

Eficiența transmisiilor de date ce utilizează protocoale de comunicație de tip ARQ

1. Principiile de lucru ale protocoalelor de tip ARQ

- Variantă simplificată care permite analiza aproximativă, dar destul de rapidă a performanțelor asigurate
- Protocoalele de tip Automat Repeat Request (ARQ) – utilizează divizarea mesajului în blocuri, 'frames'
- Controlul corectitudinii blocului recepționat se bazează pe calculul sindromului unui cod detector de erori atașat (CRC)
- În funcție de valoarea sindromului receptorul transmite un mesaj de confirmare: pozitiv (ACK – dacă $S = "0"$) sau negativ NACK – $S \neq "0"$)
- Dacă răspunsul este pozitiv, se trece la blocul următor; pentru răspuns negativ, se retransmite blocul eronat.
- Emițătorul așteaptă un interval de timp T_0 („time-out”) sosirea mesajului de confirmare
- Dacă acesta nu sosește în T_0 , emițătorul retransmite unul sau un set de blocuri, în funcție de variantă
- Dacă după un număr q de retransmisii nu se primește confirmare pozitivă emițătorul intră în starea de count time-out în care stă un interval T_T , apoi reia transmisia cu același bloc sau de la începutul mesajului. Pentru unele aplicații de tip RT în acest caz se consideră pachetul pierdut.

2. Detectarea erorilor de transmisie

- Controlul corectitudinii mesajului se face cu ajutorul codului detector de erori inclus;
- În cazul transmisiilor necodate, acest cod este un cod ciclic generat de polinomul $g(x)$.
- Codurile detectoare de erori sunt generate de polinoamele:

$$\begin{aligned} 1) g(x) &= x^{16} + x^{15} + x^2 + 1; \text{ - codul CRC-16} \\ 2) g(x) &= x^{16} + x^{12} + x^5 + 1; \text{ - codul CRC-CCITT.} \end{aligned} \quad (1)$$

- Acest cod are rolul de a detecta erorile apărute și de a nu permite trecerea blocurilor eronate către NLL
- Această cerință nu poate fi îndeplinită în totalitate, datorită probabilității ca un cuvânt de cod să fie transformat prin eronare în alt cuvânt de cod (sindromului nul îi corespund mai multe cuvinte de cod)
- Probabilitatea de apariție a unui bloc eronat nedetectat P_{ud} este egală cu raportul dintre (numărul combinațiilor de erori nedetectate \times probabilitatea de apariție a acestora) și (numărul combinațiilor posibile \times probabilitatea de apariție a acestora).
- Un cod ciclic (n, k, t) , cu $r = n - k$ biți de control, *detectează* toate combinațiile de erori $\leq r$, și *nu detectează* $2^{-(r+1)}$ din combinațiile de $r+1$ erori și respectiv 2^{-r} din combinațiile de mai mult de $r+2$ erori.
- În ipotezele unei distribuții a erorilor după legea binomială și a unei probabilități medii de eroare de bit p , expresia P_{ud} , pentru un bloc cu lungimea de L biți și $r = 16$, este dată de:

$$\begin{aligned} P_{ud} &= \frac{2^{-17} C_L^{17} p^{17} (1-p)^{L-17} + 2^{-16} \sum_{i=18}^L C_L^i p^i (1-p)^{L-i}}{\sum_{j=0}^L C_L^j p^j (1-p)^{L-j}} = \\ &= 2^{-16} \left[\frac{1}{2} \cdot C_L^{17} p^{17} (1-p)^{17} + \sum_{i=18}^L C_L^i p^i (1-p)^{L-i} \right] = 2^{-16} \cdot \left[1 - \sum_{i=1}^{16} C_L^i p^i (1-p)^{L-i} - \frac{1}{2} \cdot C_L^{17} p^{17} (1-p)^{17} \right]; \end{aligned} \quad (2)$$

- Pentru $L=1752$ biți, $P_{ud} < 2^{-17} \approx 7,5 \cdot 10^{-6}$,
- Pentru a reduce și mai mult această probabilitate se folosesc coduri detectoare de erori cu $g(x)$ de ordinul 32, CRC-32, care asigură $P_{ud} < 2^{-33} \approx 1 \cdot 10^{-10}$.
- Codul detector de erori, CRC, inclus în protocoalele de tip ARQ trebuie să asigure, pe lângă detectia erorilor, o probabilitate P_{ud} (ca un mesaj eronat să fie acceptat) mai mică decât o valoare maximă impusă.

3 Tipuri de protocoale ARQ

- După posibilitatea de a asigura transmisia și recepția simultane, protocoalele se împart în:

- a) Protocoale cu structură neechilibrată (unbalanced)
- b) Protocoale cu structură echilibrată (balanced)

- Protocoalele cu structură neechilibrată

- transmit de blocuri într-un singur sens la un moment dat;
- după transmiterea unui bloc emițătorul se oprește și așteaptă până când primește de la receptor mesajul de confirmare, STOP&WAIT (SW)
- necesită transmisii semi-duplex (HD).

Parametrii unui astfel de protocol sunt:

L - lungimea unui bloc; S -lungimea mesajului de confirmare; T_T -durata intervalului de count time-out; q - numărul de retransmisii nereușite după care protocolul intră în count time-out; T_0 - intervalul de timp în care se așteaptă sosirea mesajului de confirmare.

- Durata intervalului de time-out T_0 trebuie să îndeplinească condiția:

$$T_0 > \frac{L+S}{D} + t_c + t_p + t_i \quad (3)$$

unde: D - debitul binar al legăturii; t_c - timpul de comutare al modemului semiduplex; t_p - timpul mediu de prelucrare al blocului și mesajului de confirmare; t_i - timpul de propagare pe canal al blocului și mesajului de confirmare. Notă: atenție la setarea acestei valori în funcție de timpul de propagare prin rețea

Protocoalele cu structură echilibrată

- folosesc o „fereastră glisantă” („sliding window”) și necesită o transmisie duplex integral, FD.
- Se transmite continuu un grup de W_t blocuri cu număr de ordine, generând fereastra de emisie.
- Mesajele de confirmare sunt transmise printre blocurile de date ale transmisiei în sens opus, prin procedeul de „agățare” („piggybacking”).
- Dacă se primește confirmare negativă, blocul este reaşezat la începutul ferestrei următoare, cu același index
- Numărul maxim de retransmisii acceptate este egal cu q
- Pentru a asigura livrarea către stratul superior în ordinea de la emisie, bufferul receptorului trebuie să stocheze toate blocurile demodulate corect
- Pentru aceasta este nevoie de o fereastră de recepție de w_r blocuri.
- Deoarece transmisia este de tip FD, se poate considera că timpul de comutare a sensului de transmisie de către modemi este nul, adică $t_c = 0$.
- În cazul cel mai defavorabil, dacă primul bloc al ferestrei de emisie e eronat de q ori el obligă alte $(W_t - 1)(q+1)$ blocuri cu numere de ordine mai mari să aștepte în bufferul de recepție;
- Un criteriu care influențează numărul maxim acceptat de retransmisii q este ca numărul blocurilor stocate în buffer să nu depășească dimensiunea ferestrei de recepție:

$$W_r = (W_t - 1)(q + 1) + 1 \quad (4)$$

- Condiția de transmisie continuă, 'pipelining':

$$\frac{W_t \cdot L}{D} \geq t_i + t_p + \frac{S}{D} \quad (5)$$

- Tipuri de protocoale „sliding window”:

- SELECTIVE REPEAT (SR) - care retransmite doar blocul eronat;
- GO BACK N (GBN) - anulează N blocuri succesive recepționate, începând cu cel eronat și cere retransmiterea acestora.

- Reluarea legăturii după count time-out:

- Xmodem - nu permite recuperarea blocurilor confirmate înainte de count time-out; la reluare, mesajul este transmis de la început.
- Zmodem - permite recuperarea blocurilor confirmate corect înainte de count time-out - reluarea de la blocul de date la transmiterea căruia au avut loc cele q retransmisii. O altă variantă acestui caz constă în reluarea transmisiei cu blocul de date imediat următor celui care a generat count-timeoutul; în acest caz blocul neachitat se consideră pierdut

- Dimensionarea bufferului de emisie, pentru a reduce probabilitate de pierdere de pachete datorită umplerii sale, trebuie să țină cont de debitul efectiv al transmisiei cu ARQ (sau de timpul mediu de achitare a unui bloc), de frecvența de sosire a blocurilor care trebuie transmise și de lungimea (medie) a blocului.

4. Studiul eficienței transmisiunilor guvernate de un protocol de tip ARQ

4.1 Definirea eficienței

- O metrică ce este folosită pentru evaluarea eficienței transmisiunilor cu protocoale ARQ este valoarea medie a throughputului;

- Debitul binar efectiv (throughput) mediu - debitul binar mediu necesar pentru transmiterea și confirmarea pozitivă a tuturor biților blocului în cauză. Deoarece throughput-ul mediu este egal cu inversul duratei medii necesare transmiterii și achitării unui bit, el poate fi exprimat sub forma:

$$\Theta = 1/t_A \quad (6)$$

unde t_A este timpul mediu necesar transmiterii/confirmării reușite a unui bit.

- O altă abordare folosește ca metrică un coeficient adimensional numit eficiență, notat cu η .

- Eficiența se definește ca raportul dintre numărul mediu de biți utili confirmați pozitiv (ACK), N_u , și N_t , numărul mediu de biți efectiv transmiși sau care ar putea fi transmiși în t_i , t_c și t_p , pentru confirmarea pozitivă a celor N_u biți:

$$\eta = N_u/N_t \quad (7)$$

- Legătura între eficiență și debitele binare medii nominal. D , și efectiv Θ (throughput) este dată de relația:

$$\eta = \Theta/D ; \quad (8)$$

- Eficiența globală, adică procentul de biți utili achitați din numărul total de biți transmiși, nu depinde de D .

5. Calculul eficienței transmisiilor guvernate de protocoale de tip ARQ

- Presupunții: distribuția erorilor după legea binomială și utilizarea unui protocol de tip Zmodem;

- se consideră că lungimile blocului și mesajului de răspuns sunt constante, că sursa are tot timpul blocuri de transmis (“full que”) și că sursa are tot timpul alocate resurse de transmisie. Considerente privitoare la debitul de blocuri și la lungimea acestora vor fi prezentate ulterior. Analiza efectuată va neglija probabilitatea de pierdere a unui bloc prin congestie. Vor fi prezentate doar unele considerații privitoare la inserarea acestei probabilități în calculul eficienței tehnicii ARQ.

5.1. Protocol SW

- Eficiența este analizată pe o legătură fizică, modem semi-duplex.

- Parametrii legăturii sunt definiți mai jos, iar „algoritmul” de funcționare e reprezentat schematic în fig.1:

L - lungimea unui bloc; S - lungimea mesajului de confirmare; T_0 – durata intervalului de time-out; T_T - intervalul de count time-out; q - numărul maxim de retransmisii ale unui bloc după care protocolul intră în count time-out; t_c - timpul de comutare al modemului semiduplex; t_i - timpul de propagare pe canal al blocului și mesajului de răspuns; t_p - timpul mediu de prelucrare a blocului la recepție și mesajului de răspuns la transmițător; D - debit binar.

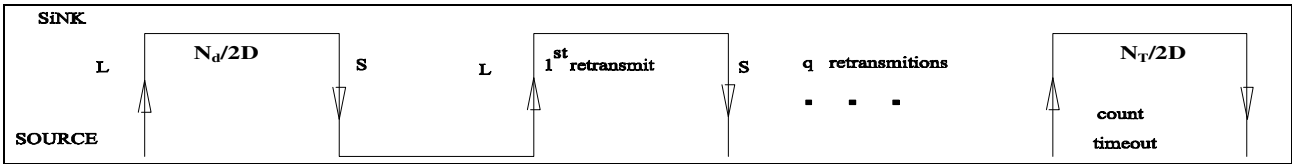


Figura 1 Modelul conexiunii SW-ARQ

- Timpii neutilizați de către protocol, d.p.d.v. al transmisiei, sunt:

$$T_d = t_c + t_i + t_p; \quad (9)$$

- Considerând un debit nominal mediu D , numerele de biți care ar fi putut fi transmise în mod ideal pe durata T_d , N_d , și pe durata T_T a count time-outului N_T sunt:

$$N_d = D \cdot T_d ; N_T = D \cdot T_T ; \quad (10)$$

- Pentru a nu provoca retransmiterea repetată a aceluiași bloc, intervalul de time-out trebuie să îndeplinească condiția (3).

- Toți biții protocolului ARQ, biți de date, antet, biți de control ai CRC, se consideră biți utili.

- N_u este numărul mediu al biților utili transmiși/confirmați pozitiv într-o transmisie, q retransmisii și în intervalul de count time-out.

- N_t - numărul total de biți fizici sau echivalenți transmiși pentru acceptarea corectă a celor N_u biți utili

- Biții echivalenți sunt biții N_d , N_T ; suplimentar se transmit și biții blocurilor confirmate negativ.

- Calculul eficienței unui protocol de tip ARQ implică evaluarea probabilității medii de eroare de bit p și a probabilității de eronare a unui bloc de date cu lungime dată.

- Probabilitatea medie de eroare de bit se poate determina în funcție de SNR din canal folosind relațiile prezentate în cursul de TM pentru modulațiile uzuale (QAM, DPSK, FSK, PAM, BB). În acest material vom considera că transmisia are o probabilitate medie a erorii de bit egală cu p pe durata tuturor celor $q+1$ încercări; această abordare are avantajul că face analiza independentă de modulația utilizată

- În ceea ce privește probabilitatea de eronare a unui bloc având L biți se va considera o distribuție binomială a erorilor

- În această ipoteză, probabilitatea ca să avem mai puțin de $(m+1)$ erori din L biți este:

$$P((m+1) < L) = \sum_{i=0}^m C_L^i p^i (1-p)^{L-i}; \quad (11)$$

- Probabilitatea ca un bloc și mesajul său de confirmare să fie recepționate corect (achitate) este:

$$P(0, L+S) = (1-p)^{L+S} = P_0 \quad (12)$$

- Dacă se ia în considerare și probabilitatea medie de congestie P_{cg} atunci probabilitatea de „achitare” (confirmare pozitivă) devine:

$$P_{0cg} = P_0 \cdot (1 - P_{cg}) \quad (12')$$

- Probabilitatea de retransmitere a perechii bloc + mesaj de confirmare P_R este exprimată de (13), iar dacă se consideră și P_{cg} ea devine (13'):

$$P_R = (1 - P_0) = 1 - ((1 - p)^{L+S}) \quad (13)$$

$$P_{Rcg} = P_R + P_{cg} - P_R \cdot P_{cg}; \quad \text{justificați relația!} \quad (13')$$

- Probabilitatea ca mesajul să fie confirmat corect (achitat) după j retransmisii, P_j :

$$P_j = P_R^j \cdot P_0 \quad (14)$$

- Probabilitatea de intrare în count time-out, care poate fi egală și cu probabilitatea de pierdere a unui pachet prin eronare în cazul în care aplicația nu acceptă count timeout, este dată de:

$$P_T = P_R^{q+1} = [1 - (1 - p)^{L+S}]^{q+1} \quad (15)$$

- pentru valori mici ale lui p ($p < 1 \cdot 10^{-5}$) și lungimi ale blocului $L < 2000$ biți, P_T poate fi aproximată prin:

$$P_T \approx [C_{L+S}^1 \cdot p - C_{L+S}^2 \cdot p^2]^{q+1} \quad (16)$$

- Numărul mediu de biți achitați după transmisia inițială, q retransmisii și intervalul de count time-out este:

$$N_u = (L + S)(P_0 + \sum_{j=1}^q P_R^j P_0) = (L + S) \sum_{j=0}^q P_R^j P_0 \quad \text{justificați relația!} \quad (17)$$

- Numărul mediu de biți transmiși pentru ca un bloc, $L+S$ biți, să fie achitat la a j -a retransmisie:

$$N_{ij} = (j + 1)(L + S + N_d) \cdot P_j = (L + S + N_d)(j + 1) \cdot P_R^j \cdot P_0; \quad \text{justificați relația!} \quad (18)$$

- Pentru a determina numărul mediu de biți transmiși N_{TO} dacă se ajunge în starea de count time-out (blocul nu a fost achitat), la numărul de biți emiși pentru achitare în cele $q+1$ încercări trebuie adăugat numărul mediu de biți transmiși în cele $q+1$ încercări nereușite înmulțit cu probabilitatea P_T a intrării în CTO, primul termen al relației (19.a), și numărul de biți care N_T care ar fi putut fi transmiși pe durata T_T a CTO, al doilea termen al relației (19.a). N_T este exprimat de relația (19.b) în care parametrul d arată numărul de blocuri care s-ar fi putut transmite pe durata count time-outului.

$$\begin{aligned} N_{TO} &= (q + 1)(L + S + N_d) \cdot P_T + N_T \cdot P_T; \text{ a.} \\ N_T &= d \cdot (L + S + N_d); \text{ d} \in \mathbb{R}; \text{ b.} \end{aligned} \quad (19)$$

- Deci numărul mediu total N_t de biți transmiși pentru achitarea unui bloc de date este:

$$\begin{aligned} N_t &= (L + S + N_d) \cdot \left(\sum_{j=0}^q (j + 1) P_0 P_R^j \right) + [(L + S + N_d)(q + 1) + N_T] \cdot P_R^{q+1} = \\ &= (L + S + N_d) \left[\sum_{j=0}^q (j + 1) \cdot P_R^j \cdot P_0 + (q + 1) \cdot P_R^{q+1} + d \cdot P_R^{q+1} \right]; \end{aligned} \quad (20)$$

- Eficiența protocolului, raportul dintre numărul mediu de biți utili achitați într-un ciclu transmisie (încercarea inițială, q retransmisii și CTO) și numărul mediu total N_t de biți transmiși pentru achitarea a N_u biți este:

$$\begin{aligned} \eta_{SWI} &= \frac{N_u}{N_t} = \frac{(L + S)}{(L + S + N_d)} \cdot \frac{P_0 + \sum_{j=1}^q P_R^j \cdot P_0}{[P_0 + \sum_{j=1}^q (j + 1) P_R^j P_0 + (q + 1) \cdot P_R^{q+1} + d P_T]} = \\ &= \frac{L + S}{L + S + N_d} \cdot \frac{P_0(1 - P_R^{q+1})}{1 - P_R^{q+1} + d \cdot P_0 \cdot P_R^{q+1}} = \eta_i \eta_t; \end{aligned} \quad \text{justificați relația!} \quad (21)$$

- **Eficienței trebuie să i se atașeze probabilitatea de neachitare a unui pachet $P_T = P_R^{q+1}$.**

- **În cazul aplicațiilor care nu acceptă count timeout, se consideră că $d = 0$, iar P_T este probabilitatea de pierdere a unui pachet (bloc de date) datorită transmisiei (nu congestiei!), $BLER = P_T$.**

- Dacă se ia în considerare și probabilitatea medie de congestie P_{cg} atunci eficiența tehnicii ARQ se calculează la fel ca mai sus, dar P_0 (12) și P_R (13), vor fi înlocuite cu P_{0cg} (12') și P_{Rcg} (13')

- În cazul în care nu avem count timeout (aplicații de tip RT) atunci $d = 0$, iar eficiența protocolului (21) nu mai depinde de numărul de retransmisii; cu creșterea lui q probabilitatea de pierdere a pachetului scade, iar întârzierea introdusă crește.

- Expresia eficienței poate fi descompusă în doi factori:

- eficiența intrinsecă η_i , care depinde de elementele fixe ale legăturii;
- eficiența transmisiei η_t , care depinde de calitatea transmisiei (de probabilitatea de eroare de bit p) și de numărul de retransmisii q , vezi considerentele de mai sus.

- Pentru un număr infinit de retransmisii:

$$\lim_{q \rightarrow \infty} \eta_t = P_0 \quad (22)$$

- Pentru un protocol cu q foarte mare: $P_T \rightarrow 0 \Leftrightarrow N_T = 0$ și $d = 0$. Rezultă că:

$$\eta_{SW_\infty} = \frac{L+S}{L+S+N_d} \cdot P_0 = \eta_i \eta_{t_\infty} \quad (23)$$

- Deci eficiența protocolului crește cu creșterea lui q , dar pentru $q > 3$ creșterile eficienței sunt neglijabile, dar e necesară creșterea bufferului la recepție pentru protocoalele SR și GBn.

- Numărul de retransmisii trebuie însă stabilit și în funcție de întârzierea introdusă, care trebuie să fie mai mică decât întârzierea maximă impusă de aplicație

- Considerând că sursa emite cu un debit binar nominal constant al transmisiei D , debitul binar efectiv (throughput) Θ se calculează ca produsul dintre D și eficiența η .

- Transmisia neguvernata de un protocol ARQ are eficiența intrinsecă unitară (nu se transmite mesaj de confirmare și nu există pauza dintre cele două blocuri), iar eficiența transmisiei este egală cu P_0 , adică:

$$\eta_{\text{non-ARQ}} = P_0 \quad (24)$$

- Comparând relațiile (21) și (23) cu (24) și ținând cont de relația (8), putem spune că throughputul unei transmisii guvernate de un protocol ARQ este mai mic decât cel al transmisiei neguvernate de acest protocol:

$$\Theta_{SW} = \eta_{SW} \cdot D < \eta_{\text{non-ARQ}} \cdot D = \Theta_{\text{non-ARQ}} \quad (25)$$

- Pe baza relației (25) rezultă că și eficiența spectrală a transmisiei ARQ e mai mică decât cea a transmisiei non-ARQ, în ipoteza în care ele au același debit nominal și deci folosesc aceeași bandă de frecvență.

- Scăderea throughputului și a eficienței spectrale este compensată creșterea probabilității de recepționare corectă a pachetelor, egală cu $(1 - P_R^{q+1})$, care poate fi adusă foarte aproape de 1, prin creșterea numărului de retransmisii, adică $q \rightarrow \infty$ (echivalent cu $P_T \rightarrow 0$).

- Timpul mediu de achitare, cu probabilitatea $(1-P_T)$ a unui bloc este: $t_m = \frac{L}{\eta_{sw} \cdot D}$ (26)

- Întârzierea medie introdusă de utilizarea protocolului ARQ se calculează ca diferența dintre timpul mediu de achitare și timpul necesar pentru efectuarea primei transmisii:

$$\tau_m = t_m - t_0 = \frac{L}{D} \left(\frac{1}{\eta_{sw}} - 1 \right) \quad (27)$$

- Variația întârzierilor în jurul valorii medii ia valori egale cu multiplii întregi ai timpului necesar transmiterii unei perechi $(L+S)$; fiecărei valori a întârzierii i se poate asocia o probabilitate de apariție, care depinde de probabilitățile P_j de achitare a unui bloc după retransmisia j . Această exprimare permite și calculul dispersiei acestei variabile aleatoare. Un alt mod de calcul a întârzierii medii va fi prezentată în capitolul H-ARQ.

- Un aspect care prezintă interes este dimensionarea bufferului de la emisie astfel încât să nu apară fenomenul de pierdere a blocurilor („discard”) datorită umplerii acestui buffer.

- Dacă presupunem că blocurile de lungime L biți sosesc cu un debit mediu de B blocuri/secundă și că transmisia cu SW-ARQ are debitul D și eficiența η_{sw} , iar mesajul de confirmare are S biți, atunci condiția pentru evitarea pierderii blocurilor este:

$$B \leq \eta_{sw} \cdot D / (L+S) \quad (28)$$

- Trebuie însă prevăzut un buffer care să permită stocarea a $q+1$ blocuri pentru cazul cel mai defavorabil în care apar q retransmisii ale aceluiași bloc.

- Dacă blocurile cu lungime L biți sosesc în salve de câte N , cu un debit de B blocuri/secundă ($B > \eta_{sw} \cdot D / L$), urmate de o pauză t_u (vezi distribuția Poisson care modelează traficul de pachete) și dacă la începutul sosirii salvei bufferul este gol, atunci dimensiunea w_t a bufferului de emisie trebuie să satisfacă condiția:

$$N - \eta_{sw} \cdot \frac{D}{L} \left(\frac{N}{B} + t_u \right) \leq w_t \quad \text{jusificați relația!} \quad (29)$$

- Deoarece valoarea de mai sus se obține considerând valoarea medie a throughputului de transmisie, pentru a nu pierde pachete în cazul intrării în count time-out, la dimensiunea minimă w_t determinată cu (29) mai trebuie adăgate „acoperitor” încă $q+1$ blocuri.

- Pentru ca relația (29) să rămână valabilă și la următoarea salvă, trebuie determinată durata minimă a pauzei t_u dintre salve, a. î. bufferul să se poată goli în cazul cel mai defavorabil când el a rămas plin la sfârșitul salvei:

$$t_u > \frac{w_i \cdot L}{\eta_{sw} \cdot D} \quad \text{jusificați relația!} \quad (30)$$

- Relația (29) permite și calcul numărului maxim de blocuri ce pot veni într-o salvă, cu debit de blocuri B și L biți/bloc, pentru a nu „umple” bufferul de dimensiune w_t , dacă debitul efectiv mediu (în biți) la nivelul fizic al legăturii studiate este $\Theta_{sw} = \eta_{sw} \cdot D$.

- pentru ca bufferul să rămână “gol” la sfârșitul pauzei t_u , trebuie impusă condiția ca debitul mediu de pachete la intrarea în buffer să fie mai mic decât (sau egal cu) debitul mediu de pachete evacuat din buffer, pe durata sosirii salvei și durata pauzei dintre salve, adică:

$$\left(\eta_{sw} \cdot \frac{D}{L} - \frac{N}{N/B + t_u} \right) \geq 0 \quad \text{jusificați relația!} \quad (31)$$

5.2. Protocol SR

- Acest tip de protocoale necesită o legătură FD, deci $t_c = 0$, $t_i = 0$, $N_d = 0$ ceea ce conduce la creșterea eficienței intrinseci și implicit a eficienței globale.

- Relația între η_{i-sw} a protocolului SW, și η_{i-sr} a protocolului SR este exprimată de (32).

$$\eta_{i-sr} = \eta_{i-sw} \frac{L + S + N_d}{L + S} = \eta_{i-sw} \left(1 + \frac{N_d}{L + S} \right) \quad (32)$$

- Valorile factorului din (32) pentru valori ale lui L și pentru $S = 32$ biți și $q=3$ sunt prezentate în tabelul 1

- Eficiența transmisie η_t nu se modifică față de protocolul SW și deci expresia eficienței protocolului de tip SR pe o transmisie FD este:

$$\eta_{SR} = \eta_{SW} \left(1 + \frac{N_d}{L + S} \right); \quad (33)$$

$N_d \setminus L \rightarrow$	1752	1016	240
50	1,02	1,05	1,18
500	1,28	1,48	2,83
5000	3,80	5,78	19,38

Tabelul 1 Coeficienții η_{iFD}/η_{iHD}

- timpul mediu de achitare a unui bloc și întârzierile introduse de ARQ-SR se calculează similar cu cele ale ARQ-SW, dar în (26) și (27) η_{sw} se înlocuiește cu η_{sr} .

5.3. Protocol GBN *studiu individual* ↓

- Eficiența acestui protocol este analizată în ipoteza unei legături FD, parametrii legăturii rămânând aceiași;

- numărul de blocuri ce se retransmit după confirmarea negativă a unui bloc este N .

- Probabilitățile P_0 , P_R , P_j , P_{Rj} , P_T sunt identice cu cele calculate pentru SW:

$$P_0 = (1 - p)^{L+S}; \quad P_R = 1 - P_0; \quad P_j = P_R^{j-1} \cdot P_0; \quad P_{Rj} = P_R^{j+1}; \quad P_T = P_R^{q+1}; \quad (34)$$

- Numărul mediu total N_t de biți transmiși pentru a achita pentru a achita $L+S$ biți:

$$\begin{aligned} N_t &= (L+S)P_0 + (N+1)(L+S)P_0P_R + \dots + (jN+1)(L+S)P_0P_R^j + \dots + (qN+1)(L+S)P_0P_R^q + P_R^{q+1}[(q+1)N+1](L+S) + N_T] = \\ &= (L+S)P_0[1 + (N+1)P_R + \dots + (jN+1)P_R^j + \dots + (qN+1)P_R^q] + P_R^{q+1}(L+S)[(q+1)N+1+d] = \\ &= \frac{(L+S)[1 + (N-1)P_R + qP_R^{q+1} - (q+N)P_R^{q+2} + d'P_0P_R^{q+1}]}{P_0}; \quad d' = \frac{N_T}{L+S}; \end{aligned} \quad (35)$$

- Considerând $N_u = L+S$ biți, eficiența protocolului GBN va fi:

$$\eta_{GBN} = \frac{L+S}{L+S} \cdot \frac{P_0}{1 + (N-1)P_R + qP_R^{q+1} - (q+N)P_R^{q+2} + d'P_0P_R^{q+1}} = \eta_i \cdot \eta_{t-GBN} \quad (36)$$

- Relațiile (35) și (36) sunt valabile dacă este asigurată condiția de continuitate a transmisiei, adică dacă timpul de propagare/procesare este mai mic decât durata necesară transmiterii unui pachet de N blocuri:

$$N_d < N(L+S); \quad (37)$$

- Dacă condiția (37) nu e îndeplinită, atunci în pentru calculul eficienței va trebui considerat și numărul de biți ce s-ar putea transmite în pauza ce apare, notat cu N'_d , și vom înlocui în (36) η_i și d , cu η'_i și respectiv d'' :

$$N'_d = N_d - N(L+S); \quad \eta'_i = \frac{L+S}{L+S+N'_d}; \quad d'' = \frac{N_T}{L+S+N'_d}; \quad (38)$$

- Variația eficienței protocolului GBN în funcție de numărul de retransmisii este similară cu cea protocolului SW, vezi (21) și (23), dar valorile între care variază această eficiență sunt mai mici

- timpul mediu de achitare a unui bloc și întârzierile introduse de ARQ-GBN se calculează similar cu cele ale ARQ-SW, dar în (26) și (27) η_{sw} se înlocuiește cu η_{GBN} .

Comparație între eficiențele celor trei tipuri de protocoale

- În figurile 2 a și b se prezintă eficiențele celor trei tipuri de protocoale pentru aceiași parametri, $L=1752$ (219 bytes), $S = 32$ (4 bytes), $q = 3$, $d = 2$, $N = 16$, în ipoteza că se asigură continuitatea transmisiei; pentru protocoalele SR și GBN, $N_d = 0$, iar pentru protocolul SW, $N_d = 50$, fig.a, și $N_d = 3500$, fig.b.

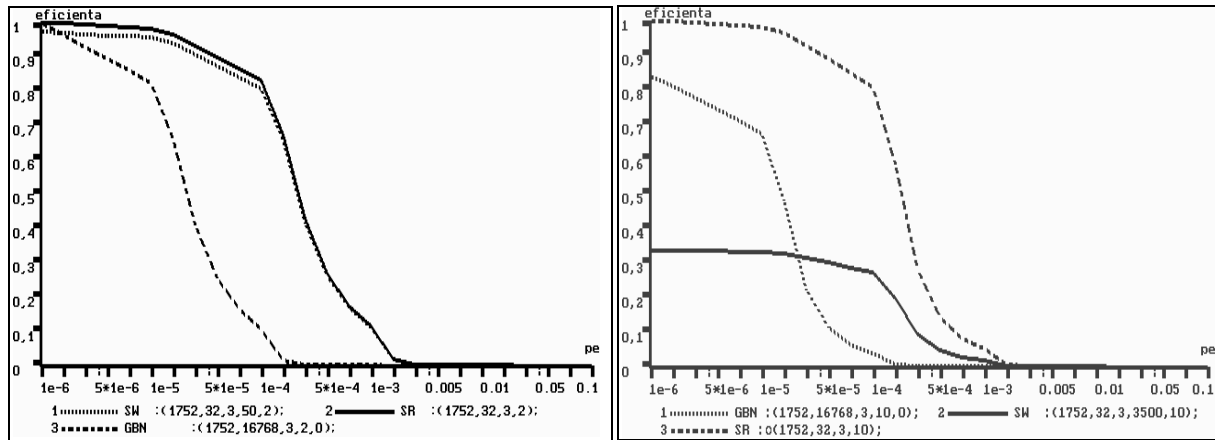


Fig.2. Eficiențele protocoalelor SR, SW și GBN pentru: a. $N_d = 50$ și $d = 2$; b. $N_d = 3500$ și $d = 10$

N_d - SW mic:

- Eficiențele $\eta_{SW} \approx \eta_{SR}$; diferența dintre ele se datorează „mascării” $N_d=0$, pentru FD-SR.
- Eficiențele $\eta_{GBN} < \eta_{SW} \approx \eta_{SR}$; la GBN retransmiterea a N blocuri la eronarea unui bloc reduce semnificativ eficiența \rightarrow GBN nerecomandabil.

- N_d - SW mare:

- $\eta_{SR} > \eta_{SW}$;
- $\eta_{SW} < \eta_{GBN}$ pentru valori mici ale lui p ;
- $\eta_{SW} > \eta_{GBN}$ pentru valori medii și mari ale lui p .

- Dar pachetul L este compus din antet H și date I , $L = I + H$, deci se poate defini o eficiență „interioară” a protocolului, η_p :

$$\eta_p = I/(I+H); \quad (39)$$

- considerând și eficiența interioară, eficiența globală a unui protocol poate fi exprimată ca produsul celor 3 eficiențe parțiale:

$$\eta = \eta_p \cdot \eta_i \cdot \eta_c; \quad (40)$$

5.4 Studiul influenței lungimii blocului asupra eficienței protocoalelor

- Studiul eficienței în funcție de p având ca parametru lungimea L a blocului, valorile S , N_d , N_T , N , q rămânând fixe, are ca scop determinarea lungimii blocului L_{opt} pentru care eficiența este maximă.

- modificarea lungimii blocului are două efecte contradictorii asupra eficienței:

- mărirea lungimii L conduce pe de o parte la creșterea lui η_i ,
- iar pe de altă parte la scăderea lui η_c , prin scăderea lui P_0 și, implicit creșterea lui P_R

- analiza teoretică a influenței lungimii blocului presupune exprimarea eficienței (21) numai în funcție de P_0 (12), folosind relația (13). Expresia obținută este:

$$\eta_{SW} = \frac{L + S}{L + S + N_d} \cdot \frac{P_0 \cdot [1 - (1 - P_0)^{q+1}]}{1 - (1 - P_0)^{q+1} + d \cdot P_0 \cdot (1 - P_0)^{q+1}}; \quad (21')$$

- Această expresie trebuie apoi derivată și înmulțită cu derivata lui P_0 în funcție de L , folosind relația (12):

$$\frac{d\eta}{dL} = \frac{d\eta}{dP_0} \cdot \frac{dP_0}{dL} = \frac{d\eta}{dP_0} \cdot P_0 \cdot \ln(1-p) \quad (41)$$

- Pentru a determina valoarea lui L pentru care eficiența e maximă, și valoarea acestui maxim, ar trebui arătat că funcția obținută are un maxim (derivata I are cel puțin o soluție, iar derivata II e negativă) și găsită soluția ecuației $d\eta/dL = 0$.

- Calculele sunt foarte laborioase iar rezolvarea ecuației menționate se poate face fie prin simplificarea formei analitice utilizând aproximări, fie prin metode numerice.

- De aceea prezentăm doar unele evaluări ale eficienței pentru câteva lungimi ale blocului de date.

- Analiza e făcută pentru protocoale de tip SW-HD, SR-FD și GBN-HD, vezi respectiv vezi figurile 3.a,b,c – S-a considerat că $\eta_{SR} = (1+N_d/(L+S))\eta_{SW}$, iar η_{GBN} e dată de (36).

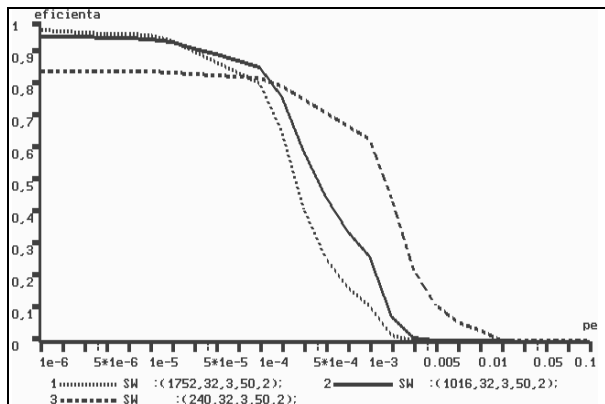


Fig.3.a Variatia eficientei SW cu L

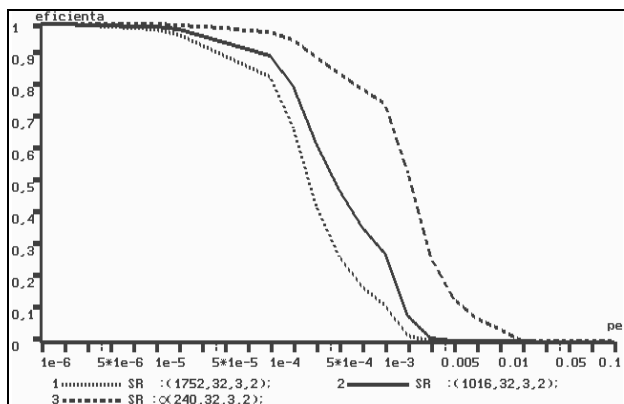
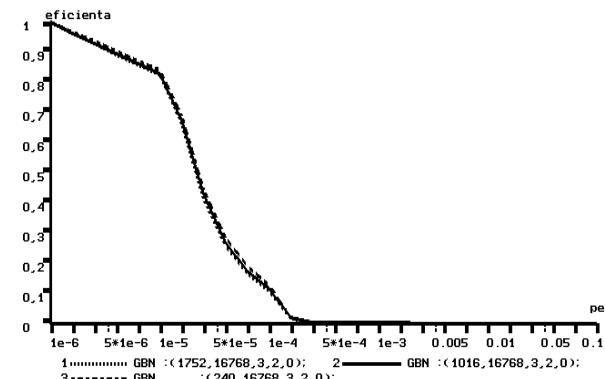


Fig.3.b Variatia eficientei SR cu L



← Fig.3.c Variatia eficientei GBN cu L

- SW-HD:

- L mare → η_{SW} mare pentru $p < p_0 (=2 \cdot 10^{-4})$;
- L mediu → η_{SW} maxim pentru $p > p_0$ - dar eficiența e mai mică.
- L mică → crește η_t , dar scade η_i (datorită lui N_d -HD); în plus η_p (39) scade și afectează semnificativ eficiența globală.

- SR-FD:

- L medii și mici sunt preferabile deoarece

- L mici trebuie ținut cont de efectul lui η_p .

- GBN: η_{GBn} depinde puțin de L; domeniul lui p e mai restrâns decât cele ale SR sau SW.

Concluzie:

- Pe canale cu probabilități medii și mari de eroare de bit, cum ar fi canalele radio fixe și mai ales mobile, utilizarea blocurilor lungi (L mare) nu este recomandabilă deoarece probabilitatea ca ele să fie recepționate corect scade odată cu creșterea lungimii blocurilor.

- Deoarece pachetele IP sunt lungi și nu pot fi transmise ca atare pe canale radio ele sunt divizate („segmentate”) la nivelul stratului fizic sau MAC în blocuri de lungimi mai mici, care sunt transmise folosind un protocol H-ARQ, care include un cod corector de erori, vezi capitolul următor. Transmisia întregului pachet IP este guvernată de un al doilea ARQ aflat în straturile superioare.

5.5. Studiul influenței timpilor de întârziere introduși de legătura de transmisie *studiu individual* ↓

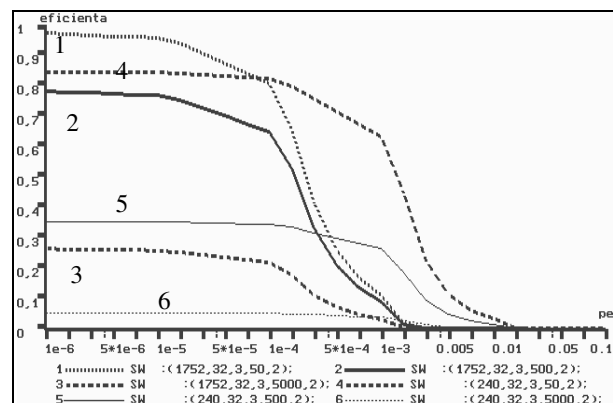


Fig. 4 Variatia eficientei SW cu N_d - L=1752 si L=240

- Pentru a studia efectul timpilor de întârziere introduși de canal vom considera doar protocolul SW;

- În cazul protocolului SR eficiența nu este afectată de timpii de întârziere, datorită transmisiei FD și datorită continuității transmisiei

- În cazul protocolului GBN, dacă este respectată condiția de continuitate, efectul lui N_d e nul;

- În figura 4 sunt prezentate curbele de variație ale η_{SW} în funcție de p cu N_d parametru pentru L=1752 sau L=240 și S=32, q=3, d=2 constante.

- η_{SW} scade odată cu creșterea lui N_d , pentru aceeași L, vezi curbele 1,2,3 și respectiv 4,5,6;

- Pentru N_d mic - lungimile mari sunt preferabile, dar trebuie ținut cont de efectul lui p.

- Pentru N_d mediu, curbele 2 si 5, pentru $p < 5 \cdot 10^{-4}$, L mare preferabil; pentru $p > 5 \cdot 10^{-4}$ - L mică de preferat, dar eficiența e mică, $\eta < 0,3$.

* N_d mare curbele 3 și 6 - L mare preferabil - η mică.

Concluzii

- Pentru toate situațiile trebuie asigurat ca $T_0 >$ timpul de propagare pe legătura utilizată.
- În ipotezele respectării condițiilor de continuitate a transmisiei, eficiențele legăturilor ce folosesc protocoale de tip SR sau GBN nu sunt influentate de timpul de propagare pe canal și de timpii de procesare. În cazul protocolului GBN, poate apărea o influență dată de (38).
- Eficiența legăturilor ce folosesc protocolul de tip SW este afectată de variația acestor timpi.
- Pentru protocoalele de tip SW, pe legăturile cu timpi mari de propagare, se folosesc blocuri de lungime mare, dacă p este mică. Dacă p este mare, se preferă scurtarea blocurilor până la găsirea unei valori optime, pentru a asigura un compromis între scăderea eficiențelor interioară și intrinsecă, pe de o parte, și creșterea eficienței transmisiei, pe de altă parte.
- Pe legăturile cu timpi mari de propagare este preferabilă utilizarea protocolului de tip SR.

Notă:

- eficiența a fost calculată în ipoteza în care fiecare retransmisie are loc cu aceeași modulație, deci debitul transmisiei rămâne constant la fiecare retransmisie; această ipoteză e valabilă pentru canale fixe.
- în cazul transmisiilor pe canale mobile modulațiile sunt utilizate adaptiv iar debitul legăturii se poate modifica la fiecare retransmisie (vezi probabilitățile de stare ale canalului), ceea ce va afecta atât eficiența transmisiei cât și întârzierile introduse de ARQ. De aceea analiza eficiența protocoalelor ARQ în cazul utilizării modulațiilor adaptive va fi analizată separat la laborator.

6. Câteva considerații privitoare la utilizarea protocoalelor de tip ARQ

- protocoalele de tip ARQ se utilizează sub forme asemănătoare cu cele descrise mai sus în cadrul mecanismului Hybrid-ARQ (H-ARQ) inclus în nivelul PHY sau în nivelul MAC al comunicațiilor pe canale radio. Aspecte privind mecanismele H-ARQ vor fi tratate în cursul dedicat protocoalelor H-ARQ.
- mecanisme de tip ARQ mai sunt utilizate pe scară largă în stratul Transport, în diversele variante ale TCP.

- mecanismele de tip ARQ utilizate în variantele TCP prezintă câteva diferențe semnificative față de cele analizate mai sus. Unele dintre acestea ar fi:

- nu se utilizează mesajul NACK. Aceasta deoarece la elaborarea lor au fost luate în considerare doar canale de transmisie de bună calitate, iar debitul solicitat nu se apropia de limita (capacitatea) canalului respectiv, ceea ce făcea ca modulația utilizată să permită asigurarea unor valori extrem de mici, neglijabile, ale PER (BLER). În această situație, principala cauză de pierdere a pachetelor era congestia în routere. Aceasta se traducea prin „nesosirea” mesajului de ACK, și producerea evenimentului de „timeout”. În consecință evenimentul de timeout se generează atât la pierderea pachetului în routere, cât și la recepționarea eronată a acestuia, care este interpretată ca o congestie, fiind de fapt o „falsă congestie”.
- modificarea ferestrei de congestie, la apariția timeout-ului, care are ca efect modificarea debitului de pachete, în funcție de frecvența de apariție a timeout-ului
- nu permit modificarea lungimii pachetului (segmentului) în cadrul aceleiași sesiuni, deoarece pe canalele fixe de foarte bună calitate configurația de transmisie este practic constantă

- calculul debitului efectiv (și al eficienței) diverselor variante ale TCP este mai dificil, dar poate fi realizat într-o manieră similară cu cea descrisă mai sus. În literatura de specialitate există mai multe abordări ale calculului acestui debit.

- datorită celor menționate mai sus, variantele TCP prezintă câteva dezavantaje pe canalele de transmisie ce asigură o probabilitatea de eroare de bit suficient de mare pentru a genera o probabilitate de eroare de pachet (segment) care nu mai poate fi neglijată, aprox. 10%, cum ar fi canalele radio fixe sau mobile, precum și unele canale pe fir care în care debitul transmisiei se apropie de capacitatea (limita) aceluși canal.

- falsa congestie provoacă micșorarea ferestrei de congestie, ceea ce conduce la scăderea nejustificată a debitului de pachete transmise și implicit la utilizarea mai puțin eficientă a legăturii respective.

- păstrarea constantă a lungimii pachetului (sau modificarea extrem de greoaie a acesteia), poate conduce la creșterea PER, cel puțin pe o perioadă de timp. fapt ce conduce la rândul lui la scăderea eficienței, vezi paragraful anterior.

- pentru eliminarea acestor neajunsuri și pentru a nu modifica structura existentă, pe canalele cu probabilitate de eronare mai ridicată (radio, radio mobile) au fost introduse protocoale H-ARQ, la nivelul fizic sau la nivelul MAC; acestea au rolul de a reduce PER la valori extrem de reduse (pentru 5G aceste probabilități pot ajunge la 10^{-5}), și astfel la straturile superioare, legătura se prezintă ca una de foarte bună calitate.

- dar mecanismele H-ARQ implică utilizarea unor resurse suplimentare de procesare, transmiterea unor biți suplimentari și introduc unele întârzieri suplimentare.
- de aceea, utilizarea combinației TCP (strat transport) + H-ARQ (strat fizic sau MAC) s-ar putea să nu fie soluția optimă, deși în prezent este răspândită pe scară largă.
- o altă abordare implică utilizarea codurilor rateless în locul retransmisiilor în cadrul protocolului de tip TCP;
- astfel se mărește eficiența, prin adaptarea redundanței la starea canalului
- această abordare este considerată în literatură ca fiind prea “agresivă” în cazul apariției congestiilor în routere