

Jakub BIAŁOGRODZKI

Politechnika Gdańska

PROBABILISTYCZNA GENERACJA SELEKTORÓW DLA ALGORYTMÓW KOMUNIKACJI GRUPOWEJ W SIECIACH RADIO- WYCH

Streszczenie. Selektory to obiekty matematyczne wykorzystywane przez kilka zaawansowanych algorytmów komunikacji grupowej dla sieci radiowych. Należy do nich algorytm Chrobak-Gąsieniec-Rytter (CGR), asymptotycznie najefektywniejszy algorytm rozsyłania dla sieci radiowych o nieznaanej topologii. Artykuł proponuje prostą, probabilistyczną metodę konstrukcji odpowiednich rodzin selektorów oraz prezentuje wyniki symulacji CGR mających na celu ustalenie, w jakim stopniu wybór konkretnej rodziny wpływa na efektywność algorytmów.

RANDOMIZED GENERATION OF SELECTORS FOR GROUP COMMUNICATION ALGORITHMS IN RADIO NETWORKS

Summary. Selectors are mathematical objects used by several advanced group communication algorithms for radio networks. Those include Chrobak-Gąsieniec-Rytter (CGR) algorithm, which is asymptotically the most efficient broadcasting algorithm for unknown topology radio networks. The article proposes randomized method of appropriate selectors and selector families construction, as well as presents the results of CGR simulations carried out to determine whether the choice of selectors family affects the efficiency of algorithms.

1. Wprowadzenie

Teoretyczne modele sieci radiowych pojawiły się po raz pierwszy w połowie lat 80 [3, 8]. Podstawowymi elementami sieci radiowych są urządzenia, mające zdolność wysyłania i odbierania, nazywane węzłami (*nodes*). W zbiorze węzłów definiujemy relację zasięgu – węzeł v jest w zasięgu węzła u , jeśli jest w stanie odebrać transmisję nadawaną przez u , przy czym relacja ta nie musi być symetryczna. Sieć modelujemy digrafem prostym, gdzie wierzchołki reprezentują węzły, a łuki opisują relację zasięgu. Komunikacja w modelu radiowym podzielona jest

na przedziały czasowe o jednostkowej długości - *rundy*. Bardziej szczegółowy opis modelu w wersji badanej przez autora można znaleźć np. w [5, 6].

Wiele uwagi poświęcono teoretycznej analizie problemów komunikacji grupowej w sieciach radiowych. Opracowano szereg algorytmów dla problemów takich jak rozsyłanie (*broadcasting*) czy plotkowanie (*gossiping*), przy czym powszechnie przyjętą miarą siły algorytmu jest pesymistyczna długość schematu komunikacyjnego (liczba rund).

Kilka zaawansowanych algorytmów korzysta z *selektorów*, stosunkowo prostych koncepcyjnie, lecz nietrywialnych w konstrukcji obiektów matematycznych. Zakres ich praktycznej przydatności obejmuje m. in. *problem pobudki* (*wake-up problem*) i *problemy ważenia monet* (*coin-weighting problems*) [9]. W kontekście sieci radiowych selektory zostały użyte po raz pierwszy w algorytmie rozsyłania Chrobak-Gąsieniec-Rytter (CGR) [6].

W rozdziale 2. opisana jest probabilistyczna metoda konstrukcji selektorów i rodzin selektorów, które można wykorzystywać np. w algorytmie CGR. Pojawia się pytanie, na ile wybór konkretnej rodziny selektorów wpływa na efektywność algorytmu. W poszukiwaniu rozstrzygnięcia przeprowadzone zostały symulacje algorytmu CGR dla różnych rozmiarów i typów sieci. Założenia symulacji oraz ich wyniki są przedstawione i omówione w rozdziale 3.

2. Probabilistyczna generacja selektorów

W rozdziale tym zaproponowany został algorytm Monte Carlo generacji selektorów. Jest on prosty i dość łatwy w implementacji. Pewnym problemem jest brak efektywnej metody sprawdzenia, czy wygenerowane obiekty są faktycznymi selektorami, co nie pozwala łatwo przekształcić tego algorytmu w algorytm Las Vegas - wymagamy ustalenia pewnego (dowolnie małego) prawdopodobieństwa, że wygenerowany obiekt nie będzie selektorem.

Definicja 1. Zbiór X *trafia* w zbiór S , jeśli $|X \cap S| = 1$. Jeśli $X \cap S = \emptyset$, to X *unika* S .

Definicja 2. Dany jest zbiór A . Rodzina zbiorów $\overline{S} = \{S_0, S_1, \dots, S_{m-1}\}$ jest ω -*selektorem* dla zbioru A , jeśli dla dowolnych rozłącznych zbiorów $X, Y \subset A$ spełniających warunki $\omega/2 \leq |X| \leq \omega$, $|Y| \leq \omega$ istnieje takie i , $0 \leq i < m$, że X *trafia* w S_i , a Y *unika* S_i . Liczba m to *długość* selektora; ω -selektor dla zbioru $\{0, 1, \dots, n-1\}$ nazywamy skrótowo (ω, n) -selektorem.

Konstrukcja selektorów o zadowalającej pod kątem zastosowań długości jest zadaniem na tyle nietrywialnym, że przez długi czas algorytmy wykorzystujące selektory podawane były w wersjach niekonstruktywnych, jedynie z probabilistycznym dowodem istnienia selektorów o określonych właściwościach.¹ W szczegól-

¹Problem istnienia algorytmu deterministycznego, generującego selektory o rozsądnej długości, pozostawał przez długi czas otwarty. Dopiero w 2002 roku Piotr Indyk podał w [9] deterministyczny algorytm generujący selektory o długości $O(\omega \log^{O(1)} n)$.

ności algorytm CGR został pierwotnie podany w wersji niekonstruktywnej, przy czym wykorzystano następujące twierdzenie:

Twierdzenie 1 ([6]). *Dla dowolnego $n > 0$ i dowolnej ω , $0 < \omega \leq n$, istnieje (ω, n) -sektor o długości $O(\omega \log n)$.*

Szkic dowodu. Rozpatrzmy następujący algorytm:

```

GENSEL( $n, \omega, m$ )
for  $i \leftarrow 0$  to  $m - 1$  do
     $S_i \leftarrow \emptyset$ ;
    for  $j \leftarrow 0$  to  $n - 1$  do
        with prob  $1/(\omega + 1)$  do
             $S_i \leftarrow S_i \cup \{j\}$ ;
return  $\bar{S} = \{S_0, S_1, \dots, S_{m-1}\}$ ;
    
```

Można wykazać, że:

$$\text{pr}[\bar{S} \text{ nie jest selektorem}] \leq (31/32)^m \omega^2 n^{2\omega}.$$

Widzimy, że chcąc uzyskać niezerowe prawdopodobieństwo tego, że generowana przez GENSEL rodzina rzeczywiście jest selektorem, wystarczy wybrać takie m , aby spełniona była nierówność

$$(31/32)^m \omega^2 n^{2\omega} < 1. \quad (1)$$

Przekształcając mamy:

$$\frac{2}{\log(32/31)} (\log \omega + \omega \log n) < m.$$

Zatem wywołując GENSEL z

$$m = \left\lceil \frac{2}{\log(32/31)} (\log \omega + \omega \log n) \right\rceil = O(\omega \log n)$$

otrzymamy z niezerowym prawdopodobieństwem faktyczny selektor. \square

Rozumowanie przeprowadzone w dowodzie twierdzenia 1 można łatwo zmodyfikować tak, aby otrzymać algorytm Monte Carlo, generujący selektor o długości $O(\omega \log n)$ z prawdopodobieństwem przynajmniej $1 - \epsilon$ dla dowolnego ustalonego z góry $\epsilon > 0$.

Twierdzenie 2. *Dla dowolnego $\epsilon > 0$ istnieje algorytm Monte Carlo generujący dla dowolnych n i ω spełniających warunki $n > 0$, $0 < \omega \leq n$ rodzinę zbiorów \bar{S} o mocy $O(\omega \log n)$, która z prawdopodobieństwem $1 - \epsilon$ jest (ω, n) -selektorem.*

Dowód. Nierówność (1) z dowodu twierdzenia 1 zastąpmy nierównością

$$(31/32)^m \omega^2 n^{2\omega} < \epsilon.$$

Po identycznych przekształceniach otrzymujemy:

$$m > \frac{2(\log \omega + \omega \log n) - \log \epsilon}{\log(32/31)}.$$

Wystarczy zatem użyć algorytmu GENSEL z

$$m = \left\lceil \frac{2(\log \omega + \omega \log n) - \log \epsilon}{\log(32/31)} \right\rceil. \quad (2)$$

□

Niestety, przekształcenie algorytmu Monte Carlo w algorytm Las Vegas nie jest w tym przypadku oczywiste, gdyż brakuje efektywnych metod weryfikacji, czy dana rodzina zbiorów jest selektorem.

Rodziny selektorów. W algorytmie CGR używa się w rzeczywistości całych rodzin selektorów o określonych właściwościach. Dla ustalonego rozmiaru sieci $n = 2^k$ interesuje nas rodzina $\overline{\mathcal{S}} = \{\overline{\mathcal{S}}_0, \overline{\mathcal{S}}_1, \dots, \overline{\mathcal{S}}_k\}$, gdzie $\overline{\mathcal{S}}_i$ jest $(2^i, n)$ -selektorem o długości $O(2^i \log n)$. Zakładając, że do generowania pojedynczych selektorów użyjemy algorytmu GENSEL, przeanalizujemy, jakie wartości parametru m są konieczne, aby cała rodzina składała się z rzeczywistych selektorów z prawdopodobieństwem przynajmniej $1 - \delta$, dla ustalonej $\delta > 0$. Niech ϵ będzie żądanym prawdopodobieństwem porażki przy generacji pojedynczego selektora. Otrzymujemy:

$$\begin{aligned} (1 - \epsilon)^{k+1} &\geq 1 - \delta \\ 1 - \epsilon &\geq (1 - \delta)^{1/(k+1)} \\ \epsilon &\leq 1 - (1 - \delta)^{1/(k+1)} \end{aligned}$$

Aby wyznaczyć minimalne długości poszczególnych selektorów, wystarczy podstawić

$$\epsilon = 1 - (1 - \delta)^{1/(k+1)} \quad (3)$$

do równania (2). Okazuje się, że dla ustalonej δ nie zwiększamy w ten sposób asymptotycznej długości generowanych selektorów: $-\log \epsilon = O(\log n)$, a długość generowanych selektorów nadal możemy szacować przez $O(\omega \log n)$ (por. (2)).

GENSELFAMILY(k, δ)

$n \leftarrow 2^k$;

$\epsilon \leftarrow 1 - (1 - \delta)^{1/(k+1)}$;

for $i \leftarrow 0$ to k do

$\omega \leftarrow 2^i$;

$$m \leftarrow \left\lceil \frac{2(\log \omega + \omega \log n) - \log \epsilon}{\log(32/31)} \right\rceil;$$

$$\bar{S}_i \leftarrow \text{GENSEL}(n, \omega, m);$$

return $\{\bar{S}_0, \bar{S}_1, \dots, \bar{S}_k\}$;

Wniosek 1. Dla ustalonych $n = 2^k > 0$ i $\delta > 0$ algorytm GENSELFAMILY generuje rodzinę $\bar{S} = \{\bar{S}_0, \bar{S}_1, \dots, \bar{S}_k\}$, gdzie \bar{S}_i jest $(2^i, n)$ -selektorem o długości $O(2^i \log n)$, z prawdopodobieństwem przynajmniej $1 - \delta$.

Tabela 1

Sumy długości selektorów w rodzinach generowanych przez algorytm GenSelFamily

δ	$n = 4$	$n = 16$	$n = 64$	$n = 256$	$n = 1024$	$n = 4096$	$n = 16384$
0.1	1062	6463	35120	181338	897833	4297269	20037549
0.01	1282	6831	35637	182004	898647	4298229	20038659
0.001	1500	7195	36145	182660	899447	4299170	20039749
0.0001	1719	7557	36653	183310	900242	4300116	20040836

W tabelicy 1 zebrane zostały łączne długości rodzin generowanych przez algorytm GENSELFAMILY, dla różnych wartości n i δ . Jak widzimy, zmniejszanie prawdopodobieństwa błędu δ ma niewielki wpływ na długość selektorów. Jest to bardzo istotne w sytuacji, gdy nie dysponujemy efektywnymi metodami sprawdzenia, czy otrzymane rodziny \bar{S}_i rzeczywiście są selektorami.

3. Symulacje, dyskusja wyników

W poprzednim rozdziale pokazano probabilistyczną metodę generacji rodzin selektorów dla algorytmu CGR. Pesymistyczne oszacowanie długości schematu $O(n \log^2 n)$ pozostaje w mocy dla dowolnej rodziny, spełniającej dyskutowane wcześniej warunki długości pojedynczych selektorów. Nie wiemy jednak, czy użycie konkretnej rodziny wpływa na rzeczywistą efektywność algorytmu (a jeśli tak, to w jakim stopniu).

W celu rozstrzygnięcia tej kwestii przeprowadzone zostały symulacje algorytmu CGR. Możemy się spodziewać, że wyniki symulacji CGR pozostaną w mocy także dla innych algorytmów wykorzystujących selektory. Założenia symulacji były następujące:

- Badane rozmiary sieci: 64, 128, 256, 512, 1024.
- Dla każdego rozmiaru sieci wygenerowanych zostało 10 rodzin selektorów, spełniających warunki algorytmu CGR, oznaczonych A, B, \dots, J . Do generacji wykorzystany został algorytm GENSELFAMILY, z prawdopodobieństwem błędu $\delta = 0,001$. Łącznie wygenerowano 50 rodzin.

- Dla każdego rozmiaru sieci symulacje przeprowadzone zostały na sześciu digrafach: drodze symetrycznej (*path*), cyklu symetrycznym (*cycle*), digrafie pełnym (*complete*) oraz trzech digrafach modelujących rzeczywiste sieci radiowe (*reala*, *realb*, *realc*). Wykorzystano gotowe digrafy używane przez autora we wcześniejszych badaniach. Procedura konstruowania digrafów modelujących rzeczywiste sieci została opisana w [1]. Tutaj wspomnijmy jedynie, że digrafy tej klasy generowane były według następujących szablonów:
 - typ *a*: mała średnica, duże stopnie wejściowe,
 - typ *b*: średnica rzędu $n^{1/2}$, umiarkowane stopnie wejściowe,
 - typ *c*: średnica rzędu n , małe stopnie wejściowe.
- Dla każdej badanej pary (*digraf*, *rodzina selektorów*) przeprowadzono serię symulacji. W każdej symulacji w ramach serii losowana była numeracja wierzchołków i identyfikator źródła. Długość serii ustalana była dynamicznie, na podstawie otrzymanywnych wyników i zadanych parametrów ufności. Konkretnie:
 - Przyjęto poziom ufności $1 - \alpha$ przy $\alpha = 0,05$ i pożądaną względną długość przedziału ufności $\epsilon = 0,01$.
 - Po każdych 100 symulacjach wyznaczana była średnia długość schematu L_{avg} oraz aktualny przedział ufności $\langle U_{min}, U_{max} \rangle$. Seria kończyła się, jeśli długość przedziału ufności w stosunku do wartości średniej nie przekraczała ϵ , jednak nie później niż po przeprowadzeniu 20000 symulacji.
- Wszystkie symulacje zostały przeprowadzone na domowym komputerze PC autora.

Dla każdego badanego digrafu G zdefiniujemy dwie wartości: $L_{min}(G)$ i $L_{max}(G)$, odpowiednio średnią liczbę rund uzyskaną na tym digrafie dla "najlepszej" i "najgorszej" rodziny selektorów. Przyjmijmy $\lambda(G) = L_{max}(G)/L_{min}(G)$. Tak zdefiniowany parametr λ przyjmuje wartości z zakresu $\langle 1, \infty \rangle$ i dobrze opisuje wrażliwość algorytmu CGR na zmianę rodziny selektorów dla danego digrafu.

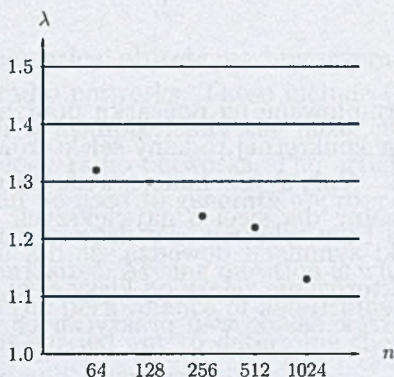
- Największą wartość $\lambda(2,05)$ otrzymujemy dla digrafu *complete128* ($L_{max} = 9,93$ dla rodziny G , $L_{min} = 4,85$ dla rodziny J). Wynik ten oznacza, że w pewnych przypadkach możemy przynajmniej dwukrotnie zwiększyć (lub zmniejszyć) efektywność algorytmu CGR, zmieniając rodzinę selektorów.
- Najmniejszą wartością $\lambda(1,18)$ charakteryzuje się digraf *reala1024* ($L_{max} = 46,58$ dla rodziny D , $L_{min} = 39,33$ dla rodziny F), jednak nawet w tym przypadku wartość jest istotnie większa od 1. Stąd wniosek, że algorytm CGR wykazuje sporą wrażliwość na wybór rodziny selektorów w każdych warunkach, przynajmniej dla badanych tutaj rozmiarów sieci.

- Wrażliwość jest istotnie różnicowana przez klasę sieci (typ digrafu). Ilustrującą to wartości zostały przedstawione w tabelicy 2. Jak widzimy, algorytm CGR jest najbardziej wrażliwy na zmianę rodziny selektorów dla sieci pełnych, natomiast najmniej dla sieci rzeczywistych typu *b*. Pozostałe klasy grafów charakteryzują się podobną wrażliwością.

Tabela 2
Wrażliwość algorytmu CGR na
wybór rodziny selektorów dla
poszczególnych klas sieci

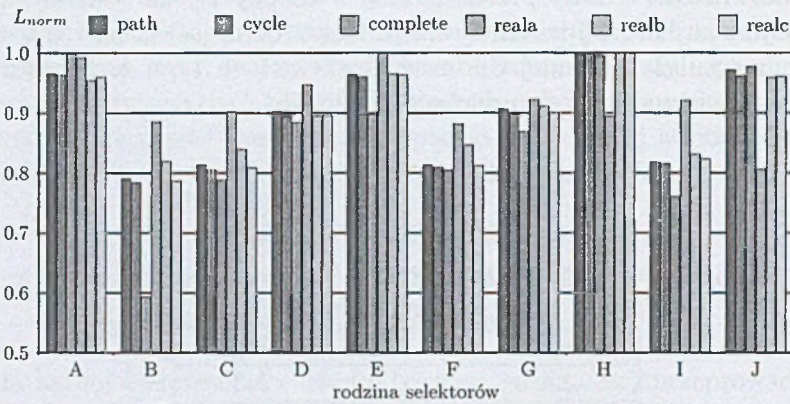
klasa	λ_{min}	λ_{max}
path	1,19	1,39
cycle	1,23	1,38
complete	1,66	2,05
real <i>a</i>	1,18	1,38
real <i>b</i>	1,13	1,32
real <i>c</i>	1,22	1,39

- Dla wszystkich klas sieci, z wyjątkiem sieci pełnych, obserwujemy wyraźny spadek wartości λ wraz ze wzrostem rozmiaru sieci. Oznacza to, że wybór "dobrego" selektora jest bardziej kluczowy w sieciach o mniejszym rozmiarze. Zależność tę ilustruje rysunek 1.



Rys. 1. Sieci rzeczywiste typu *b*: spadek wartości λ wraz ze wzrostem rozmiaru sieci

- Obserwujemy silną korelację w wynikach poszczególnych rodzin selektorów dla różnych klas sieci (przy ustalonym rozmiarze). Przykładową zależność tego typu przedstawia rysunek 2. Oznacza to, że jeśli dla pewnej klasy sieci rodzina *A* jest lepsza od rodziny *B*, to prawdopodobnie będzie lepsza również



Rys. 2. Znormalizowane średnie długości schematu osiągnięte dla poszczególnych rodzin selektorów (A-J) w poszczególnych klasach sieci, przy rozmiarze $n = 512$. Obserwujemy wyraźną korelację

dla innych sieci o tym samym rozmiarze. Wynik ten implikuje interesujące perspektywy w dziedzinie zastosowań praktycznych. Możemy sobie wyobrazić pulę starannie dobranych rodzin selektorów, zapewniających możliwie krótkie schematy dla poszczególnych rozmiarów sieci. Rodziny z takiej puli mogłyby być wykorzystywane wielokrotnie, w sieciach dowolnych typów.

4. Podsumowanie

Odpowiedź na sformułowane na początku poprzedniego rozdziału pytanie jest jednoznaczna. Wybór konkretnej rodziny selektorów ma duży wpływ na wydajność algorytmu CGR. Wpływ ten maleje wraz ze wzrostem rozmiaru sieci, pozostaje jednak zauważalny dla sieci o największych badanych tu rozmiarach (1024). Co więcej, wyniki symulacji dowodzą, że dla ustalonego rozmiaru sieci efektywność rodziny selektorów nie zależy od klasy sieci, co prowadzi do interesujących konkluzji w kontekście zastosowań praktycznych.

LITERATURA

1. Białogrodzki J.: Implementacja algorytmów rozsyłania i plotkowania w sieciach radiowych. Politechnika Gdańska 2004. Praca magisterska.
2. Bruschi D., Del Pinto M.: Lower bounds for the broadcast problem in mobile radio networks. *Distributed Computing*, 10, 1997, p. 129–135.
3. Chlamtac I., Kutten S.: On broadcasting in radio networks – problem analysis

- and protocol design. *IEEE Transactions on Communications*, 33, 1985, p. 1240–1246.
4. Chlebus B.S., Gąsieniec L., Ostlin A., Robson J.M.: Deterministic radio broadcasting. *Proc. 27th International Colloquium on Automata, Languages and Programming (ICALP)*, 2000, p. 717–728.
 5. Chlebus B.S., Gąsieniec L., Gibbons A.M., Pelc A., Rytter W.: Deterministic broadcasting in unknown radio networks. *Proc. 11th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2000, p. 861–870.
 6. Chrobak M., Gąsieniec L., Rytter W.: Fast broadcasting and gossiping in radio networks. *Proc. 41st IEEE Symposium on Foundations of Computer Science (FOCS)*, 2000, p. 575–581.
 7. De Marco G., Pelc A.: Faster broadcasting in unknown radio networks. 1999. Nie publikowane.
 8. Gallager R.: A perspective on multiaccess channels. *IEEE Transactions on Information Theory*, 31, 1985, p. 124–142.
 9. Indyk P.: Explicit constructions of selectors and related combinatorial structures with applications. *Proc. 13th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2002, p. 697–703.

Recenzent: Dr hab. inż. Adam Czornik

Abstract

Selectors are mathematical objects used by several advanced group communication algorithms for radio networks. Those include Chrobak-Gąsieniec-Rytter (CGR) algorithm, which is asymptotically the most efficient broadcasting algorithm for unknown topology radio networks. The article proposes simple Monte Carlo approach which can be used to generate appropriate selectors and selector families effectively. That approach was actually used to generate some selector families for various network sizes. Arising question is whether the arbitrary choice of selector family affects the performance of algorithms. Generated families were used in CGR simulations carried out to determine the answer. Results of those simulations are presented and discussed.