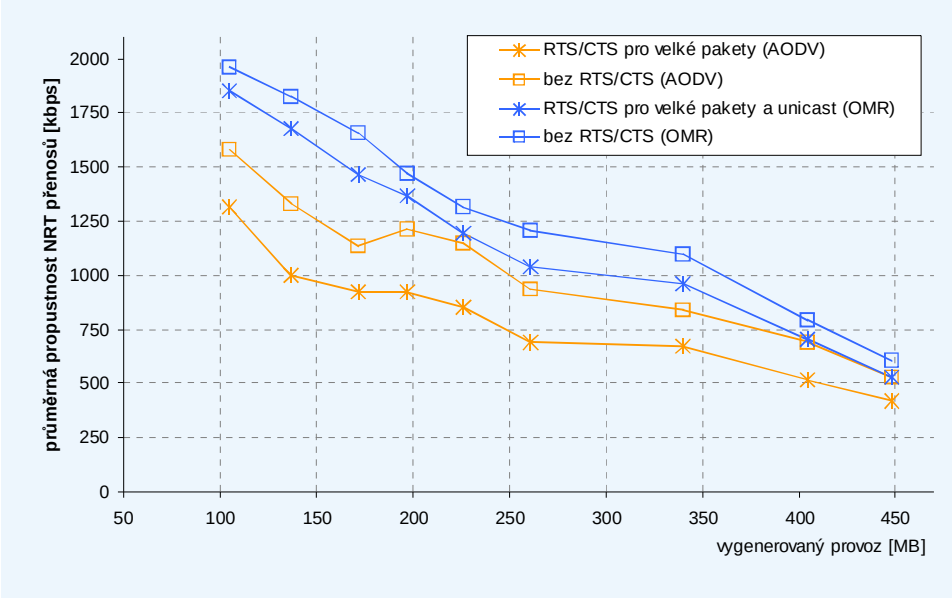


Graf č. 8. Průměrná propustnost jednotlivých NRT přenosů při použití AODV a OMR (5. simulační běh).

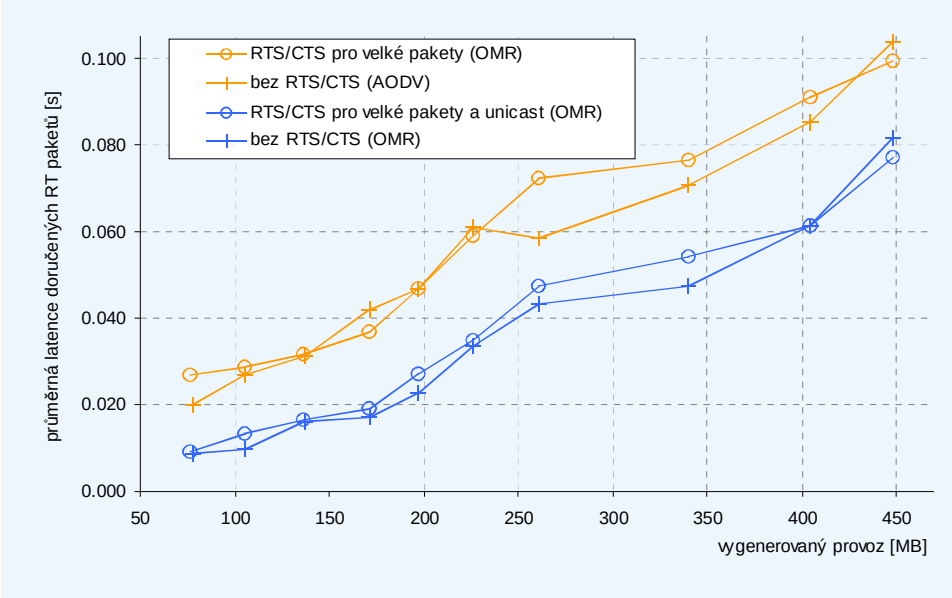
Z grafu č.7 je patrný vliv zvyšující se zátěže na samotné FTP přenosy (jejich průměrnou propustnost). Z důvodu malého vzorku dat (pouze 3 FTP přenosy) nejsou uváděny výsledky týkající se propustnosti v prvním simulačním běhu (platí i pro další experimenty). Nezávisle na zátěži zde OMR dosahuje o 15 až 30 % vyšší propustnost než AODV. Graf č. 8 hodnotí propustnost jednotlivých NRT přenosů v 5. simulačním běhu v závislosti na použití AODV či OMR. Až na výjimky, kde nejsou rozdíly výrazné, dosahuje OMR vyšší propustnosti NRT přenosů.

4.2. Důsledky použití mechanismu RTS/CTS

Druhý experiment ("A.02") je založen na stejné topologii, stejné mobilitě uzlů i stejném provozu v síti jako první experiment. Zkoumá se zde možný přínos mechanismu Request To Send / Clear To Send, jak jej definuje přístupová metoda standardu IEEE 802.11. Parametr RTSTreshold je při tomto experimentu nastaven na hodnotu 512 bytů. Ani v tomto experimentu se tedy RTS/CTS nepoužívá u VoIP přenosů, protože délka jejich paketů je nižší než 512 bytů (konkrétně 128 bytů).



Graf č. 9. Vliv zátěže na průměrnou propustnost NRT přenosů s/bez použití mechanismu RTS/CTS.



Graf č. 10. Závislost průměrné latence doručených RT paketů na zátěži s použitím a bez použití mechanismu RTS/CTS.

Jak již bylo uvedeno v popisu simulačního modelu a definici samotného protokolu, OMR používá ve většině případů anycast (v případech, kdy považuje více svých sousedů za vhodné kandidáty pro další přenos paketu). Při anycastu se přitom mechanismus RTS/CTS nepoužívá. U OMR se tedy mechanismus RTS/CTS (v případě dostatečně velkých paketů) použije pouze na těch bezdrátových přeskocích podél cesty, kdy uzel zná jediného vhodného následníka.

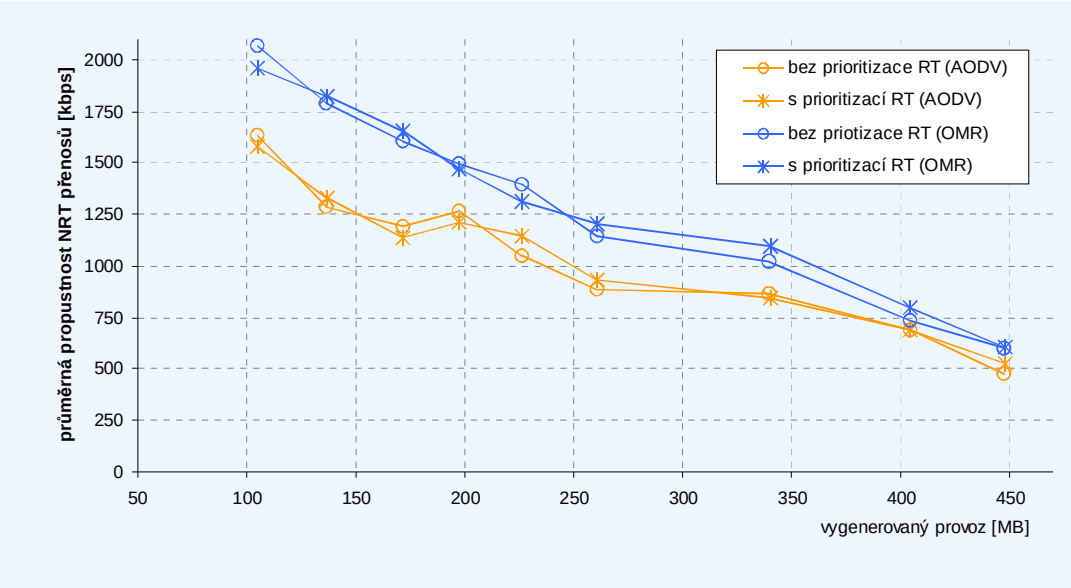
Graf č. 9 ukazuje vliv zátěže na průměrnou propustnost NRT přenosů při použití mechanismu RTS/CTS. Pro srovnání obsahuje také již dříve uvedené výsledky bez použití RTS/CTS (resp. s RTSTreshold nastaveným na hodnotu 5000 bytů, což tento mechanismus efektivně vyřazuje z provozu). Propustnost je při použití RTS/CTS u AODV zhruba o 20% nižší, a to bez ohledu na zátěž. Při použití OMR je vliv RTS/CTS podle očekávání nižší, ale také znatelný (a také negativní). Uzly, které jsou od vysílače fyzicky v opačném směru než zamýšlený příjemce a uslyší RTS, si potenciálně zcela zbytečně znemožní přistupovat k médiu po dobu, během které se má uskutečnit vysílání datového paketu. Problém předsunuté stanice se tak vlastně ještě rozšiřuje. Výsledkem je nižší dosažitelná utilizaci bezdrátového média.

U VoIP přenosů, které v simulaci tvoří převážnou část RT provozu se RTS/CTS nepoužívá. Jak vyplývá z grafu č. 10, použití RTS/CTS u video-streamingových a FTP přenosů se na latenci doručených RT paketů jednoznačným způsobem neprojevilo. Pro srovnání graf č. 10 opět obsahuje již dříve uvedené výsledky s vyšším RTSTreshold.

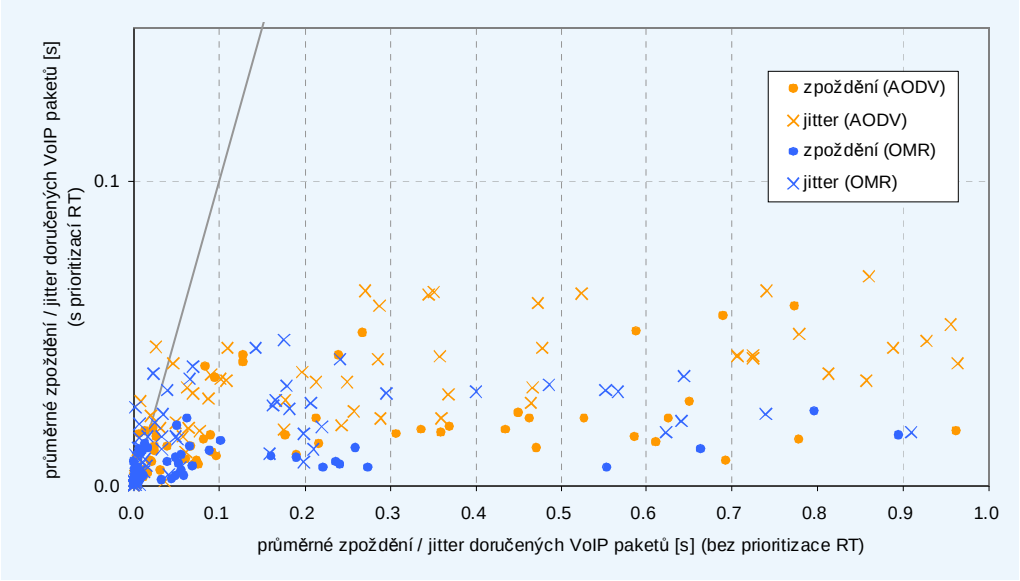
4.3. Přínos použitých QoS mechanismů na linkové úrovni

Jak již bylo uvedeno v definici modelu MeshQoS, na linkové úrovni se používají dva QoS mechanismy. Předně je to prioritizace RT paketů nad NRT pakety podle poměru parametrů RTPriority a NRTPriority. Ve všech experimentech se používají hodnoty 85/15 (ve prospěch real-time). Druhým QoS mechanismem je prioritizace podle relativního stáří paketu (vztaženého k jeho TTL).

Přínos těchto mechanismů se snaží objasnit další experiment ("A.03"). Opět je použita stejná topologie a stejné parametry jako v prvním experimentu (mechanismus RTS/CTS se tedy nepoužívá). Jediným rozdílem jsou použité generátory provozu. Ty generují přenosy, které jsou identické co do doby vzniku, zdrojových a cílových uzlů, přenosových rychlosti atd., ale všechny jsou označeny jako real-time (k potlačení RT/NRT prioritizace) a u všech je definováno velmi vysoké TTL (10s, k potlačení prioritizace starších paketů nad mladšími). Experiment by tak měl zhruba vystihovat prostředí best-effort sítě. Prioritizace paketů směrovacího protokolu nad běžnými datovými pakety zde zůstává.



Graf č. 11. Vliv zátěže na průměrnou propustnost NRT přenosů s/bez použití QoS mechanismů na linkové úrovni.



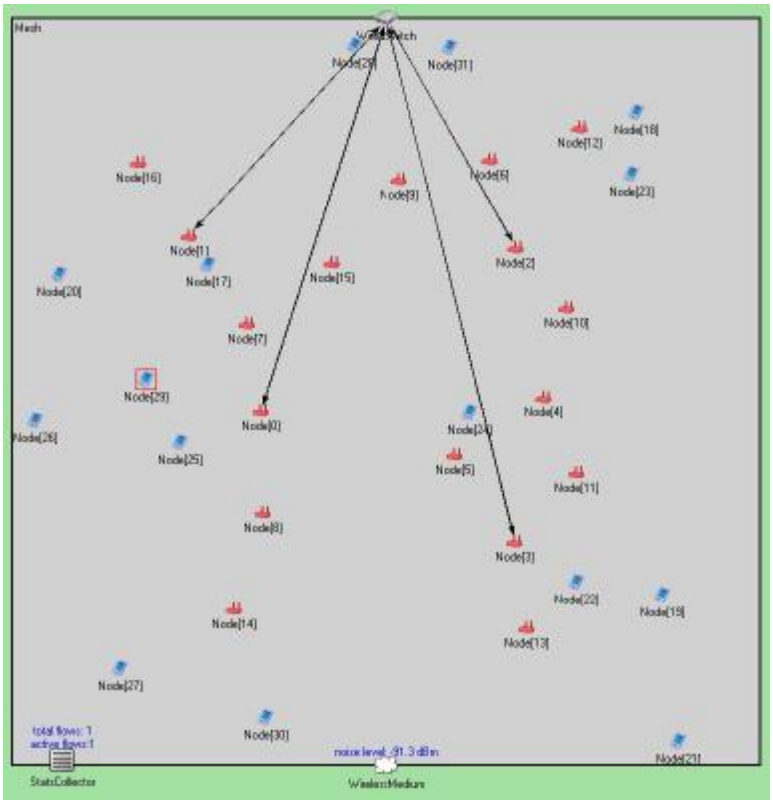
Graf č. 12. Průměrná latence a jitter doručených paketů u jednotlivých VoIP přenosů s/bez použití QoS mechanismů na linkové úrovni (1. simulační běh).

Jak ilustruje graf č. 11, na průměrnou propustnost NRT přenosů nemají použité QoS mechanismy linkové vrstvy prakticky žádný dopad. Jednoznačně se neprojevil ani vliv na ztrátovost (jak původně RT přenosů, tak původně NRT přenosů). Naopak velmi dramatický dopad mají zmíněné mechanismy na časové charakteristiky. Ve všech simulačních bězích vzrostla průměrná latence i jitter doručených paketů zhruba o řád (na desetinásobek). Pro ilustraci je v grafu č.10 znázorněna latence a jitter u konkrétních VoIP přenosů s použitím a bez použití zmíněných QoS mechanismů v 1. simulačním běhu (tj. minimální zkoumané zátěži 78 MB).

Následující experimenty jsou založeny topologii B1 (viz obr. 8 a tab. 1). Třináct bezdrátových směrovačů zde rozšiřuje pokrytí čtyř přístupových bodů propojených drátovou infrastrukturou. Síť dále obsahuje patnáct mobilních uzlů, které se (až na jeden experiment, který zkoumá vliv mobility) pohybují "pěší" rychlostí. Celkem je v tomto scénáři 32 aktivních uzlů, tedy přesně dvojnásobek než u dosud popsaných experimentů. Prostor vymezený pro pohyb mobilních uzlů je 1000 x 1000 m.

Cílem těchto experimentů bylo studovat rozsáhlejší konfigurace s delšími přenosovými cestami. Z tohoto důvodu (a také kvůli tomu, aby se mohl v posledním experimentu objektivně zhodnotit přínos samotného drátového propojení přístupových bodů) se upustilo od předpokladu, že velká část provozu v mesh sítích originuje nebo terminuje v drátové části sítě. U všech dále popsaných experimentů proto všechny datové přenosy vznikají v mobilních uzlech a jejich cílem je vždy některý jiný mobilní uzel. Průměrný počet přeskoků všech doručených paketů se v jednotlivých simulačních bězích těchto experimentů pohybuje nejčastěji v rozmezí 3.4 až 3.8. Pro jednotlivé přenosy přitom má průměrný hop-count přirozeně velmi vysoký rozptyl.

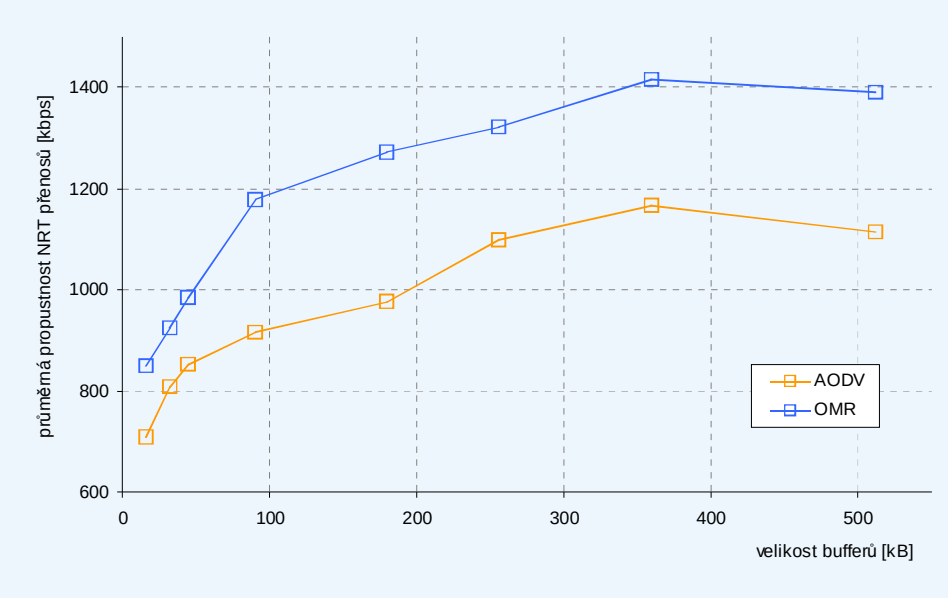
Delší cesty znamenají více bezdrátových přeskoků a úměrně tomu větší vytížení bezdrátového média. Při stejném objemu vygenerovaného provozu by tedy následující experimenty představovaly pro síť podstatně těžší problémy. Zpravidla je v nich proto generováno méně provozu než dosud.



Obrázek č. 8. Topologie B1.

4.4. Vliv velikosti bufferů

První experiment založený na topologii B1 ("B1.01") zkoumá vliv celkové velikosti paketových bufferů, které má k dispozici každý uzel (ať už je využije na jakékoli vrstvě) na dosaženou průměrnou propustnost NRT přenosů. Zátěž je konstantní – celkem 96 FTP přenosů vygeneruje cca 275 MB dat. Přínos větší velikosti bufferů na časové charakteristiky RT přenosů se nepředpokládá a ani se zde nehodnotí (v použitém scénáři jsou pouze FTP přenosy).

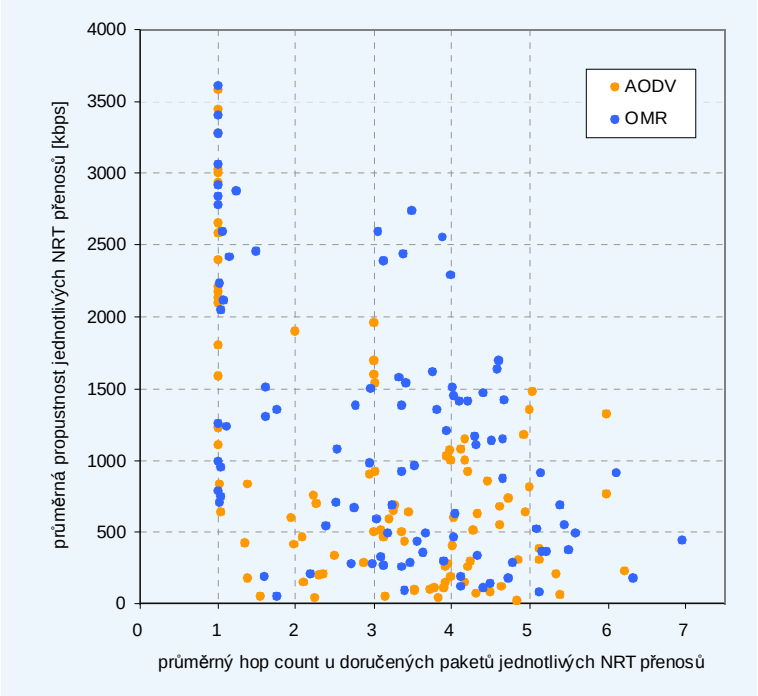


Graf č. 13. Závislost průměrné dosažené propustnosti NRT přenosů na velikosti dostupných bufferů.

Graf č. 13 dokazuje vliv velikost bufferů na dosaženou průměrnou propustnost FTP přenosů. OMR zde dosahuje konstantně o cca 25% lepších výsledků. Pravděpodobně to způsobuje implicitní load balancing, který OMR díky paralelnímu používání více přenosových cest obsahuje. NRT pakety se tedy v průměrném případě "rozprostřou" mezi buffery více uzlů než při použití unicastového AODV. Graf také naznačuje, že po dosažení určité velikosti bufferů se jejich dalším zvětšováním může propustnost naopak snižovat (křivky zde mají vrchol pravděpodobně někde mezi 300 a 400 kB). Řízení toku zde přestává spolehlivě fungovat – zpětná vazba při ztrátě paketu zřejmě omezí rychlost vysílání zdrojového uzlu příliš pozdě.

4.5. Vliv délky přenosových cest

Na výsledky předchozího experimentu je nahlédnuto ještě z jiného pohledu.



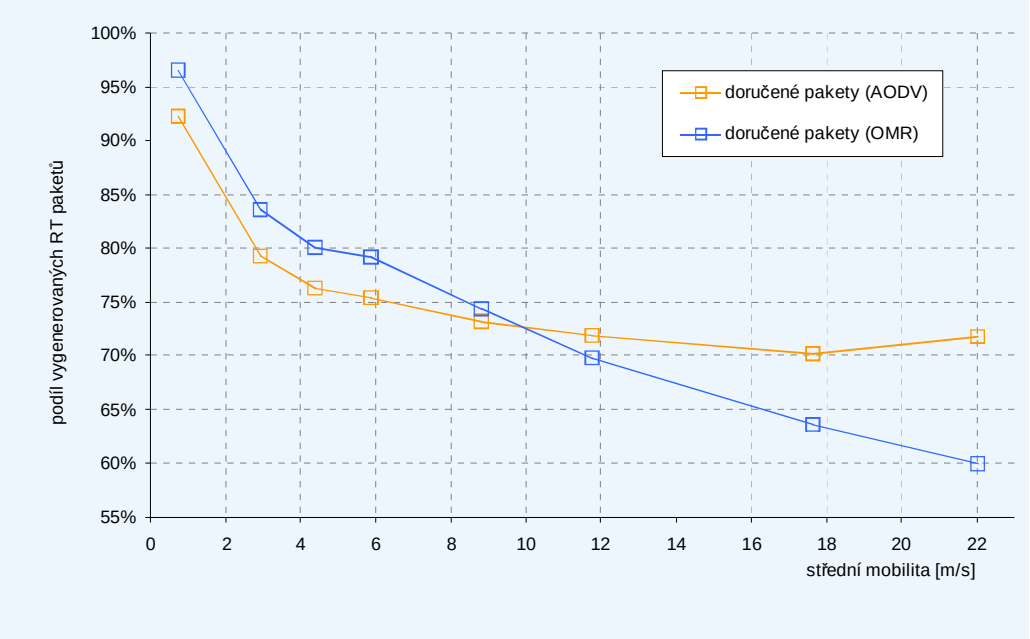
Graf č. 14. Propustnost jednotlivých NRT přenosů v závislosti na délce přenosové cesty (4. simulační běh).

Graf č. 14 zobrazuje závislost propustnosti jednotlivých FTP přenosů pro 4. simulační běh experimentu (velikost bufferů 90 kB) na délce přenosové cesty. Potvrzuje se očekávání, že propustnost v zásadě nepřímo úměrně klesá s rostoucí délkou cesty (což je pro multihop bezdrátové přenosy charakteristické). OMR v několika případech (hop count 3 až 4) dosahuje velmi vysoké propustnosti. To je způsobeno pravděpodobně tím, že tyto přenosy (jako samozřejmě i mnohé další) používají drátovou infrastrukturu a tedy zákonitost nepřímé úměry omezují.

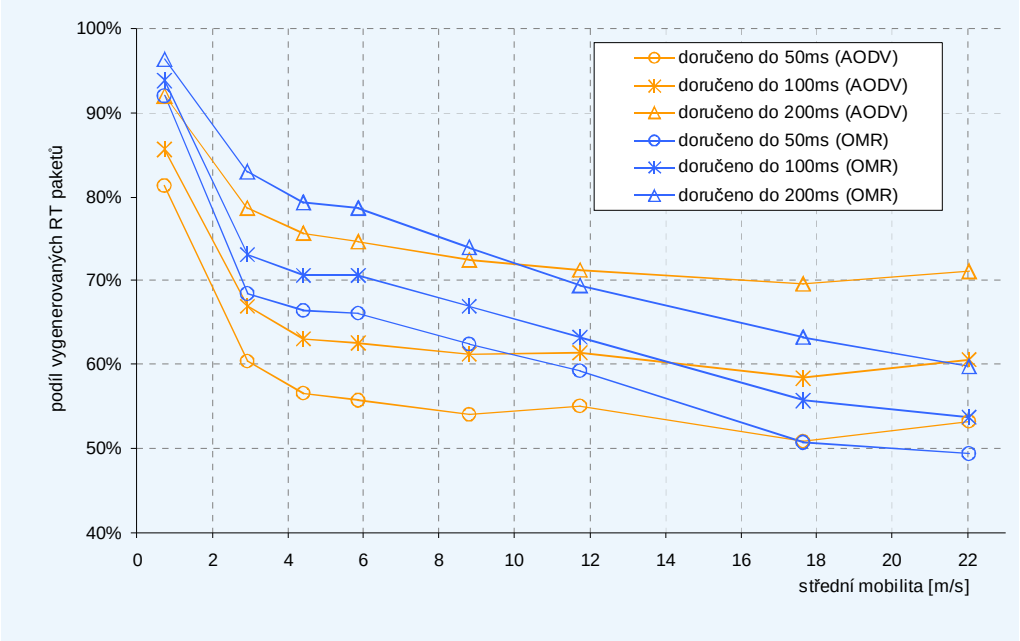
Vliv délky přenosových cest na časové charakteristiky je studován (při různých úrovních mobility) v dalším experimentu – ten na rozdíl od tohoto experimentu obsahuje RT přenosy.

4.6. Vliv mobility

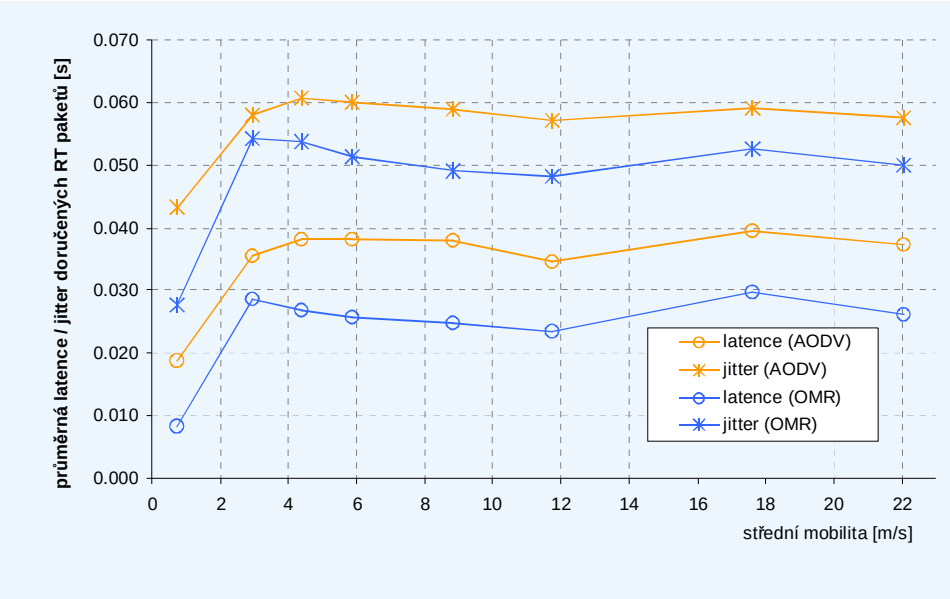
Další experiment ("B1.02") studuje dopady různé úrovně mobility na RT přenosy (při konstantní nižší zátěži). V každém simulačním běhu uskuteční přibližně stejný počet VoIP přenosů (v průměru 90), které vygenerují zhruba 40 MB časově kritických dat. Vzhledem k mobilitě (resp. různé konektivitě způsobené mobilitou) se přitom jedná o různé přenosy (v jinou dobu mezi jiným párem uzlů). Video-streamingové ani FTP přenosy se v tomto experimentu neuvažují.



Graf č. 15. Vliv mobility na úspěšnost doručování RT paketů.



Graf č. 16. Vliv mobility na podíl RT paketů doručených do 50, 100 a 200 ms.



Graf č. 17. Závislost latence a jitteru doručených RT paketů na mobilitě.

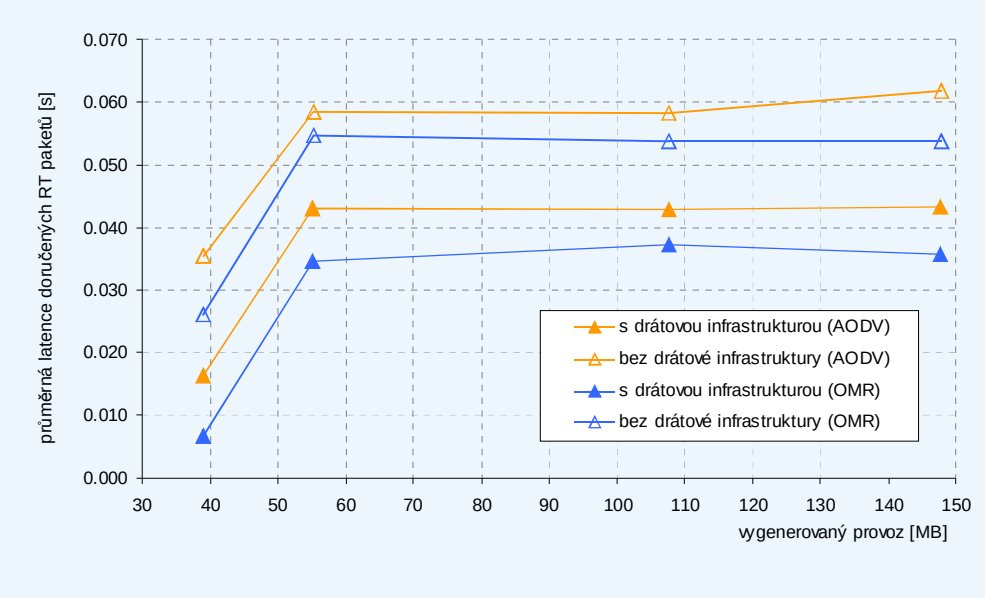
Graf č.15 ukazuje DR VoIP přenosů při různých středních rychlostech mobilních uzlů. Graf č.16 detailněji zobrazuje podíl paketů doručených do 50, 100, resp. 200 ms od odeslání. Ukazuje se, že při nižší a střední mobilitě (cca do 10 m/s) podává lepší výsledky OMR. Při vyšší mobilitě je naopak lepší AODV – čistě reaktivní povaha tohoto protokolu je pro vysoce mobilní uzly evidentně vhodnější. AODV si přitom i při vysokých rychlostech uzlů udržuje poměrně stabilní DR.

Latence a jitter doručených paketů je při vyšších úrovních mobility poměrně stabilní (viz graf č.17), nicméně násobně vyšší než v 1. simulačním běhu (střední rychlost mobilních uzlů 0.7 m/s, tj. během minutového VoIP přenosu urazí uzel v průměru pouze 44 metrů). OMR zde dosahuje lepších výsledků latence i jitteru než AODV.

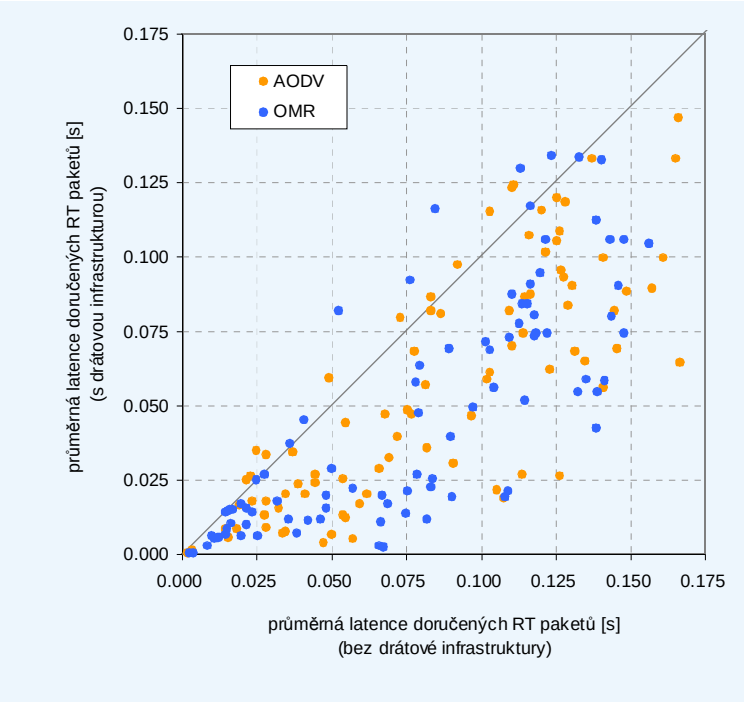
Zajímavé je, že při žádné úrovni mobility se u konkrétních přenosů jednoznačným způsobem neprojevil vliv průměrného počtu přeskoků doručených paketů (tj. střední délka cesty) na průměrnou latenci, jitter ani celkové DR přenosu. Poměrně velká část "dlouhých" cest (v podmínkách simulace cca 6 přeskoků) ještě poskytuje DR nad 90%. Často zmiňovaná neférová výhoda kratších spojení nad delšími je zde pravděpodobně podstatně eliminována prioritizací paketů podle jejich relativního stáří (vztaženého k TTL).

4.7. Přínos drátové infrastruktury

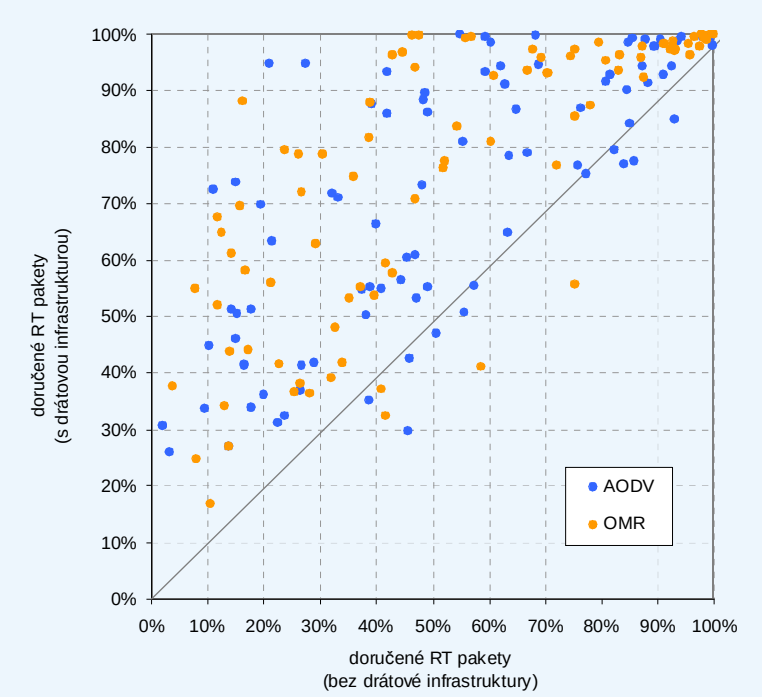
Poslední experiment (konfigurace "B1.03" a "B0.01") zkoumá přínosy drátové části infrastruktury při čtyřech konkrétních úrovních zatížení. Při nich je uskutečněno 38, 64, 114, resp. 162 přenosů, z nichž je vždy zhruba polovina VoIP a polovina FTP. Vygenerováno je tak celkem 39, 55, 108 a 148 MB. Pro srovnání jsou vždy uvedeny výsledky s použitím i bez použití drátového propojení přístupových bodů. Topologie B0 bez drátového propojení přitom vychází z topologie B1, jediný rozdíl je v tom, že přístupové body zde fungují pouze jako bezdrátové směrovače.



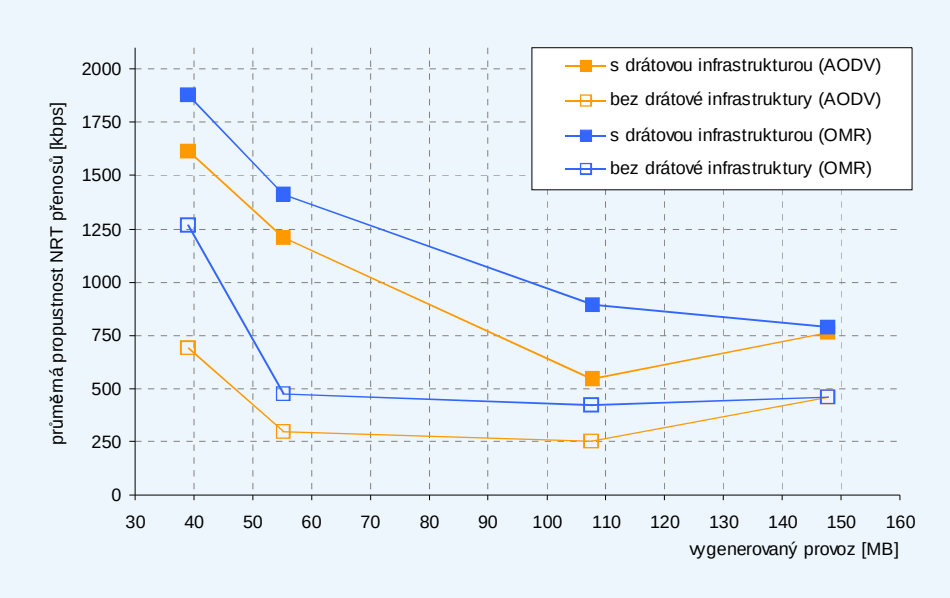
Graf č. 18. Průměrné latence doručených RT paketů při různých úrovních zátěže s použitím a bez použití drátové infrastruktury.



Graf č. 19. Vliv použití či nepoužití drátové infrastruktury na průměrné latence jednotlivých RT přenosů (4. simulační běh).



Graf č. 20. Vliv použití či nepoužití drátové infrastruktury na DR jednotlivých RT přenosů (4. simulační běh).



Graf č. 21. Průměrná propustnost NRT přenosů při různých úrovních zátěže s použitím a bez použití drátové infrastruktury.

Grafy č.18 a 19 ilustrují vliv použití drátové infrastruktury na latenci – graf č. 18 na průměrnou latenci doručených RT paketů a graf č. 19 na průměrnou latenci jednotlivých RT přenosů ve 4. simulačním běhu. Pro 90% RT přenosů a v souhrnu i pro všechny doručené RT pakety je s drátovou infrastrukturou latence nižší. Drátová infrastruktura má pozitivní vliv i na DR většiny RT přenosů (výsledky 4. simulačního běhu viz graf č.20).

Přínos pro NRT přenosy (jejich propustnost) je patrný z grafu č.21. Drátová infrastruktura znamená radikální zlepšení propustnosti (v podstatě až na 4. simulační běh s nejvyšší zátěží jde o násobné zlepšení). Významná část provozu je totiž "svedena" do drátové infrastruktury. Tím se sníží vytížení bezdrátového média. Navíc se každý přenos, který drátovou infrastrukturu používá, fakticky rozdělí na dva kratší, z hlediska bezdrátových interferencí nezávislé přenosy. Vyjma 4. simulačního běhu, kde jsou výsledky vyrovnané, dosahuje OMR lepších výsledků.

5. Závěr

V teoretické části této práce byl podán ucelený přehled možností, technik i celkových "filozofií" zajišťování kvality služeb v bezdrátovém prostředí. Pozornost byla přitom zaměřena na sítě s více bezdrátovými přeskoky (multihop sítě), kde je problém mnohem obecnější a těžší, než u přístupových sítí. Některé zajímavé mechanismy byly ilustrovány popisem konkrétních technik a protokolů. Cílem bylo nezaběhnout do přílišných technických detailů jejich specifikací, ale soustředit se na základní principy.

Praktická část práce měla za cíl detailněji studovat užší problém – vlivy konkrétních mechanismů linkové a síťové vrstvy na dosažitelné QoS charakteristiky end-to-end přenosů u konkrétní kategorie bezdrátových sítí. Zvolena přitom byla kategorie mesh sítí (tj. sítí s bezdrátovou infrastrukturou) založených na technologii WiFi (resp. IEEE 802.11b). Důvod této volby je zřejmý – právě mesh konfigurace představují moderní trend rozšiřování v současné době velice populárních přístupových sítí na bázi WiFi.

Pro účely tohoto studia byl navržen a na bázi simulačního systému OMNeT++ implementován a odladěn diskrétní simulační model MeshQoS. Jeho popis je uveden ve 3. kapitole této práce. MeshQoS je přitom poměrně obecný a do vysoké míry konfigurovatelný model. V rámci MeshQoS byly implementovány dva konkrétní směrovací protokoly – institucí IETF standardizovaný protokol AODV pro mobilní ad-hoc sítě a nově navržený oportunistický protokol OMR. Na bázi modelu MeshQoS bylo provedeno množství experimentů. Mimo dokumentace toho, jaké kvalitativní parametry je vůbec možné v mesh sítích očekávat bylo jejich hlavním cílem zjistit možné přínosy oportunistického směrování.

Výsledky experimentů, které jsou analyzovány ve 4. kapitole této práce, jednoznačně prokázaly pozitivní efekty použití oportunistického principu směrování na výslednou kvalitu služby. V zásadě lze říct, že tam, kde je k tomu prostor, OMR zlepšuje všechny charakteristiky (často významně). Tam kde prostor není, alespoň neublíží. Výjimkou byl experiment studující vlivy mobility uzlů. V konkrétním scénáři se ukázalo, že pro střední rychlost pohybu uzlů větší než cca 10 m/s AODV dosahuje vyšší úspěšnosti doručování paketů. Souvisí to s tím, že protokol AODV je čistě reaktivní, kdežto centrální část OMR je proaktivní.

Relativně lepší výsledky OMR ve srovnání s AODV naznačují, že oportunistické přístupy mají potenciál zajišťovat vysokou kvalitu služeb. OMR přitom není ani zdaleka "maximum možného". Nejdůležitější oblasti, kde ještě zbývá značný prostor pro optimalizaci OMR jsou:

- proaktivní povaha protokolu, která představuje poměrně vysokou režii (ve srovnání s AODV násobně vyšší)
- OMR se musí vypořádávat s poměrně velkým množstvím duplicitních paketů, které v některých situacích anycastové vysílání přináší (až cca 10%)
- mechanismy proměřování průchodnosti jednotlivých bezdrátových spojů příliš nereflektují dynamiku sítě
- použitá metrika ETXC nezohledňuje možné korelace ztrát u různých kandidátů na další přenos paketů

6. Seznam citované literatury

[1] International Telecommunication Union Recommendation ITU-T E.800: Overall network operation, telephone service, service operation and human factors: Terms and definitions related to quality of service and network performance including dependability. Approved in 08/1994.

[2] Hardy W.C.: QoS Measurement and Evaluation of Telecommunications Quality of Service. ISBN: 0-471-49957-9, John Wiley & Sons, June 2001.

[3] International Telecommunication Union Recommendation ITU-T G.114: Transmission systems and media, digital systems and networks:One-way Transmission Time. Approved in 05/2003.

[4] Alghannam A.N., Woodward M.E., Mellor J.E.:Security As A QoS Routing Issue. Proceedings of the 2005 ACM conference on Emerging network experiment and technology, pages 222 - 223, Toulouse, France, 2005.

[5] Wang Z., Crowcroft J.:QoS Routing for Supporting Resource Reservation. IEEE Journal on Selected Areas in Communications, vol. 14, no. 7, pp. 1228–1234, Sept. 1996.

[6] IEEE 802.15 WPAN Working Group: IEEE Standard for Wireless Personal Area Networks. http://standards.ieee.org/.

[7] IEEE 8022.11 WLAN Working Group: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY). Specifications, IEEE 802.11 Standard, 1999.

[8] IEEE 802.16 WMAN Working Group: IEEE Standard for Local and Metropolitan Area Networks. http://standards.ieee.org/.

[9] Truong H.L., Vannuccini G.: The IEEE 802.11e MAC for Quality of Service in Wireless LANs, in Proceedings of SSGRR 2003w, L'Aquila, Italy, Jan.6-12, 2003.

[10] Haas Z., Deng J., Liang B., Papadimitratos P., Sajama S.:Wireless Ad Hoc Networks. Encyclopedia of Telecommunications. John Wiley, December 2002

[11] Zhu C.,Corson M.:QoS routing for mobile ad hoc networks. Proceedings of INFOCOM 2002, New York.

[12] Chen S., Nahrstedt K.:Distributed Quality-of-Service Routing in Ad-Hoc Networks. IEEE Journal on Special Areas in Communications, Special Issue on Ad-Hoc Networks, 1999.

[13] Bruno R., Conti M., Gregori E.:Mesh Networks: Commodity Multihop Ad Hoc Networks. IEEE Communications Magazine, March 2005.

[14] Draves R., Padhye J., Zill B.: Routing in Multi–Radio, Multi–Hop Wireless Mesh Networks. Proc. ACM MobiCom '04, Philadelphia, PA, Sept. 26–Oct. 1, 2004, pp. 114–128.

[15] Chambers B.A.: The Grid Roofnet - a Rooftop Ad Hoc Wireless Network, Master's Thesis at the Massachusetts Institute of Technology, June 2002.

[16] Champaign-Urbana Community Wireless Network (CUWiN). http://www.cuwireless.net/.

[17] IEEE 802.11s MWLAN (Mesh Wireless Local Area Networks) Working Group. http://www.ieee.org/.

[18] Hung-Yun H., Sivakumar, R.:IEEE 802.11 over multi-hop wireless networks: problems and new perspectives. Vehicular Technology Conference, 2002. Proceedings. VTC 2002-Fall. pp.748-752.

[19] Xu S., Saadawi T.:Does the IEEE 802.11 MAC Protocol Work Well in Multihop Wireless Ad Hoc Networks? Journal of the Brazilian Computer Society vol.9 no.1, Campinas, November 2003.

[20] Haas Z.J., Deng J.:Dual Busy Tone Multiple Access (DBTMA) - A Multiple Access Control Scheme for Ad Hoc Networks. IEEE Transactions of Communications, to appear.

[21] Udani S., Smith J.:Power management in mobile computing. Technical report, University of Pennsylvania, MS-CIS-98-26, 1998.

[22] Gehrmann C., Nikander P.:Securing ad hoc services, a Jini view. Proceedings of the 1st ACM international symposium on Mobile ad hoc networking & computing, Boston, MA, august 2000, pp. 135-136.

[23] Luo H., Zerfos P., Kong J., Lu S., Zhang L:Self-securing Ad Hoc Wireless Networks. ISCC (IEEE Symposium on Computers and Communications) 2002, Italy.

[24] Wayne Radinsky: Obey The Law! http://www.waynerad.com/laws.php.

[25] Lee D.: Enhanced IP Services for Cisco Networks. Indianapolis, Cisco Press, 1999.

[26] Yanella A, Miorandi D., Pupolin S.,Riamondi P.: On Providing Soft-QoS in Wireless Ad-Hoc Networks. Proceedings of WPMC'03, pp. 91-95, Yokosuka, Kanagawa, Japan, October 19-22 2003.

[27] Mirhakkak M., Schult N., Thomson D.:Dynamic Quality-of-Service for Mobile Ad Hoc Networks. MobiHoc 2000, Boston, Massachusetts.

[28] Morgan Y.L.,Kunz T.:PYLON: An Architectural Framework for Ad-hoc QoS Interconnectivity with Access Domains. HICSS'03 pres, Hawaii, USA, Jan. 2003.

[29] Braden R., Clark D., Shenker S.:Integrated Services in the Internet Architecture: an Overview. IETF Informational RFC 1663, June 1994.

[30] Shenker S., Partridge C., Guerin R.:Specification of Guaranteed Quality of Service. IETF Standards Track 2212, September 1997.

[31] Wroclawski J.:Specification of the Controlled-Load Network Element Service. IETF Standards Track 2211, September 1997.

[32] Wroclawski J.:The Use of RSVP with IETF Integrated Services. IETF Standards Track 2210, September 1997.

[33] Xiao S., Ni L.N.:Internet QoS: A Big Picture. IEEE Network Magazine, March 1999.

[34] Fodor G., Persson F., Williams B.: Application of Integrated Services on Wireless Accesses. INTERNET-DRAFT draft-fodor-intserv-wireless-issues-01.txt, IETF, January 2002.

[35] Xiao H., Seah W., Lo A., Chua K.C.:A Flexible Quality of Service Model for Mobile Ad-Hoc Networks. Proceedings of IEEE Vehicular Technology Conference 2000-spring, Tokyo, Japan, May 2000.

[36] Blake S., Black D., Carlson M., Davies E., Wang Z., Weiss W.:An Architecture for Differentiated Service. IETF Informational FRC 2475, December 1998.

[37] Nichols K., Blake S., Baker F., Black D.: Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers. IETF Standards Track RFC 2474, December 1998.

[38] Nichols K., Carpenter B.:Definition of Differentiated Services Per Domain Behaviors and Rules for their Specification. IETF Informational RFC 3086, April 2001.

[39] Nichols K., Jacobson V., Zhang L.:A Two-bit Differentiated Services Architecture for the Internet. IETF Informational FRC 2638, July 1999.

[40] Johnson D.B., Maltz D.A., Hu Y.C.:The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks (DSR). INTERNET-DRAFT draft-ietf-manet-dsr-10.txt, IETF MANET Working Group, July 2004.

[41] Chen K., Shah S.H., Nahrstedt K.:Cross-Layer Design for Data Accessibility in Mobile Ad Hoc Networks. Journal of Wireless Personal Communications, vol. 21, pp. 49–76, 2002.

[42] Serban R., Gara S., Dabbous W.:QoS Signaling Protocols. IEEE Communications 2000, December, 2000

[43] Braden R., Zhang L., Berson S., Herzog S., Jamin S.:Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification. IETF Standards Track 2205, September 1997.

[44] Talukdar A.K., Badrinath B.R., Acharya A.:MRSVP - A resource reservation protocol for an integrated services network with mobile hosts. ACM Wireless Networks Journal, Vol. 7, No. 1, pp. 5-19, 2001.

[45] Tseng C., Lee G., Liu R., Wang T.:HMRSVP: a hierarchical mobile RSVP protocol. ACM Wireless Networks Journal, Vol. 9, No. 2, pp. 95-102, 2003.

[46] Perkins C.E.:IP Mobility Support. IETF Network Working Group Standards Track RFC 2002, October 1996.

[47] Manner J., Suihko T., Kojo M., Liljeberg M., Raatikainen K.:Localized RSVP. INTERNET DRAFT draft-manner-lrsvp-04.txt, IETF Transport Area Working Group, September 2004.

[48] Lee S.B., Campbell A. T.:INSIGNIA: In-band signaling support for QoS in mobile ad hoc networks. Proceedings of the 5th International Workshop on Mobile Multimedia Communication, Berlin, October 1998.

[49] Xue J., Stuedi P., Alonso G.:ASAP: An Adaptive QoS Protocol for Mobile Ad Hoc Networks. Proceedings of the 14th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC2003), Beijing, China, September 2003.

[50] Chen T., Gerla M., Kazantzidis M., Romanenko Y., Slain I.:Experiments on QoS Adaptation for Improving End User Speech Perception Over Multi-hop Wireless Networks. In Proceedings of QoS Mini Conference in conjunction with IEEE ICC'99, Vancouver, Canada, Jun. 1999.

[51] Chan R.:Channel Prediction for Adaptive Modulation in Wireless Communications. IEEE Transactions on Communications, vol. 52, pp. 307-316, February 2004.

[52] Qian L., Jones D.L., Ramchandran K., Appadwedula S.:A General Joint Source-Channel Matching Method for Error Resilient Wireless Video Transmission. To appear in the IEEE Transactions on Image Processing

[53] Gannoune L., Robert S., Rodellar D.: A Survey of QoS Techniques and Enhancements for IEEE 802.11 Wireless LANs. EIVD-Swisscom Inno. report - May 2003.

[54] Lin C.R., Gerla M.:MACA/PR: An Asynchronous Multimedia Multihop Wireless Network. in Proceedings of IEEE INFOCOM '97.

[55] Sobrinho J.L., Krishnakumar A.S.:Quality-of-Service in Ad Hoc Carrier Sense Multiple Access Wireless Networks. IEEE Journal on Selected Areas in Communications, 17(8):1353--1368, August 1999.

[56] Lindgren A., Almquist A., Schelén O.:Quality of Service Schemes for IEEE 802.11 - A Simulation Study. Proceedings of the 9th International Workshop on Quality of Service, p.281-287, June 06-08, 2001.

[57] Liao W.H., Tseng S.L., Sheu J.P.: A Multi-Path QoS Routing Protocol in a Wireless Mobile Ad Hoc Network. IEEE Int'l Conf. on Networking (ICN), 2001.

[58] Perkins C.E., Belding-Royer E., Das S.: Ad hoc On-Demand Distance Vector (AODV) Routing. IETF Experimental RFC 3561, July 2003.

[59] Perkins C.E.: Highly Dynamic Distance Vector (DSDV) Routing for Mobile Computers. ACM SIGMOD 1994, Conference on Communications Architectures, Protocols and Applications p224-234 1994.

[60] Sinha P., Sivakumar R., Bharghavan V.:CEDAR: a Core-Extraction Distributed Ad hoc Routing algorithm. IEEE Journal on Selected Areas in Communications, Vol. 17, No. 8, pp. 1454–1465, 1999.

[61] Tseng Z.C., Ni S.Y., Chen Y.S., Sheu J.P.:The Broadcast Storm Problem in a Mobile Ad Hoc Network. In ACM MOBICOM '99, August 1999.

[62] Sinha P., Sivakumar R., Bharghavan V.:MCEDAR: Multicast Core-Extraction Distributed Ad hoc Routing. Wireless Communications and Networking Conference, 1999, pp.1313-1318.

[63] Guha S., Khuller S.:Approximation algorithms for connected dominating sets. Tech. Rep. 3660, Inst. for Adv. Computer Studies, Dept. of Computer Sci., Univ. of Maryland, College Park, June 1996.

[64] Schulzrinne H., Casner S., Frederick R., Jacobson V.:RTP: A Transport Protocol for Real-Time Applications. IETF Standards Track RFC 3550, July 2003.

[65] Friedman T., Caceres R., Clark A.:RTP Control Protocol Extended Reports (RTCP XR). IETF Standards Track RFC 3611, November 2003.

[66] Kopparty S., Krishnamurthy S.V., Faloutsos M., Tripathi S.K.:Split TCP for Mobile Ad Hoc Networks. GLOBECOM 2002 - IEEE Global Telecommunications Conference, vol. 21, no. 1, pp. 139-143, November 2002.

[67] Biaz S., Vaidya N. et al:TCP over wireless networks using multiple acknowledgements. Texas A&M University, Technical Report 97-001, 1997.

[68] Banerjee S., Goteti J.:Extending TCP for wireless networks. Department of Computer Science, University of Maryland, College Park, May 1997.

[69] Allman M, Paxson V., Stevens W.:TCP Congestion Control. IETF Standards Track RFC 2581, April 1999.

[70] Liu C., Jain R.:Approaches of Wireless TCP Enhancement and A New Proposal Based on Congestion Coherence. the 36th Hawaii International Conference on System Sciences, Quality of Service in Mobile and Wireless Network minitrack, Big Island, Hawaii, January 5-9, 2003.

[71] Mehta M.N., Vaidya N.H.:Delayed Duplicate-Acknowledgements. A proposal to Improve Performance of TCP on Wireless Links.Texas A&M University, December 24, 1997.

[72] Liu L., Singh S.:ATCP: TCP for Mobile Ad Hoc Networks. IEEE Journal on Selected Areas in Communication, 19(7):1300--1315, 2001.

[73] Ramakrishnan K., Floyd S., Black D.:The Addition of Explicit Congestion Notification (ECN) to IP. IETF Standards Track RFC 3168, September 2001.

[74] Varga A.: OMNeT++ - Objective Modular Network Testbed in C++. http://www.omnetpp.org/.

[75] Janson M., Karlsson M.:WOK - A Simulation Model for DFS and Link Adaptation in IEEE 802.11a WLAN.

[76] Biswas S., Morris R.:Opportunistic Routing in MultiHop Wireless Networks. MIT Laboratory for Computer Science, 2005.

[77] Xiuchao W.:Simulate 802.11b Channel within NS2, 2004, available online at http://www.comp.nus.edu.sg/ wuxi-ucha/Sim80211ChNS2.pdf

[78] Perkins C.E., Belding-Royer E.M. Chakeres I.D.: Ad hoc On-Demand Distance Vector (AODV) Routing. INTERNET DRAFT draft-perkins-manet-rfc3561bis-01.txt, September 2004.

[79] Biswas S.:Opportunistic Routing in Multi-Hop Wireless Networks. Master of Science Thesis at the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, March 2005.

[80] De Couto D., Aguayo D., Bicket ., Morris R.:A High Throughput Path Metric for MultiHop Wireless Routing. M.I.T. Computer Science and Artificial Intelligence Laboratory, 2003.

[81] Zaumen J.J., Garcia-Luna-Aceves J.J:Loop-Free Multipath Routing Using Generalized Diffusing Computations. Proc. IEEE INFOCOM, March 1998.