

Tehnike za izboljšanje delovanja VoIP aplikacij

Borut Klepec
Zavarovalnica Triglav d.d.
Ljubljana
Lendavska 5, 9000 Murska Sobota
bklepec@attglobal.net

Anton Kos
Univerza v Ljubljani
Fakulteta za elektrotehniko
Tržaška 25, 1000 Ljubljana
anton.kos@fe.uni-lj.si

Abstract

In order to effectively compete in today's telecommunication market, service providers and enterprises are turning towards voice over IP (VoIP) technology. Initially, the goal of using VoIP technologies in enterprises was better utilizing internal packets networks and avoiding long-distances toll charges. On the other side service providers need to quickly rollout new revenue generating services. There are many compelling reasons to migrate voice traffic to IP networks, but certain challenges must first be overcome. This paper will introduce important concepts surrounding performances of VoIP applications in corporate networks and in internet. We will explain end-to-end delays, delay jitter, packet reordering and frame erasures and also emerging techniques for solving such problems. Other important issues for VoIP, such as voice coders, signalisation, protocols, security and compatibility with PSTN network will not be covered while this is outside of the scope of this paper. This paper will focus on performances of VoIP applications in wide area networks.

1. Uvod

IP omrežja niso najbolj primerna za prenos podatkov v realnem času, kot je denimo prenos govora. Razlog za to leži predvsem v nepovezavnem načinu delovanja IP omrežij in tudi v nizki kvaliteti storitve, ki jo ta lahko nudijo. IP omrežja so bila narejena za prenos podatkov, kjer zakasnitve IP paketov in zanesljivost dostave nista kritična, vendar pa se danes uporabljajo tudi za prenos performančno zahtevnejšega prometa. Če želimo, da bodo IP omrežja temu kos, moramo proučiti in vpeljati mehanizme, ki bodo omogočali dovolj kvaliteten prenos podatkov za izvajanje aplikacij v realnem času. Za kvaliteto prenašanja govora v vseh omrežjih s paketno komutacijo so pomembni štirje performančni vidiki: skupna zakasnitev od konca do konca (*total end-to-end delay*), spremenljivost zakasnitev (*packet delay variation, delay jitter*), stopnja izgube paketov (*loss rate*) in dostava paketov v napačnem vrstnem redu (*out-of-order packet delivery*). V poglavjih 2 do 5 bomo navedli razloge, zakaj je v IP omrežjih tako težko zagotoviti kvaliteten prenos govora, kako lahko učinkovito rešujemo nastale težave.

Predstavili pa bomo tudi nekaj praktično uporabnih rezultatov. V članku ne bomo obravnavali posebnosti posameznih standardov kodiranja govora; na tem mestu naj le omenimo, da je izbira načina kodiranja pomembna in vpliva tudi na občutljivost kodiranega signala na navedene performančne kriterije (različni standardi kodiranja govora so denimo različno občutljivi na izgube paketov pri prenosu). S poglavjem 6 bomo zaključili prispevek in povzeli najpomembnejše ugotovitve.

2. Zakasnitve od konca do konca

Zakasnitve od konca do konca ali *end-to-end delay* so pri interaktivni komunikaciji med dvema govorcema zelo pomembne, če ne najpomembnejši dejavnik kvalitete govornega signala. Kadar so zakasnitve od konca do konca prevelike, začne tekoč pogovor zastajati, saj nastopi t.i. efekt *half-duplex* povezave (poslušalec predvideva, da sogovornik na nasprotni strani ni začel z govorom, zato prevzame besedo, medtem pa prispe tudi govorni signal z nasprotnega konca). Pri zakasnitvah od konca do konca lahko opazujemo dva vidika:

- zakasnitev govornega signala od ust govorca do ušes poslušalca, kjer so vštete prav vse zakasnitve, ki v nastopajo pri prenosu in obdelavi digitalnih govornih informacij ter
- zakasnitve prenosa IP paketov od izvirnega do ciljnega sistema, kjer so vključene samo zakasnitve obdelave IP paketov v obeh končnih sistemih ter zakasnitve prenosa in procesiranja v omrežju. Tem zakasnitvam ponavadi pravimo tudi zakasnitve v omrežju (*network delay*).

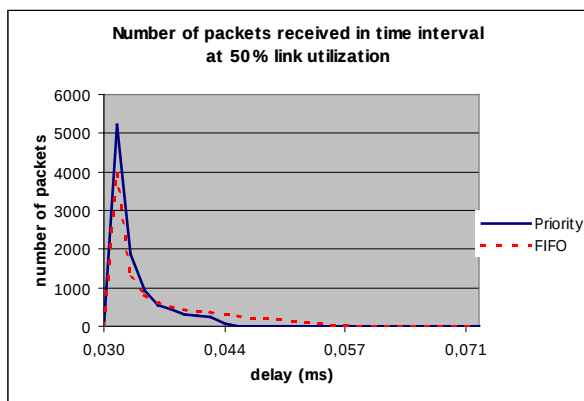
Zakasnitve v pri prenosu paketov v IP omrežju delimo na fiksne in spremenljive. V našem prispevku bomo opazovali le zakasnitve pri prenosu IP paketov preko omrežja in tako izpustili obravnavo zakasnitev paketiranja, procesiranja in zakasnitev kodirnih algoritmov v končnih sistemih, ki so ponavadi fiksne.

Fiksne zakasnitve so tiste, ki za dani paket nastopajo tudi ob času ničelne obremenitve omrežja, t.j. ob odsotnosti vsega ostalega prometa. Odvisne so od števila usmerjevalnikov, preko katerih potuje IP paket, njihovih zmogljivosti, razdalje od izvora do cilja ter hitrosti povezav:

- Vsi paketi so deležni zakasnitev pri prehodu skozi usmerjevalnik zaradi obdelave paketa ter (pogostokrat) tudi zaradi potrebe, da usmerjevalnik sprejme celotni paket, preden ga začne obdelovati. Obdelava IP paketa v usmerjevalniku obsega še določanje poti na podlagi zapisov v usmerjevalni tabeli ter spreminjanje nekaterih polj v IP paketu (npr. polja TTL).
- Zakasnitev razširjanja (*propagation*) je posledica končne hitrosti električnega ali optičnega signala; za optični kabel znaša ta približno tretjino svetlobne hitrosti v vakuumu. Zakasnitev razširjanja znaša približno 5 μ s na kilometer razdalje.
- Zakasnitev prenosa (*transmission*) dobimo z razmerjem dolžine IP paketa in bitne hitrosti povezave.

Zakasnitve v čakalnih vrstah omrežnih elementov so pri nespremenjeni poti in konstantni dolžini IP paketa pogostokrat edini spremenljivi del skupnih zakasnitev od konca do konca. Težko jih napovemo in so odvisne od trenutne obremenitve omrežja, lahko pa z upoštevanjem nekaterih izhodišč dovolj natančno določimo njihovo zgornjo mejo [6]. Zmanjšujemo jih lahko z uporabo principov in arhitektur za kontrolo kvalitete storitve oz. z naprednimi načini strežbe čakalnih vrst.

Pa si na kratko pogledimo, kaj je možno narediti glede zakasnitev pri prenosu IP paketov. Če predpostavimo omrežje kot vnaprej določeno, lahko fiksne zakasnitve zmanjšamo edino tako, da pošiljamo krajše pakete in tako znižamo zakasnitev prenosa.



Spremenljive zakasnitve lahko zmanjšamo z uporabo mehanizmov za kontrolo kvalitete storitve, konkretno pa zadeve rešujemo z različnimi načini strežbe čakalne vrste. Dokument RFC denimo 2598 opisuje storitev *expedited forwarding*, ki omogoča dostavo paketov z minimalnimi zakasnitvami. Eden od možnih mehanizmov je uporaba strežbe po prioriteti (*priority queuing*), ki učinkovito zmanjšuje zakasnitve za govorni promet, kadar ta nosi visoko prioriteto [2]. Podobno lahko v lokalnih omrežjih uporabljamo standard 802.1Q/p za IP telefonijo. Poleg tega obstaja še cela vrsta mehanizmov strežbe čakalnih

vrst (in s tem kvalitete storitev), ki pa jih tu ne bomo obravnavali.

Za primer prikazujemo rezultat simulacij strežbe po prioriteti omrežja s štirimi usmerjevalniki in 50% obremenitvijo hrbtenične povezave. Iz slike levo spodaj je razvidno, da zakasnitev paketov za govorni promet pri uporabi strežbe po prioriteti učinkovito zmanjšamo tudi pri nižjih obremenitvah povezav. Učinek je viden v skrajšanju repa porazdelitve zakasnitev.

Naslednja možnost za zmanjšanje zakasnitev je uporaba tehnike sekanja in prepletanja povezave (*link fragmentation and interleaving*) za WAN povezave, katerih bitne hitrosti so nižje od 768 kbps [4]. S to tehniko preprečimo obdelovanje in odpošiljanje daljših paketov v celoti ter omogočimo pošiljanje kratkih VoIP paketov med daljšimi. To izvedemo tako, da dolge pakete razsekamo in na tem segmentu poti njihove dele obravnavamo kot posamezne pakete.

3. Spreminjanje zakasnitev

Spreminjanje zakasnitev (pogosto omenjeno tudi kot *delay jitter*) ovira pravilno rekonstrukcijo paketov govornega toka, ki ga izvor pošilja v zaporednem in periodičnem vzorcu, v govorni signal. Spreminjanje zakasnitev je največkrat interpretirano kot razlika v zakasnitvi od konca do konca dveh (včasih se posebej poudarja zaporednih) paketov v istem govornem toku. Zakasnitve paketov v Internetu izkazujejo časovno odvisnost: če je paket n deležen velike zakasnitve, obstaja velika verjetnost, da bo tudi paket $n+1$ zelo zakasnen [5].

Zmanjšanje vpliva spremenljivih zakasnitev na prenos govora v IP omrežjih je eden od pomembnejših izzivov za proizvajalce in razvijalce VoIP opreme. Odstranjevanje zakasnitev v sprejemniku pomeni zbiranje paketov in njihovo hranjenje dovolj dolgo, da lahko sprejmemo tudi pakete govora z veliko zakasnitvijo in jih vključimo v rekonstrukcijo govornega signala.

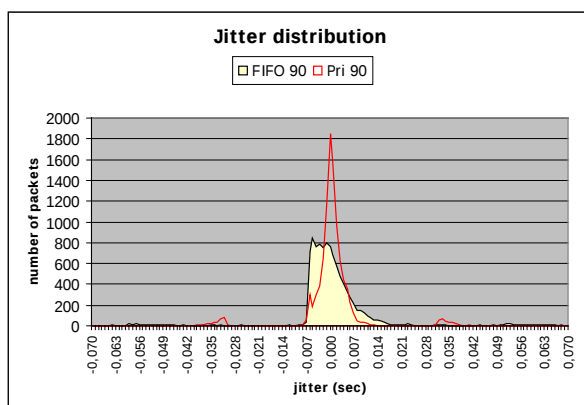
Za tvorbo enakomernega toka digitaliziranega govora najpogosteje uporabimo medpomnilnik v sprejemni napravi, ki absorbira spremembe zakasnitev, preden pakete pošljemo v dekodirnik. Kadar uporabljamo to tehniko, se govorni paketi ne predvajajo takoj po njihovem prispetju, ampak se hranijo v medpomnilniku, dokler ne nastopi njihov čas predvajanja (*playout time*). Čeprav tak način vnaša dodatno zakasnitev v prenos govora, pa lahko po drugi strani vključimo v rekonstrukcijo govora tudi bolj zakasnjene pakete, ki bi jih drugače morali zavreči. Z uporabo medpomnilnika torej izbiramo optimum med velikostjo dodatne zakasnitve, ki jo vnašamo v pogovor (*buffering delay*) in količino paketov, ki jih moramo zavreči, ker so zamudili svoj čas predvajanja (*late loss*) [3].

Danes poznamo tri načine uporabe medpomnilnika za odstranjevanje spremenljivosti zakasnitev [3]:

- vpeljava fiksnega medpomnilnika v sprejemnem sistemu, katerega velikosti ne spreminjamo (*fixed playout time*); obstajata dve različici fiksnega medpomnilnika: fiksni medpomnilnik in statični medpomnilnik, katerega velikost na novo določimo na začetku vsake seje,
- vpeljava spremenljivega medpomnilnika v sprejemnem sistemu, katerega velikost korigiramo med obdobji tišine na podlagi izmerjene zakasnitve omrežja (*adaptive playout time, between talk-spurts adjustment*)
- vpeljava medpomnilnika, katerega velikost sproti spreminjamo ob vsakem sprejetem paketu (*within talk-spurts adjustment*), kar pa zahteva spreminjanje velikosti govornih paketov [3]

Kadar želimo prilagajati dolžino medpomnilnika v IP omrežjih (točka b), ponavadi uporabimo naslednjo tehniko: v sprejemnem sistemu štejemo pozne pakete in računamo njihovo razmerje napram uspešno obdelanim paketom. Na podlagi tega razmerja nastavimo dolžino medpomnilnika tako, da dosežemo vnaprej določen dovoljen delež poznih paketov. Pri prilagajanju medpomnilnika, kot je opisano v točki b, moramo torej pazljivo izbirati razmerje med deležem izgubljenih paketov in zakasnitvami, ki jih vnaša medpomnilnik.

Medpomnilnik se lahko vgradi v digitalni signalni procesor (DSP), kjer se izvaja tudi kodiranje in dekodiranje signala, v splošno namenski RISC procesor ali pa kar v programsko kodo aplikacije. Ponavadi je dolžina medpomnilnika tolikšna, da lahko ta hrani med 20 ms in 50 ms govora.



Spreminjanje zakasnitev se lahko občutno zmanjša tudi v nekaterih primerih, kadar za prenos govora uporabljamo omrežje z omogočeno kvaliteto storitve. Zgornja slika predstavlja spreminjanje zakasnitev kot razliko med zakasnitvijo od konca do konca za dva zaporedna paketa govora. Vrednosti so izračunane iz rezultatov simulacij preprostega IP omrežja s štirimi

usmerjevalniki in 90% obremenitvijo hrbtnične povezave. Opazovan govorni izvor je pošiljal en paket s koristno vsebino govora 10 oktetov vsakih 10 ms. Iz slike je razvidno, da so razlike med zakasnitvami zaporednih paketov pri uporabi strežbe po prioriteti (špica v grafu) precej manjše kot v primeru, ko uporabljamo FIFO strežbo čakalne vrste.

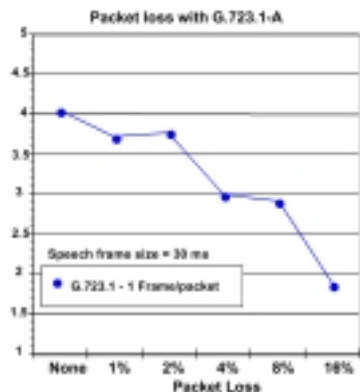
4. Izguba govornih okvirjev

Izguba govornih okvirjev (*frame erasure*) nastopi, kadar IP paket, ki prenaša govorno informacijo, ne prispe na cilj pravočasno. Razlogov za to je lahko več: paket se pri prenosu pokvari, paket se zavrže zaradi obremenjenosti usmerjevalnika (torej zaradi polne čakalne vrste), paket se izgubi zaradi napake v delovanju omrežja ali pa preprosto prispe na cilj tako zakasnen, da zamudi svoj čas predvajanja in ga moramo zavreči. Izguba govornih paketov je lahko posamična ali pa lahko izgubimo zaporedoma več govornih okvirjev. V Internetu so izgube paketov časovno odvisne: če se izgubi paket n obstaja večja verjetnost tudi za izgubo paketa $n+1$ [5]. Kadar v enem IP paketu prenašamo več govornih okvirjev istega pogovora, takrat ob izgubi IP paketa vselej izgubimo več zaporednih okvirjev. Proti izgubi govornih okvirjev se največkrat uporabljata naslednji tehniki:

- vnaprejšnje odpravljanje napak (*forward error correction, FEC*) deluje tako, da na izvoru tvorimo in pošiljamo redundantne informacije,
- prikrivanje izgub (*loss concealment*) lahko uporabimo samostojno ali v povezavi z FEC, deluje pa tako, da ob izgubi okvirja govora ponovno predvajamo zadnji uspešno sprejeti okvir.

Uporaba metode vnaprejšnjega odpravljanja napak zahteva procesiranje tako v izvornem kot ciljnem sistemu, toda tudi njena uspešnost je odvisna od stopnje in porazdeljenosti izgub. Poleg tega pa uporaba FEC povečuje potrebno pasovno širino in še dodatno povečuje zakasnitve zaradi potrebnega procesiranja. Tehnika prikrivanja izgub je učinkovita, kadar so izgube posamične in se pojavljajo redko.

Obstaja pa še soodvisnost med stopnjo izgub in velikimi zakasnitvami paketov (izguba paketa običajno sledi velikim zakasnitvam paketov), kar lahko privede do združevanja obeh učinkov in posledično do daljšega izpada ali popačenja govora [5]. Zaznavna kvaliteta govora temelji na končnih izgubah paketov (po morebitnem FEC algoritmu in medpomnilniku), ta pa je lahko precej različna od samih izgub v omrežju. Uporaba vnaprejšnjega odkrivanja napak lahko poleg tega zaradi potrebnega časa procesiranja povzroči še dodatne izgube zaradi poznih paketov. Za primer navajamo odvisnost MOS vrednosti od stopnje izgube paketov za koder G.723.1-A. (Vir: Nortel Networks).



Iz zgornje slike je razvidno, da sam koder dobro prenaša stopnjo izgub do približno 2%. Tudi za zmanjšanje izgub okvirjev govora pri prenosu oz. v splošnem proti izgubi IP paketov uporabljamo mehanizme QoS v IP omrežjih. Dokument RFC 2597 denimo opisuje storitev *assured forwarding*, ki preprečuje izgubo paketov zaradi polnih čakalnih vrst v omrežnih elementih.

5. Sprejem paketov v napačnem vrstnem redu

Sprejem paketov v napačnem vrstnem redu ali *out-of-order packet delivery* je pojav do katerega pride pri omrežjih s kompleksno topologijo, kjer je možnih več poti od izvora do cilja (tipičen primer je Internet). V zasebnih IP omrežjih, ki so običajno topološko enostavna in kjer je od izvora do cilja možna le ena pot, tega pojava običajno ne zasledimo.

V sprejemnem sistemu, kjer želimo rekonstruirati originalni govorni signal, moramo razvrstiti sprejete IP pakete (ali natančneje govorne okvire) v pravilni vrsti red. To ponavadi opravljamo kar v medpomnilniku, ki ga uporabimo za odstranjevanje spreminjanja zakasnitve. Medpomnilnik torej razvršča prejete govorne okvire na podlagi zaporedne številke (*sequence number*), ki jo prebere iz RTP glave.

6. Zaključek

Vpeljava kontrole kvalitete storitve v IP omrežja vpliva na vse štiri omenjene performančne faktorje in zato je pravzaprav razumljivo, zakaj proizvajalci omrežne opreme in ponudniki omrežnih storitev postavljajo pričakovanja glede uporabe QoS tako visoko. Z uvedbo diferenciacije prometa na realno-časovni promet in ostali promet lahko učinkovito zmanjšamo zakasnitve paketov, znižamo spremenljivost zakasnitve in odpravimo izgube zaradi zasičenja omrežja [1], [2]. Uporaba tehnik QoS v omrežjih bo omogočila niz novih storitev, med drugim tudi dovolj kvalitetno IP telefonijo. Proizvajalci opreme pa vidijo korist predvsem v konvergenci različnih medijev v eni sami IP napravi. Pri uvedbi QoS lahko pravzaprav govorimo o dveh pristopih: absolutna zagotovila kvalitete (kamor spadajo tehnologije ATM, RSVP in MPLS) ter relativna zagotovila kvalitete (kamor spadajo različne

variante diferenciacije prometa, kot npr. TOS, IP precedence in *Frame Relay*).

Mimo navedenih tehnik pa lahko uporabljamo nekatere prijeme, s katerimi izboljšamo delovanje aplikacij za prenos govora v IP omrežjih, vendar z le omejenim učinkom. Te tehnike so: uporaba naprednih načinov za odstranjevanje spreminjanja zakasnitve v končnih sistemih (torej vsaj prilagajanje dolžine medpomnilnika glede na razmere v omrežju), uporaba vnaprejšnjega popravljanja napak in prikrivanja izgub okvirjev govora, uporaba tehnike sekanja in prepletanja povezav. V zasebnih omrežjih korporacij je mogoče razmeroma enostavno uporabiti strežbo po prioriteti v vseh usmerjevalnikih in dodeliti govornemu prometu najvišjo prioriteto, vendar moramo pred tem poznati vzorce obremenitve, lastnosti prometa, predvsem pa delež prometa visoke prioritete.

Tehnike za izboljšanje delovanja aplikacij za prenos govora bi torej lahko razdelili v dve skupini: v prvo skupino spadajo mehanizmi za zagotavljanje QoS, v drugo pa vse ostale tehnike. Uvedba QoS v Internet je izjemno kompleksna naloga, predvsem z vidika definiranja storitev in njihovega zagotavljanja od konca do konca v omrežju več ponudnikov. Podpora kakovosti storitev pa se z veliko hitrostjo prebija v lokalna omrežja, kjer proizvajalci opreme že ponujajo omrežna stikala, ki prepoznajo vrsto prometa in ob upoštevanju nastavljenih pravil omogočajo kvalitetno izvajanje aplikacij v realnem času.

Literatura

- [1] A. Kos, B. Klepec: Supporting Real-time Applications in Differentiated Services Network. International Symposium on Telecommunications VITEL 2000, Ljubljana
- [2] B. Klepec, A. Kos: Performance of VoIP Applications in a Simple Differentiated Service Network Architecture, International Conference on Trends in Communications EUROCON 2001, Bratislava
- [3] Y. J. Liang, N. Färber, B. Girod: Adaptive Playout Scheduling and Loss Concealment for Voice Communication over IP Networks
- [4] Cisco Systems: Cisco IP Telephony QoS Design Guide, available at <http://www.cisco.com>
- [5] W. Jiang, H. Schulzrinne: Modeling of Packet Loss and Delay and Their Effect on Real-Time Multimedia Service Quality, Proceedings of NOSSDAV 2000
- [6] M. Mandjes, K. van der Wal, R. Kooij, H. Bastiaansen: End-to-end delay models for interactive services on a large-scale IP network, Proceedings of 7th IFIP workshop of modeling and evaluation of ATM/IP networks, Antwerp, Belgium