ホワイトリスト順序問題の計算困難性

高知工科大学 情報学群　原田 崇司 [1]
神奈川大学 理学部　田中 賢 [2]
新潟大学 情報基盤センター　三河 賢治 [3]

[1]Takashi Harada
School of Information, Kochi University of Technology
[2]Ken Tanaka
Department of Science, Kanagawa University
[3]Kenji Mikawa
Center for Academic Information Service, Niigata University

1　はじめに

パケット分類に関して，ルールリストの線型探索によって生じる遅延を最小化するルールの順序を求める最適化問題がある．この問題はルール順序最適化問題 (Optimal Rule Ordering, **ORO**) とよばれており，**NP** 困難であることが証明されている [3, 4]．また，この最適化問題に対して，種々の発見的解法や厳密解法が提案されている [6, 7, 1, 8, 5]．

本稿では，パケットに適用するアクションの種類を 1 種類だけに制限した最適化問題，ホワイトリスト順序最適化問題を導入する．更に，この問題の判定問題版が **NP** 困難であることを，**NP** 完全であることが知られている判定問題 **EXACT COVER BY 3-SETS** (**XC3**) [2] からの多項式時間帰着を示すことによって証明する．

2　パケット分類

ネットワーク機器に到着するパケットに対して，分類ポリシーに従ってアクションを適用することをパケット分類という．線型探索によるパケット分類は，図 1 のようにモデル化される．パケットはルールリストの上位のルールから順番に照合され，最初に合致したルールのアクションが適用される．

本稿で扱うホワイトリスト順序問題を説明するために，初めにルール順序最適化問題を説明する．

パケットを長さ l のビット列とする．ルールをルール番号 $i \in \{ 1, 2, \ldots, n \}$, 長さ l の条件式 $c \in \{0, 1, *\}^l$, パケットに適用するアクション $a \in \{A_1, A_2, \ldots, A_m\}$ の三つ組とする．ただし，n はルールリストが含むルールの数，m はアクションの種類数である．ルールリストの例を表 1 に示す．ルールの P と D は，それぞれ $Permit$ と $Deny$ を意味しパケット通過の許可と拒否とを表す．

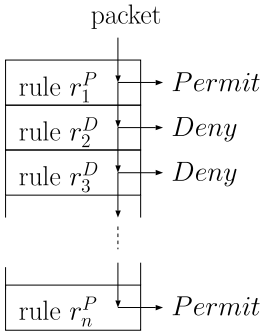

図 1: パケット分類モデル

表 1: ルールリスト

Filter $\mathcal{R}$	$\lvert E(\mathcal{R},i)\rvert_{\mathcal{U}}$
$r_1^D = 0\ 0\ 0\ 0$	1
$r_2^D = 0\ 0\ 0\ 1$	1
$r_3^D = 1\ 0\ 1\ 0$	1
$r_4^P = *\ 0\ 0\ *$	2
$r_5^P = 0\ 0\ *\ 1$	1
$r_6^P = 1\ *\ *\ 0$	2
$r_7^D = *\ *\ *\ *$	8
$L(\mathcal{R},\mathcal{U}) = 79$	

ルールリスト $\mathcal{R} = \langle\ r_1^{a_1}, r_2^{a_2}, \ldots, r_n^{a_n}\ \rangle$ は，パケットの集合 $\{0,1\}^l$ からアクションの集合 $\{A_1, A_2, \ldots, A_m\}$ への函数と見なせる．これより，$\mathcal{R}(p)$ によって，パケット p に対してルールリスト $\mathcal{R}$ が与えるアクションを表す．例えば，表 1 のルールリストによってパケット 0001 に与えられるアクションは D であり，$\mathcal{R}(0001) = D$ と表す．

ルールリスト $\mathcal{R}$ とルールの順序 σ が与えられると，ルール r_i^a によって適用されるアクションが決まるパケットの集合が定まる．この集合を $E(\mathcal{R}_\sigma, i)$ と表す．例えば，表 1 のルールリストと順序 id において，$E(\mathcal{R}, 4) = \{1000, 1001\}$ となる．0000 と 0001 は r_1^D と r_2^D によってアクションが決まるので $E(\mathcal{R}, 4)$ に含まれないことに注意されたい．

ネットワーク機器に到着するパケットの頻度分布をパケット全体の集合 $\{0,1\}^l$ から自然数全体の集合 $\mathbb{N}$ への函数とし，$\mathcal{F}$ で表す．

ルールリスト $\mathcal{R}$，ルールの順序 σ，頻度分布 $\mathcal{F}$ により，ルール r_i^a によってアクションが決まるパケットの数が定まる．この数を $\lvert E(\mathcal{R}_\sigma, i)\rvert_{\mathcal{F}}$ と表し，「評価パケット数」または「ルール重み」という．ルール重みはルール順序 σ に依存することに注意されたい．

ルールとパケットとの照合を遅延 1 と見做し，ルールリスト $\mathcal{R}$, 頻度分布 $\mathcal{F}$, ルール順序 σ におけるパケット分類による遅延を以下のように定義する．

定義 1.（パケット分類の遅延）

$$L(\mathcal{R}_\sigma, \mathcal{F}) = \sum_{i=1}^{n-1} i|E(\mathcal{R}_\sigma, \sigma^{-1}(i))|_{\mathcal{F}} + (n-1)|E(\mathcal{R}_\sigma, \sigma^{-1}(n))|_{\mathcal{F}}. \tag{1}$$

パケット分類による遅延を最小化するルールの順序を求める問題，ルール順序最適化問題を以下のように定義する．

定義 2.（ルール順序最適化問題）

入力: ルールリスト $\mathcal{R}$，頻度分布 $\mathcal{F}$
出力: 遅延 $L(\mathcal{R}_\sigma, \mathcal{F})$ が最小となる順序 σ
ただし，任意の p について $\mathcal{R}(p) = \mathcal{R}_\sigma(p)$.

更に，デフォルトルール $r_n^{a_n}$ 以外のアクションを 1 種類だけに限った最適化問題を考える．特に $Permit$ だけに限った以下の最適化問題を考える．この問題のことをホワイトリスト順序最適化問題とよぶ.

定義 3.（ホワイトリスト順序最適化問題）

入力: ルールリスト $\mathcal{R}$，頻度分布 $\mathcal{F}$
出力: 遅延 $L(\mathcal{R}_\sigma, \mathcal{F})$ が最小となる順序 σ
ただし，任意の p について $\mathcal{R}(p) = \mathcal{R}_\sigma(p)$．また，$a_n$ 以外のアクションは全て $Permit$.

この問題の判定問題版，ホワイトリスト順序問題を以下のように定義する．

定義 4.（ホワイトリスト順序問題, ***WHITELIST ORDERING, WO***）

入力: ホワイトリスト $\mathcal{R}$，頻度分布 $\mathcal{F}$,
　　正の整数 K
　問: $L(\mathcal{R}_\sigma, \mathcal{F}) \leq K$ となる順序 σ はあるか.
ただし，任意の p について $\mathcal{R}(p) = \mathcal{R}_\sigma(p)$.

本稿では，ホワイトリスト順序問題 (**WO**) が **NP** 困難であることを証明する．

3 EXACT COVER BY 3-SETS (XC3)

判定問題 **EXACT COVER BY 3-SETS** (**XC3**) とは次の問題である．

定義 5. ***EXACT COVER BY 3-SETS (XC3)***

入力: 集合 $S = \{s_1, s_2, \ldots, s_n\}$ と S の部分集合の族 $\mathcal{C} = \{C_1, C_2, \ldots, C_m\}$
ただし，S のサイズは 3 の倍数であり，任意の $1 \leq i \leq m$ について $|C_i| = 3$ である．
　問: 以下を満たす $\mathcal{C}$ の部分集合 $\mathcal{D}$ があるか？
　　$\bigcup_{C_i \in \mathcal{D}} C_i = S$, 任意の $C_i, C_j \in \mathcal{D} (i \neq j)$ について $C_i \cap C_j = \emptyset$.

この判定問題は **NP** 完全であることが知られている．以下に **XC3** の具体例を挙げる．

$$\begin{aligned} S &= \{1,2,3,4,5,6,7,8,9\} \\ \mathcal{C} &= \{\{5,7,9\},\{4,5,9\},\{3,6,7\},\{1,4,8\},\{7,8,9\},\{2,3,6\}\} \end{aligned} \tag{2}$$

このインスタンスに対しては，$\mathcal{D} = \{\ \{1,4,8\},\ \{2,3,6\},\ \{5,7,9\}\ \}$ と $\mathcal{C}$ の部分集合を取れるので，答えは Yes となる．これに対して，$\mathcal{C} = \{\ \{5,7,9\},\ \{4,5,9\},\ \{3,6,7\},\ \{3,4,8\},\ \{7,8,9\},\ \{1,2,3\}\ \}$ とすると，答えは No となる．

4　XC3 からホワイトリスト順序問題への帰着

定理 1. ホワイトリスト順序問題 ***(WO)*** は ***NP*** 困難である．

証明. **XC3** から **WO** への多項式時間帰着アルゴリズム f を示す．

f は，**XC3** のインスタンス $S, \mathcal{C}$ を入力として受け取り，以下のように定義される，**WO** のインスタンス $\mathcal{R}, \mathcal{F}, K$ を出力する．

ホワイトリストに含まれるルールのルール長 l を $|S|$ とし，ルール数 n を $|\mathcal{C}|+1$ とする．デフォルトルール r_n を除くルール $r_i = b_1 b_2 \cdots b_l$ の各ビットを

$$b_j = \begin{cases} `*' & \text{if } j \in C_i \\ `0' & \text{otherwise} \end{cases}$$

とする．$\mathcal{R}$ は，このようにして生成される $r_1, r_2, \ldots, r_n$ から成る．そして，頻度分布 $\mathcal{F}$ を式 (3) とする．

$$\mathcal{F}(p) = \begin{cases} 1 & \text{if } \exists! \, i \in \{1,\ldots,l\}\ p_i = `1' \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

ここで，p_i はパケット p の i ビット目のビットを表す．更に，$K = \sum_{i=1}^{N} 3i = \frac{3N(N+1)}{2}$ とする．ただし，$N = |S|/3$ である．

f が多項式時間で計算可能であることは明らかである．よって，以下では **XC3** のインスタンスを x と表したとき $x \in \mathbf{XC3} \iff f(x) \in \mathbf{WO}$ が成り立つことを示す．

(⇒) 集合 S の部分集合の族 $\mathcal{C}$ の部分集合 $\mathcal{D} = \{\ D_1, D_2, \ldots, D_m\ \}$ が S の厳密被覆になっていると仮定する．このとき，f によって生成されるルールリスト $\mathcal{R} = \langle\ r_1, r_2, \ldots, r_m, r_{m+1}\ \rangle$ と頻度分布 $\mathcal{F}$ に対して，頻度が 1 となる全てのパケットをルールリストの上位 $|S|/3$ 個のルール $r_{\sigma^{-1}(1)}, r_{\sigma^{-1}(2)}, \ldots, r_{\sigma^{-1}(|S|/3)}$ で評価するような順序 σ が存在する．それらのルールの評価パケット数はそれぞれ丁度 3 となるので，そのような順序での遅延は

$$L(\mathcal{R}_\sigma, \mathcal{F}) = \sum_{i=1}^{|S|/3} 3i = \frac{|S|(|S|/3+1)}{2} = K$$

となる．よって，$x \in \mathbf{XC3} \Rightarrow f(x) \in \mathbf{WO}$ が成り立つ．

(⇐) $f(x) \in$ **WO** $\Rightarrow x \in$ **XC3** の対偶，$x \notin$ **XC3** $\Rightarrow f(x) \notin$ **WO** を示す．

集合 S の部分集合の族 $\mathcal{C}$ の部分集合 $\mathcal{D} = \{ D_1, D_2, \ldots, D_m \}$ に対して，S の厳密被覆が存在しないと仮定する．このとき，f によって生成されるルールリスト $\mathcal{R} = \langle r_1, r_2, \ldots, r_m, r_{m+1} \rangle$ と頻度分布 $\mathcal{F}$ に対して，頻度が 1 となる全てのパケットをルールリストの上位 $|S|/3$ 個のルール $r_{\sigma^{-1}(1)}$,$r_{\sigma^{-1}(2)}$, $\ldots$, $r_{\sigma^{-1}(|S|/3)}$ で評価するような順序 σ は存在しない．即ち，任意の順序 σ において，$|S|/3+1$ 番目以降のルールによって評価されるパケットが少なくとも 1 つ存在する．そのような順序においては，以下の関係が成り立つ．

$$
\begin{aligned}
K &= \frac{|S|(|S|/3+1)}{2} \\
&= \sum_{i=1}^{|S|/3} 3i \\
&< L(\mathcal{R}_\sigma, \mathcal{F}) = \sum_{i=1}^{|S|/3} i|E(\mathcal{R}_\sigma, \sigma^{-1}(i))|_{\mathcal{F}} + \sum_{i=|S|/3+1}^{m+1} i|E(\mathcal{R}_\sigma, \sigma^{-1}(i))|_{\mathcal{F}}
\end{aligned}
$$

これより，$x \notin$ **XC3** $\Rightarrow f(x) \notin$ **WO** が成り立つので $f(x) \in$ **WO** $\Rightarrow x \in$ **XC3** が成り立つ．

以上より，**XC3** から **WO** への多項式時間帰着アルゴリズム f が存在するので **WO** は **NP** 困難である． □

XC3 から **WO** への帰着の具体例を示す．

XC3 のインスタンス (2) に対する **WO** のインスタンスは表 2 のルールリストと式 (4) の頻度分布，$K = \frac{3 \cdot 3(3+1)}{2} = 18$ となる．ただし，w_i はルール r_i の表で示される順序における重みを表す．ルールリストを順序 $\sigma = (3\ 4\ 5\ 1\ 6\ 2\ 7)$ で並び替えると表 3 のルールリストになり，このルールリストの遅延は 18 である．

表 2: ルールリスト

Filter $\mathcal{R}$
$r_1^P = 0\ 0\ 0\ 0\ *\ 0\ *\ 0\ *$
$r_2^P = 0\ 0\ 0\ *\ *\ 0\ 0\ 0\ *$
$r_3^P = 0\ 0\ *\ 0\ 0\ *\ *\ 0\ 0$
$r_4^P = *\ 0\ 0\ *\ 0\ 0\ 0\ *\ 0$
$r_5^P = 0\ 0\ 0\ 0\ 0\ 0\ 0\ *\ *\ *$
$r_6^P = 0\ *\ *\ 0\ 0\ *\ 0\ 0\ 0$
$r_7^D = *\ *\ *\ *\ *\ *\ *\ *\ *\ *$

表 3: 並び替えたルールリスト

Filter $\mathcal{R}_{(3\ 4\ 5\ 1\ 6\ 2)}$	w_i
$r_4^P = * \ 0 \ 0 * 0 \ 0 \ 0 * 0$	3
$r_6^P = 0 * * 0 \ 0 * 0 \ 0 \ 0$	3
$r_1^P = 0 \ 0 \ 0 \ 0 * 0 * 0 *$	3
$r_2^P = 0 \ 0 \ 0 * * 0 \ 0 \ 0 *$	0
$r_3^P = 0 \ 0 * 0 \ 0 * * 0 \ 0$	0
$r_5^P = 0 \ 0 \ 0 \ 0 \ 0 \ 0 * * *$	0
$r_7^D = * * * * * * * * *$	0

$$\mathcal{F}(p) = \begin{cases} 1 & \text{if } p = 100000000 \\ 1 & \text{if } p = 010000000 \\ 1 & \text{if } p = 001000000 \\ 1 & \text{if } p = 000100000 \\ 1 & \text{if } p = 000010000 \\ 1 & \text{if } p = 000001000 \\ 1 & \text{if } p = 000000100 \\ 1 & \text{if } p = 000000010 \\ 1 & \text{if } p = 000000001 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

5 おわりに

本稿では，ルール順序最適化問題において，デフォルトルール以外のアクションを *Permit* に限定した最適化問題，ホワイトリスト順序最適化問題の判定問題が **NP** 困難であることを証明した．この事実より，既に知られていることであるが，ルールのアクションを限定しない一般のルール順序最適化問題は **NP** 困難である．

Hamed らは，重み変動を考慮しないルール順序最適化問題が，単一機械ジョブスケジューリング問題から帰着することによって **NP** 困難であることを示した．筆者らは，重み変動を考慮するルール順序最適化問題が **NP** 困難であることを示している [4]．更に，本稿において，デフォルトルール以外のアクションを *Permit* に限定した最適化問題，ホワイトリスト順序最適化問題の判定問題版も **NP** 困難であることを証明した．即ち，重み変動を考慮する一般のルール順序最適化問題は，デフォルトルール以外のアクションを Permit に限ってもなお難しい問題である．

一般のルール順序最適化問題を解く際に，部分問題としてホワイトリスト順序最適化問題を解きたい場合がある．これより，ホワイトリスト順序最適化問題に対する発見的アルゴリズムや近似アルゴリズムを提案する必要がある．更に，これらの解法を利用した一般のルール順序最適化問題に対する効率的なアルゴリズムを考案することが今後の課題である．

参考文献

[1] T. Fuchino, T. Harada, K. Tanaka, and K. Mikawa. Acceleration of packet classification using adjacency list of rules. In *28th International Conference on Computer Communication and Networks (ICCCN)*, pages 1–9, July 2019.

[2] Michael R. Garey and David S. Johnson. *Computers and Intractability; A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., USA, 1990.

[3] Hazem Hamed and Ehab Al-Shaer. Dynamic rule-ordering optimization for high-speed firewall filtering. In *Proceedings of the 2006 ACM Symposium on Information, Computer and Communications Security*, ASIACCS '06, pages 332–342, New York, NY, USA, 2006. ACM.

[4] T. Harada and K. Tanaka. Computational complexity of relaxed optimal rule ordering. *IEICE technical report*, 119(340):47–54, Dec 2019.

[5] T. Harada, K. Tanaka, and K. Mikawa. Simulated annealing method for relaxed optimal rule ordering, Mar 2020 (accepted).

[6] G. Misherghi, L. Yuan, Z. Su, C. N. Chuah, and H. Chen. A general framework for benchmarking firewall optimization techniques. *IEEE Transactions on Network and Service Management*, 5(4):227–238, December 2008.

[7] 日景 喬一 and 山田 敏規. D-1-6 ルール間の依存関係を保持したファイアウォールの負荷最小化のためのアルゴリズム (d-1. コンピュテーション, 一般セッション). 電子情報通信学会総合大会講演論文集, 2016(1):6, mar 2016.

[8] 文岩 涼祐 and 山田 敏規. ファイアウォールルール整列問題に対する厳密アルゴリズム. 情報処理学会研究報告アルゴリズム *(AL)*, 2019-AL-175(15):1–6, nov 2019.