Mai lasă şi pe mîine din ce poţi face azi

Mihai Budiu — mailto:mihaib+@cs.cmu.edu
http://www.cs.cmu.edu/~mihaib/

ianuarie 2001

Subiect: discipline de planificare (pentru transmisia pachetelor de către rutere) care nu conservă lucrul

Cunoştinţe necesare: cunoştinţe elementare despre funcţionarea reţelelor de calculatoare

Cuvinte cheie: ruter, planificare, QoS, pachet, conservarea lucrului (work-conserving), streaming

Cuprins

Chiar dacă în urmă cu zece ani se ducea o luptă acerbă între mai multe standarde diferite pentru interconectarea calculatoarelor, în ziua de azi a fost stabilit un cîştigător indisputabil: standardul care defineşte funcţionarea Internetului. Am scris de mai multe ori în PC Report încercînd să explic succesul protocoalelor de comunicaţie folosite în Internet (cele mai importante dintre acestea fiind TCP, Transmission Control Protocol şi IP, Internet Protocol); în secţiunea "Alte surse de informaţie" puteţi găsi legături on-line spre unele din aceste texte.

În textul de faţă vom vedea însă că unele din trăsăturile care fac aceste protocoale atît de eficiente sunt şi grave slăbiciuni, care fac problematică desfăşurarea unor servicii importante în reţea, cum ar fi transmisiunea de sunet şi imagine în timp real (streaming). Vom discuta apoi pe scurt o clasă de soluţii paradoxale propuse de cercetători pentru a remedia slăbiciunile, care constau toate în a... sta cu mîinile încrucişate chiar şi cînd e multă treabă de făcut. Vom vedea deci că înţelepciunea populară care ne îndeamnă să "nu lăsăm pe mîine ce putem face azi" nu este întotdeauna potrivită.

1 Funcţionarea Internetului

Una din principalele calităţi ale Internetului este robusteţea; aşa cum este concepută, reţeaua poate funcţiona chiar în prezenţa unor defecţiuni majore într-o mulţime din calculatoarele participante. Robusteţea este o trăsătura definitorie a Internetului, care a fost proiectat pentru a servi drept reţeaua forţelor armate americane, trebuind să continue să funcţioneze şi în prezenţa unor atacuri nucleare.

Robusteţea însă a avut nişte consecinţe foarte importante chiar în absenţa calamităţilor: orice sistem complex suferă de malfuncţii şi defecţiuni; legea numerelor mari afirmă ca accidentele sunt norma. Robusteţea protocoalelor este esenţială pentru funcţionarea oricărei reţele mari; Internetul a devenit cea mai mare reţea pentru că era cea mai robustă, şi a fost singura care a putut fi extinsă la scară mondială.

Robusteţea Internetului provine din doi factori:

- Serviciul oferit de reţea este minimalist: reţeaua nu oferă nici un fel de garanţii cu privire la fiabilitatea, timpul de livrare sau corectitudinea datelor transmise.
- Serviciile complexe, inclusiv garantarea fiabilităţii, sunt implementate în întregime în cele două puncte terminale ale comunicaţiei[1].

Acest din urmă factor este o sabie cu două tăişuri. Pe de o parte, dacă cele două puncte care doresc să comunice vor să aibă o discuţie decentă, în pofida eventualelor lipsuri din reţea, printr-un efort de calcul se pot înţelege foarte bine. Pe de altă parte sănătatea întregii reţele depinde de o bună colaborare între punctele terminale; cum am arătat şi în alte articole din PC Report, dacă un punct terminal nu respectă regulile de bună purtare, poate răni şi pe ceilalţi participanţi la trafic.

2 Multiplexare statistică şi rezervare

Serviciul oferit de Internet este numit şi "cît de bine poţi fără să te speteşti", sau, în terminologia englezească, "best effort". Un astfel de serviciu este satisfăcător pentru multe din aplicaţiile folosite la ora actuală, cum ar fi transferul de fişiere, poşta electronică, sau "plimbarea" pe web. Nici una dintre aceste aplicaţii nu are cerinţe foarte stringente de viteză şi fiabilitate, şi ca atare, faptul că din cînd în cînd un pachet se pierde sau vine mai tîrziu decît era aşteptat, poate fi compensat TCP prin retransmisie, fără a irita prea mult utilizatorul.

Situaţia se schimbă însă dramatic atunci cînd vrem să transferăm date care sunt destinate consumului direct de către oameni: fie că transmitem sunete sau imagini, simţurile noastre vor fi foarte neplăcut deranjate chiar de mici imperfecţiuni în transmisie. De exemplu, urechea este sensibilă la pauze mai mari de 100 de milisecunde în conversaţie; ca atare, aplicaţiile de telefonie digitală au cerinţe stringente de performanţă. Cerinţele sunt dificil de satisfăcut pentru că însăşi

[1] Internetul a fost proiectat aproape exclusiv pentru comunicaţie punct-la-punct, adică între doi participanţi, spre deosebire, de exemplu, de reţeaua de televiziune, care transmite informaţii multor receptori simultan.

durata propagării semnalului electromagnetic între două puncte îndepărtate de pe glob este de ordinul zecilor de milisecunde (într-o milisecundă lumina în vid străbate numai 300 de kilometri). Protocoalele nu mai pot compensa pierderea pachetelor folosind retransmisia, pentru că un semnal livrat prea tîrziu face conversaţia foarte neplăcută.

Reţeaua poate fi proiectată pentru a oferi suficiente resurse (capacitate) pentru a face astfel de aplicaţii stringente posibile. Cea mai mare parte din cercetarea curentă în domeniul reţelelor de calculatoare este orientată în direcţia proiectării şi implementării de servicii care garantează o anumită calitate; numele tehnic este "Quality of Service", abreviat şi QoS.

Cea mai importantă problemă din QoS este managementul resurselor: cînd ai la dispoziţie o reţea folosită de 500 de milioane de utilizatori, orice resursă este potenţial insuficientă. Pentru a putea garanta calitatea serviciilor resursele necesare *trebuie să fie rezervate* dinainte de către cei care o vor folosi.

Ideea de rezervare este însă complet opusă întregii filozofii a Internetului; Internetul este o reţea "fără memorie", care, de îndată ce primeşte un pachet, încearcă să scape de el trimiţindu-l în direcţia în care crede că se află destinaţia, şi uitînd totul despre existenţa acelui pachet. Termenul tehnic pentru o astfel de arhitectură este "stateless". Ideea de rezervare implică însă reţinerea de informaţii care caracterizează traficul: pentru cine este rezervarea, ce resurse sunt rezervate şi pentru cît timp, cît din resursele rezervate sunt folosite, etc.

Noţiunea de rezervare nu este nouă; reţelele telefonice o folosesc ca un ingredient de bază. Pentru că o convorbire telefonică are cerinţe foarte stringente de întîrziere şi trafic (bandwidth), reţeaua telefonică funcţionează complet diferit decît Internetul. Stabilirea unei convorbiri telefonice are trei etape:

- În prima etapă, după formarea numărului, reţeaua încearcă să rezerve resursele necesare pentru transmisiunea informaţiei de la un capăt la altul, de la telefonul apelant la cel apelat, trecînd prin toate centralele telefonice intermediare. Dacă una singură din aceste centrale telefonice nu are resursele necesare, apelul este refuzat, şi utilizatorul primeşte un ton de ocupat.

- În a doua etapă, utilizatorul transmite datele; în telefonia digitală, vocea este transformată în biţi, care sunt apoi transmişi de-a lungul canalelor rezervate pînă la destinaţie.

- Cînd puneţi receptorul în furcă, un semnal de terminare a conexiunii traversează toate centralele, eliberînd resursele alocate.

Cantitatea de date transmisă prin vocea umană este însă foarte variabilă; o mare parte din timp tăcem la telefon (nu uitaţi că o sută de milisecunde reprezintă o durată importantă pentru reţea). Reţeaua telefonică însă foloseşte aceleaşi resurse, indiferent că vorbim sau că tăcem, pînă eliberaţi conexiunea.

Prin contrast, cînd folosim o aplicaţie pentru a trimite informaţie codificînd voce prin Internet, numai pachetele cu informaţie sunt trimise; liniştea nu consumă resurse ale reţelei. Din această cauza transmisiunea prin Internet este inerent mai *ieftină*: consumi numai resursele de care ai strictă nevoie. Reţeaua telefonică consumă mult mai multe resurse decît necesar, pentru că le rezervă pentru cazul cel mai defavorabil. Din cauza aceasta convorbirile telefonice sunt mult mai scumpe decît transmisiunea de date, şi companiile care au apărut în ultima vreme, care oferă telefonie prin Internet (numită şi "voce peste IP", "voice over IP", sau VoIP) pot oferi preţuri mult mai reduse.

Internetul face ceea ce se numeşte "multiplexare statistică", adică foloseşte resursele în comun pentru toată lumea ("a multiplexa" înseamnă a pune o resursă în comun; "statistic" vine din faptul că atunci cînd multă lume foloseşte o resursă în mod intermitent, ea poate fi utilizată mult mai eficient). Reţeaua telefonică foloseşte rezervare.

Este important de realizat că noţiunea de rezervare este inseparabilă de cea de *control al accesului* (access control); dacă nu sunt destule resurse pentru rezervare, acţiunea de transmisiune va fi refuzată (ton de ocupat). Internetul, la ora actuală, nu refuză niciodată pachete, indiferent cît de "ocupat" este.

Trebuie să remarcăm că noţiunile de rezervare şi multiplexare statistică nu sunt mutual exclusive, în sensul că pot fi amîndouă folosite în aceeaşi reţea simultan. Toate propunerile care încearcă să aducă QoS în Internet lasă loc şi pentru vechea funcţionare, best effort: resursele rezervate, sau cele care sunt rezervate, dar nu sunt folosite, pot fi folosite pentru trafic de "prioritate scăzută".

Cercetarea din domeniul QoS încearcă să construiască o infrastructură pentru rezervarea resurselor folosind arhitectura existentă a Internet-ului. În acest articol nu vom vorbi despre cum se face rezervarea, ci despre un alt aspect: presupunînd că poţi rezerva resurse, cum trebuie să le exploatezi cît mai eficient? Altfel spus, cît de puţine resurse sunt necesare pentru a oferi un serviciu garantat?

3 Resurse în Internet

Internetul este o colecţie de reţele mici conectate una de alta. Reţelele adiacente sunt conectate una de alta folosind calculatoare specializate numite *rutere*. Ruterele îşi cheltuiesc toată puterea de calcul pentru două scopuri:

- Să calculeze topologia Internetului, pentru a şti în ce direcţie se află fiecare destinaţie posibilă. Cum se întîmplă asta am discutat pe larg într-unul din articolele mele anterioare din PC Report.
- Să primească pachete cu date şi să le trimită mai departe, spre destinaţie. Pentru a alege ruta potrivită, ruterele folosesc informaţia despre topologie, pe care au calculat-o.

Din punct de vedere al puterii de calcul, a două activitate a ruterelor este cea mai importantă, şi cea care pune cele mai mari constrîngeri. Asta se întîmplă pentru că topologia Internetului nu variază prea des şi prea mult, aşa că informaţia mai veche despre topologie este în general corectă. În schimb, pachetele care vin au la dispoziţie doar cîteva milisecunde pentru a fi procesate[2].

Cum manipulează un ruter un pachet? Ce fel de resurse consumă? Figura 1 ilustrează schematic funcţionarea unui ruter.

Resursele consumate sunt deci: timp de procesor, pentru analiza pachetului şi deciderea rutei, lărgime de bandă (bandwidth) la intrare, la ieşire şi în reţeaua internă de interconectare care leagă intrările de ieşiri şi memorie, pentru stocare în vederea prelucrării.

[2]Ruterele cele mai performante pot prelucra acum pînă la un terabit de informaţie pe secundă, cu viteze de pînă la 20 de gigabiţi pe secundă pe fiecare interfaţă. Dacă un pachet are în medie 500 de octeţi, asta îi oferă 25 de *nanosecunde* de timp pentru a fi procesat. Chiar pentru un procesor de 1 Ghz, asta înseamnă 25 de instrucţiuni.

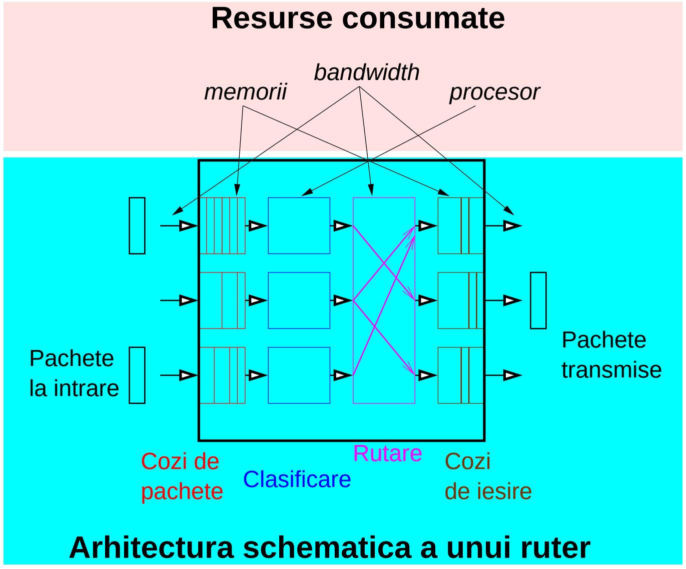

Figura 1: Procesarea pachetelor de date de către rutere. Fiecare pachet este citit de la interfaţa de intrare într-o memorie (buffer), unde este analizat. Destinaţia şi alţi parametri sunt extrase şi tabela cu topologia reţelei este consultată pentru a decide încotro pachetul trebuie trimis (această operaţie se numeşte clasificare). Pachetul este apoi trimis spre interfaţa de ieşire. Dacă interfaţa de ieşire este ocupată cu alte pachete, cel curent stă în memorie într-o coadă, pînă îi vine rîndul.

Ruterele comerciale la ora actuală folosesc algoritmi foarte simpli pentru procesarea pachetelor: ele sunt prelucrate în ordinea sosirii, puse în cozi la interfeţele de ieşire în aceeaşi ordine (FIFO: first in, first out) şi trimise de îndată ce transmisia pachetului anterior este terminată.

4 Ne-conservarea lucrului

Să zicem că sunteţi un ruter care primeşte o grămadă de pachete pe toate interfeţele, şi că aveţi ceva memorie în care puteţi gestiona pachetele care nu au fost încă transmise. Ce părere aveţi: e mai bine (A) ca de îndată ce o interfaţă de ieşire e liberă să căutaţi un pachet care poate fi trimis pe acolo, sau (B) să vă relaxaţi un pic şi să nu trimiteţi nimic, chiar dacă ar fi pachete-candidat?

Dacă veţi răspunde (A), veţi proceda la fel ca toţi cercetătorii din reţele pînă acum 10 ani. Ultima decadă însă a adus în atenţie politici din categoria (B), demonstrînd că au unele calităţi superioare.

Algoritmul care decide *cînd* fiecare pachet trebuie transmis se numeşte algoritm de planificare (scheduling). Despre algoritmii din categoria (A) se spune că "conservă lucrul", pe cînd cei din categoria (B), nu. În engleză disciplinele de planificare de tip (A) se numesc "work conserving", iar cele din categoria (B) "non-work conserving".

Aparent o disciplină care nu conservă lucrul nu are nici un sens: iroseşte timp, în care ar putea face treabă utilă. De ce să întîrzii emiterea unui pachet gata? Răspunsul este subiectul central al acestui articol: pentru că în felul acesta poţi face traficul *mai predictibil*, ducînd la o *economie substanţială de resurse* pentru ruterele traversate de trafic.

Pentru a înţelege în ce fel predictibilitatea traficului depinde de disciplina de planificare, să privim figura 2. Presupunînd că pachetele sunt prelucrate în ordinea venirii, o rafală de pachete pe una din conexiuni poate întîrzia pachetele pe cealaltă conexiune. În acest fel, trafic independent poate distorsiona regularitatea unui trafic care intră foarte regulat în reţea.

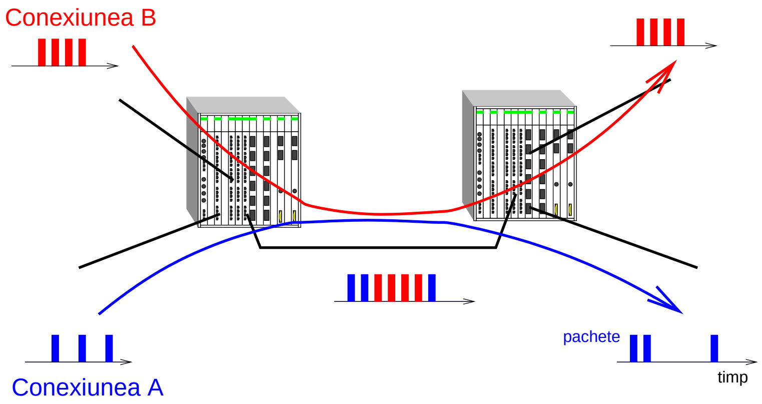

Figura 2: Două conexiuni independente, A şi B, care folosesc în comun un canal de comunicaţii pentru un timp, folosind disciplina de planificare FIFO. Chiar dacă conexiunea A trimite pachete în mod foarte regulat, o rafală de pachete (burst) care soseşte pe conexiunea B poate întîrzia următorul pachet al conexiunii A. Ca atare, ruterul S3 va vedea pachetele de la A inegal distanţate.

Doi parametri ne vor ajuta să înţelegem importanţa regularităţii: rata medie şi rata instantanee a traficului. Dacă sursa A trimite cîte un pachet pe secundă timp de zece secunde, rata medie şi cea maximă de transmisiune sunt de 1 pachet pe secundă. Dacă sursa A trimite însă 10 pachete în prima secundă şi apoi 9 secunde nimic, rata medie este tot de 1 pachet pe secundă, dar rafala de pachete din prima secundă face rata instantanee de 10 pachete pe secundă.

Chiar dacă traficul respectă în medie rezervarea, neuniformitatea poate duce la probleme:

- Prelucrarea unei rafale cere mai multe resurse decît a unui trafic regulat cu aceeaşi rata medie de transmisie. De exemplu, trebuie mai mult spaţiu de stocare pentru pachete înainte de a fi prelucrate.
- Dacă se foloseşte o disciplină care conservă lucrul, rafalele sosite perturbă traficul independent care trece prin acelaşi ruter. Cu alte cuvinte, apariţia unei rafale crează noi rafale.
- Prezenţa rafalelor duce la variaţii mari ale întîrzierii pachetelor. Dacă pachetele au plecat echidistant în timp dar au format o rafală, înseamnă că cele din urmă le-au prins pe cele dinainte, deci primele au o întîrziere de propagare mai mare decît cele din urmă.

Ruterele din reţea trebuie deci să prevină apariţia neregularităţilor în trafic. Ruterele au suficiente resurse pentru a prelucra traficul atunci cînd acesta vine cu rata medie, declarată la stabilirea conexiunii, în procesul de rezervare (dacă nu ar avea destule resurse, ar fi respins rezervarea). Cheia este ca traficul să aibă o rată instantanee egală cu cea medie, şi atunci resursele vor fi suficiente. În caz contrar, se vor crea cozi de pachete, care vor cauza şi mai multe neregularităţi.

Soluţia este oferită de algoritmii de planificare care nu conservă lucrul; aceşti algoritmi vor încerca să re-creeze la ieşire un trafic regulat. Chiar dacă pachetele nu vor sosi echidistant, din cauza interferenţelor între diferitele conexiuni care străbat acelaşi ruter, ruterele vor întîrzia unele din pachete, pentru a recrea un trafic regulat la ieşire.

Partea frumoasă a acestei scheme este că, dacă toate ruterele colaborează, traficul rămîne predictibil şi regulat în mare măsură, deci toate ruterele au nevoie de mai puţine resurse, şi pot satisface mult mai bine cerinţele de calitate.

4.1 Avantaje şi dezavantaje

Cercetătorii au ridicat o seamă de obiecţii la adresa protocoalelor care nu conservă lucrul. Unele dintre ele sunt nefondate, dar totuşi interesante de analizat:

- **Destinaţia finală poate elimina variabilitatea întîrzierii cu un buffer mai mare.** Acest lucru este adevărat: nu trebuie neapărat ca reţeaua să ofere un trafic în care toate pachetele au acelaşi timp de tranzit. Însă pentru unele aplicaţii interactive, variabilitatea întîrzierii poate cauza neplăceri la receptor. Mai important, rafalele în reţea cer memorii mai mari din partea ruterelor, pentru a fi stocate; ruterele folosesc memorii de mare viteză, care sunt mult mai costisitoare decît cele din calculatoarele obişnuite.

- **Disciplinele care nu conservă lucrul măresc întîrzierea medie a pachetelor.** Şi acest lucru e adevărat: unele pachete sunt întîrziate în mod artificial în reţea, pentru a recrea uniformitatea. Dar acest lucru nu e grav: ceea ce-l interesează pe receptor în aplicaţii gen streaming audio/video nu este întîrzierea medie, ci întîrzierea *maximă*. Receptorul trebuie oricum să întîrzie pachetele care vin prea devreme (pentru că altfel reproducerea semnalului ar fi ne-uniformă). Disciplinele care nu conservă lucrul nu măresc întîrzierea maximă, ci doar pe cea medie.

- **Disciplinele care nu conservă lucrul irosesc capacitatea reţelei, lăsînd libere canalele.** Într-un scenariu real, nici aceasta nu este o problemă: vă reamintiţi că în Internet capacitatea suplimentară poate fi folosită pentru trafic tradiţional, de tip best-effort. Dacă ruterul nu are de trimis nici un pachet dintr-o conexiune garantată, poate trimite eventual pachete best-effort aflate în cozi. Dacă aceste pachete crează rafale la ruterele următoare, ele pot fi pur şi simplu aruncate, pentru că reţeaua nu promite nici un fel de fiabilitate pentru legături de tip best-effort.

- **Disciplinele care nu conservă lucrul vor pedepsi o sursă cu rezervaţie care transmite la o rată mai mare decît cea rezervată, chiar dacă reţeaua este nefolosită.** Cercetătorii din reţele încă nu au găsit o soluţie satisfăcătoare pentru această problemă. Dar considerente de altă natură vor spune dacă aceasta este sau nu o problemă reală. De exemplu, pentru traficul de voce, capacitatea maximă este clar limitată de spectrul vocii umane.

- **Disciplinele care nu conservă lucrul sunt mai greu de implementat.** Este adevărat că operaţiile cerute de o astfel de disciplină de planificare sunt mult mai complicate decît cele ale unei simple discipline FIFO; cea din urmă se poate implementa simplu şi rapid, cu o listă înlănţuită. Dar, în general, pentru a oferi servicii garantate, operaţiile făcute de ruter vor deveni toate mai complexe (nu numai planificarea, ci şi clasificarea sau decizia rezervării). Dintre aceste trei operaţii, complexitatea planificatorului este cea mai uşor de redus, iar impactul ei asupra performanţei nu este atît de mare ca al celorlalte două.

5 Concluzii

În acest articol am discutat pe scurt despre cercetarea curentă din reţele de calculatoare care încearcă să adauge Internetului servicii cu calitate garantată, necesare pentru transmisiuni de sunet sau imagini video. Am văzut că aceasta este o problemă foarte grea, din cauză că Internetul funcţionează în mod minimalist, şi reţine foarte puţine informaţii despre datele care-l tranzitează.

Am discutat apoi pe scurt despre efectul disciplinelor de planificare, folosite pentru a alege ordinea şi momentul la care pachetele primite de un ruter sunt trimise mai departe spre destinaţie. Am văzut de asemenea că a termina treaba cît se poate de repede nu e întotdeauna optim, ci că lucrul regulat este mai sănătos pentru întreaga reţea.

Deocamdată aceste idei sunt testate doar în laboratoare; rămîne de văzut care va fi succesul lor în Internet.

6 Alte surse de informaţie

- Acest articol discută pe larg politicile de planificare care nu conservă lucrul: H. Zhang, "Providing End-to-End Performance Guarantees Using Non-Work-Conserving Disciplines", Computer Communications: Special Issue on System Support for Multimedia Computing, 18(10), Oct 1995.
  `http://redriver.cmcl.cs.cmu.edu/~hzhang-ftp/ComputerCommunication95.ps.gz`
- O discuţie excelentă a disciplinelor de planificare care nu conservă lucrul (şi nu numai) se găseşte în: Srinivasan Keshav "An Engineering Approach to Computer Networking", 1997 Addison-Wesley, ISBN 0201634422, paginile 224–229.
- Informaţii complementare despre funcţionarea Internet-ului puteţi găsi în mai multe din articolele mele anterioare din PC Report; vedeţi de pildă:
  - Funcţionarea ruterelor în Internet, din PC Report din martie 1999: `http://www.cs.cmu.edu/~mihaib/articles/articles.html#internet`.
  - Principii arhitecturale în Internet în introducerea articolului meu din PC Report din aprilie 1999: `http://www.cs.cmu.edu/~mihaib/articles/articles.html#tunel`.
  - Interferenţa între traficul bine-intenţionat şi cel perfid în Internet este discutată în PC Report din mai 2000: `http://www.cs.cmu.edu/~mihaib/articles/articles.html#tcp`.
- Două documente ale Internet Engineering Task Force (IETF) descriu arhitecturi posibile pentru implementarea de servicii diferenţiate:
  - RFC 2998: A Framework for Integrated Services Operation over Diffserv Networks `http://www.ietf.org/rfc/rfc2998.txt`
  - RFC 1633: Integrated Services in the Internet Architecture: an Overview `http://www.ietf.org/rfc/rfc1633.txt`
- PC Magazine România din ianuarie 2001 a fost consacrat subiectului VoIP (voice over IP).