

KÓDOK FELBONTHATÓSÁGA

Egyetemi doktori (PhD) értekezés

Szerző: Falucskai János

Témavezető: Dr. Fazekas Gábor

DEBRECENI EGYETEM

Természettudományi Doktori Tanács

Matematika- és Számítástudományok Doktori Iskola

Debrecen, 2013

Ezen értekezést a Debreceni Egyetem Természettudományi Doktori Tanács Matematika- és Számítástudományok Doktori Iskola Számítástudomány és alkalmazásai programja keretében készítettem a Debreceni Egyetem természettudományi doktori (PhD) fokozatának elnyerése céljából.

Debrecen, 2013. február 20.

.....
Falucskai János
jelölt

Tanúsítom, hogy Falucskai János doktorjelölt 2012 - 2013 között a fent megnevezett Doktori Iskola Számítástudomány és alkalmazásai programjának keretében irányításommal végezte munkáját. Az értekezésben foglalt eredményekhez a jelölt önálló alkotó tevékenységével meghatározóan hozzájárult. Az értekezés elfogadását javasolom.

Debrecen, 2013. február 20.

.....
Dr. Fazekas Gábor
témavezető

KÓDOK FELBONTHATÓSÁGA

Értekezés a doktori (Ph.D.) fokozat megszerzése érdekében
a matematika- és számítástudományok tudományágban.

Írta: Falucskai János okleveles matematika, ábrázoló geometria és
számítástechnika tanár.

Készült a Debreceni Egyetem Matematika– és Számítástudományok
doktori iskolája (Számítástudomány és alkalmazásai programja) keretében.

Témavezető: Dr. Fazekas Gábor

A doktori szigorlati bizottság:

elnök: Dr.

tagok: Dr.

Dr.

A doktori szigorlat időpontja: 20... ..

Az értekezés bírálói:

Dr.

Dr.

Dr.

A bírálóbizottság:

elnök: Dr.

tagok: Dr.

Dr.

Dr.

Dr.

Az értekezés védésének időpontja: 20... ..

Tartalomjegyzék

1	Bevezetés	1
2	Szakirodalmi előzmények	3
2.1	A kódokkal kapcsolatos alapfogalmak	5
2.2	Az automatákkal kapcsolatos alapfogalmak	7
2.3	Felbonthatósági kritériumok	8
2.4	A Sardinas–Patterson tétel	10
2.5	A „virág” automata	11
2.6	Automaták összefűzött szorzata	20
2.7	A Tsuji-féle konstrukció	25
3	Eredmények	31
3.1	Egy, a felbonthatóságot eldöntő új algoritmus	31
3.1.1	Az algoritmus megadása	31
3.1.2	Az algoritmus és a Sardinas–Patterson algoritmus kapcsolata	40
3.2	A felbonthatóság egy új értelmezése	45
3.2.1	Kvázikódok	46
3.2.2	Jelölés rendszer	47
3.2.3	Kvázikódok felbonthatósági kritériumai	50
3.2.4	A Sardinas–Patterson algoritmus kvázikódokra való alkalmazhatósága	57

3.2.5	A kvázikódok dekompozíciója	59
3.2.6	A kvázikódot elfogadó véges automata	63
3.2.7	Kvázikódok nemfelbonthatóságának egy szükséges feltétele	65
3.2.8	A kvázikód egy lehetséges alkalmazása	74
4	Összegzés és további kutatási feladatok	77
5	Summary and further research tasks	81
	Irodalomjegyzék	88
	Publikációs lista	89
	Előadások	90
	Egyéb publikációk	92

1. fejezet

Bevezetés

A kódot tekinthetjük a kódszavakat tartalmazó formális nyelvnek, s véges jellege miatt ez a nyelv hármastípusú nyelvtannal leírható, így megadható például véges automatával. A célom az volt, hogy egy adott rendszerbeli problémát egy másik rendszerben oldjak meg, nevezetesen kódelméleti problémákat véges automatával, kapcsolati mátrixszal, reguláris kifejezéssel modellezzek.

Egy ilyen probléma a kód felbonthatósága, azaz egy adott kód esetén annak eldöntése, hogy van-e két különböző sorozata a kódszavaknak, mely ugyanazon üzenetet hordozza. Terveim között szerepelt a kódolás általánosítása (melyet kvázikódoknak nevezek a dolgozatban), valamint az általánosított kódok felbonthatóságának vizsgálata a fentebb említett eszközök valamelyikével. Melléktermékként cél volt a létrejövő algoritmus hagyományos kódok esetén való alkalmazása is.

A 2. fejezetben bemutatom az értekezés témájának legfontosabb fogalmait és tételeit, a szakirodalmi előzményeket. A további fejezetekben saját eredményeimet tárgyalom. A 3.1. fejezetben leírok egy saját –automataelméleti alapú felbonthatóságot eldöntő– algoritmust, majd megmutatom, hogy algoritmus ekvivalens a

Sardinas–Patterson algoritmussal. A 3.2. fejezetben a felbonthatóság egy új értelmezését tárgyalom. Bevezetem a kvázikód és a kvázikódot elfogadó véges automata fogalmát. Ezután megadok néhány kritériumot a kvázikódok felbonthatóságára, majd belátom, hogy a Sardinas–Patterson algoritmus kvázikódokra nem alkalmazható. Megmutatom, hogy a felbonthatóság problémája kvázikódokra eldönthető véges halmazok esetén, majd megadom a kvázikódok nemfelbonthatóságának egy szükséges feltételét, s végül a kvázikódok egy lehetséges kriptográfiai alkalmazását ismertetem.

2. fejezet

Szakirodalmi előzmények

A kódelmélet az információelmélet egyik legfontosabb alkalmazási területe. Kialakulásának időpontját ezért elsősorban az információelmülethez szokás kötni, ami az 1940-es években Shannon munkásságával vette kezdetét. Az első cikk ebben a témakörben az „A mathematical theory of communication” [Sha48] volt 1948-ban, melyet egy évvel később követett a szakterület alap szakirodalmának számító könyv, a „The mathematical theory of communication” [SW49], amit Shannon egyik kollégájával, Weaver-rel közösen írt. Ez a könyv magyarul is megjelent [SW86]. A könyv az eredeti cikktől lényegében csupán annyiban tér el, hogy Weaver írt egy bevezetőt hozzá, valamint megváltoztatták a címét.

Az információelmélettel együtt megszületett kódelmélet alapvetően három területtel foglalkozik:

- Egyértelmű dekódolhatóság
- Optimalizálás, vagy más néven adattömörítés
- Hibafelismerés és hibajavítás

Jelen értekezés a kódok felbonthatóságával foglalkozik, mely a felsorolt területek közül mindegyikkel összefüggésben van, ugyanis ha egy kód nem bontható fel, akkor nem lehet dekódolni az általa hordozott üzenetet, s emiatt használhatatlan. Látható, hogy a felbonthatóság kérdése alapvető a kódelmélet területén, gyakran nevezik a felbontható kódokat egyértelműen megfejthető (*uniquely decipherable*), vagy röviden *UD* kódoknak.

A kódelmélet fejlődése során az azonos hosszúságú kódszavakból álló kódokat elsősorban a hibajavítással kapcsolatban vizsgálták (lsd. [MS77]), a változó hosszúságú kódok elmélete az adattömörítésekhez, kód optimalizálásokhoz kötődik. A későbbiekben látni fogjuk, hogy a konstans hosszúságú kódok nem esnek az érdeklődésikörünkbe, ugyanis minden konstans hosszúságú kódszavakból álló kód prefix, s emiatt mindig felbontható.

A változó hosszúságú kódokkal mélyebben először *Schützenberger* [Sch55], valamint *Gilbert és Moore* [GM59], majd később *Shyr* [Shy01] foglalkozott. *Schützenberger* vizsgálatai nyomán került kapcsolatba a kódelmélet a nemkommutatív algebra elméletével, melynek korai eredményeit *Nivat* [Niv66] foglalta össze.

A változó hosszúságú kódok és az automaták kapcsolatának alapos összefoglalóját találhatjuk *Berstel és Perrin* [BP85] könyvében, melynek folyamatosan frissített verziója szabadon elérhető a világhálón [BPR09].

A felbonthatóságra problémájára először *Sardinas és Patterson* [SP53] adott megoldást, ez általában *Sardinas–Patterson* algoritmusként ismert. A felbonthatóság kérdésével ezek után többen is foglalkoztak, pl. *Markov* [Mar62], *Bandyopadhyay* [Ban63], *Levenshtein* [Lev64], *Riley* [Ril67], *de Luca* [dL76]. A felbonthatóságot eldöntő algoritmusok optimalizálásával kapcsolatban *Spehner* [Spe76], *Rodeh* [Rod82], *Apostolico és Giancarlo* [AG84], *McCloskey* [McC96] végzett kutatásokat. A felbonthatóságot eldöntő algoritmus pontos bonyolultsága jelenleg még nem ismert, de a kérdéssel kapcsolatban

Hoffmann [Hof84], *Galil* [Gal85] és *McCloskey* [McC96] valamint *König* [Kön94] értékes eredményeket értek el. Dolgozatomban nem vizsgálom az általam kifejlesztett algoritmusok bonyolultságait.

Az automatákat alkalmazó algoritmusokra fogok részletesebben kitérni, mivel a dolgozat keretei nem teszik lehetővé a felbonthatósággal kapcsolatos szakirodalmi eredmények teljes ismertetését. A kódok egyértelmű megfejthetősége az automataelméletben is jelenlévő reguláris kifejezések többértelműségi problémájának speciális esete, melyre vannak eldöntő algoritmusok, amik megtalálhatóak például *Eilenberg* [Eil74], vagy *Aho* és társai [AHU75] munkáiban. A kódok és a reguláris kifejezések kapcsolatára *Brzozowsky* [Brz67] mutatott rá.

2.1. A kódokkal kapcsolatos alapfogalmak

1. Definíció. Az A ábécé egy véges nem üres halmaz, elemei betűk, vagy más néven szimbólumok.

2. Definíció. Az A ábécé feletti w szó az A ábécé véges számú elemeiből képzett szimbólumsorozat.

$$w = (a_1, a_2, \dots, a_n), \quad a_i \in A.$$

3. Definíció. Az A ábécé feletti teljes nyelv az A feletti összes szó által alkotott halmaz, jelölése A^* .

Az A feletti nemüres szavak halmazának jelölése A^+ . A^* -on értelmezve van a konkatenáció (összefűzés) asszociatív művelete, mellyel két szimbólumsorozatot egyesíthetünk

$$(a_1, a_2, \dots, a_n)(b_1, b_2, \dots, b_m) = (a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_m).$$

Az asszociativitás miatt használható a

$$w = a_1 a_2 \dots a_n$$

alak a $w = (a_1, a_2, \dots, a_n)$ forma helyett, ahol minden egyes $a_i \in A$ szimbólumot megfeleltetünk az (a_i) jelsorozatnak.

4. Definíció. Az üres jelsorozatot *üres szónak* nevezzük és λ -val jelöljük.

Az üres jelsorozatról könnyű belátni, hogy a konkatenációra nézve az egységelem szerepét tölti be.

5. Definíció. Az A feletti C kód az A^+ véges részhalmaza.

Használni fogom a $C^0 = \{\lambda\}$ és a $C^i = C^{i-1}C$ ($i > 0$) jelölést.

6. Definíció. A C kód elemeit *kódszavaknak*, a C^* elemeit pedig *üzeneteknek* nevezzük.

7. Definíció. [BR05] Egy C kódot *UD* (*uniquely decipherable*) kódnak, azaz felbonthatónak nevezünk, ha minden egyes üzenetnek egyértelmű felbontása van, tehát ha teljesül az

$$x_1x_2 \cdots x_n = y_1y_2 \cdots y_m \quad (2.1)$$

egyenlőség, ahol ha $x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m \in C$, akkor $n = m$ és $x_1 = y_1, \dots, x_n = y_n$.

A felbonthatóságnak természetesen több definíciója is ismert, a következő bináris kódok esetén használatos, de könnyen átfogalmazható tetszőleges ábécé feletti kódra.

8. Definíció. [DDP99] Egy K kódot felbonthatónak nevezünk, ha tetszőleges bináris jelsorozat legfeljebb egyféleképpen bontható kódszavak szorzatára.

9. Definíció. [DDP99] Egy kódot prefixmentesnek, vagy röviden prefixnek nevezünk, ha egyetlen kódszó sem valódi kezdőszelete egy másiknak.

1. Tétel. [DDP99] Bármely prefix kód felbontható.

2.2. Az automatákkal kapcsolatos alapfogalmak

10. Definíció. Automata alatt egy véges automatát, azaz egy $\mathcal{A} = (Q, I, Q_F, A, \delta)$ ötöst értünk, ahol a Q egy nem üres és véges halmaz, $I \subseteq Q$ a kezdő állapotok halmaza, $Q_F \subseteq Q$ az elfogadó állapotok halmaza, A egy nem üres és véges halmaz a bemenő jelek halmaza és

$$\delta = \{(p, q, a) | p, q \in Q, a \in A\} \subseteq Q^2 \times A$$

a mozgási szabályok halmaza.

Amennyiben $I = \{i\}$, azaz I egyelemű halmaz, akkor az $\mathcal{A}$ automatát időnként $\mathcal{A} = (Q, i, Q_F, A, \delta)$ formában adom meg. Ha valamely $p, q \in Q, a \in A$ hármásra $(p, q, a) \in \delta$, akkor (p, q, a) helyett használom a $p \xrightarrow{a} q$, illetve a $q \in \delta(p, a)$ jelölést is, és ha $\delta(p, a) = \{q\}$, akkor esetenként a $\delta(p, a) = q$ jelöléssel élek.

$(p, q, a) \in \delta$ esetén a $p \xrightarrow{a} q$ kifejezést az $\mathcal{A}$ automata egy élének hívjuk, s azt is mondjuk, hogy az $\mathcal{A}$ automatában a $p \xrightarrow{a} q$ szerepel. Amennyiben valamely $p_1, \dots, p_k \in Q, a_1, \dots, a_{k-1} \in A$ esetén az $\mathcal{A}$ automatában $p_1 \xrightarrow{a_1} p_2, p_2 \xrightarrow{a_2} p_3, \dots, p_{k-1} \xrightarrow{a_{k-1}} p_k$ szerepelnek, úgy azt mondjuk, hogy létezik egy p_1 állapotból p_k állapotba vezető $a_1 \cdots a_{k-1}$ címkéjű út. Ilyenkor használom a $p_k \in \delta(p_1, a_1 \cdots a_{k-1})$, illetve a $p_1 \xrightarrow{a_1 \cdots a_{k-1}} p_k$ jelölést is. Ha $p_1 \in I$ és $p_k \in Q_F$, akkor azt is mondjuk, hogy az automata felismeri az $a_1 \cdots a_{k-1}$ bemenő jelsorozatot.

Az $\mathcal{A}$ által felismert bemenő jelsorozatok halmazát az $\mathcal{A}$ automata által felismert nyelvnek, vagy röviden az $\mathcal{A}$ automata nyelvének hívjuk. Két automatát ekvivalensnek nevezünk, ha ugyanazt a nyelvet ismerik fel.

Amennyiben minden egyes $p \in Q$, $a \in A$ párra teljesül, hogy $|\{(p, q, a) | (p, q, a) \in \delta\}| \leq 1$, akkor determinisztikus automatáról beszélünk. Ha viszont van olyan $p \in Q$, $a \in A$ pár, hogy $|\{(p, q, a) | (p, q, a) \in \delta\}| > 1$, úgy az $\mathcal{A}$ automatát nemdeterminisztikus automatának hívjuk.

Egy $\mathcal{A} = (Q, I, Q_F, A, \delta)$ automatához tartozó determinisztikus automatának nevezzük azt a $\mathcal{B} = (Q', I', Q'_F, A, \delta')$ automatát, melyre

- $Q' = \{D \in 2^Q | \exists w \in A^*, i \in I : D = \{q \in Q | q \in \delta(i, w)\}\}$
- $I' = I \in Q'$
- $Q'_F = \{D \in Q' | D \cap Q_F \neq \emptyset\}$
- $\delta' = \{(P_1, P_2, a) | P_1, P_2 \in Q', a \in A\} \subseteq Q'^2 \times A$ és tetszőleges $P_1 \in Q'$ és $a \in A$ esetén $P_2 = \bigcup_{q \in P_1} \delta(q, a)$

2. Tétel. [Sal73] Bármely véges automata ekvivalens a hozzátartozó determinisztikus véges automatával.

2.3. Felbonthatósági kritériumok

Legyen adva egy pozitív egész számokból álló $u(n)_{n \geq 1}$ sorozat, valamint a k darab szimbólumot tartalmazó A ábécé. Az A ábécé feletti C kódra teljesüljön az $u_n = |(C \cap A^n)|$ egyenlőség.

3. Tétel. [McM56] Az A ábécé feletti adott hosszúságú kódszavakkal rendelkező felbontható kód akkor, és csak akkor létezik, ha

$$\sum_{n \geq 1} u_n k^{-n} \leq 1$$

PÉLDA: A tételben u_n jelöli az n hosszúságú kódszavak számát a C halmazban. Legyen adva a $C = \{00, 012, 011, 02, 10, 110, 1111, 12, 2\}$ kód. Alkalmazzuk rá a fenti képletet, ekkor mivel $u_1 = 1, u_2 = 4, u_3 = 3, u_4 = 1, k = 3$, kapjuk, hogy

$$\begin{aligned} \sum_{n \geq 1} u_n k^{-n} &= \sum_{n=1}^4 u_n k^{-n} = 3^{-1} + 4 \cdot 3^{-2} + 3 \cdot 3^{-3} + 3^{-4} = \\ &= \frac{1}{3} + 4 \cdot \frac{1}{9} + 3 \cdot \frac{1}{27} + \frac{1}{81} = \frac{27}{81} + \frac{36}{81} + \frac{9}{81} + \frac{1}{81} = \frac{73}{81} \leq 1, \end{aligned}$$

akkor a tétel szerint lehet, hogy felbontható, ugyanis a tétel csupán szükséges feltételt fogalmaz meg a felbonthatóság tekintetében. Valóban, hiszen ha megvizsgáljuk a $C = \{0, 00\}$ kódot, akkor egyértelmű, hogy a 00 jelsorozat egyrészt előáll a második kódszóból, másrészt pedig előállítható az első kódszó kétszeri ismétlése által is, tehát nem felbontható a kód: α_i -vel jelölve az i . kódszót, adódik, hogy $\alpha_1 \alpha_1 = \alpha_2$. A tétel alapján kiszámolt érték kisebb lesz, mint egy:

$$\sum_{n \geq 1} u_n k^{-n} = \sum_{n=1}^2 u_n 2^{-n} = 2^{-1} + 2^{-2} = \frac{1}{2} + \frac{1}{4} = \frac{3}{4} \leq 1.$$

Tehát a tétel csupán akkor alkalmazható egy adott kód felbonthatóságának az eldöntésére, ha $\sum_{n \geq 1} u_n k^{-n} > 1$. Ellenkező esetben csak azt tudjuk biztosan, hogy létezik felbontható kód, melyben u_n darab n hosszú kódszó van. Egy ilyen –nem csak felbontható, hanem még prefix tulajdonsággal is bíró– kód megalkotását teszi lehetővé a következő tétel bizonyításában megadott algoritmus:

4. Tétel. (Kraft)[GGV02] Ha az $l_1, l_2, \dots, l_n$ pozitív egész számokra

$$\sum_{i=1}^n k^{-l_i} \leq 1,$$

akkor létezik olyan $C = \{w_1, \dots, w_n\}$ prefix kód, melyre $|w_1| = l_1, \dots, |w_n| = l_n$.

1. Következmény. [GGV02] A 3. és a 4. tételeket összevetve levonhatjuk azt a fontos következtetést, hogy minden felbontható kódhoz létezik vele ekvivalens (azonos kódszóhosszú) prefix kód.

2.4. A Sardinas–Patterson tétel

A Sardinas–Patterson [SP53] algoritmusnak a következő az alapgondolata: határozzuk meg az összes lehetségesnek tűnő többszörös felbontás összes maradványát. Egy valódi többszörös felbontást abból a tényből vesz észre az algoritmus, hogy az üres szó egy lesz a maradványok közül.

Legyen X és Y két részhalmaza az A^+ halmaznak, valamint legyen $x \in X$ és $y \in Y$. Jelölje $X^{-1}Y$ a következő halmazt: w egy eleme a $X^{-1}Y$ halmaznak, ha $\exists x \in X, y \in Y$ úgy, hogy $xw = y$. Amennyiben $X = \{x\}$, $Y = \{y\}$, azaz X és Y egyelemű halmazok, akkor $x^{-1}y$ jelöli azt a w szót, melyre $xw = y$.

Legyen a C halmaz a A^+ halmaz részhalmaza, és legyen

$$\begin{aligned} U_1 &= C^{-1}C \setminus \{\lambda\} \\ U_2 &= C^{-1}U_1 \cup U_1^{-1}C \\ &\vdots \\ U_n &= C^{-1}U_{n-1} \cup U_{n-1}^{-1}C \end{aligned} \tag{2.2}$$

A fenti jelöléseket alkalmazva igaz a következő tétel:

5. Tétel (Sardinas – Patterson). [BP85] Legyen $C \subset A^+$ és legyen $(U_n)_{n \geq 1}$ definiálva a fentiek szerint. Bármely $n \geq 1$ esetén λ eleme az

U_n halmaznak akkor, és csak akkor, ha létezik valamely $k \in \{1, \dots, n\}$ -re egy $u \in U_k$ szó és $i, j \geq 0$ egészek úgy, hogy

$$uC^i \cap C^j \neq 0, \quad i + j + k = n$$

Ennek segítségével belátható a *Sardinas-Patterson* algoritmus alapjául szolgáló tétel:

6. Tétel. [BP85] A $C \subset A^+$ halmaz felbontható kód akkor, és csak akkor, ha a fentebb definiált U_n halmazok közül egyetlenegy sem tartalmazza az üres szót.

2.5. A „virág” automata

Legyen X egy tetszőleges részhalmaza az A^+ halmaznak, ahol A egy ábécé. Defináljuk az

$$\mathcal{A}_D(X) = (Q, I, Q_F, A, \delta)$$

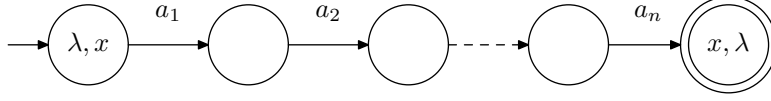
automatát, ahol

$$Q = \{(u, v) \in A^* \times A^* \mid uv \in X\}, I = \lambda \times X, Q_F = X \times \lambda$$

és az automata δ mozgási szabályaira igaz, hogy $(u, v) \xrightarrow{a} (u', v')$ él lesz az automatában akkor és csak akkor, ha $ua = u'$ és $v = av'$. Más szavakkal az $\mathcal{A}_D(X)$ élei a következők lesznek:

$$(u, av) \xrightarrow{a} (ua, v), \text{ ahol } uav \in X, a \in A.$$

Másképpen úgy is mondhatjuk, hogy az $\mathcal{A}_D(X)$ automata a 2.1. ábrán megadott formájú nem összefüggő élek halmazának az uniója, ahol egy-egy összefüggő „élhalmaz” tartozik minden $x = a_1 \cdots a_n \in X$ kódszóhoz. Az X halmazhoz tartozó *virág automata* definíció szerint



2.1. ábra. Az $x = a_1 \cdots a_n$ kódszóhoz tartozó $\mathcal{A}_D(X)$ automata részlet

az $\mathcal{A}_D(X)$ automata iteráltja. Általában $\mathcal{A}_D^*(X)$ a jelölése, esetleg szokás még $(\mathcal{A}_D(X))^*$ formában is írni. Az $\mathcal{A}_D^*(X)$ automatát az $\mathcal{A}_D(X)$ automatából két lépésben nyerhetjük. Induljunk ki az $\mathcal{A}_D(X)$ automatából, adjunk hozzá egy új ω állapotot, valamint az

$$\begin{array}{lll} \omega & \xrightarrow{a} & (a, v) \quad \text{éleket } av \in X \text{ esetén,} \\ (u, a) & \xrightarrow{a} & \omega \quad \text{éleket } ua \in X \text{ esetén,} \\ \omega & \xrightarrow{a} & \omega \quad \text{éleket } a \in X \text{ esetén.} \end{array}$$

Legyen ω az egyetlen elfogadó, és egyben az egyetlen kezdő állapota az $\mathcal{A}_D^*(X)$ automatának. Emiatt az eredeti $\mathcal{A}_D(X)$ automatában lévő $\lambda \times X$ kezdő állapotok a továbbiakban nem érhetők el, illetve az $X \times \lambda$ állapotok zsákutcák lesznek, emiatt nem is jelennek meg az $\mathcal{A}_D^*(X)$ automatában. Az ω állapotnak általában (λ, λ) a jelölése. Emiatt a következő módon definiálható az $\mathcal{A}_D^*(X)$ automata:

$$\mathcal{A}_D^*(X) = (P, (\lambda, \lambda), (\lambda, \lambda), A, \delta'),$$

ahol

$$P = \{(u, v) \in A^+ \times A^+ \mid uv \in X\} \cup \{(\lambda, \lambda)\},$$

és a δ' mozgási szabályok az alábbi négy típusúak lehetnek:

$$\begin{array}{lll} (u, av) & \xrightarrow{a} & (ua, v) \quad \text{ahol } uav \in X, (u, v) \neq (\lambda, \lambda) \\ (\lambda, \lambda) & \xrightarrow{a} & (a, v), \quad \text{ahol } av \in X, v \neq \lambda \\ (u, a) & \xrightarrow{a} & (\lambda, \lambda), \quad \text{ahol } ua \in X, u \neq \lambda \\ (\lambda, \lambda) & \xrightarrow{a} & (\lambda, \lambda), \quad \text{ahol } a \in X. \end{array}$$

Az automata elnevezését a grafikus reprezentáció indokolja, minden $x \in X$ kódszó esetén egy-egy „virágszirom” alakul ki, hiszen minden kódszó egy-egy egyszerű

$$(\lambda, \lambda) \xrightarrow{x} (\lambda, \lambda)$$

utat definiál az $\mathcal{A}_D^*(X)$ automatában. Ha $x = a \in A$, akkor ez a

$$(\lambda, \lambda) \xrightarrow{a} (\lambda, \lambda)$$

útként jelenik meg. Ha $x = a_1 \cdots a_n$ és $n \geq 2$, akkor az út a

$$(\lambda, \lambda) \xrightarrow{a_1} (a_1, a_2 \cdots a_n) \xrightarrow{a_2} (a_1 a_2, a_3 \cdots a_n) \xrightarrow{a_3} \cdots \longrightarrow (\lambda, \lambda)$$

út lesz.

PÉLDA: [BP85] Legyen $X = \{aa, ba, bb, baa, bba\}$ egy kód, határozzuk meg a hozzátartozó virág automatát. Ehhez először az $\mathcal{A}_D(X)$ automatát kell felírunk, emiatt definiálásához elsőnek megadom az automata belső állapotainak Q halmazát. A felírásban egy-egy sor az egy-egy kódszóból származó állapotokat jelenti.

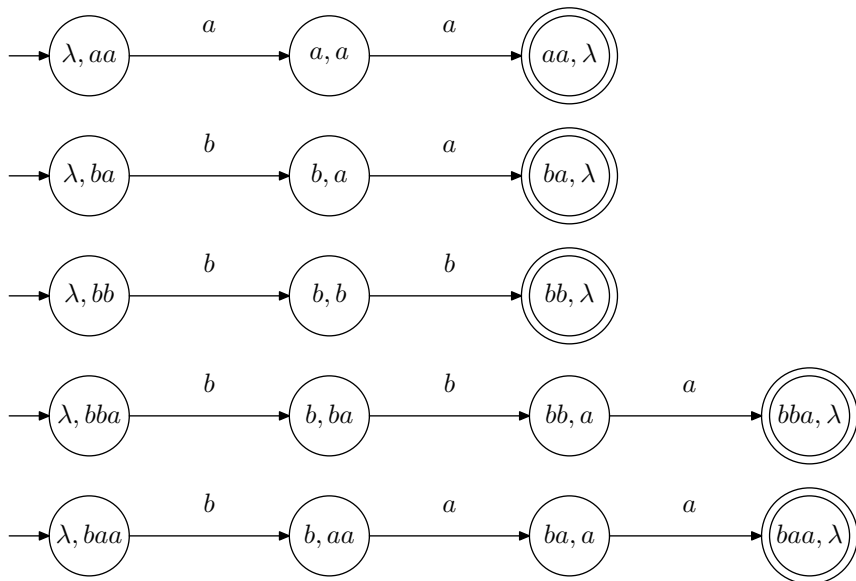
$$\begin{aligned} Q = \{ & (\lambda, aa), (a, a), (aa, \lambda), \\ & (\lambda, ba), (b, a), (ba, \lambda), \\ & (\lambda, bb), (b, b), (bb, \lambda), \\ & (\lambda, baa), (b, aa), (ba, a), (baa, \lambda), \\ & (\lambda, bba), (b, ba), (bb, a), (bba, \lambda) \} \end{aligned}$$

Az $\mathcal{A}_D(X)$ automata I kezdő állapotainak halmaza az

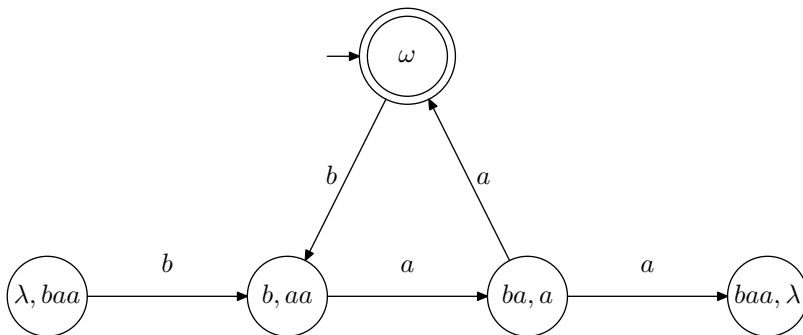
$$I = \{(\lambda, aa), (\lambda, ba), (\lambda, bb), (\lambda, baa), (\lambda, bba)\}$$

halmaz, az elfogadó állapotok Q_F halmaza pedig a

$$Q_F = \{(aa, \lambda), (ba, \lambda), (bb, \lambda), (baa, \lambda), (bba, \lambda)\},$$



2.2. ábra. Az $X = \{aa, ba, bb, baa, bba\}$ kódhoz tartozó $\mathcal{A}_D(X)$ automata

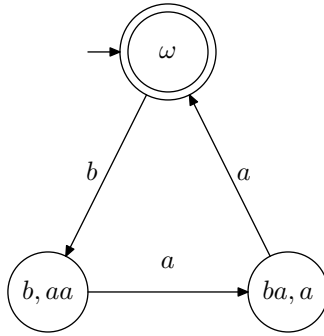


2.3. ábra. Az $\mathcal{A}_D^*(X)$ automata baa kódszóhoz tartozó „részautomatája”

halmaz. Ezekkel a jelölésekkel az $\mathcal{A}_D(X)$ automata ábrája a 2.2. ábrán látható. Bevezetem az új ω állapotot, ez például a baa kódszóhoz tartozó „részautomata” esetében a 2.3 ábrán látható.

Könnyű észrevenni, hogy az ω kezdő, és egyben elfogadó állapot jellege miatt az eredeti (λ, baa) kezdő állapot és a (baa, λ) elfogadó állapot elveszti kezdő illetve elfogadó jellegét. Mivel a (λ, baa) állapotba a továbbiakban nem tudunk eljutni, illetve a (baa, λ) állapotból nem tudunk továbbmenni elfogadó állapotba, ezeket elhagyhatjuk. Így kapjuk a 2.4. ábrán látható, a baa kódszóhoz tartozó úgynevezett $\mathcal{A}_D^*(X)$ trim „részautomatát”, ahol a trim jelző a következőt jelenti:

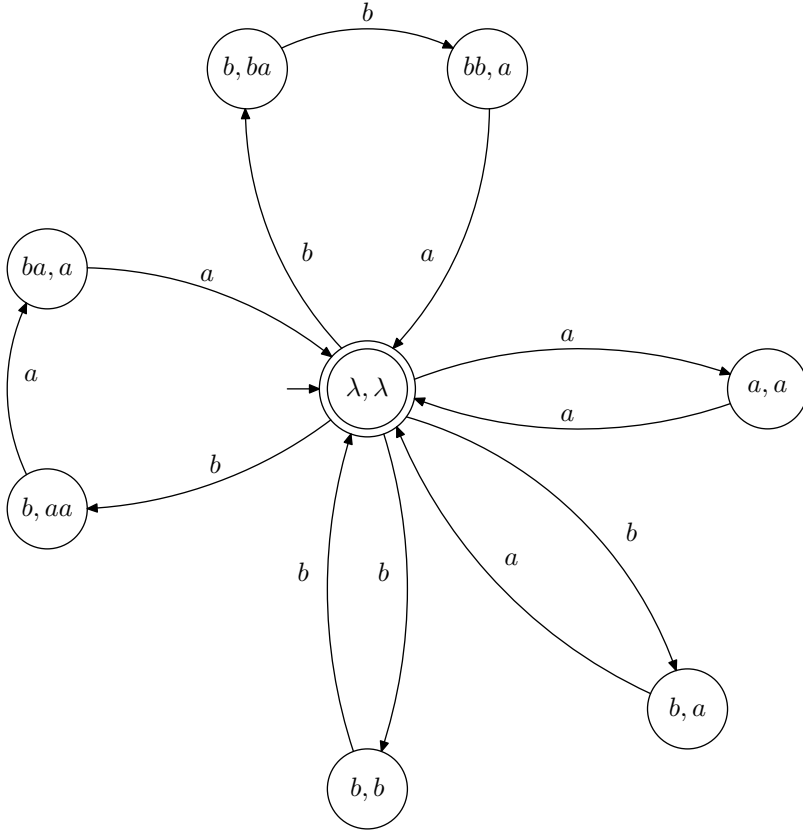
11. Definíció. [BP85] Egy automatát *trim* automatának nevezünk, ha minden állapota elérhető a kezdő állapottól, és minden állapotából el tudunk érni valamely elfogadó állapotba.



2.4. ábra. A baa kódszóhoz tartozó trim automata részlet

Az $X = \{aa, ba, bb, baa, bba\}$ kódhoz tartozó $\mathcal{A}_D^*(X)$ virág automata a 2.5. ábrán látható.

12. Definíció. [BP85] Az A ábécé feletti $\mathcal{A} = (Q, I, Q_F, A, \delta)$ automatát egyértelmű automatának nevezzük, ha bármely $p, q \in Q$



2.5. ábra. Az $X = \{aa, ba, bb, baa, bba\}$ kódhoz tartozó $\mathcal{A}_D^*(X)$ virág automata

és $w \in A^*$ esetén a p állapotból a q állapotba w címkével legfeljebb egy út létezik.

13. Definíció. [BP85] Legyen az $\mathcal{A} = (Q, I, Q_F, A, \delta)$ automata az A ábécé felett értelmezve. Az $\mathcal{A}$ automata *négyzet automatáján* azt az

$\mathcal{S}(\mathcal{A}) = (Q \times Q, I \times I, Q_F \times Q_F, A, \delta')$ automatát értjük, amelyben a

$$(p_1, p_2) \xrightarrow{a} (q_1, q_2)$$

él pontosan akkor szerepel, ha a

$$p_1 \xrightarrow{a} q_1 \text{ és a } p_2 \xrightarrow{a} q_2$$

élek szerepeltek az $\mathcal{A}$ automatában.

7. Tétel. [BP85] Az $\mathcal{A} = (Q, I, Q_F, A, \delta)$ automata akkor, és csak akkor egyértelmű automata, ha nincs

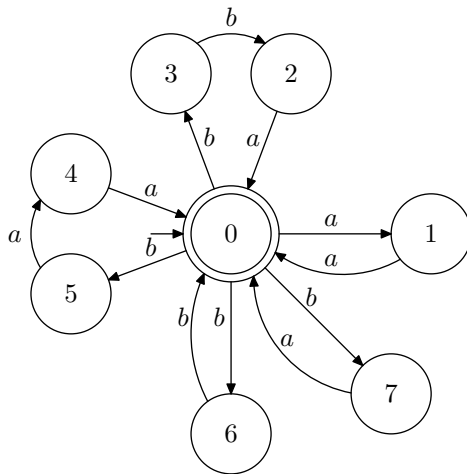
$$(p, p) \xrightarrow{u} (r, s) \xrightarrow{v} (q, q) \quad r \neq s \quad (2.3)$$

formájú út az $\mathcal{S}(\mathcal{A})$ automatában.

A tétel lehetőséget ad arra, hogy egy egyértelmű automatával felismerhető X kódról eldöntsük a felbonthatóságot. Ha a kód az $\mathcal{A}(X)$ automatával ismerhető fel, akkor határozzuk meg az $\mathcal{A}^*(X)$ automatát. Döntsük el, hogy az $\mathcal{A}^*(X)$ automata egyértelmű-e, azaz alkossuk meg az $\mathcal{S}(\mathcal{A}^*(X))$ automatát, majd ebben az automatában keressünk a 2.3 egyenletben leírt formájú utakat. Ha létezik legalább egy ilyen út, akkor a kód nem lesz felbontható, ellenben igen. Ezen az elgondoláson alapszik a következő tétel, tekintve azt a tényt, hogy az $\mathcal{A}_D(X)$ virág automata a definíciójából adódóan mindig egyértelmű automata.

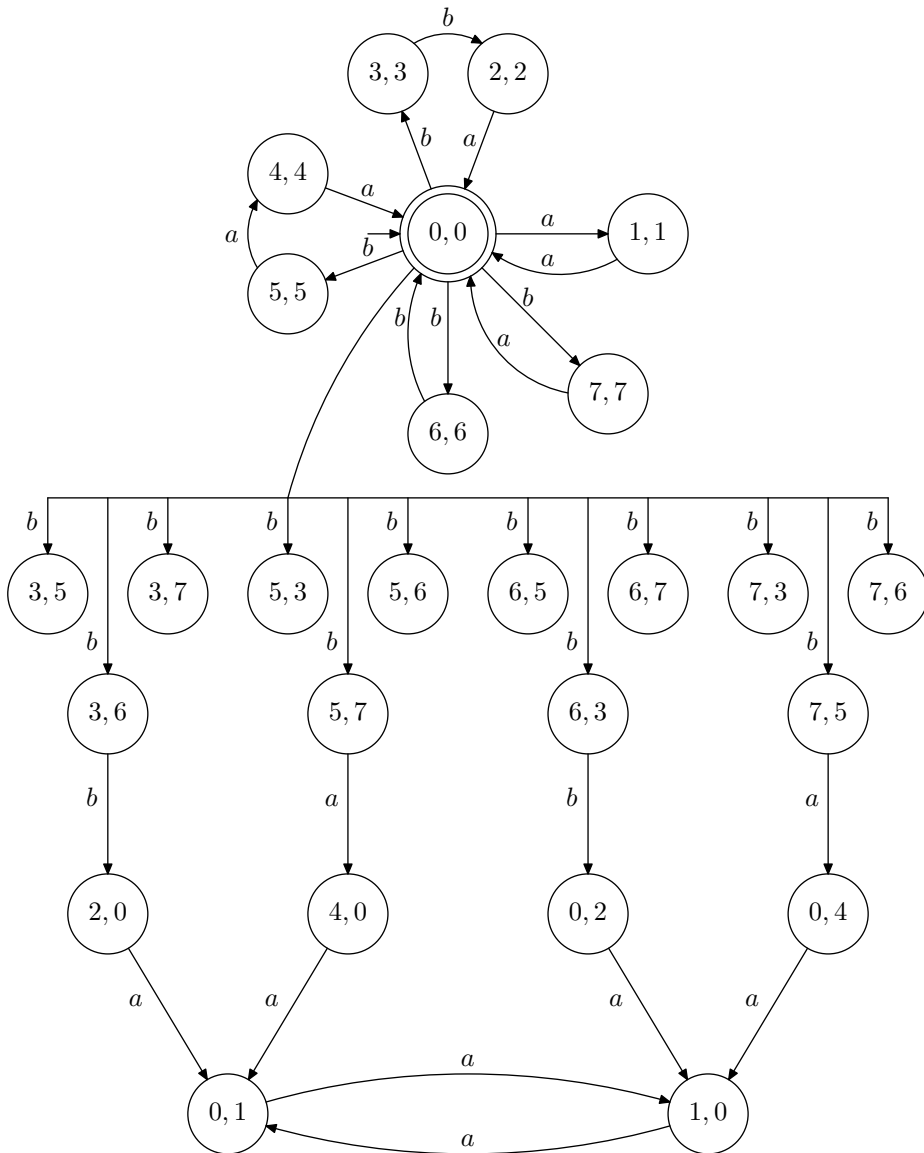
8. Tétel. [BP85] Egy X kód akkor és csak akkor felbontható, ha az $\mathcal{A}_D^*(X)$ virág automata egyértelmű automata.

PÉLDA: [BP85] Jelöljük át a könnyebb kezelhetőség miatt a 2.5. ábrán látható $\mathcal{A}_D^*(X)$ automata állapotait, az eredmény a 2.6. ábrán van megadva. Ennek készítettem el az $\mathcal{S}(\mathcal{A}_D^*(X))$ négyzet automatáját,



2.6. ábra. Az $X = \{aa, ba, bb, baa, bba\}$ kódhoz tartozó átjelölt $\mathcal{A}_D^*(X)$ virág automata

mely a 2.7. ábrán van megadva. Csak a $(0, 0)$ kezdő állapotból elérhető állapotok kerültek megrajzolásra. Mivel nincsenek benne a (2.3.) egyenletben megadott formájú utak, az $X = \{aa, ba, bb, baa, bba\}$ kód felbontható. A 2.7. ábrát én készítettem el, [BP85]-ben nem szerepel.



2.7. ábra. Az $X = \{aa, ba, bb, baa, bba\}$ kódhoz tartozó átjelölt $\mathcal{S}(\mathcal{A}_D^*(X))$ virág automata

2.6. Automaták összefűzött szorzata

Az automatáknak egy olyan szorzatát fogjuk definiálni, amelynek segítségével eldönthető lesz egy kódról, hogy felbontható-e. Az ötlet az, hogy készítsük el a kódot elfogadó automatának az önmagával való Descartes szorzatát úgy, hogy amikor az automata elfogadó állapotba kerül, akkor egyúttal vigyük át kezdő állapotba is. Ez nemdeterminisztikus úton tehető meg. Az automata ily módon elkészített, úgynevezett összefűzött szorzat automatája akkor kerül elfogadó állapotba, ha minden komponense ugyanakkor éri el az elfogadó állapotot.

14. Definíció. Egy automatát normalizáltnak nevezünk, ha

1. A kezdő állapot nem fordul elő egyetlen mozgási szabály jobb oldalán sem, így egy nem üres jelsorozat olvasása után az automata a kezdő állapottól különböző állapotban lesz mindig.
2. Az automatának egyetlen elfogadó állapota van és ez az elfogadó állapot nem szerepel egyetlen szabály baloldalán sem.

15. Definíció. Sikeres útnak nevezzük a kezdő állapotból valamely elfogadó állapotba vezető utat. Az út során érintett jelsorozat az út címkéje.

Ismert, hogy bármely reguláris nyelvhez megadható vele ekvivalens minimálautomata. A minimálautomata determinisztikus, így egyértelmű is. Ebből létrehozva a normalizált automatát a determinisztikusság eltűnhet, de az egyértelműség megmarad.

16. Definíció. [Kön94] Legyen az $(\mathcal{A}_i)_{i \in I}$ az $\mathcal{A}_i = (Q_i, q_0^{(i)}, \{t_i\}, \delta_i)$ normalizált automaták egy halmaza az A ábécé felett értelmezve. A jelölésben a Q_i a belső állapotok halmaza, $q_0^{(i)}$ a kezdő állapot, t_i az elfogadó állapot, valamint δ_i -vel jelöljük a mozgási szabályok

halmazát. Értelmezzük ezeknek a normalizált automatáknak az összefűzött szorzatát, mely az

$$\mathcal{A} = \left(\prod_{i \in I} Q_i, (q_0^{(i)})_{i \in I}, \{t_i\}_{i \in I}, A, \delta \right)$$

automata lesz a következő mozgási szabályokkal:

A

$$((p_i)_{i \in I}, (q_i)_{i \in I}, a) \in \delta$$

akkor, és csak akkor igaz, ha

$$\forall i \in I \quad (p_i, q_i, a) \in \delta_i, \text{ és } q_i = t_i$$

vagy

$$\exists J \subset I : \forall i \in J : (p_i, t_i, a) \in \delta \text{ és } q_i = q_0^{(i)}$$

$$\text{és } \forall i \in I - J : (p_i, q_i, a) \in \delta_i \text{ és } q_i \neq t_i.$$

A kódok felbonthatóságának vizsgálatához csak olyan összefűzött szorzat automatát fogok vizsgálni, mely összesen két tényezőtől áll, emiatt $I = \{1, 2\}$, s ebben az esetben ha az $\mathcal{A}_1 = (Q_1, q_0^{(1)}, \{t_1\}, A, \delta_1)$ automatának vesszük az összefűzött szorzatát az $\mathcal{A}_2 = (Q_2, q_0^{(2)}, \{t_2\}, A, \delta_2)$ automatával, akkor az $\mathcal{A}_1 \boxtimes \mathcal{A}_2 = (Q_1 \times Q_2, \{(q_0^{(1)}, q_0^{(2)})\}, \{(t_1, t_2)\}, \delta)$. A mozgási szabályokra teljesül, hogy

$$((p_1, p_2), (q_1, q_2), a) \in \delta$$

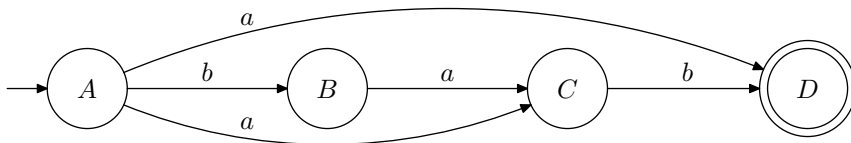
akkor és csak akkor igaz, ha következők közül valamelyik feltétel teljesül:

- $(p_1, q_1, a) \in \delta_1, (p_2, q_2, a) \in \delta_2$ és $((q_1, q_2) = (t_1, t_2) \text{ vagy } (q_1 \neq t_1 \text{ és } q_2 \neq t_2))$.

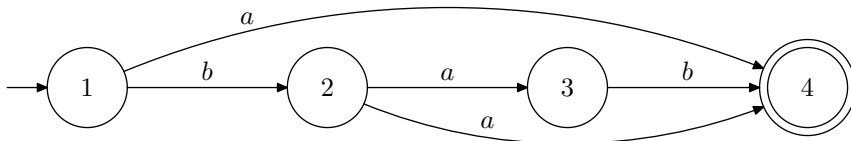
- $(p_1, q_1, a) \in \delta_1, q_1 \neq t_1$ és $(p_2, t_2, a) \in \delta_2, q_2 = q_0^{(2)}$.
- $(p_1, t_1, a) \in \delta_1, q_1 = q_0^{(1)}$ és $(p_2, q_2, a) \in \delta_2, q_2 \neq t_2$.

Tehát az összefűzött szorzat automatában a mozgási szabályok a szokásosak, ha mindkét automatában nem elfogadó, vagy mindkét automatában elfogadó állapotba megyünk az a szimbólummal. Ellenkező esetben az elfogadó állapot helyett a kezdő állapotot fogjuk írni.

PÉLDA: [Kön94] Legyenek adottak a 2.8. és a 2.9. ábrán látható automata.



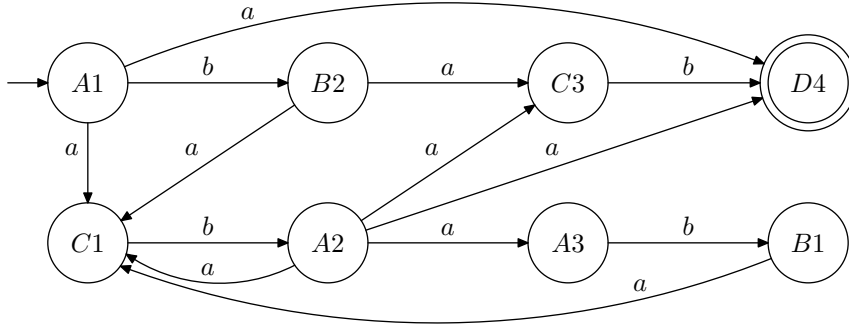
2.8. ábra. Az $\mathcal{A}_1$ automata



2.9. ábra. Az $\mathcal{A}_2$ automata

Ekkor megalkotva az összefűzött szorzatukat, a 2.10 automatát kapjuk. Látható, hogy az $\mathcal{A}_3$ automata normalizált, és a konstrukció miatt ez igaz minden összefűzött szorzat automatára.

17. Definíció. [Kön94] Egy összefűzött szorzat automata egy állapotát *reset* állapotnak nevezzük, ha nem kezdőállapot és legalább egy komponense kezdőállapot.

2.10. ábra. Az $\mathcal{A}_3 = \mathcal{A}_1 \boxtimes \mathcal{A}_2$ automata

18. Definíció. [Kön94] A két komponensű összefűzött szorzat automatában egy sikeres út elnevezése

- *Nulla oldalú*, ha nem érint egyetlen reset állapotot sem.
- *Jobb oldalú*, ha létezik legalább egy reset állapot az út során érintve, és az összes érintett reset állapot $(q, q_0^{(2)})$ alakú.
- *Bal oldalú*, ha létezik legalább egy reset állapot az út során érintve, és az összes érintett reset állapot $(q_0^{(1)}, q)$ alakú.
- *Egy oldalú*, ha vagy bal, vagy jobb oldalú.
- *Két oldalú*, ha az út során van érintve legalább egy $(q_0^{(1)}, q)$ és legalább egy $(q, q_0^{(2)})$ alakú reset állapot.

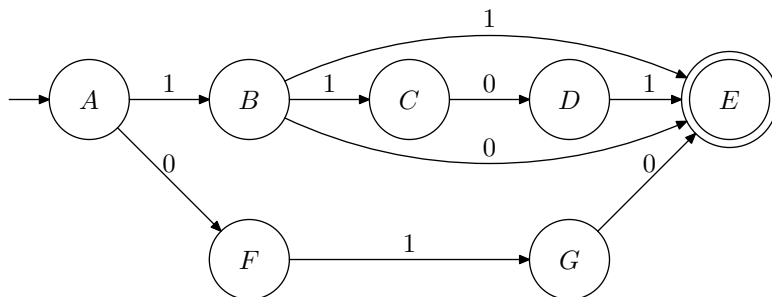
Bal oldalú reset állapotnak nevezzük a $(q_0^{(1)}, q)$ alakú állapotot, és hasonlóan jobb oldalú reset állapotnak nevezzük a $(q, q_0^{(2)})$ alakú állapotot.

Legyenek az X és Y halmazok az A^+ halmaz részhalmazai, valamint az $\mathcal{A}(X)$ és $\mathcal{A}(Y)$ automaták, melyek az adott halmazt ismerik fel. Tekintetbe véve a sikeres utakat, a következő halmazokat definiáljuk az $i = 0, 1, 2$ esetekben:

19. Definíció. [Kön94] $W_i(X, Y) = \{w | w \text{ egy } i \text{ oldalú sikeres útnak címkéje az } \mathcal{A}(X) \boxtimes \mathcal{A}(Y) \text{ automatában}\}$, és $W_i(X) = W_i(X, X)$

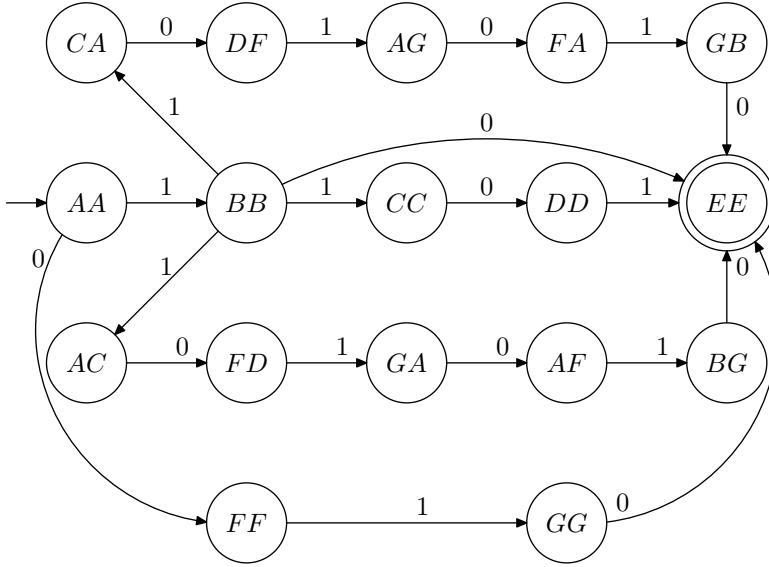
9. Tétel. [Kön94] Egy X kód akkor és csak akkor felbontható, ha $W_1(X) = W_2(X) = \emptyset$

PÉLDA: Vizsgáljuk meg az $X = \{010, 10, 11, 1101, \}$ kódot a fenti algoritmus szerint. A normalizált $\mathcal{A}(X)$ automata a 2.11. ábrán van megadva. Megkonstruáltam az $\mathcal{A}(X) \boxtimes \mathcal{A}(X)$ automatát, az



2.11. ábra. A normalizált $\mathcal{A}(X)$ automata

eredmény a 2.12. ábrán található. Ekkor a bal oldalú reset állapotok az AG, AC, AF állapotok lesznek, a jobb oldalú reset állapotok pedig a CA, FA, GA állapotok. Határozzuk meg a $W_1(X)$ halmazt. Egy oldalú sikeres utakat keresünk, azaz vagy bal oldalú, vagy jobb oldalú utakat. Ha figyelembe vesszük a 2.12. ábrát, láthatjuk, hogy mindegyik bal oldalú reset állapotot vagy megelőzi, vagy követi legalább egy jobb oldalú reset állapot, azaz AG -t megelőzi CA , AC -t követi GA , AF -t megelőzi GA , emiatt nincs bal oldalú sikeres út. Hasonló okok miatt jobb oldalú sincs, ugyanis CA -t követi AG , FA -t megelőzi AG , GA -t pedig követi AF . Tehát nincs egyoldalú út, azaz $W_1(X) = \emptyset$. A továbbiakban meghatározzuk a $W_2(X)$ halmazt,

2.12. ábra. Az $\mathcal{A}(X) \boxtimes \mathcal{A}(X)$ automata

kétoldalú utakat keresünk. Az $\mathcal{A}_3$ automatában két ilyen út van:

$$AA \xrightarrow{1} BB \xrightarrow{1} CA \xrightarrow{0} DF \xrightarrow{1} AG \xrightarrow{0} FA \xrightarrow{1} GB \xrightarrow{0} EE$$

$$AA \xrightarrow{1} BB \xrightarrow{1} AC \xrightarrow{0} FD \xrightarrow{1} GA \xrightarrow{0} AF \xrightarrow{1} BG \xrightarrow{0} EE$$

Ezek szerint $W_2(X) = \{1101010\}$. A 9. tétel miatt tehát a kód nem lesz felbontható, ugyanis bár $W_1(X) = \emptyset$, de $W_2(X) \neq \emptyset$.

2.7. A Tsuji-féle konstrukció

Az A ábécé feletti szavakat tartalmazó véges X halmaz esetén közzismert, hogy van olyan algoritmus, mely alapján eldönthető, hogy X felbontható kódot ad-e [BP85]. Defináljuk többértelmű szóként

azokat az X^* halmazba tartozó szavakat, melyeknek minimum két felbontása létezik, és konstruáljunk egy $\mathcal{A}_X$ automatát oly módon, hogy az $\mathcal{A}_X$ által elfogadott szavak $L(\mathcal{A}_X)$ halmaza egyenlő az X^* halmazban lévő összes többértelmű szó halmazával. Megmutatjuk, hogy az A ábécé feletti szavakat tartalmazó X halmaz felbontható kód akkor, és csak akkor, ha az $\mathcal{A}_X$ automata által felismert $L(\mathcal{A}_X)$ halmaz üres halmaz.

Az A ábécé feletti szavak X véges halmaza esetén legyen a $P(X)$ halmaz a következőképpen definiálva: $P(X) = \{p \in X \mid p \text{ valódi kezdőszelete valamely } X \text{ halmazbeli szónak}\}$, azaz $P(X) = \{p \in X \mid p \cdot q = r \in X, p \neq r, q \neq r\}$. Jelölje c_1 a $P(X)$ halmaz számosságát. Ekkor létezik egy φ injektív leképezés, mely $P(X)$ -et a természetes számok halmazára képezi le úgy, hogy $1 \leq \varphi(p) \leq c_1$ bármely $p \in P(X)$ esetén.

Hasonlóan az előbbiekhez, legyen $S(X)$ halmaz a következőképpen definiálva: $S(X) = \{s \in A^+ \mid s \text{ valódi végzőszelete valamely } X \text{ halmazbeli szónak}\}$, azaz $S(X) = \{s \in A^+ \mid q \cdot s = r \in X, s \neq r\}$. Ekkor létezik egy ψ injektív leképezés, mely $S(X)$ -et a természetes számok halmazára képezi le úgy, hogy $c_1 + 1 \leq \psi(s)$ bármely $s \in S(X)$ esetén.

Az A feletti $\mathcal{A}_X$ automatát induktív módon konstruáljuk. $\mathcal{A}_X$ élei, azaz a mozgási szabályai és $\mathcal{A}_X$ csomópontjai, azaz belső állapotai a következő szabályok által kerülnek definiálásra.

Legyen q_0 az $\mathcal{A}_X$ automata egyetlen kezdőállapota. A $\varphi(P(X))$ halmaz minden eleme $\mathcal{A}_X$ egy-egy állapota lesz és minden $p \in P(X)$ szó esetén

$$q_0 \xrightarrow{p} \varphi(p)$$

egy út lesz az $\mathcal{A}_X$ automatában. Mivel minden $p \in P(X)$ szó valamely $x \in X$ jelsorozatnak valódi prefix része, az $u = p^{-1}x$ szó szuffix része az x jelsorozatnak, tehát u eleme az $S(X)$ halmaznak. Ekkor a $\psi(u)$

az $\mathcal{A}_X$ automata belső állapota lesz és

$$\varphi(p) \xrightarrow{u} \psi(u)$$

egy út lesz az $\mathcal{A}_X$ automatában. Ha $\psi(u)$ az $\mathcal{A}_X$ automata egy állapota és $v \in S(X)$, akkor az uv konkatenációt írhatjuk a $uv = x_1 \cdots x_m$ alakban is, ahol $x_1, \dots, x_m \in X$ és v az x_m valódi szuffixuma. Ekkor $\psi(v)$ az $\mathcal{A}_X$ automata állapota lesz és

$$\psi(u) \xrightarrow{v} \psi(v)$$

egy út lesz az $\mathcal{A}_X$ automatában. Mivel $\varphi(P(X)) \cup \psi(S(X))$ és $S(X)$ végesek, az eljárás véges számú lépésből áll. Legyen Q a az $\mathcal{A}_X$ automata belső állapotainak halmaza. Az $\mathcal{A}_X$ automata elfogadó állapotainak halmaza a $\psi(S(X) \cap X^*) \cap Q$ halmaz lesz.

20. Definíció. [Tsu01] Egy w szót többértelműnek nevezünk az X halmazra nézve, ha léteznek $x_1, \dots, x_m, y_1, \dots, y_n$ elemei az X halmaznak, továbbá olyan, hogy $w = x_1 \cdots x_m = y_1 \cdots y_n$ és $x_1 \cdots x_i \neq y_1 \cdots y_j$, ha $1 \leq i \leq m-1$ és $1 \leq j \leq n-1$.

10. Tétel. [Tsu01] Legyen $\mathcal{A}_X$ a fentiek szerint definiálva. Egy A ábécé feletti X halmaz esetén az X halmazhoz tartozó $\mathcal{A}_X$ automata által elfogadott $L(\mathcal{A}_X)$ nyelv az X^* halmazban lévő többértelmű szavak halmaza.

11. Tétel. [Tsu01] Egy A ábécé feletti X halmaz felbontható kód akkor, és csak akkor, ha az X halmazhoz tartozó $\mathcal{A}_X$ automata által elfogadott $L(\mathcal{A}_X)$ nyelv az üres halmaz.

PÉLDA: [Tsu01] Legyen a vizsgálandó kódunk a következő: $\{aab, aabb, aabbab, aba, baa, bbaaba\}$. Ekkor a $P(X)$ halmaz elemei azok a kódszavak lesznek, melyek valódi prefixumai valamely más kódszónak. Tehát $aab \in P(X)$, mivel aab valódi

prefixuma az $aabb$ kódszónak (valamint az $aabbab$ kódszónak is, de ez nem befolyásoló tényező jelenleg). További elem még az $aabb$ kódszó is, mivel valódi prefixuma az $aabbab$ kódszónak. Tehát $P(X) = \{aab, aabb\}$. Így legyen $\varphi(aab) = 1$ és $\varphi(aabb) = 2$. Meghatározzuk az összes valódi postfixumát a kódszavaknak, tehát megkeressük az $S(x)$ elemeit. A kódszavak fenti sorrendje alapján kapjuk, hogy $S(X) = \{ab, b, abb, bb, abbab, bbab, bab, ba, a, aa, baaba, aaba, aba\}$. Nem feltétlenül szerepel a hozzájuk tartozó számok mindegyike az automatában, csak azokat tüntetjük fel, amelyek elérhetők a kezdő állapotból. Az elfogadó állapot definíciója szerint $Q_F = \psi(S(X) \cap X^*) \cap Q$, ebből már ismerjük az $S(X) \cap X^*$ halmazt, melynek egyetlen eleme az aba kódszó, tehát amennyiben szerepel az automatában, $\psi(aba) = 15$ lesz az elfogadó állapot, ugyanis a $\psi(S(X))$ a következőképpen van definiálva: $\psi(b) = 3, \psi(bab) = 4, \psi(ab) = 5, \psi(aa) = 6, \psi(baaba) = 7, \psi(bbab) = 8, \psi(a) = 9, \psi(ba) = 10, \psi(abb) = 11, \psi(bb) = 12, \psi(aaba) = 13, \psi(abbab) = 14, \psi(aba) = 15$. Mivel a $q_0 \xrightarrow{p} \varphi(p)$ alakú szabályokkal kezdjük az automata konstruálását, ezért a kezdő állapotból két mozgási szabállyal léphetünk tovább, nevezetesen $q_0 \xrightarrow{aabb} \varphi(aabb) = 2$ és a $q_0 \xrightarrow{aab} \varphi(aab) = 1$ szabályok által. A továbbiakban a $\varphi(p) \xrightarrow{u} \psi(u)$ alakú szabályokat kell alkalmaznunk. Ez három mozgási szabályt fog eredményezni:

- $\varphi(aabb) \xrightarrow{ab} \psi(ab)$ miatt $2 \xrightarrow{ab} 5$
- $\varphi(aab) \xrightarrow{b} \psi(b)$ miatt $1 \xrightarrow{b} 3$
- $\varphi(aab) \xrightarrow{bab} \psi(bab)$ miatt $1 \xrightarrow{bab} 4$

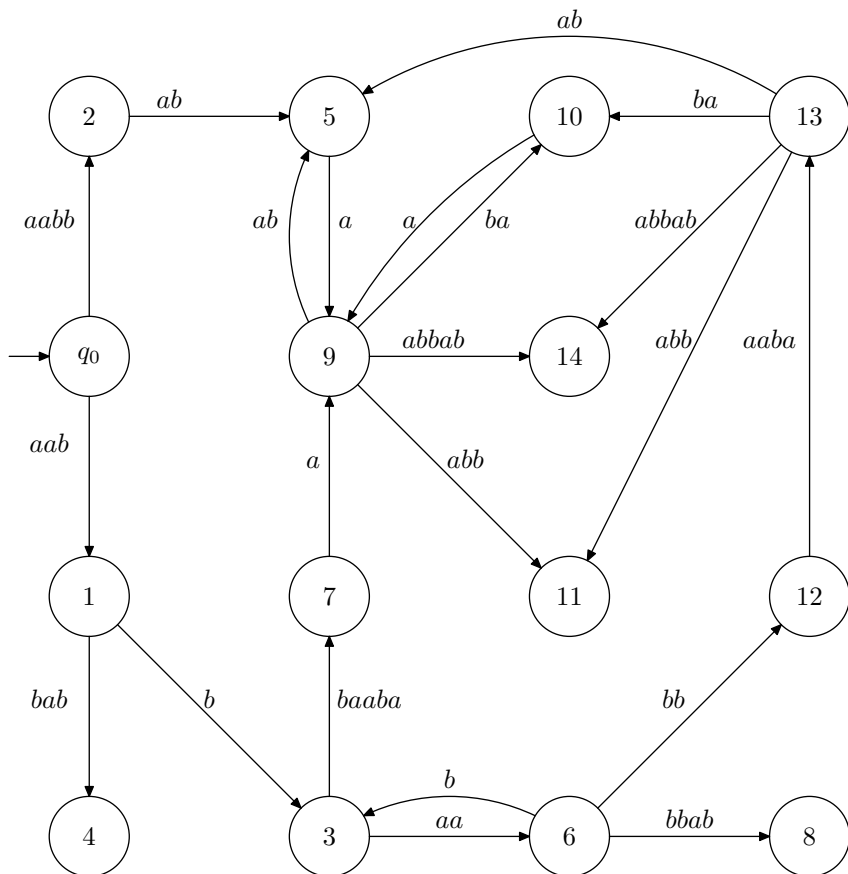
Ezzel kimerítettük a $\varphi(p) \xrightarrow{u} \psi(u)$ alakú szabályok által nyújtott lehetőségeket, mivel a prefixumok $P(X)$ halmazából nem tudunk más szuffixummal kiegészíteni egyetlen elemet sem úgy, hogy kódszót

kapjunk. A továbbiakban a $\psi(u) \xrightarrow{v} \psi(v)$ alakú szabályokat fogjuk alkalmazni minden $\psi(S(X))$ állapotra. A szabály alkalmazásával újabb $\psi(S(X))$ állapotok kerülnek be az automatába, de nyilván véges lesz az eljárásunk. Tehát például az 5 állapotra alkalmazva a szabályt, kapjuk, hogy $\psi(ab) \xrightarrow{a} \psi(a)$, azaz $5 \xrightarrow{a} 9$. További magyarázatot mellőzve megadom az $\mathcal{A}_X$ automata grafikus reprezentációját a 2.13. ábrán.

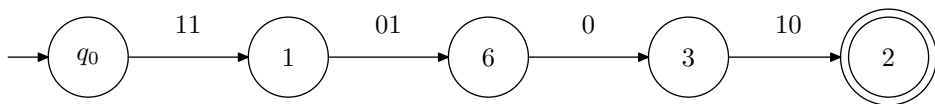
Látható, hogy a remélt $\psi(aba) = 15$ állapot nem szerepel az automatában, mivel nem érhető el a q_0 állapotból. Emiatt $Q_F = \psi(S(X) \cap X^*) \cap Q = \emptyset$. Így elfogadó állapotok hiányában az $\mathcal{A}_X$ automata $L(\mathcal{A}_X)$ nyelve üres, tehát a 11. tétel miatt az $X = \{aab, aabb, aabbab, aba, baa, bbaaba\}$ kód felbontható.

1. Megjegyzés. Helyre igazítottam két mozgási szabályt, ugyanis a cikkben szereplő megoldásban tévesen voltak megadva. Az itt megadott ábra az általam elvégzett korrekciókat tartalmazza.

Mutatok arra is egy saját példát, amikor a kód nem lesz felbontható. PÉLDA: Legyen adva a $K = \{010, 10, 11, 1101\}$ kód. Vizsgáljuk meg felbonthatóság szempontjából, azaz konstruáljuk meg az automatát. $P(K) = \{11\}$, így definíció szerint $\varphi(11) = 1$. $S(X) = \{10, 0, 1, 101, 01\}$. Legyen $\psi(S(K))$ a következőképpen adott: $\psi(10) = 2, \psi(0) = 3, \psi(1) = 4, \psi(101) = 5, \psi(01) = 6$. A végállapot a $\psi(S(K) \cap K^*) = \{\psi(10)\} = \{2\}$ lehet, amennyiben megjelenik az automatában, azaz a q_0 állapotból elérhető. Az automata ábrája a 2.14. ábrán látható. Az esélyes $\psi(10) = 2$ állapot szerepel az automatában, emiatt $Q_F = \psi(S(K) \cap K^*) \cap Q = \{\psi(10)\} = \{2\}$. Így az $\mathcal{A}_X$ automata által elfogadott jelsorozatok halmaza nem üres halmaz, ugyanis $L(\mathcal{A}_X) = \{1101010\}$. Tehát a nyelv nem üres, így az 11. tétel miatt a $K = \{010, 10, 11, 1101\}$ kód nem lesz felbontható. Valóban, látható, hogy a 1101010 jelsorozat előáll kétféle felbontásban a K kód kódszavainak szorzataként: $11\ 010\ 10 = 1101\ 010$



2.13. ábra. Az $X = \{aab, aabb, aabbab, aba, baa, bbaaba\}$ kód $\mathcal{A}_X$ automatája



2.14. ábra. A $K = \{010, 10, 11, 1101\}$ kód $\mathcal{A}_K$ automatája

3. fejezet

Eredmények

3.1. Egy, a felbonthatóságot eldöntő új algoritmus

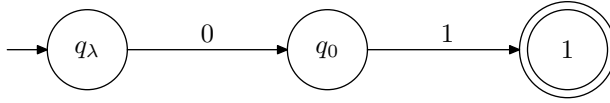
3.1.1. Az algoritmus megadása

Az én algoritmusom is automataelméleti alapokon nyugszik. Az A ábécé feletti C kódot elfogadó automatát a kódszavakat elfogadó automaták segítségével konstruálom meg. Ha a w_i kódszó $x_1x_2 \dots x_n$, $x_j \in A$ alakú, akkor a hozzá tartozó $\mathcal{A}(\{w_i\})$ automata $\mathcal{A}(\{w_i\}) = (Q^{(i)}, q_\lambda, Q_F^{(i)}, A, \delta^{(i)})$ lesz, ahol $Q^{(i)}$ az állapotok halmaza, q_λ az $\mathcal{A}(\{w_i\})$ kezdőállapota és $Q_F^{(i)}$ az elfogadó állapotot tartalmazó egy elemű halmaz. Esetünkben $Q_F^{(i)} = \{i\}$ és $|Q^{(i)}| = \text{length}(w) + 1$,

mivel az $\mathcal{A}(\{w_i\})$ automata mozgási szabályai a következők:

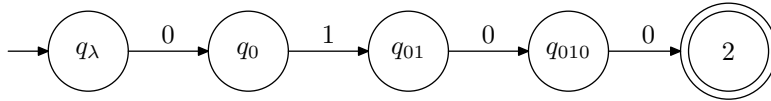
$$\begin{aligned}
 \delta(q_\lambda, x_1) &= \{q_{x_1}\} \\
 \delta(q_{x_1}, x_2) &= \{q_{x_1x_2}\} \\
 &\vdots \\
 \delta(q_{x_1x_2\dots x_{n-2}}, x_{n-1}) &= \{q_{x_1x_2\dots x_{n-2}x_{n-1}}\} \\
 \delta(q_{x_1x_2\dots x_{n-1}}, x_n) &= \{i\}
 \end{aligned}$$

Így $\mathcal{A}(\{w_i\})$ képes felismerni a w_i kódszót. A 3.1. ábrán a $w_1 = 01$ kódszó automatája látható.



3.1. ábra. Az $\mathcal{A}(\{w_1\}) = \mathcal{A}(\{01\})$ automata

Ha w_1 prefix része a w_2 jelsorozatnak, akkor az automatáiknak lesznek közösen jelölt állapotai, pontosabban $(Q^{(1)} \setminus \{1\}) \subset Q^{(2)}$. Ezt a tulajdonságot figyelhetjük meg a 3.1. és a 3.2. ábra összehasonlításánál, ahol $w_1 = 01$ és $w_2 = 0100$.



3.2. ábra. Az $\mathcal{A}(\{w_2\}) = \mathcal{A}(\{0100\})$ automata

Mivel w_1 prefix része a w_2 kódszónak, használhatjuk a következő jelölést:

$$w_1 = x_1x_2 \cdots x_n; \quad w_2 = x_1x_2 \cdots x_nx_{n+1} \cdots x_m$$

Jelölje $\delta^{(1)}$ az $\mathcal{A}(\{w_1\})$ automata mozgási szabály halmazát, s hasonlóan $\delta^{(2)}$ az $\mathcal{A}(\{w_2\})$ automata mozgási szabály halmazát. Az $\mathcal{A}(\{w_1\})$ és az $\mathcal{A}(\{w_2\})$ automaták mozgási szabályai a következők lesznek általános esetben, ha w_1 prefixuma a w_2 jelsorozatnak.

A $\delta^{(1)}$ mozgási szabály halmaz:

$$\begin{aligned} \{\delta(q_\lambda, x_1) &= \{q_{x_1}\}, \\ \delta(q_{x_1}, x_2) &= \{q_{x_1x_2}\}, \\ &\vdots \\ \delta(q_{x_1x_2\dots x_{n-2}}, x_{n-1}) &= \{q_{x_1x_2\dots x_{n-2}x_{n-1}}\}, \\ \delta(q_{x_1x_2\dots x_{n-1}}, x_n) &= \{1\}\} \end{aligned}$$

A $\delta^{(2)}$ mozgási szabály halmaz:

$$\begin{aligned} \{\delta(q_\lambda, x_1) &= \{q_{x_1}\}, \\ \delta(q_{x_1}, x_2) &= \{q_{x_1x_2}\}, \\ &\vdots \\ \delta(q_{x_1x_2\dots x_{n-2}}, x_{n-1}) &= \{q_{x_1x_2\dots x_{n-2}x_{n-1}}\}, \\ \delta(q_{x_1x_2\dots x_{n-1}}, x_n) &= \{q_{x_1x_2\dots x_{n-1}x_n}\}, \\ &\vdots \\ \delta(q_{x_1x_2\dots x_{m-2}}, x_{m-1}) &= \{q_{x_1x_2\dots x_{m-2}x_{m-1}}\}, \\ \delta(q_{x_1x_2\dots x_{m-1}}, x_m) &= \{2\}\} \end{aligned}$$

Látható, hogy

$$(\delta^{(1)} \setminus \{\delta(q_{x_1x_2\dots x_{n-1}}, x_n) = 1\}) \subset \delta^{(2)},$$

emiatt

$$\delta^{(1)} \cup \delta^{(2)} = \delta^{(2)} \cup \{\delta(q_{x_1x_2\dots x_{n-1}}, x_n) = \{1\}\}.$$

Legyen

$$\mathcal{A}(\{w_1, w_2\}) = (Q, q_\lambda, Q_F, A, \delta^{(1,2)}),$$

ahol

$$Q^{(1,2)} = Q^{(1)} \cup Q^{(2)}, Q_F^{(1,2)} = \{1, 2\},$$

továbbá

$$\delta^{(1,2)} = \delta^{(1)} \cup \delta^{(2)} \cup \{\delta(1, x_1) = \{q_{x_1}\}, \delta(2, x_1) = \{q_{x_1}\}\}.$$

Mivel

$$\begin{aligned} \delta(q_{x_1 x_2 \dots x_{n-1}}, x_n) &= \{1\} && \in \delta^{(1)} \\ \delta(q_{x_1 x_2 \dots x_{n-1}}, x_n) &= \{q_{x_1 x_2 \dots x_{n-1} x_n}\} && \in \delta^{(2)}, \end{aligned}$$

emiatt $\mathcal{A}(\{w_1, w_2\})$ nemdeterminisztikus véges automata, ugyanis a szabályok bal oldalai azonosak. Látható, hogy az $\mathcal{A}(\{w_1, w_2\})$ automata éppen a $\{w_1, w_2\}^+$ nyelvet fogadja el.

Figyelembe véve a nemfelbontható jelsorozatokot, egyértelmű, hogy a lehetséges különböző felbontásokban léteznie kell legalább két olyan kódszónak, hogy az egyik prefix része a másiknak. Tehát, ha $u_1 \dots u_n = w_1 \dots w_m$, akkor létezik egy k küszöbszám úgy, hogy bármely $i < k$ esetén $u_i = w_i$, de $u_k \neq w_k$ ahol $1 \leq k \leq \min\{n, m\}$.

Egyesítsük a fentebb bemutatott módon a kódszavakhoz tartozó automatákat, ekkor az $\mathcal{A}(\{w_1, \dots, w_n\})$ automatát kapjuk a $C = \{w_1, \dots, w_n\}$ kód esetén. Használhatjuk a kódhoz tartozó automatára az $\mathcal{A}(C)$ rövidebb jelölést is. Jelölje $x_1^{(l)}$ az l . kódszó első szimbólumát. Tehát

$$\mathcal{A}(C) = (q, Q_F = \{1, \dots, n\}, Q = Q^{(1)} \cup \dots \cup Q^{(n)}, A, \delta),$$

ahol

$$\delta = \delta^{(1)} \cup \dots \cup \delta^{(n)} \cup \{\delta(1, x_1^{(1)}) = q_{x_1^{(1)}}, \dots, \delta(1, x_1^{(n)}) = q_{x_1^{(n)}}, \dots,$$

$$\delta(n, x_1^{(1)}) = q_{x_1^{(1)}}, \dots, \delta(n, x_1^{(n)}) = q_{x_1^{(n)}}\}.$$

A konstrukcióból következik, hogy az $\mathcal{A}(C)$ automata pontosan a C^+ nyelvet fogja elfogadni.

12. Tétel. [Fal06][Fal11b] Ha az $\mathcal{A}(C)$ automata determinisztikus, akkor a C kód felbontható.

BIZONYÍTÁS: Emlékezzünk vissza arra, hogy mindegyik prefix kód felbontható az 1. tétel miatt. Emiatt elegendő azt megmutatni, hogy a C kód prefix tulajdonsággal bír, amennyiben az $\mathcal{A}(C)$ automata determinisztikus. Tegyük fel –az állításunkkal ellentétben–, hogy az $\mathcal{A}(C)$ automata determinisztikus, és a C kód nem prefix tulajdonságú. Ekkor létezik a w_i és a w_j kódszó, és teljesül rájuk a $w_i = w_j\alpha$, ahol $\alpha \in \Delta^+$. Részletesebben kifejtve, léteznek az $x_1, \dots, x_{|w_j|}, \dots, x_{|w_i|} \in \Delta$ szimbólumok úgy, hogy $w_j = x_1 \cdots x_{|w_j|}$ és $w_i = x_1 \cdots x_{|w_j|} \cdots x_{|w_i|}$. A konstrukciónk miatt ebből viszont következik, hogy $\delta(q_{x_1 \cdots x_{|w_j|-1}}, x_{|w_j|}) = q_{x_1 \cdots x_{|w_j|}} \notin Q_F$ és létezik $j \in Q_F$ úgy, hogy $\delta(q_{x_1 \cdots x_{|w_j|-1}}, x_{|w_j|}) = j$ és $j \neq q_{x_1 \cdots x_{|w_j|}}$. De akkor $\mathcal{A}(C)$ nemdeterminisztikus. Emiatt, ha $\mathcal{A}(C)$ determinisztikus, akkor a C kód prefix, ahogy állítottuk. $\square$

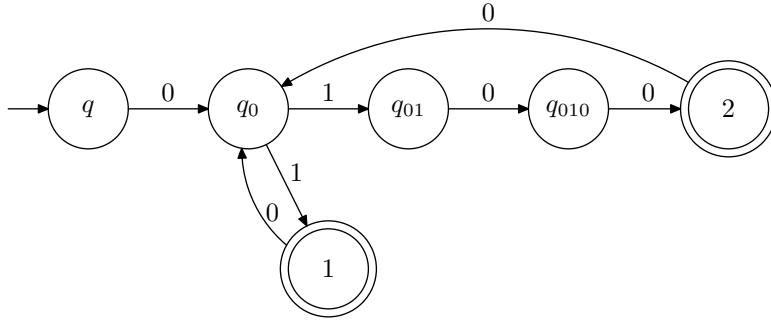
2. Megjegyzés. Léteznek nemprefix kódok, melyek felbonthatók, például a $C = \{01, 0100\}$ kód. Tehát felbontható kódhoz tartozhat nemdeterminisztikus automata, azaz nem fordítható meg a tétel. A $C = \{01, 0100\}$ kódhoz tartozó $\mathcal{A}(C)$ automata grafikus reprezentációja a 3.3. ábrán látható.

Az automata a

$$\delta(q_0, 1) = \{1, q_{01}\}$$

szabály miatt nemdeterminisztikus, de a kód felbontható. Természetesen a nemfelbontható kódokhoz tartozó automaták nemdeterminisztikusak lesznek, ha a fenti módszerrel konstruáljuk meg őket.

A 12. tétel csupán elégséges kritériumot szolgáltat, a következőkben megadom a felbonthatóság egy szükséges és elégséges feltételét. A konstrukció a 2. tételben megadott kapcsolaton alapszik.

3.3. ábra. Az $\mathcal{A}(\{01, 0100\})$ automata

Ha a v jelsorozat előállítható a C kód segítségével, akkor az $\mathcal{A}(C)$ automata elfogadja a v jelsorozatot, azaz $\mathcal{A}(C)$ elolvassa és valamely elfogadó állapotban áll meg. Ha v több, mint egyféleképpen előállhat C kódszavai konkatenációjaként, akkor különböző állapotszekvenciákat kapunk a különböző elolvasások során. Ezeket a különböző utakat egyesítem a determinisztikus automata által.

Jelölje $\mathcal{N}(w)$ a w kódszó indexét a C halmazban, ekkor $\mathcal{N}(w)$ elfogadó állapot lesz az $\mathcal{A}(C)$ automatában a konstrukció miatt. Legyen továbbá $u_1 \cdots u_n = v_1 \cdots v_m$, ahol $u_i, v_j \in C$. Jelölje k a legkisebb olyan indexet, ahol $u_k \neq v_k$. Ekkor $u_{k+1} \cdots u_n \neq v_{k+1} \cdots v_m$. Legyen a d_1 és d_2 számokra igaz, hogy $\forall 1 \leq i < d_1$ és $\forall 1 \leq j < d_2$ esetén $u_k \cdots u_{k+i} \neq v_k \cdots v_{k+j}$, de $u_k \cdots u_{k+d_1} = v_k \cdots v_{k+d_2}$. Ekkor d_1 és d_2 definíciója miatt $u_{k+d_1} \neq v_{k+d_2}$, mert ellenkező esetben $u_k \cdots u_{k+d_1-1} = v_k \cdots v_{k+d_2-1}$ teljesülne, ami ellentmond a feltevésünknek. Jelölje q_λ a kezdő állapotot. Így a $q_\lambda \xrightarrow{u_1 \cdots u_{k+d_1}} \mathcal{N}(u_{k+d_1})$ út és a $q_\lambda \xrightarrow{v_1 \cdots v_{k+d_2}} \mathcal{N}(v_{k+d_2})$ különböző utak lesznek az $\mathcal{A}(C)$ automatában. Megalkotva az $\mathcal{A}(C)$ automatához tartozó $\mathcal{A}_D(C)$ determinisztikus véges automatát, ez a két út egyesítve lesz, és egy olyan állapotba fognak vezetni, melyben szerepel $\mathcal{N}(u_{k+d_1})$ és

a $\mathcal{N}(v_{k+d_2})$ állapot is.

Így egy jelsorozat két (vagy több) különböző felbontása csak úgy lehetséges, ha a determinisztikus automata valamely állapotában több, mint egy elfogadó állapota szerepel a nemdeterminisztikus automatának. Jelölje $Q_F^{\mathcal{A}(C)}$ az $\mathcal{A}(C)$ automata elfogadó állapotainak halmazát.

13. Tétel. [Fal04, Fal06, Fal11b] Egy kód felbontható akkor és csak akkor, ha az $\mathcal{A}_D(C)$ automatában legfeljebb egy – az $\mathcal{A}(C)$ automatához tartozó – elfogadó állapot jelenik meg bármely szabály jobboldalán, azaz bármely szabályra igaz az, hogy ha $\delta(\{q_{i_1}, \dots, q_{i_n}\}, x) = \{q_{j_1}, \dots, q_{j_m}\}$ az $\mathcal{A}_D(C)$ automatában, akkor nem létezik $l \neq k$ úgy, hogy $q_{j_l} \in Q_F^{\mathcal{A}(C)}$ és $q_{j_k} \in Q_F^{\mathcal{A}(C)}$ is egyszerre teljesül.

BIZONYÍTÁS: A bizonyítás indirekt úton történik. Tegyük fel, hogy a kód felbontható, és a determinisztikus automata valamely állapotában legalább két elfogadó állapota szerepel a nemdeterminisztikus automatának, azaz létezik $\delta(\{q_{i_1}, \dots, q_{i_n}\}, x) = \{q_{j_1}, \dots, q_{j_m}\}$ és $\exists l \neq k$ úgy, hogy $q_{j_l} \in Q_F^{\mathcal{A}(C)}$ és $q_{j_k} \in Q_F^{\mathcal{A}(C)}$. Jelölje v a q kezdő állapotból a $\{q_{j_1}, \dots, q_{j_m}\}$ állapotba vezető út során olvasott szimbólum sorozatot, azaz $q \xrightarrow{v} \{q_{j_1}, \dots, q_{j_m}\}$. Ekkor a nemdeterminisztikus automatában $q \xrightarrow{v} q_{j_l}$ és $q \xrightarrow{v} q_{j_k}$ két különböző sikeres út lesz, azaz két különböző felbontása van a v jelsorozatnak, tehát a kód nemfelbontható. Ellentmondásra jutottunk, emiatt igaz a tétel egyik irányban.

A másik irány bizonyításához tegyük fel, hogy nincs olyan állapot a determinisztikus automatában, melyben legalább két különböző állapot szerepel a nemdeterminisztikus automata elfogadó állapotai közül és a kód nem felbontható. Viszont ha nem felbontható, akkor létezik a kódszavaknak legalább két olyan sorozata, hogy $w_{i_1} \dots w_{i_n} = w_{j_1} \dots w_{j_m}$ és $w_{i_n} \neq w_{j_m}$. Az automata determinisztikus,

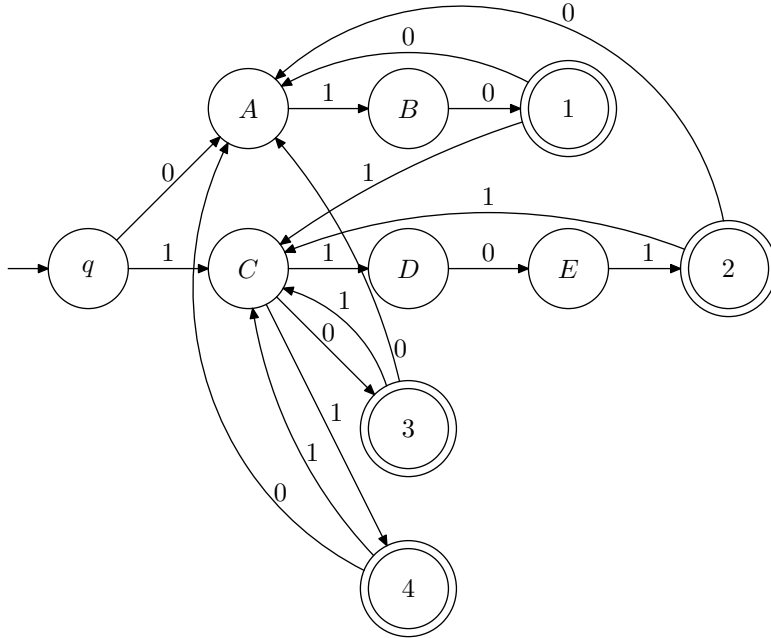
ezért ugyanazon állapotba kerülünk mindkét jelsorozat olvasása után a determinisztikus automatában. Ekkor viszont ez az állapot tartalmazni fogja a nemdeterminisztikus automata i_n és a j_m elfogadó állapotait a w_{i_n} és a w_{j_m} kódszavak miatt. Ellentmondásra jutottunk, tehát igaz a tétel. $\square$

PÉLDA: Legyen $C = \{010, 1101, 10, 11\}$.

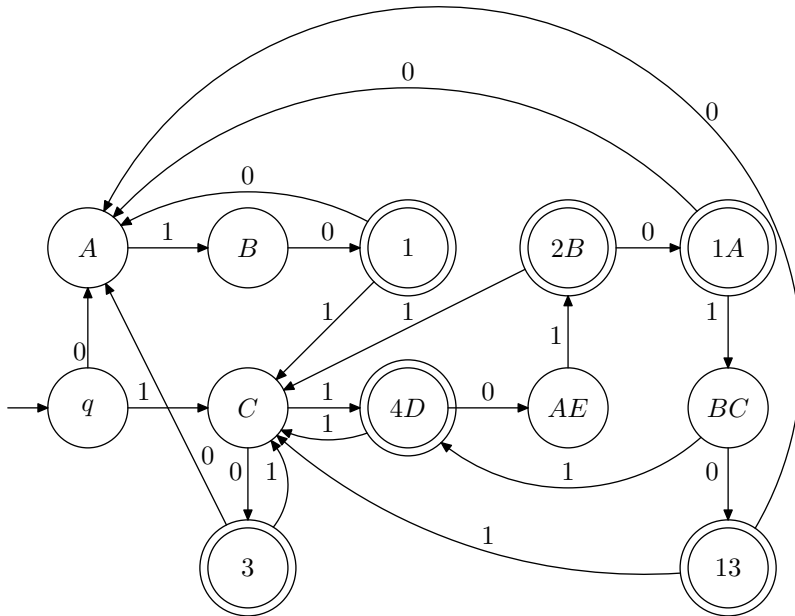
A kódhoz tartozó $\mathcal{A}(\{010, 1101, 10, 11\})$ automata a 3.4. ábrán látható. Konstruáljuk meg a hozzá tartozó $\mathcal{A}_D(\{010, 1101, 10, 11\})$ determinisztikus automatát. Az eredmény az 3.5. ábrán van megadva. Látható, hogy

$$\delta(\{B, C\}, 0) = \{1, 3\},$$

tehát a kód nemfelbontható.



3.4. ábra. Az $\mathcal{A}(\{010, 1101, 10, 11\})$ automata

3.5. ábra. Az $\mathcal{A}_D(\{010, 1101, 10, 11\})$ automata

3.1.2. Az algoritmus és a Sardinas – Patterson algoritmus kapcsolata

Amint már megjegyeztem, a Sardinas – Patterson féle algoritmus a többszörös felbontásnál jelentkező maradványok vizsgálatán alapszik. A többszörös felbontásnak szükséges feltétele, hogy különböző kódszavakkal kezdődjenek a felbontások, tehát a kód ne legyen prefixmentes. Valamely kódszó tehát valódi kezdőszelete valamely másik kódszónak, emiatt lesz nemdeterminisztikus az $\mathcal{A}(C)$ automata.

A (2.2) egyenletek alapján evidens, hogy az $U = \bigcup_{i \geq 1} U_i$ halmaz minden maradványt tartalmaz. Az U_i halmaz definíciója rekurzív, de véges halmazról van szó, mivel bármely többszörös felbontásból adódó maradvány hossza nem lehet nagyobb, mint a leghosszabb kódszó. Tekintetbe véve a Sardinas – Patterson algoritmus alapjául szolgáló többszörös felbontásokat, legyen $u_1 \cdots u_n \alpha = v_1 \cdots v_m$, ahol $u_i, v_j \in C$ és $\text{length}(u_1 \cdots u_k) \neq \text{length}(v_1 \cdots v_l)$ bármely $k \leq n, l \leq m$ esetén, kivéve az egyetlen $\alpha = \lambda, k = n, l = m$ esetet. Ekkor az α maradvány szuffix részé a v_m kódszónak úgy, hogy $\alpha \neq v_m$. Az α maradvány nem nagyobb a leghosszabb kódszónál, azaz lehet λ , vagy valamely kódszó is. A többszörös felbontásokban jelentkező maradványok meghatározásához tehát az $u_1 \cdots u_n \alpha = v_1 \cdots v_m$ jelsorozatokot kell figyelemmel kísérni az automatákban. Jelölje x_i az automata input szimbólumait, így $u_1 \cdots u_n = x_1 x_2 \cdots x_p$, $v_1 \cdots v_m = x_1 x_2 \cdots x_p x_{p+1} \cdots x_r$, és $\alpha = x_{p+1} \cdots x_r$. Nyilván $q_\lambda \xrightarrow{u_1 \cdots u_n} \mathcal{N}(u_n)$ és $q_\lambda \xrightarrow{v_1 \cdots v_m} \mathcal{N}(v_m)$, ahol $\mathcal{N}(u_n)$ és $\mathcal{N}(v_m)$ elemei a $Q_F^{\mathcal{A}(C)}$ halmaznak. Ekkor az $\mathcal{A}(C)$ nemdeterminisztikus automatában egy α maradvány a következőképpen jelenik meg:

- Ha $\alpha \neq \lambda$, akkor egyrészt $q_\lambda \xrightarrow{x_1 x_2 \cdots x_p} \mathcal{N}(u_n)$ a $q_\lambda \xrightarrow{u_1 \cdots u_n} \mathcal{N}(u_n)$ út miatt, másrészt pedig $q_\lambda \xrightarrow{x_1 x_2 \cdots x_p} q$ a $q_\lambda \xrightarrow{v_1 \cdots v_m} \mathcal{N}(v_m)$ út miatt, ahol $q \notin Q_F^{\mathcal{A}(C)}$, mert α valódi szuffixuma a v_m kódszónak. Tehát az $\mathcal{A}_D(C)$ automatában $\mathcal{N}(u_n)$ és q együtt fog szerepelni

valamely állapotban.

- Ha $\alpha = \lambda$, akkor egyrészt $q_\lambda \xrightarrow{x_1 x_2 \cdots x_p} \mathcal{N}(u_n)$ a $q_\lambda \xrightarrow{u_1 \cdots u_n} \mathcal{N}(u_n)$ út miatt, másrészt pedig $q_\lambda \xrightarrow{x_1 x_2 \cdots x_p} \mathcal{N}(v_m)$ a $q_\lambda \xrightarrow{v_1 \cdots v_m} \mathcal{N}(v_m)$ út miatt. Tehát az $\mathcal{A}_D(C)$ automatában $\mathcal{N}(u_n)$ és $\mathcal{N}(v_m)$ együtt fog szerepelni valamely állapotban.

A q által indukált maradványokat megkapjuk, ha az $\mathcal{A}(C)$ automatában megkeressük azokat a jelsorozatok, melyekkel a q állapotból –csak elutasító állapotokat érintve– elfogadó állapotba kerülünk.

2. Következmény. [Fal08a] Minden α maradvány eleme az $\mathcal{A}_D(C)$ automata valamely állapotában valamely $\mathcal{N}(w)$ állapottal együtt szereplő q állapot által indukált maradvány halmaznak, ahol $\alpha \in U$, $\mathcal{N}(w)$ és q eleme a $Q^{\mathcal{A}(C)}$ halmaznak.

Legyen c_1 és c_2 eleme a C^* halmaznak, továbbá legyen az iménti jelölések mellett $\mathcal{N}(u_n)$ és q eleme a $Q^{\mathcal{A}(C)}$ halmaznak, valamint az $\mathcal{A}(C)$ automatában legyen a $q_\lambda \xrightarrow{\beta} q$ út olyan, hogy útközben legfeljebb csak utolsó lépésben érintünk elfogadó állapotot. Ha az $\mathcal{A}_D(C)$ automata valamely q_D állapotában a $\mathcal{N}(u_n)$ állapot együtt van a q állapottal, akkor létezik $c_1 = c_2$ úgy, hogy mivel $\mathcal{N}(u_n) \in q_D$, ezért $q_\lambda \xrightarrow{c_1 u_1 \cdots u_n} q_D$, és mivel $q \in q_D$, ezért $q_\lambda \xrightarrow{c_2 v_1 \cdots v_{m-1} \beta} q_D$. Emiatt $c_1 u_1 \cdots u_n = c_2 v_1 \cdots v_{m-1} \beta$. Legyen α a q által indukált valamely maradvány. Mivel $q \in Q^{\mathcal{A}(C)}$, legalább egy ilyen α biztos létezik. Így $c_1 u_1 \cdots u_n \alpha = c_2 v_1 \cdots v_{m-1} \beta \alpha$, ahol $\beta \alpha \in C$. Jelölje a $\beta \alpha$ kódszót v_m , ekkor elhagyva a $c_1 = c_2$ jelsorozatokat kapjuk, hogy $u_1 \cdots u_n \alpha = v_1 \cdots v_m$, amiből következik a Sardinas–Patterson algoritmust figyelembe véve, hogy $\alpha \in U$. Megjegyezzük, hogy c_1 és c_2 nem feltétlenül azonos módon bomlik kódszavak szorzatára.

3. Következmény. [Fal08a] Ha az $\mathcal{A}_D(C)$ automata egy q_D állapotában valamely $\mathcal{N}(w)$ állapottal együtt szerepel egy q állapot, akkor a q által indukált maradvány halmaz minden α eleme benne van az U halmazban, ahol $\alpha \in U$, $\mathcal{N}(w)$ és q eleme a $Q^{\mathcal{A}(C)}$ halmaznak.

Jelölje $R(q)$ az $\mathcal{A}(C)$ automata q állapota által indukált maradványok halmazát, és legyenek az U_i halmazok a Sardinas–Patterson algoritmus alapján létrehozott halmazok. Ekkor a 2. és a 3. következményekből adódik a következő állítás.

4. Következmény. [Fal07, Fal08a] $\bigcup_{i \geq 1} U_i = \bigcup_{\{\mathcal{N}(w), q\} \subseteq q_D} R(q)$, ahol $\mathcal{N}(w) \in Q_F^{\mathcal{A}(C)}$, $q_D \in Q_F^{\mathcal{A}_D(C)}$

PÉLDA: Legyen $C = \{010, 1101, 10, 11\}$. A hozzá tartozó $\mathcal{A}(010, 1101, 10, 11)$ automata a 3.4. ábrán látható. Kiszámolva az $R(q)$ halmazokat az $\mathcal{A}(C)$ automata állapotaira a 3.1. számú táblázatot kapjuk. Egy q állapot által indukált maradványt úgy határozzunk meg, hogy követjük az összes olyan utat, mely során csak utolsó lépésben érintünk elfogadó állapotot. Minden elfogadó állapot a λ maradványt indukálja egyedül. Ha az $\mathcal{A}(C)$ automata q állapotából több elfogadó állapot érhető el az elébb említett módon, akkor a q állapot több maradványt indukál. Erre láthatunk példát a C állapot esetén.

Konstruáljuk meg az $\mathcal{A}_D(010, 1101, 10, 11)$ automatát. Az eredmény a 3.5. ábrán van megadva. Láthatjuk, hogy

$$\{1, 3\} \in \delta(\{B, C\}, 0),$$

így λ egy a maradványok közül (ebben az esetben a kód nemfelbontható).

Végezetül megadjuk az $\mathcal{A}_D(C)$ automata alapján azokat az állapotokat, melyek valamely elfogadó állapottal közös állapotban jelentek meg: $\{A, B, D, 1, 3\}$. Így a lehetséges maradványok a

Állapot	Maradványok
A	10
B	0
C	1, 0, 101
D	01
E	1
1	λ
2	λ
3	λ
4	λ

3.1. táblázat. Az $\mathcal{A}(010, 1101, 10, 11)$ automata maradványai

következők lesznek: $\{10, 0, 01, \lambda\}$. Ugyanerre az eredményre jutunk a Sardinas–Patterson algoritmus alkalmazásával is:

$$U_1 = \{01\} \quad U_2 = \{0\} \quad U_3 = \{10\} \quad U_4 = \{\lambda\},$$

azaz $\bigcup U_i = \{01, 0, 10, \lambda\}$.

Megmutatom, hogy még szorosabb kapcsolat áll fent az automata és a Sardinas–Patterson algoritmus között. Az U_i halmaz definíciójából kitűnik, hogy az elemei két olyan sztringből származnak, melyek együttesen i darab teljes kódszót tartalmaznak, kivéve a λ esetben, ahol a kódszavak száma $i + 1$. Az $\mathcal{A}_D(C)$ automatában a kódszavak számát az érintett elfogadó állapotok számából kapjuk meg. Ez nem pontosan egyenlő, ugyanis tekintsük a következő sztringeket:

$$c_1 u_1 \cdots u_n \alpha = c_2 v_1 \cdots v_m,$$

ahol c_1 és c_2 eleme a C^* halmaznak, $u_i, v_j \in C$ és $\text{length}(u_1 \cdots u_k) \neq \text{length}(v_1 \cdots v_l)$ bármely $k \leq n, l \leq m$ esetén, kivéve az egyetlen $\alpha = \lambda$, $k = n$, $l = m$ esetet. Ha elhagyjuk a sztringek $c_1 = c_2$ prefix

részeit, akkor valóban megkapjuk U helyes indexét. Formálisan ez a következőt jelenti az $u_1 \cdots u_n \alpha = v_1 \cdots v_m$ tekintetében:

$$i = \begin{cases} n + m & \text{ha } \alpha \neq \lambda \\ n + m - 1 & \text{ha } \alpha = \lambda \end{cases} \quad (3.1)$$

PÉLDA: Ha $C = \{010, 1101, 10, 11\}$, akkor a 3.2. táblázatot kapjuk az $\mathcal{A}_D(C)$ automata segítségével. Azok az $\mathcal{A}(C)$ automatában lévő állapotok kerülnek egy sorba, melyek együtt szerepelnek az $\mathcal{A}_D(C)$ automata valamely állapotában, s egyikük elfogadó állapot. Az indexet a 3.1. egyenlet alapján számoltam ki, a maradványokat pedig a 3.1. táblázatból kapjuk meg.

Állapot	Érintett kódszó sorozat	Index	Maradvány
1	11 010	3	10
A	1101 0		
2	1101	2	0
B	11 01		
4	11	1	01
D	11		
1	1101 010	4	λ
3	11 010 10		
3	11 010 10	4	λ
1	1101 010		

3.2. táblázat. Az U_i halmazok indexeinek meghatározása az $\mathcal{A}_D(\{010, 1101, 10, 11\})$ automatában

Tehát

$$U_1 = \{01\} \quad U_2 = \{0\} \quad U_3 = \{10\} \quad U_4 = \{\lambda\},$$

melyek azonosak a Sardinas–Patterson algoritmus szerinti U_i halmazokkal.

Meg kell említeni, hogy több indexet is kaphatunk az állapotpárookra.

3.2. A felbonthatóság egy új értelmezése

A kódolás során az elsődleges közlés betűihez rendeltünk különböző jelsorozatok, azaz betűszerinti kódolással foglalkoztam mélyebben, itt vizsgáltam a felbonthatóság kérdését. Természetesen lehetne vizsgálni ugyanezt a problémát nem betűszerinti kódolás esetén is, de láthattuk, hogy a felbonthatóság kérdése független a kódolás mikéntjétől, kizárólag a kódszavak halmazától függ. Alapvetően a kódolásnak, mint függvénynek az értelmezési tartományát vizsgáltam, (a Δ^+ egy véges részhalmazát), azaz az $f : \Sigma \rightarrow \Delta^+$ injektív leképezés volt az érdeklődés tárgya. Felbonthatónak neveztünk egy ilyen kódolást, amennyiben teljesül az

$$f(x_1) \dots f(x_n) = f(y_1) \dots f(y_m) \Rightarrow n = m \text{ és } f(x_i) = f(y_i) \text{ } (x_i = y_i)$$

összefüggés, ahol $x_i, y_i \in \Sigma$, $f(x_i), f(y_i) \in \Delta^+$.

A betűszerinti kódolás definíciója és a konkatenáció tulajdonsága miatt a leképezés homomorf, emiatt $g : \Sigma^+ \rightarrow \Delta^+$ alakban is írható. Ekkor a felbonthatóság definíciója az alábbi összefüggés teljesülését jelenti, azaz izomorfának kell teljesülni a leképezésre:

$$g(x_1 \dots x_n) = g(y_1 \dots y_m) \Rightarrow n = m \text{ és } \forall i \leq n : x_i = y_i$$

$$\text{ahol } x_i, y_i \in \Sigma, g(x_1 \dots x_n), g(y_1 \dots y_m) \in \Delta^+$$

3.2.1. Kvázikódok

Most a felbonthatóságot egy olyan leképezésen értelmezem, ahol egy $x \in \Sigma$ szimbólumhoz nem csak egy, hanem tetszőleges véges számú

$$\{p_1, \dots, p_m\} \in 2^{\Delta^*} \setminus \emptyset$$

jelsorozat tartozik, azaz $f(x) = \{p_1, \dots, p_m\}$.

21. Definíció. [Fal08b, Fal11a] A Δ feletti H kvázikód a $2^{\Delta^*} \setminus \emptyset$ egy véges részhalmaza.

22. Definíció. [Fal08b, Fal11a] A H kvázikód elemeit *kódhalmazoknak*, a H^+ elemeit pedig *üzeneteknek* nevezzük.

23. Definíció. [Fal08b, Fal11a] Egy $\bar{f} : \Sigma \rightarrow H$ bijektív függvényt a Σ halmazhoz tartozó H kvázikód megadásának nevezzük.

24. Definíció. [Fal08b, Fal11a] Tetszőleges $\bar{f} : \Sigma \rightarrow H$ kvázikód megadása esetén az $f(x_1 \cdots x_n) = \bar{f}(x_1) \cdots \bar{f}(x_n)$, $x_1, \dots, x_n \in \Sigma$ feltételnek eleget tevő $f : \Sigma^+ \rightarrow H^+$ függvényt kvázikódolásnak nevezzük.

A felbonthatóságot a szokásos módon értelmezem, tehát az $f : \Sigma^+ \rightarrow H^+$ kvázikódolás felbontható, ha az

$$f(x_1) \cdots f(x_n) = f(y_1) \cdots f(y_m)$$

egyenlőségből következik, hogy

$$n = m \text{ és } f(x_i) = f(y_i) \text{ (} x_i = y_i \text{)}$$

25. Definíció. [Fal08b, Fal11a] Egy H kvázikódot *UD (uniquely decipherable)* kvázikódnak, azaz felbonthatónak nevezzük, ha minden egyes üzenetnek egyértelmű felbontása van, tehát ha teljesül az

$$f(x_1) \cdots f(x_n) = f(y_1) \cdots f(y_m),$$

egyenlőség, ahol $f(x_1), \dots, f(x_n), f(y_1), \dots, f(y_m) \in H$, akkor $n = m$ és $f(x_1) = f(y_1), \dots, f(x_n) = f(y_n)$.

3.2.2. Jelölés rendszer

A következőkben használatos definíciók és jelölések megtalálhatók a [HU79],[BP85], és [RS97] művekben, a fontosabbakat az alábbiakban részletezem. A Σ feletti *szó* a Σ véges nem üres halmaz elemeinek egy véges sorozata. Ha nem lehet félre érteni, néha elhagyom a " Σ feletti" kifejezést. A Σ halmazt *ábécének* nevezzük, Σ elemeit *betűknek*. A Σ halmaz *triviális*, ha csupán egy eleme van, különben azt mondjuk, hogy a Σ halmaz *nemtriviális*. Az egyetlen betűt sem tartalmazó szót *üres szónak* nevezzük és a λ szimbólummal jelöljük. A w szó $|w|$ *hossza* a w szó betűinek száma, ahol minden betű annyiszor van megszámlolva, ahányszor szerepel. Így $|\lambda| = 0$. Ha az $u = x_1 \cdots x_k$ és a $v = x_{k+1} \cdots x_\ell$ a Σ ábécé feletti szavak ($x_1, \dots, x_k, x_{k+1}, \dots, x_\ell \in \Sigma$) akkor a *szorzatuk*, vagy másnéven *konkatenáltjuk* az $uv = x_1 \cdots x_k x_{k+1} \cdots x_\ell$ szó, mely szintén Σ feletti. Továbbá minden Σ feletti $u = x_1 \cdots x_k$ szó esetén $u\lambda = \lambda u = u$ ($= x_1 \cdots x_k$), valamint $\lambda\lambda = \lambda$. Nyilván minden $u, v \in \Sigma^*$ szóra igaz, hogy $|uv| = |u| + |v|$. Ezek után világos, hogy minden u, v, w Σ feletti szó esetén $u(vw) = (uv)w$. Más szavakkal az $uu' = w'w$ egyenlőség mindig teljesül, ha $u' = vw$ és $w' = uv$. Tehát a konkatenáció asszociatív művelet, az egység elem szerepét pedig a λ látja el. Kiterjesztjük ezt a szavakon értelmezett műveletet halmazokra: Legyenek az U és a V szavakból álló halmazok. Ekkor a két halmaz *konkatenációja* (vagy *szorzata*) az $UV = \{uv : u \in U, v \in V\}$ halmaz. Az u szó *faktora* (vagy *rész szava*) egy v szónak, ha $v = v_1 u v_2$ igaz valamely v_1 és v_2 szavak esetén. Ha a v_1 szó üres szó, azt mondjuk, hogy az u szó a v szó *prefix* része, és ha v_2 üres, akkor azt mondjuk, hogy az u szó *suffix* része a v szónak. Mivel a konkatenáció asszociatív művelet, tetszőleges $X_1, \dots, X_n \subseteq \Sigma^*$ halmazok esetén az $X = X_1 \cdots X_n$

halmaz egyértelműen definiált. Azt mondjuk, hogy az $X_1 \cdots X_n$ az X halmaz egy lehetséges *dekompozíciója*. Ha az u szó a v szó rész szava (prefix, suffix) olyan, hogy $u \neq v$ akkor *valódi rész szóról* (*valódi prefix, valódi suffix*) beszélünk. Ha a nem üres $X \subseteq \Sigma^*$ halmaz zárt az elemei faktoraira nézve, akkor az X halmazt *faktoriál* halmaznak nevezzük. Definíció szerint az üres halmaz nem faktoriál halmaz, és minden faktoriál halmaz tartalmazza legalább a λ üres szót. Hasonlóan, ha egy nem üres $X \subseteq \Sigma^*$ halmaz tartalmazza az elemeinek az összes prefixumát (azaz zárt a prefixumokra), azt mondjuk, hogy *prefixiál* halmaz. Analóg módon, ha zárt a suffixumokra, azt mondjuk, hogy *suffixiál* halmaz. Nyilván minden faktoriál halmaz prefixiál és suffixiál halmaz is egyben.

Használni fogom a következő két állítást:

14. Tétel. Véges sok prefixiál (suffixiál) halmaz minden konkatenáltja szintén prefixiál (suffixiál) halmaz.

BIZONYÍTÁS: Nyilván elegendő belátni két prefixiál (suffixiál) halmaz konkateáltjára az állítást. Legyen adva két $X_1, X_2 \subseteq \Sigma^*$ nem üres prefixiál (suffixiál) halmaz és legyen $u \in X_1$ és $v \in X_2$. Elegendő bizonyítani, hogy uv minden prefixuma (suffixuma) benne van az $X_1 X_2$ halmazban. Legyen r az uv prefix (suffix) része és $|u| \leq |r|$ ($|v| \leq |r|$). Ekkor létezik egy $r = ur'$ ($r = r'v$) dekompozíció, ahol r' egy prefix része a v szónak (suffix része az u szónak). De X_1 prefixiál (suffixiál) halmaz. Emiatt $r' \in X_2$ ($r' \in X_1$). Így $r \in X_1 X_2$. Tegyük fel, hogy r prefix (suffix) része az uv szónak és $|u| > |r|$ ($|v| > |r|$). Akkor r prefix része az u -nak (suffix része a v -nek) és így $r \in X_1$ ($r \in X_2$). Mivel λ minden szónak prefix (suffix) része és $X_2 \neq \emptyset$ ($X_1 \neq \emptyset$), emiatt $\lambda \in X_2$ ($\lambda \in X_1$). Így $r \in X_1 X_2$ ismételten. $\square$

3. Megjegyzés. [AF05] Analóg állítás igaz a faktoriál halmazokra: Véges sok faktoriál halmaz minden konkatenáltja szintén faktoriál halmaz.

15. Tétel. Legyen adott egy véges, nem üres $X \subset \Sigma^*$ halmaz. Bármely $X_1, X_2 \subseteq \Sigma^*$ nem üres suffixiál halmaz esetén ha $XX_1 = XX_2$, akkor $X_1 = X_2$.

BIZONYÍTÁS: Tegyük fel az állításunkkal ellentétben, hogy léteznek az $X \subset \Sigma^*$ véges nem üres halmaz, valamint az $X_1, X_2 \subseteq \Sigma^*$ nem üres suffixiál halmazok úgy, hogy $XX_1 = XX_2$ és $X_1 \neq X_2$. Ebben az esetben az X_1 és az X_2 halmazok közül valamelynek lenni kell egy elemének, amely nincs benne a másikban. Mondjuk $r \in X_2$, de $r \notin X_1$.

Az X halmaz végessége miatt létezik egy $u \in X$ szó, amely nem valódi prefix része egyetlen egy X halmazbeli szónak sem. (Például u ilyen tulajdonságú, ha az egyik leghosszabb szó az X halmazban.) $XX_1 = XX_2$ és $ur \in XX_2$ miatt léteznek az $u_1 \in X, r_1 \in X_1$ szavak és $u_1r_1 = ur$.

A kiindulási feltételek miatt u nem valódi prefixuma u_1 -nek. Így $u_1r_1 = ur$ miatt r suffix része az r_1 szónak. Emlékezzünk vissza, hogy X_1 suffixiál halmaz. Emiatt, ellentétben a feltevésünkkel $r_1 \in X_1$ miatt $r \in X_1$, mivel r suffix része r_1 -nek. Ezzel teljes a bizonyítás. $\square$

4. Megjegyzés. A fenti állítást nem terjeszthetjük ki tetszőleges végtelen $X \subseteq \Sigma^*$ halmazokra. Vegyük például az $X_1 = \{x_1\}, X_2 = \{x_2\}, x_1, x_2 \in \Sigma, x_1 \neq x_2$ halmazokat és legyen $X \subseteq \Sigma^*$ egy tetszőleges végtelen halmaz a következő tulajdonsággal:

$$\forall r \in X : rx_1, rx_2 \in X.$$

Ekkor $XX_1 = XX_2$, de $X_1 \neq X_2$.

Egy $X \subseteq \Sigma^*$ faktoriál halmazt *indekompozícionálhatónak* nevezünk, ha $X = AB$ maga után vonja, hogy $X = A$ vagy $X = B$ minden $A, B \subseteq \Sigma^*$ faktoriál halmaz esetén; egyébként azt mondjuk, hogy X *dekompozícionálható*. Egy adott $X \subseteq \Sigma^*$ halmaz és $X_1, \dots, X_n \subseteq \Sigma^*$ indekompozícionálható faktoriál halmazok esetén azt

mondjuk, hogy az $X = X_1 \cdots X_n$ az X egy *kanonikus dekompozíciója*, ha teljesül a következő két eset közül valamelyik::

- $X = X_1 = \{\lambda\}, n = 1$;
- $X \neq \{\lambda\}, X_i \neq \{\lambda\}, i \in \{1, \dots, n\}$, továbbá, $X \neq X_1 \cdots X_{i-1} X'_i X_{i+1} \cdots X_n$ bármely $i \in \{1, \dots, n\}$ esetén és egy faktoriál halmazra igaz, hogy $X'_i \subsetneq X_i$.

16. Tétel. [AF05] Bármely X faktoriál halmaznak létezik faktoriál halmazokból álló egyértelmű kanonikus dekompozíciója.

3.2.3. Kvázikódok felbonthatósági kritériumai

A következő definícióban a betűszerinti kódolásnál ismertetett prefixmentes tulajdonsággal analóg tulajdonságot adok meg a kvázikódokra.

26. Definíció. [Fal11a] Egy A halmazt λ -prefixmentesnek nevezünk egy B halmazra nézve, ha $\exists a \in A$ ($a \neq \lambda$) úgy, hogy $\forall \alpha \in \Sigma^*$ és $\forall b \in B$ ($b \neq \lambda$) esetén $a\alpha \neq b$ és $b\alpha \neq a$, azaz $a \in A$ nem kezdőszelete egyetlen $b \in B$ -nek sem és egyetlen $b \in B$ sem kezdőszelete a -nak.

PÉLDA: Legyen $A = \{baa, ab, b\}$ $B = \{baa, bba\}$, ekkor A λ -prefixmentes B -re nézve, mert $ab \in A$ nem kezdőszelete B egyetlen elemének sem és B egyetlen eleme sem kezdőszelete ab -nek.

PÉLDA: Legyen $A = \{baa, ab, b\}$ $B = \{baa, a, bba, aa\}$, ekkor A nem λ -prefixmentes B -re nézve, mert $ab \in A$ nem kezdőszelete ugyan B egyetlen elemének sem, de $a \in B$ kezdőszelete $ab \in A$ -nak.

17. Tétel. [Fal11a] A λ -prefixmentesség egy halmazra nézve, mint reláció

- irreflexív
- nem szimmetrikus

- nem tranzitív

BIZONYÍTÁS:

- *Irreflexív:* A definícióból következik, ugyanis egy elem mindig prefix része önmagának, így egyetlen halmaz sem λ -prefixmentes önmagára nézve
- *Nem szimmetrikus:* Például az $A = \{ab, b\}$ és a $B = \{a, bba, aa\}$ halmazok esetében az A halmaz nem λ -prefixmentes a B halmazra nézve (ab -nek kezdőszelete a és b kezdőszelete bba -nak), viszont a B halmaz λ -prefixmentes az A halmazra nézve ($aa \in B$ miatt). Megemlítjük, hogy lehetnek olyan A és B speciális halmazok, amikor az A halmaz λ -prefixmentes a B halmazra és viszont. Például az $A = \{a\}, B = \{b\}$ esetben az A halmaz λ -prefixmentes a B halmazra nézve és a B halmaz is λ -prefixmentes az A halmazra nézve.
- *Nem tranzitív:* Például az $A = \{a\}; B = \{b\}; C = \{aa\}$ halmazok esetében látható, hogy az A halmaz λ -prefixmentes a B halmazra nézve, valamint a B halmaz is λ -prefixmentes a C halmazra nézve, viszont az A halmaz nem lesz λ -prefixmentes a C halmazra nézve.

Megemlítjük, hogy lehetnek olyan A, B és C speciális halmazok, amikor az A halmaz λ -prefixmentes a B halmazra nézve, a B halmaz λ -prefixmentes a C halmazra nézve és teljesül az is, hogy az A halmaz λ -prefixmentes a C halmazra nézve. Például tekintve az $A = \{ab, b\}; B = \{bba\}; C = \{a\}$ halmazokat, teljesül, hogy az A halmaz λ -prefixmentes a B halmazra nézve, a B halmaz λ -prefixmentes a C halmazra nézve és az A halmaz λ -prefixmentes a C halmazra nézve.

□

27. Definíció. [Fal11a] Két A, B halmazt (egymásra nézve) λ -prefixmentesnek nevezünk, ha vagy A λ -prefixmentes B -re nézve, vagy B λ -prefixmentes A -ra nézve. A *vagy* nem kizáró vagyként értendő.

PÉLDA: Az alábbi halmazok egymásra nézve λ -prefixmentesek:

1. Az $A = \{ab\}$ halmaz és a $B = \{a, aa\}$ halmazok egymásra nézve λ -prefixmentesek, ugyanis bár az A halmaz nem λ -prefixmentes a B halmazra nézve, de a B halmaz λ -prefixmentes az A halmazra nézve (az aa elem miatt).
2. Az $A = \{a, aa\}$ halmaz és a $B = \{ab\}$ halmazok egymásra nézve λ -prefixmentesek, ugyanis bár a B halmaz nem λ -prefixmentes az A halmazra nézve, de az A halmaz λ -prefixmentes a B halmazra nézve (az aa elem miatt).
3. Az $A = \{a\}$ halmaz és a $B = \{b\}$ halmazok egymásra nézve λ -prefixmentesek, ugyanis az A halmaz λ -prefixmentes a B halmazra nézve, valamint a B halmaz is λ -prefixmentes az A halmazra nézve.

18. Tétel. [Fal11a] Az egymásra nézve való λ -prefixmentesség, mint reláció

- irreflexív
- szimmetrikus
- nem tranzitív

BIZONYÍTÁS:

- *Irreflexív:* Mivel egy halmaz önmagára nézve nem λ -prefixmentes, a definíció miatt ugyanazon halmazok egymásra nézve nem λ -prefixmentesek.

- *Szimmetrikus:* A definícióban szereplő *vagy* kapcsolat miatt szimmetrikus a reláció.
- *Nem tranzitív:* Például az $A = \{a\}; B = \{b\}; C = \{a\}$ halmazok esetében látható, hogy az A és B halmaz λ -prefixmentes egymásra nézve, ahogy a B és a C halmaz is, Viszont az A és a C halmazok nem lesznek λ -prefixmentesek egymásra nézve. Megemlítjük, hogy lehetnek olyan A, B és C speciális halmazok, amikor az A és a B halmaz λ -prefixmentesek egymásra nézve, ahogy a B és a C valamint az A és a C is λ -prefixmentesek egymásra nézve. Ilyen speciális halmazok például az $A = \{a\}; B = \{b\}; C = \{c\}$ halmazok.

□

28. Definíció. [Fal11a] Egy halmazokból álló halmazt λ -prefixmentesnek nevezünk, ha bármely két eleme egymásra nézve λ -prefixmentes.

PÉLDA: Például a $H = \{\{a\}, \{b\}, \{c\}\}$ kvázikód kódhalmazai páronként egymásra nézve λ -prefixmentesek, így a H kvázikód λ -prefixmentes. A $G = \{\{a\}, \{b\}, \{ac\}\}$ kvázikód esetében az a prefix része ac -nek, ezért létezik két kódhalmaz, melyek egymásra nézve nem λ -prefixmentesek. Emiatt a G kvázikód nem lesz λ -prefixmentes.

19. Tétel. [Fal11a] Ha egy nem üres, a $\{\lambda\}$ halmazt nem tartalmazó suffixiál halmazokból álló kvázikód λ -prefixmentes, akkor a kvázikód felbontható.

BIZONYÍTÁS: A bizonyítás indirekt úton történik. Tegyük fel, hogy a kvázikód suffixiál halmazokból áll és λ -prefixmentes, de nem felbontható. Mivel nem felbontható, létezik egy olyan G halmaz,

mely legalább kétféleképpen előáll a kódból. Jelöljük H_i -vel az $f(x_i)$ halmazt. Tekintsük ekkor a következő két felbontást:

$$G = H_{i_1} \dots H_{i_s} \text{ és } G = H_{j_1} \dots H_{j_t}$$

Ekkor létezik egy pozitív egész l szám az indirekt feltevés miatt, hogy $\forall k < l$ esetén $H_{i_k} = H_{j_k}$, de $H_{i_l} \neq H_{j_l}$. Ha $l = 1$, akkor $H_{i_1} \neq H_{j_1}$ és $H_{i_1} \dots H_{i_s} = H_{j_1} \dots H_{j_t}$. Másrészt, használva a 14. tételt, minden $H_{i_1} \dots H_{i_{l-1}}$, $H_{j_1} \dots H_{j_{l-1}}$, $H_{i_l} \dots H_{i_s}$, $H_{j_l} \dots H_{j_t}$ halmaz suffixiál. Ezért, alkalmazva a 15. tételt, a

$$H_{i_1} \dots H_{i_{l-1}} H_{i_l} \dots H_{i_s} = H_{j_1} \dots H_{j_{l-1}} H_{j_l} \dots H_{j_t}$$

és a

$$H_{i_1} \dots H_{i_{l-1}} = H_{j_1} \dots H_{j_{l-1}}$$

egyenletekből azt kapjuk, hogy

$$G' = H_{i_l} \dots H_{i_s} = H_{j_l} \dots H_{j_t},$$

és az l fenti tulajdonsága miatt $H_{i_l} \neq H_{j_l}$. Így, $\forall p \in G'$ a $p = x\beta = y\gamma$ alakban is írható, ahol $x \in H_{i_l}$, $y \in H_{j_l}$. Azaz, $x\alpha = y$, vagy $x = y\alpha$, ahol $\alpha \in \Delta^*$.

Könnyű belátni, hogy létezik $p \in G'$ olyan, hogy $p = x\beta = y\gamma$ és $x \neq \lambda$, $y \neq \lambda$ a halmazok konkatenációs tulajdonságai és amiatt, hogy a kvázikód nem tartalmazza a $\{\lambda\}$ halmazt. Emiatt létezik $\alpha \in \Delta^*$ valamely $x \in H_{i_l}$ esetén úgy, hogy $x\alpha = y$, vagy $x = y\alpha$ igaz valamely $y \in H_{j_l}$ -re és $x \neq \lambda$, $y \neq \lambda$.

Tehát H_{i_l} nem λ -prefixmentes H_{j_l} -re nézve (analóg módon azt kapjuk, hogy H_{j_l} sem λ -prefixmentes H_{i_l} -re nézve). Tehát a H_{i_l} és a H_{j_l} halmazok nem λ -prefixmentesek egymásra nézve, ezért a kvázikód nem λ -prefixmentes. Ellentmondásra jutottunk, s így a tétel bizonyítást nyert.. $\square$

5. Megjegyzés. A fejezet eredményei analóg módon megfogalmazhatók prefixiál, illetve λ -suffixmentes kvázikódokra.

1. Probléma. Természetes módon felmerül a kérdés, hogy a fenti tétel kiterjeszthető-e, másként fogalmazva, létezik-e olyan nem üres, a $\{\lambda\}$ halmazt nem tartalmazó halmazokból álló kvázikód, mely λ -prefixmentes, de nem felbontható, illetve létezik-e olyan nem üres, a $\{\lambda\}$ halmazt nem tartalmazó suffixiál halmazokból álló kvázikód, mely nem felbontható.

Salomaa és Mateescu [MSY98]-beli konstrukcióját felhasználva az első kérdésre az alábbi megoldását kapjuk:

20. Tétel. Létezik λ -prefixmentes kvázikód, mely nem felbontható.

BIZONYÍTÁS: A bizonyításban felhasználjuk Salomaa és Mateescu [MSY98] következő konstrukcióját:

$$L = \{b, ba, baa, c, caa, caaa, caaaa\}$$

$$L_1 = \{ab, aaab, aaaab, c\}, L_2 = \{ab, aaaab, c\}$$

Egyszerű ellenőrzéssel igazolható, hogy a fenti halmazok λ -prefixmentes kvázikódot alkotnak. Másrészt, amint Salomaa és Mateescu is kimutatták [MSY98], a fenti konstrukciókat alkalmazva $LL_1 = LL_2$ és $L_1 \neq L_2$. $\square$

A probléma második kérdését válaszolja meg a következő eredményem:

21. Tétel. Létezik olyan nem üres, a $\{\lambda\}$ halmazt nem tartalmazó suffixiál halmazokból álló kvázikód, mely nem felbontható.

BIZONYÍTÁS: Tekintsük a következő konstrukciót:

$$L_1 = \{\lambda, a\}, L_2 = \{\lambda, a, aa\}$$

Egyszerű ellenőrzéssel igazolható, hogy a fenti halmazok suffixiál (és egyben prefixiál) halmazok, másrészt $L_1L_2 = L_2L_1$. $\square$

22. Tétel. [Fal11a] Létezik nem λ -prefixmentes felbontható kvázi kód.

BIZONYÍTÁS: A bizonyításhoz megadunk egy, a tételben szereplő tulajdonságokkal bíró egyszerű kvázikódot. Legyen adva az $f(1) = \{a\}, f(2) = \{ab\}$ kvázi kódolás. Tegyük fel, hogy ez a kvázi kód nem felbontható. Akkor a 19. tételhez tartozó bizonyítás jelölésrendszerét és eredményeit felhasználva létezni kell a $G' = H_{i_1} \dots H_{i_s} = H_{j_1} \dots H_{j_t}$ és $H_{i_l} \neq H_{j_l}$ halmaznak. Ekkor viszont $\forall p \in G'$ felírható $p = x\beta = y\gamma$ alakban, ahol $x \in H_{i_l}, y \in H_{j_l}$, tehát vagy $x\alpha = y$, vagy $x = y\alpha$, ahol $\alpha \in \Delta^*$. Jelen esetben ez azt jelenti, hogy $p = f(1)\beta = f(2)\gamma$, azaz $p = a\beta = ab\gamma$. A konstrukcióból kitűnik, hogy mind β , mind γ előállítható a kvázi kód elemeinek konkatenációjaként, viszont a $p = a\beta = ab\gamma$ egyenlőségből adódik, hogy β -nak b -vel kell kezdődnie, ami tekintve a kvázi kódot, nem teljesülhet. Emiatt ellentmondásra jutottunk, tehát a $H = \{\{a\}, \{ab\}\}$ felbontható. Vizsgáljuk meg λ -prefixmentesség szempontjából a $H_1 = \{a\}, H_2 = \{ab\}$ halmazokat. H_1 nem lesz λ -prefixmentes H_2 -re nézve, hiszen $a \in H_1$ prefix része az $ab \in H_2$ -nek, és ugyanezen ok miatt H_2 sem lesz λ -prefixmentes H_1 -re nézve. Így H_1 és H_2 egymásra nézve sem lesznek λ -prefixmentesek, emiatt H nem lesz λ -prefixmentes. Tehát a $H = \{\{a\}, \{ab\}\}$ kvázi kód felbontható, de nem λ -prefixmentes. $\square$

23. Tétel. [Fal11a] Ha A és A^k eleme a H kvázikódnak, akkor a H kvázikód nem lesz felbontható.

BIZONYÍTÁS: Legyen a kvázikódolásban $f(x) = A; f(y) = A^k$, ekkor $\underbrace{f(x) \dots f(x)}_k = \underbrace{A \dots A}_k = A^k$ és $f(y) = A^k$, tehát

$$\exists n \neq m : f(x_{i_1}) \dots f(x_{i_n}) = f(x_{j_1}) \dots f(x_{j_m}),$$

tehát a kvázikód nem lesz felbontható. $\square$

A tétel általánosabban is kimondható:

24. Tétel. [Fal11a] Ha $\exists A = \prod_{i=1}^m A_i^{k_i} \in H$ ($m > 1, k_i \geq 1, A_1, A_2, \dots, A_m \in H$), akkor a kvázikód nem felbontható.

BIZONYÍTÁS: Legyen a kvázikódolásban $f(x) = A = \prod_{i=1}^m A_i^{k_i}; f(y_1) = A_1, \dots, f(y_m) = A_m$, ekkor $f(x) = A = \prod_{i=1}^m A_i^{k_i}$ és $\underbrace{f(y_1) \dots f(y_1)}_{k_1} \dots \dots \underbrace{f(y_m) \dots f(y_m)}_{k_m} = \prod_{i=1}^m A_i^{k_i}$, tehát

$$\exists n \neq m : f(x_{i_1}) \dots f(x_{i_n}) = f(x_{j_1}) \dots f(x_{j_m}),$$

tehát a kvázikód nem lesz felbontható. $\square$

3.2.4. A Sardinas–Patterson algoritmus kvázikódokra való alkalmazhatósága

A kódoknál a felbonthatóság kérdésére a Sardinas–Patterson algoritmus adott választ, adódott az ötlet, hogy megpróbáljam kvázikódokra is alkalmazni. Az algoritmus a következő hatványhalmaz rendszer létrehozását jelenti:

Legyen X és Y két részhalmaza az $2^{\Delta^+} \setminus \emptyset$ halmaznak, valamint legyen $A \in X$ és $B \in Y$. Jelölje $X^{-1}Y$ a következő halmazokból álló halmazt: a C halmaz egy eleme a $X^{-1}Y$ halmaznak, ha $AC = B$.

Legyen a H halmaz a $2^{\Delta^+} \setminus \emptyset$ halmaz részhalmaza, és legyen

$$\begin{aligned}
U_1 &= H^{-1}H \setminus \{\lambda\} \\
U_2 &= H^{-1}U_1 \cup U_1^{-1}H \\
&\vdots \\
U_{n+1} &= H^{-1}U_n \cup U_n^{-1}H
\end{aligned} \tag{3.2}$$

A fenti jelöléseket alkalmazva a sejtésem a következő volt a Sardinas–Patterson tétel [SP53] alapján:

Amennyiben $\exists i$ úgy, hogy $\{\lambda\} \in U_i$, akkor H nem lesz felbontható. Sajnos ez az állítás nem igaz, ugyanis a halmazok kissé másképpen viselkednek a konkatenáció műveletére nézve, mint a jelsorozatok. A jelsorozatoknál ugyanis egyértelmű a következő összefüggés:

Legyen $x, y, z \in \Sigma^* \setminus \{\lambda\}$, ekkor $xy = xz$ csak úgy lehetséges, ha $y = z$. A Sardinas–Patterson algoritmus erre az egyértelműsége épít, maradékokat keres, a halmaz sorozat megalkotásában is ezt használja ki. Minden halmaznak van két triviális felbontása, azaz $X = AB$ teljesül bármely X halmazra, ha

1. $A = \{\lambda\}, B = X$
2. $A = X, B = \{\lambda\}$

29. Definíció. [AF05] A csak triviális felbontással rendelkező halmazokat *prímhalmazoknak* nevezzük.

Analóg módon azt várnánk, hogy ha A, B, C prímhalmazok, akkor $AB = AC$ egyenlőségből következik, hogy $B = C$. Ez nem így van, például [MSY98]-ban L, L_1, L_2 prímhalmazok, mégis $LL_1 = LL_2$. Nevezetesen, a 20. tétel bizonyításában felhasznált Salomaa és Mateescu által megalkotott konstrukciót alkalmazva legyen

$$L = \{b, ba, baa, c, caa, caaa, caaaa\}$$

$$L_1 = \{ab, aaab, aaaab, c\}, L_2 = \{ab, aaaab, c\}$$

Ez számunkra azt jelenti, hogy ha vesszük ezekből a halmazokból álló kvázikódot, azaz $H = \{L, L_1, L_2\}$ és alkalmazzuk a kvázikódokra módosított Sardinias–Patterson algoritmust, akkor a halmazok prímtulajdonságai miatt már az első lépésben $U_1 = \emptyset$, s természetesen itt véget is ér az algoritmus anélkül, hogy $\{\lambda\} \in U_i$ valamely i -re teljesülne. Az algoritmus szerint tehát a kód felbontható, de láthattuk, hogy nem az, ugyanis legyen adva a következő kvázikódolás:

$$f(x_1) = L, f(x_2) = L_1, f(x_3) = L_2.$$

Ekkor $f(x_1)f(x_2) = f(x_1)f(x_3)$, azaz a kód nemfelbontható, hiszen $f(x_2) \neq f(x_3)$.

3.2.5. A kvázikódok dekompozíciója

A dekompozíció meghatározásához a halmazok prímtulajdonságát eldöntő algoritmust [MSY98] fogom használni:

Legyen R egy reguláris nyelv a Σ ábécé felett, és legyen $\mathcal{A} = (Q, q_0, Q_F, \Sigma, \delta)$ az R -hez tartozó minimálautomata, mely természetesen determinisztikus véges automata. (Q a belső állapotok halmaza, q_0 a kezdőállapot, Q_F az elfogadó állapotok halmaza és δ az állapotátmeneteket leíró mozgási szabályok halmaza.) Kiterjesztjük a δ -t a Σ feletti szavakra, azaz $\delta(q, w) = q'$ azt jelenti, hogy a w szó az $\mathcal{A}$ automatát a q állapotból a q' állapotba viszi. Két nyelvet definiálunk egy $P \subseteq Q$ halmaz esetén a következőképpen:

$$R_1^P = \{w \mid \delta(q_0, w) \in P\},$$

$$R_2^P = \{w \mid \delta(p, w) \in Q_F, p \in P\}$$

25. Tétel. [MSY98] Legyen R és $\mathcal{A}$ a fentiek szerint definiálva. Tegyük fel, hogy $R = L_1 L_2$, ahol L_1 és L_2 tetszőleges nyelvek. Definiáljuk a $P \subseteq Q$ halmazt a következő módon:

$$P = \{p \in Q \mid \delta(q_0, w) = p, \text{ valamely } w \in L_1 \text{ esetén}\}.$$

Ekkor $R = R_1^P R_2^P$, továbbá $L_1 \subseteq R_1^P$ és $L_2 \subseteq R_2^P$.

Definíció szerint egy Q halmaz nem üres P részhalmaza dekompozíciós halmaz az R reguláris nyelvre nézve, ha $R = R_1^P R_2^P$. Az $R = R_1^P R_2^P$ dekompozíciót az R halmaz P dekompozíciós halmaz által indukált dekompozíciójának is nevezik. Azt mondjuk, hogy egy L halmaz $L = L_1 L_2$ dekompozícióját tartalmazza az $L = L'_1 L'_2$, ha $L_i \subseteq L'_i, i = 1, 2$.

26. Tétel. [MSY98] Egy R reguláris nyelv minden dekompozíciója benne van R valamely dekompozíciós halmaza által indukált dekompozíciójában. Reguláris nyelvek tekintetében eldöntött a prím tulajdonság problémája.

Felhasználva a fenti eredményeket, megkonstruálom a következő automatát:

Legyen $\mathcal{A} = (Q, q_0, Q_F, \Delta, \delta)$ a $H = \{f(x_1), \dots, f(x_n)\}$ kvázikódból származó H^* valamely véges X részhalmazához tartozó minimálautomata. (Q a belső állapotok halmaza, q_0 a kezdőállapot, Q_F az elfogadó állapotok halmaza és δ az állapotátmeneteket leíró mozgási szabályok halmaza.) Hasonló módon kiterjesztjük a δ -t, azaz $\delta(x, w) = y$ azt jelenti, hogy a w szó az $\mathcal{A}$ automatát az x állapotból az y állapotba viszi. Nyelveket definiálunk $P_1, P_2 \subseteq Q$ halmazok esetén a következőképpen:

$$R_{P_1, P_2} = \{w | \delta(p, w) \in P_2, p \in P_1\}$$

27. Tétel. [Fal11a] Legyen H és $\mathcal{A}$ a fentiek szerint definiálva. Tegyük fel, hogy $X = f(x_{i_1}) \cdots f(x_{i_k})$, ahol $f(x_{i_l}) \in H$. Definiáljuk a $P_0, \dots, P_k \subseteq Q$ halmazokat a következő módon:

$$P_0 = \{q_0\}$$

$$P_1 = \{p \in Q | \delta(q_0, w) = p, \text{ valamely } w \in f(x_{i_1}) \text{ esetén}\}.$$

(Ekkor nyilván $f(x_{i_1}) \subseteq R_{q_0, P_1}$.)

$$P_2 = \{p \in Q \mid \delta(q_0, w) = p, \text{ valamely } w \in f(x_{i_1})f(x_{i_2}) \text{ esetén}\}.$$

$\vdots$

$$P_k = Q_F = \{p \in Q \mid \delta(q_0, w) = p, \text{ valamely } w \in f(x_{i_1}) \cdots f(x_{i_k})\text{-ra}\}.$$

Ekkor $X = R_{P_0, P_1} \cdots R_{P_{k-1}, P_k}$, és $f(x_{i_1}) \subseteq R_{P_0, P_1}, \dots, f(x_{i_k}) \subseteq R_{P_{k-1}, P_k}$.

BIZONYÍTÁS: Először a részhalmaz relációkat látjuk be. Az $f(x_{i_m}) \subseteq R_{P_{m-1}, P_m}$ bebizonyítását indirekt módon végezzük, azaz tegyük fel, hogy valamely $w_{i_m} \in f(x_{i_m})$ és $p \in P_{m-1}$ esetén $\delta(p, w_{i_m}) \notin P_m$. Válasszuk a $w \in f(x_{i_1}) \cdots f(x_{i_{m-1}})$ szót úgy, hogy $\delta(q_0, w) = p$. Mivel $ww_{i_m} \in f(x_{i_1}) \cdots f(x_{i_m})$, adódik, hogy $\delta(q_0, ww_{i_m}) \in P_m$. Ekkor viszont $\delta(q_0, ww_{i_m}) = \delta(p, w_{i_m})$, ami az indirekt feltevésünk szerint nem eleme a P_m halmaznak. Az ellentmondás miatt tehát igaz az $f(x_{i_m}) \subseteq R_{P_{m-1}, P_m}$ tartalmazás.

A következőkben belátjuk az $X = R_{P_0, P_1} \cdots R_{P_{k-1}, P_k}$ összefüggést. Legyen a $w_1 \cdots w_k$ tetszőleges szó úgy kiválasztva, hogy $w_m \in R_{P_{m-1}, P_m}$ teljesül minden $1 \leq m \leq k$ esetén. Mivel $w_1 \in R_{P_0, P_1}$, teljesül, hogy $\delta(q_0, w_1) = p_1$. P_1 definíciója miatt adódik, hogy $p_1 \in P_1$. R_{P_1, P_2} definíciója miatt pedig igaz, hogy $\delta(p_1, w_2) = p_2 \in P_2$. R_{P_{m-1}, P_m} definíciója miatt hasonlóan kapjuk minden $1 \leq m \leq k$ esetén, hogy $\delta(p_{m-1}, w_m) = p_m \in P_m$. Így $\delta(q_0, w_1 \cdots w_k) = \delta(p_1, w_2 \cdots w_k) = \cdots = \delta(p_{k-1}, w_k) = p_k \in Q_F$, és emiatt $w_1 \cdots w_k \in X$. Ezért $X \supseteq R_{P_0, P_1} \cdots R_{P_{k-1}, P_k}$, tehát az egyik irányú tartalmazást sikerült belátni.

Vizsgáljuk meg a másik irányú tartalmazás is. Legyen w egy tetszőleges szó az X halmazban. Ekkor X definíciója miatt a w szó $w = w_1 \cdots w_k$ alakban is írható, ahol $w_m \in f(x_{i_m}), 1 \leq m \leq k$. A már bebizonyított $f(x_{i_m}) \subseteq R_{P_{m-1}, P_m}$ tartalmazás miatt kijelenthető, hogy $w_m \in R_{P_{m-1}, P_m}$, ahol $1 \leq m \leq k$. Emiatt a w szóra teljesül, hogy

$w = w_1 \cdots w_k \in R_{P_0, P_1} \cdots R_{P_{k-1}, P_k}$. Ezért $X \subseteq R_{P_0, P_1} \cdots R_{P_{k-1}, P_k}$, amit belátni szerettünk volna.

Figyelembe véve a kölcsönös részhalmaz relációk teljesülését, kapjuk, hogy $X = R_{P_0, P_1} \cdots R_{P_{k-1}, P_k}$. $\square$

30. Definíció. [Fal11a] A $\{P_1, \dots, P_{k-1}\}$ ($P_m \subseteq Q$) nem üres részhalmazokból álló halmazt dekompozíciós halmaznak mondjuk az X halmazra és a $H = \{f(x_1), \dots, f(x_n)\}$ kvázikódra nézve, ha $X = R_{q_0, P_1} \cdots R_{P_{k-1}, Q_F}$. Az $X = R_{q_0, P_1} \cdots R_{P_{k-1}, Q_F}$ dekompozíciót az X halmaz $\{P_1, \dots, P_{k-1}\}$ dekompozíciós halmaz által indukált dekompozíciójának nevezzük. Azt mondjuk, hogy a véges X halmaz $X = X_1 \cdots X_k$ dekompozíciója benne van az $X = X'_1 \cdots X'_k$ if $X_m \subseteq X'_m$, $m = 1, 2, \dots, k$ dekompozícióban.

28. Tétel. [Fal11a] Egy véges X halmaz minden dekompozíciója benne van X valamely dekompozíciós halmaza által indukált dekompozíciójában. A felbonthatóság problémája eldönthető véges halmazok esetén.

BIZONYÍTÁS: A 28. tétel első része a 27. tétel következménye. Egy véges X halmaz összes lehetséges dekompozíciójának meghatározásához végig kell ellenőriznünk az összes $\{P_1, \dots, P_{k-1}\}$ nem üres részhalmazokból álló halmazt, ahol $P_m \subseteq Q$. Ha közülük több, mint egy indukál nem triviális dekompozíciót, akkor a H kvázikód nem lesz felbontható az adott X halmazra nézve. $\square$

A tétel alkalmazásával a következő algoritmust nyertem: Vegyük az összes lehetséges $P_1, \dots, P_k$ részhalmazát az X halmazhoz tartozó minimálautomata állapotainak. Határozzuk meg az R_{p_i, p_j} halmazokat, majd hagyjuk el azokat, melyekre nem létezik olyan l , hogy $f(x_l) \subseteq R_{p_i, p_j}$. Ha létezik legalább két különböző $X = R_{q_0, P_1} \cdots R_{P_{k-1}, Q_F}$ sorozata az R halmazoknak, akkor a H kvázikód X -re nézve nem lesz felbontható.

Megjegyezem, hogy van olyan nemfelbontható kvázikód, melynek csak egy (vagy nulla) dekompozíciója van egy adott halmazra nézve. Például a $H = \{\{a\}, \{aa\}, \{b\}\}$ kvázikódnak egy dekompozíciója van az $X = \{ba\}$ halmazra vonatkoztatva.

3.2.6. A kvázikódot elfogadó véges automata

A következő fejezetekben annak eldöntésére, hogy egy kvázikód felbontható-e, véges automatát konstruálok. A jobb használhatóság érdekében felteszem, hogy a λ üresszó nem eleme a kódhalmazok egyikének sem, tehát a továbbiakban a Δ feletti H kvázikód a $2^{\Delta^+} \setminus \emptyset$ hatványhalmaz egy véges részhalmaza. Amennyiben egy Σ halmazhoz tartozó H kvázikód az $\bar{f} : \Sigma \rightarrow H$ függvénnyel van megadva, akkor a leképezést meghatározó automata az az $\mathcal{A}_{\bar{f}} = (Q, q_0, Q_F, A, \delta)$ automata lesz, ahol

- q_0 az $\mathcal{A}_{\bar{f}}$ kezdőállapota
- $Q_F = \Sigma = \{x_1, \dots, x_n\}$ az elfogadó állapotok halmaza
- Q az állapotok halmaza
- $A = \Delta$ a szalagábécé
- δ a mozgási szabályok halmaza

A δ mozgási szabályok halmazának elemeit az A feletti szavakra kiterjesztjük. Így $\delta(q, w) = \{q'\}$ azt jelenti, hogy a q állapotból a w szó a q' állapotba viszi az automatát és közben nem érintettünk egyetlen elfogadó állapotot sem. Legyen $\bar{f}(x_i) = H_i$, ahol $x_i \in \Sigma$ és legyen $H_i = \{p_{i,1}, \dots, p_{i,m_i}\} \in 2^{\Delta^+} \setminus \emptyset$. Ezekkel a jelölésekkel leírva az

$\mathcal{A}_{\bar{f}}$ automata kiterjesztett mozgási szabályai a következők:

$$\begin{array}{llll}
x_1 \in \delta(q_0, p_{1,1}), & x_2 \in \delta(q_0, p_{2,1}), & \dots, & x_n \in \delta(q_0, p_{n,1}), \\
\vdots & \vdots & & \vdots \\
x_1 \in \delta(q_0, p_{1,m_1}), & x_2 \in \delta(q_0, p_{2,m_2}), & \dots, & x_n \in \delta(q_0, p_{n,m_n}), \\
\vdots & \vdots & & \vdots \\
x_1 \in \delta(x_1, p_{1,1}), & x_2 \in \delta(x_1, p_{2,1}), & \dots, & x_n \in \delta(x_1, p_{n,1}), \\
\vdots & \vdots & & \vdots \\
x_1 \in \delta(x_1, p_{1,m_1}), & x_2 \in \delta(x_1, p_{2,m_2}), & \dots, & x_n \in \delta(x_1, p_{n,m_n}), \\
\vdots & \vdots & & \vdots \\
x_1 \in \delta(x_n, p_{1,1}), & x_2 \in \delta(x_n, p_{2,1}), & \dots, & x_n \in \delta(x_n, p_{n,1}), \\
\vdots & \vdots & & \vdots \\
x_1 \in \delta(x_n, p_{1,m_1}), & x_2 \in \delta(x_n, p_{2,m_2}), & \dots, & x_n \in \delta(x_n, p_{n,m_n}).
\end{array}$$

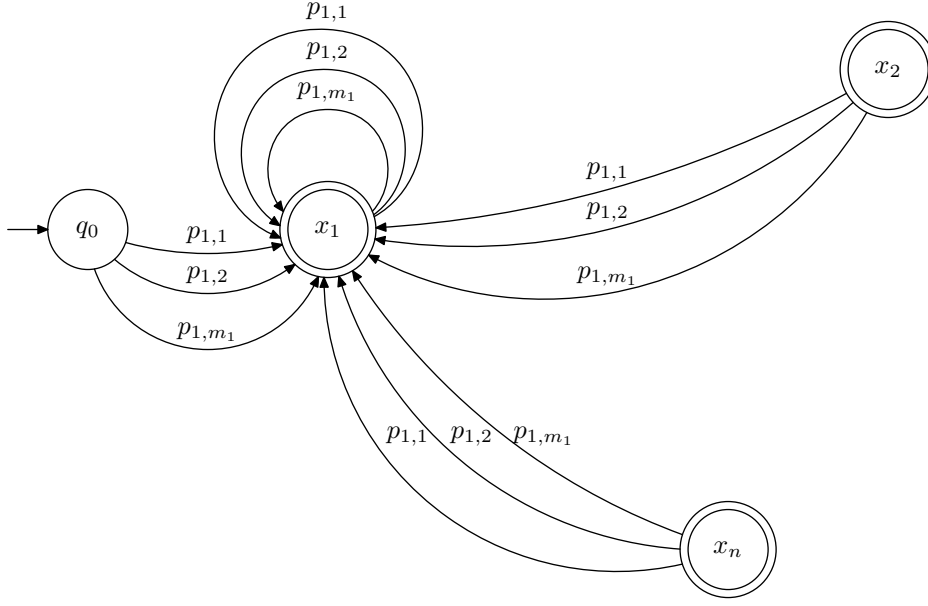
29. Tétel. [Fal11a] Ha $\exists i : H_i$ nem prefix halmaz, akkor $\mathcal{A}_{\bar{f}}$ nemdeterminisztikus.

BIZONYÍTÁS: Ha H_i nem prefix halmaz, akkor $\exists \gamma \in \Delta^+, \alpha, \beta \in H_i : \beta = \alpha\gamma$, emiatt egyrészt $q_0 \xrightarrow{\alpha} x_i$, mivel $\alpha \in H_i$, másrészt $q_0 \xrightarrow{\alpha\gamma} x_i$, s ekkor $q_0 \xrightarrow{\alpha} x \neq x_i$, tehát $\mathcal{A}_{\bar{f}}$ nemdeterminisztikus. $\square$

30. Tétel. [Fal11a] Ha $\exists i, j : H_i \cap H_j \neq \emptyset$, akkor $\mathcal{A}_{\bar{f}}$ nemdeterminisztikus.

BIZONYÍTÁS: Ha $H_i \cap H_j \neq \emptyset$, akkor $\exists \alpha \in H_i \cap H_j$, emiatt egyrészt $\alpha \in H_i$ miatt $q_0 \xrightarrow{\alpha} x_i$, másrészt $\alpha \in H_j$ miatt $q_0 \xrightarrow{\alpha} x_j$, tehát $\mathcal{A}_{\bar{f}}$ nemdeterminisztikus. $\square$

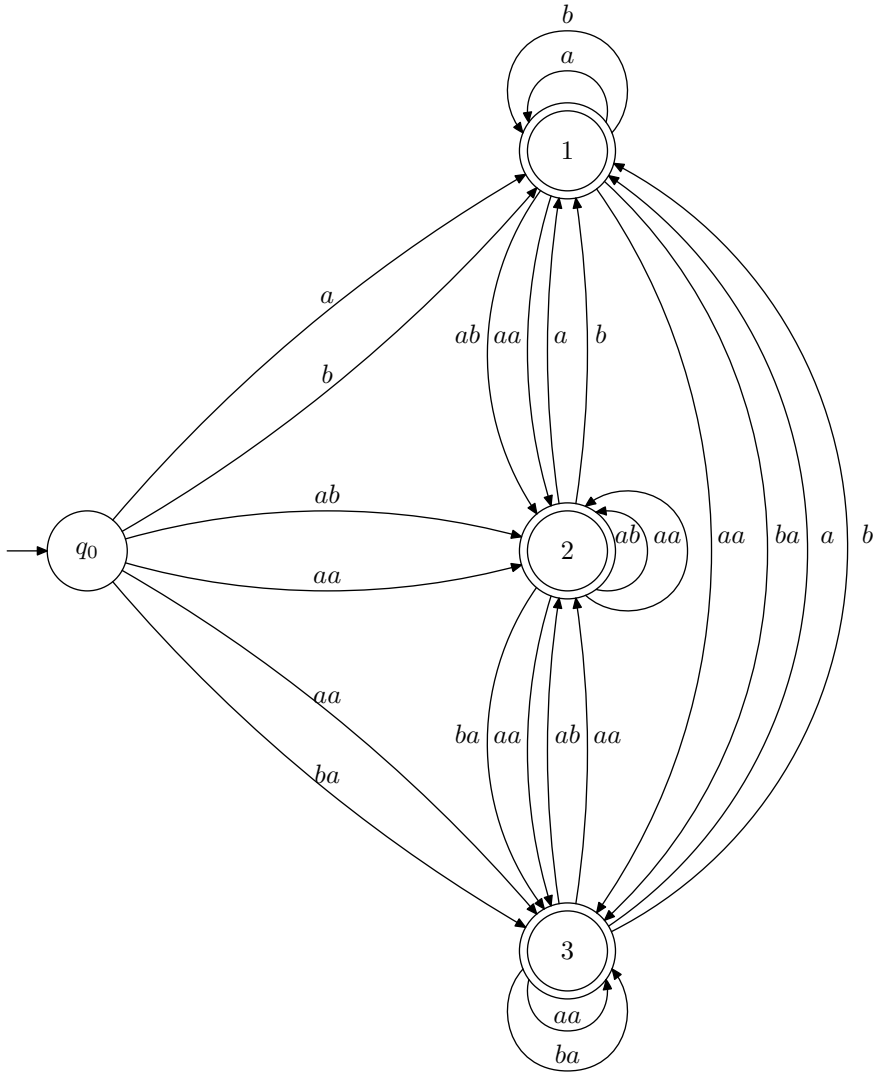
31. Tétel. [Fal11a] Ha $\bigcup_{i=1}^n H_i$, nem prefix halmaz, akkor $\mathcal{A}_{\bar{f}}$ nemdeterminisztikus.

3.6. ábra. Az automata x_1 -re vonatkozó részlete

BIZONYÍTÁS: Ha $\bigcup_{i=1}^n H_i$ nem prefix halmaz, akkor $\exists \gamma \in \Delta^+, \alpha \in H_{i_1}, \beta \in H_{i_2} : \beta = \alpha\gamma$, emiatt egyrészt $q_0 \xrightarrow{\alpha} x_{i_1}$, mivel $\alpha \in H_{i_1}$, másrészt $q_0 \xrightarrow{\alpha\gamma} x_{i_2}$, s ekkor $q_0 \xrightarrow{\alpha} x \neq x_{i_1}$, tehát $\mathcal{A}_{\bar{f}}$ nemdeterminisztikus. $\square$

3.2.7. Kvázikódok nemfelbonthatóságának egy szükséges feltétele

Amennyiben $A = f(x_{i_1}) \dots f(x_{i_k}) = f(x_{j_1}) \dots f(x_{j_l})$; $i_k \neq j_l$, akkor bármely $w \in A$ -ra igaz, hogy felírható $w = \alpha_{i_1} \dots \alpha_{i_k} = \beta_{j_1} \dots \beta_{j_l}$ alakban, ahol $\alpha_{i_m} \in f(x_{i_m})$ ahol $1 \leq m \leq k$ és $\beta_{j_m} \in f(x_{j_m})$ ahol $1 \leq m \leq l$.



3.7. ábra. A $H = \left\{ \{a, b\}, \{ab, aa\}, \{aa, ba\} \right\}$ kvázikód automatája

Eszerint, ha egy kód nem felbontható, akkor léteznie kell a fenti A halmaznak, és az A halmaz elemeinek teljesíteni kell az előbbieken ismertetett tulajdonságot. A következőkben ennek a tulajdonságnak a megjelenését fogom vizsgálni a kvázikódkhoz tartozó determinisztikus automatában. Mivel az automata determinisztikus, ezért egy jelsorozat olvasása által maximum egy állapotba kerülhet. Ha a jelsorozat eleme a kvázikód által generálható valamely halmaznak, akkor valamely végállapotba fog kerülni az automatám.

31. Definíció. [Fal11a] Ha egy nemdeterminisztikus automatában létezik a $\delta(p, a) = \{q_1, \dots, q_n\}$ szabály, akkor a $q_1, \dots, q_n$ állapotokat a determinisztikus automatában megjelenő $\delta(p, a) = \{q\}$ szabály generátor állapotainak nevezzük, és q -t $q_1 \dots q_n$ -nel jelöljük.

32. Definíció. [Fal11a] Egy determinisztikus automatát jelöltnek nevezünk, ha az állapotainak jelölései tartalmazzák a nemdeterminisztikus automata generátor állapotainak jelölését. A generátor állapotok közül a végállapotokat az utolsó kódszóval indexeljük. Egy állapot jelölthalmazán a generátor állapotainak halmazát értjük.

PÉLDA: Amennyiben a determinisztikus automatában például a $\delta(p_1 \dots p_n, a) = \{q_1 \dots q_m\}$ szabály úgy adódott, hogy a w kódszó olvasása véget ért a q_l állapotban, akkor q_l -t q_{l_w} -vel fogjuk jelölni.

32. Tétel. [Fal11a] Egy kvázikód nemfelbonthatóságának egy szükséges feltétele, hogy a hozzátartozó jelölt determinisztikus automatában létezen legalább egy olyan állapot, mely jelölthalmazában szerepel az eredeti nemdeterminisztikus automata legalább két végállapota.

BIZONYÍTÁS: Tekintsük az eredeti, még nem determinisztikussá tett automatáját a kvázikódnak. Ekkor a jelsorozat egyrészt az x_{i_k}

állapotba visz a $w = \alpha_{i_1} \dots \alpha_{i_k}$ egyenlőség miatt, másrészt eljuthatunk vele az x_{j_l} állapotba is a $w = \beta_{j_1} \dots \beta_{j_l}$ egyenlőség miatt. A determinisztikussá tétel során tehát legalább egy olyan állapot létre fog jönni a determinisztikus automatában, mely jelölthalmazában az x_{i_k} állapot együtt szerepel az x_{j_l} állapottal. $\square$

A tétel nem ad igazi eszközt a kezünkbe, ugyanis a gyakorlatban általában teljesül a tétel, ritka az, hogy egy kvázikód esetén ne lenne olyan jelsorozat, mely többféleképpen előáll a kvázikód halmazából származó elemek konkatenációjaként.

Ettől erősebb feltételt találunk, ha észrevevessük azt, hogy nem csak egy jelsorozatra kell teljesülnie ennek, hanem –ahogy említettük az imént– bármely $w \in \mathcal{L}(A)$ jelsorozat esetén.

33. Definíció. [Fal11a] Legyenek adva az $A, B, C_1, \dots, C_m$ halmazok. Azt mondjuk, hogy az A és a B halmaz kölcsönösen párosítva van a $C_1, \dots, C_m$ halmazokban, ha az összes $a \in A$ esetén létezik legalább egy $b \in B$ és $1 \leq i \leq m$ úgy, hogy $C_i \supseteq \{a, b\}$ és viszont, azaz az összes $b \in B$ esetén létezik legalább egy $a \in A$ és $1 \leq i \leq m$ úgy, hogy $C_i \supseteq \{a, b\}$.

Az $f(x) = \{w_1, \dots, w_n\}$ esetén jelölje $\tilde{f}(x)$ az $\{x_{w_1}, \dots, x_{w_n}\}$ halmazt.

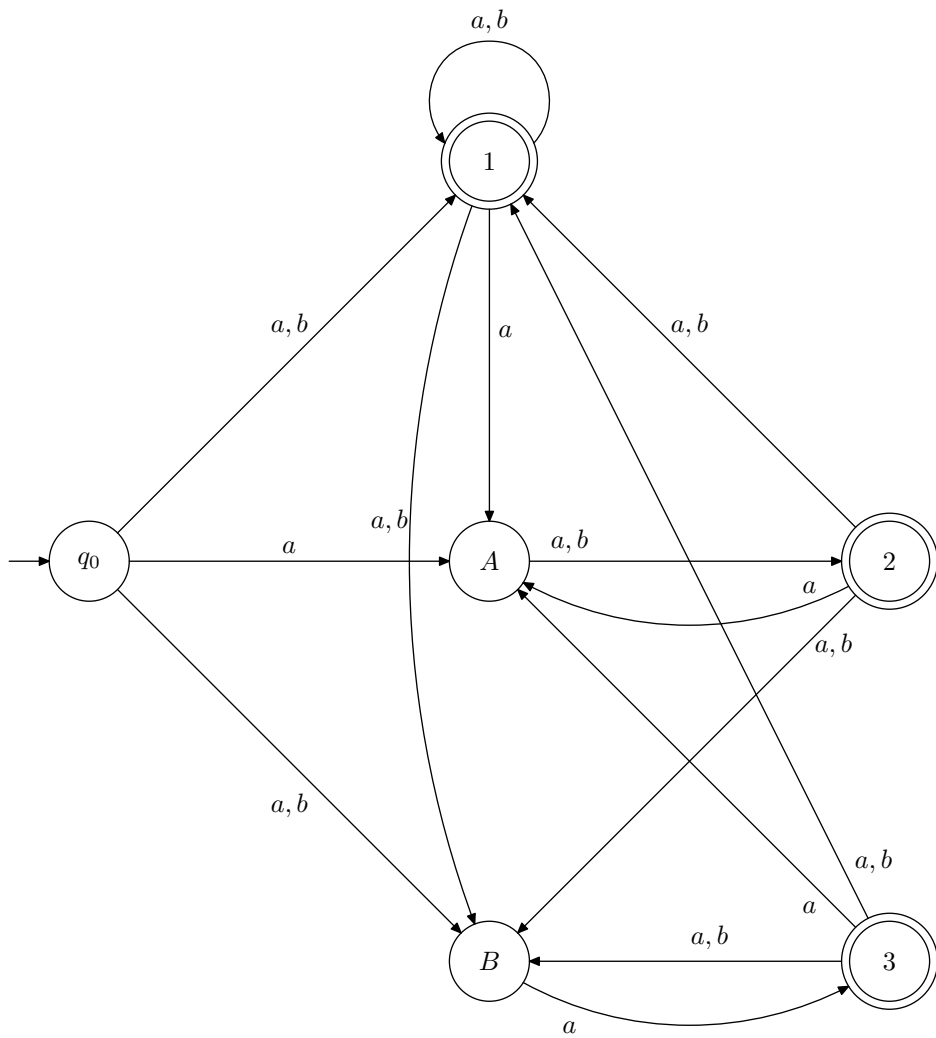
33. Tétel. [Fal08b, Fal11a] Ha egy $H = \{f(x_1), \dots, f(x_n)\}$ kvázikód nemfelbontható, akkor a H kvázikódhoz tartozó jelölt determinisztikus automatában léteznek a $C_1, \dots, C_m$ állapotok és az x_i, x_j generátor állapotok úgy, hogy $\tilde{f}(x_i)$ és $\tilde{f}(x_j)$ kölcsönösen párosítva van a $C_1, \dots, C_m$ állapotokhoz tartozó jelölthalmazokban.

BIZONYÍTÁS: Mivel a $H = \{f(x_1), \dots, f(x_n)\}$ kvázikód nemfelbontható, létezik az $A = f(x_{i_1}) \dots f(x_{i_k}) = f(x_{j_1}) \dots f(x_{j_l})$ halmaz, és $i_k \neq j_l$. Jelöljük az A halmaz számosságát $|A|$ -val. Ekkor az A halmazt felírhatjuk $A = \{w_1, \dots, w_{|A|}\}$ alakban is, ahol A minden

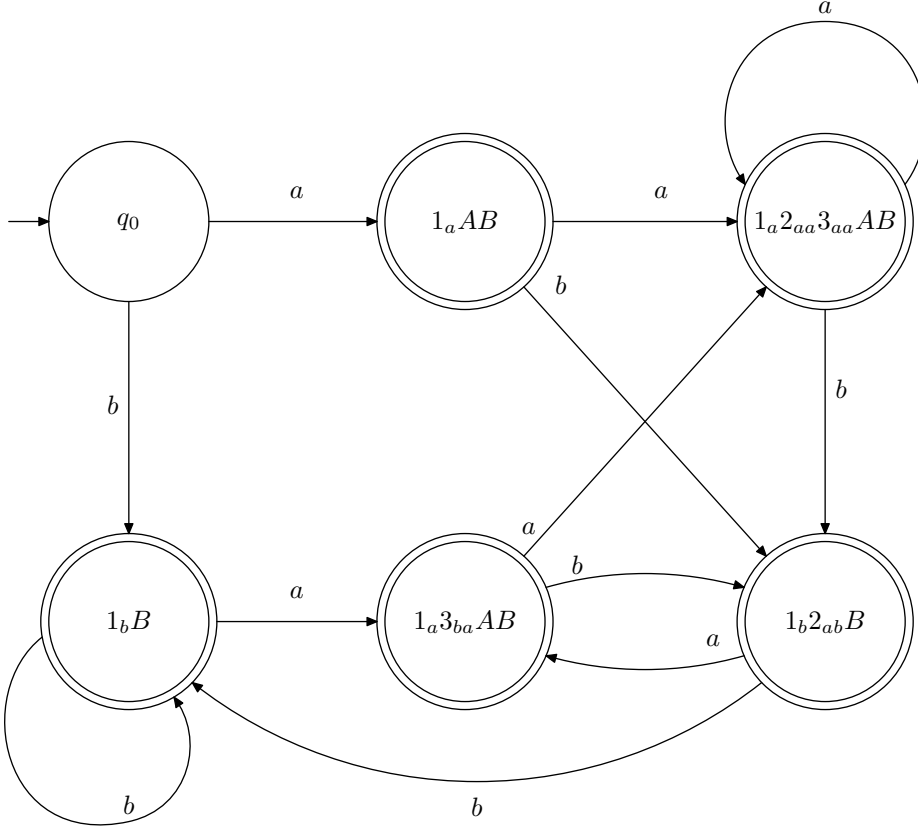
eleme írható $w_r = \gamma\alpha_{i_k} = \delta\beta_{j_l}$ formában. Tehát A minden eleme egyrészt az $\alpha_{i_k} \in f(x_{i_k})$ jelsorozatra, másrészt pedig a $\beta_{j_l} \in f(x_{j_l})$ jelsorozatra végződik. Ha egy A halmazra igaz, hogy $A = B_1 \cdots B_n$, akkor a B_n halmaz minden eleme az A halmaz valamely elemének suffix része lesz. Emiatt mindegyik $\alpha_{i_k} \in f(x_{i_k})$ és $\beta_{j_l} \in f(x_{j_l})$ jelsorozat szerepel valamely $w_i \in A$ jelsorozat szuffix részeként. Ezért minden $\alpha_{i_k} \in f(x_{i_k})$ jelsorozathoz fog tartozni legalább egy $\beta_{j_l} \in f(x_{j_l})$, és viszont, azaz minden $\beta_{j_l} \in f(x_{j_l})$ jelsorozathoz fog tartozni legalább egy $\alpha_{i_k} \in f(x_{i_k})$ úgy, hogy $w_r = \gamma\alpha_{i_k} = \delta\beta_{j_l}$ valamely $1 \leq r \leq |A|$ esetén. $\square$

Ahhoz, hogy ezt követni tudjuk a determinisztikussá tett automátában, az x_i állapotokat ellátjuk egy alsó indexszel, mely megmutatja, hogy melyik jelsorozat ért véget az x_i állapotban, tehát például ha $x_i = 2$ és $f(x_i) = f(2) = \{ab, aa\}$, akkor ha az ab jelsorozattal jutottunk a determinisztikussá tétel során a 2-es állapotba, akkor ezt a 2_{ab} jelöléssel jelezzük. A következő példa a $H = \{\{a, b\}, \{ab, aa\}, \{aa, ba\}\}$ kvázikód esetén mutatja a kiinduló automatát: Vizsgáljuk meg, hogy teljesül-e a szükséges feltétele a nemfelbonthatóságnak, azaz létezik-e két olyan állapot, melyek kölcsönös párosítással vannak jelen.

- Az 1-es és a 2-es kölcsönös párosításának vizsgálata: $f(1) = \{a, b\}$; $f(2) = \{ab, aa\}$. Az 1_a együtt van a 2_{aa} -val és az 1_b együtt van a 2_{ab} -vel, tehát kölcsönös párosítása van 1-nek és 2-nek, teljesül a nemfelbonthatóság szükséges feltétele. Bár felesleges tovább vizsgálni, mégis nézzük meg a feladat demonstratív jellege miatt a többi lehetséges kölcsönös párosításokat is.
- Az 1-es és a 3-as kölcsönös párosításának vizsgálata: $f(1) = \{a, b\}$; $f(3) = \{aa, ba\}$. Az 1_a együtt van a 3_{aa} -val és az 3_{ba} -val, de az 1_b nincs együtt egyetlen $f(3)$ beli elemmel sem, emiatt nincs kölcsönös párosítása az 1-nek és 3-nak.
- A 2-es és a 3-as kölcsönös párosításának vizsgálata: $f(2) =$



3.8. ábra. A $H = \left\{ \{a, b\}, \{ab, aa\}, \{aa, ba\} \right\}$ kvázikód kiinduló automatája



3.9. ábra. A $H = \{\{a, b\}, \{ab, aa\}, \{aa, ba\}\}$ kvázikód determinisztikus automatája

$\{ab, aa\}; f(3) = \{aa, ba\}$. Az 2_{aa} együtt van a 3_{aa} -val, de a 2_{ab} nincs együtt egyetlen $f(3)$ -beli elemmel sem, ahogy 3_{ba} sincs együtt egyetlen $f(2)$ -beli elemmel sem, tehát nincs kölcsönös párosítása a 2-nek és 3-nak.

Tehát teljesül a szükséges feltétel, azaz lehet, hogy a kvázikód

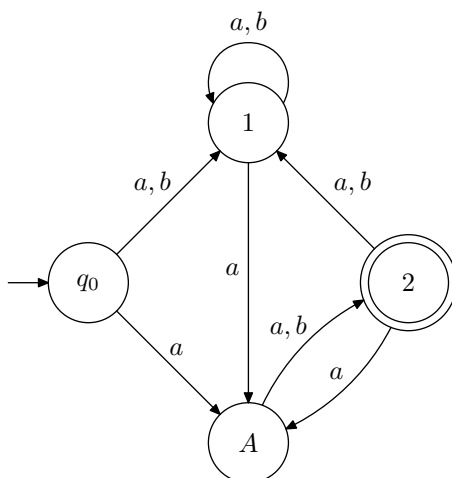
nemfelbontható. Valóban, ha megnézzük, $f(1)f(2) = f(3)f(1)$, hiszen

$$\{a, b\}\{ab, aa\} = \{aab, aaa, bab, baa\}$$

és

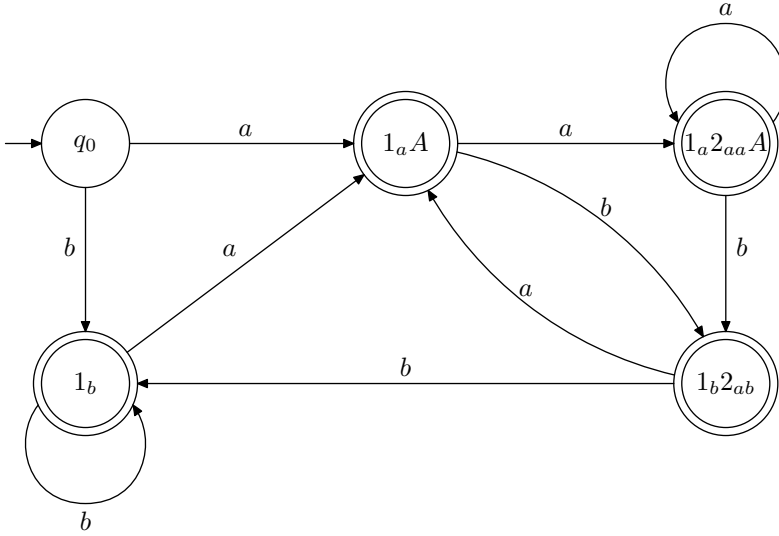
$$\{aa, ba\}\{a, b\} = \{aaa, aab, baa, bab\}.$$

Amint látható, tényleg létezik a két konkatenáció $f(1)$ és $f(2)$ szuffixummal. Ha megvizsgáljuk az $f(1) = \{a, b\}$, $f(2) = \{aa, ab\}$



3.10. ábra. A $H = \{\{a, b\}, \{aa, ab\}\}$ kvázikód kiinduló automatája

kvázikódolást, akkor látható, hogy valóban csak szükséges a feltétel és nem elegendő, ugyanis a kódolás által meghatározott $H = \{\{a, b\}, \{aa, ab\}\}$ kvázikódhoz tartozó 3.10. ábrán látható automatát tegyük determinisztikussá. Az eredmény a 3.11. ábrán van megadva. Vizsgáljuk meg, hogy teljesül-e a szükséges feltétele a nemfelbonthatóságnak, azaz a két állapot kölcsönös párosítással van-e jelen.



3.11. ábra. A $H = \{\{a, b\}, \{aa, ab\}\}$ kvázikód determinisztikus automatája

Az 1-es és a 2-es kölcsönös párosításának vizsgálata: $f(1) = \{a, b\}$; $f(2) = \{aa, ab\}$. Az 1_a együtt van a 2_{aa} -val és az 1_b együtt van a 2_{ba} -val, tehát kölcsönös párosítása van 1-nek és 2-nek, teljesül a nemfelbonthatóság szükséges feltétele. Azonban tekintve egy tetszőleges $X = X_1 \cdots X_n$ konkatenációját a halmazoknak, ahol $X_i = \{a, b\}$ vagy $X_i = \{aa, ab\}$, mindig eldönthető a sorozat első halmazáról, hogy az $\{a, b\}$ vagy az $\{aa, ab\}$ halmazzal egyenlő. Ugyanis ha X minden eleme a -val kezdődik, akkor csak az $\{aa, ab\}$ halmaz lehet az első tényező, különben pedig az $\{a, b\}$. Hagyjuk el az első elemet, de akkor ugyanezen érvelés alapján eldönthető a második halmazról is a hovatartozása és így tovább, tehát X felbontható. Tehát találtunk egy kvázikódot, melyre teljesül a szükséges feltétel a 33. tétel alapján, azaz a feltétel csupán szükséges, de nem elégséges.

3.2.8. A kvázikód egy lehetséges alkalmazása

Informatikai és általában minden kommunikációs rendszerben alapvető feladat a felhasználók azonosítása, melyet legtöbbször felhasználó neve és jelszava alapján oldanak meg. Ismertek más technikák is, például biometrikus adatok alapján való felismerés, egyedi adat alapján való azonosítás (megerősítő SMS). A témának bőséges szakirodalma van, például Kaufman és mások [KPS95], vagy Nemetz és Vajda [NV91] írt a témáról összefoglaló jellegű művet. A legtöbb azonosítási rendszerben problémát jelent az úgynevezett „beékelődéses” támadás ellen való védekezés. A probléma abból adódik, hogy az adataink a hálózaton sok gépen jutnak keresztül mire eléri az adott célt, s a közbülső gépek mindegyike potenciálisan lehetőséget kínál arra, hogy az áthaladó adatforgalmat tárolja. A gyakorlatban emiatt titkosítottan küldik az adatokat a csatornán, lehetőleg olyan kódolást választva, melynek feltörése hosszú időt vesz igénybe, így a küldött adatok a kívánt ideig biztonságban vannak. Az értekezésben megadok egy olyan megoldást ennek a veszélynek a kivédésére, mely a kvázikódokat használja. A technika elméleti jellegű, a gyakorlatban még nem került kipróbálásra, implementálásra.

Tegyük fel, hogy az U ügyfél szeretné magát azonosítani az S rendszerben. Az azonosítási algoritmus kiválasztásánál figyelembe vesszük, hogy beékelődéses támadásra kell felkészülni, emiatt nincs lehetőség egy jelszót többször használni, tehát a felhasználónak minden egyes belépés után más és más jelszóval kell belépni. Lehetőségként a következő módszerek merülnek fel, melyekkel szembeni kifogást is megadom:

- Valamilyen függvénnyel generálódjon az új jelszó a régeből.
Probléma: A függvény kiismerhető egy elég sokáig tartó lehallgatás esetén.
- Minden egyes belépés után úgynevezett biztonságos csatornán

új jelszót kapjon az ügyfél.

Probléma: A biztonságos csatorna általában csak személyes találkozással oldható meg, nem kényelmes az ügyfél számára sem.

- Kapjon induláskor az ügyfél egy számára „dedikált” jelszó halmazt.

Probléma: Ezt biztos nem tudja az ügyfél fejben tartani, azaz rögzíteni fogja, ami már nem biztonságos.

Egy lehetséges megoldást kínál a kvázikódolás, melynek működését a következő példában szemléltetem:

Legyen adva a $H = \{f(0), \dots, f(9)\}$ kvázikód. A kvázikód által indukált kódolást mind az ügyfél, mind a rendszer egy alkalmazással oldja meg, amely képes rögzíteni a kiküldésre került, illetve beérkezett jelszavakat, s ezeket többször nem küldi be, illetve nem fogadja el. Az ügyfél a rendszertől biztonságos csatornán megkapja a fent említett H kvázikódot. Azonosításkor az ügyfél például egy négyjegyű számot ír be, melyet az ügyféloldali alkalmazás kódol át –a H kvázikód alapján– egy jelsorozattá. Feltéve azt, hogy a H kvázikód minden halmaza 10 elemű és alkalmasan megválasztott, a lehetséges jelsorozatok száma négyjegyű kód esetén 10000, mely bőségesen elegendőnek tűnik, s célszerűen emelhető egyrészt a számjegyek számával, másrészt a kvázikód halmazainak számosságával. A négyjegyű számból létrejött jelsorozatot az ügyféloldali alkalmazás elküldi a rendszernek, egyben letiltja az újbóli elküldését. A fogadó oldalon a rendszer fogadja a jelsorozatot, beazonosítja, hogy származhat-e az adott felhasználótól, majd pozitív válasz esetén azonosítja az ügyfelet és egyben letiltja a beérkezett jelsorozat újbóli elfogadását. Így a beékelődött támadó hiába szerzi meg a jelszót és a felhasználói nevet, nem tudja felhasználni újbóli belépésre. A problémát jelen esetben az okozza, hogy a halmazokat hogyan tudjuk jól megválasztani, azaz mennyi az esélye annak, hogy úgynevezett „brute-force” támadással

rendszerünk feltörhető. Az ilyen támadókkal szembeni védekezés egyrészt a sikertelen belépési kísérletek számának korlátozásával, másrészt pedig emberi tényező beiktatásával lehetséges, például a gyakran alkalmazott karakter felismerés technikájával (CAPTCHA).

Az ismertetett technológiával kapcsolatban a kvázikód felbonthatósága azért fontos, mert nemfelbonthatóság esetén több négyjegyű számhoz tartoznának pontosan ugyanazon jelsorozatok. Természetesen így is lehetséges, hogy egy-egy esetben két különböző számhoz ugyanazon jelsorozat generálódik némely esetben, de a halmazok alkalmas megválasztásával ez kikerülhető, például mindegyik halmaz azonos hosszúságú jelsorozatokot tartalmaz és diszjunktak.

4. fejezet

Összegzés és további kutatási feladatok

Ismertettem azokat a főbb eredményeket, melyek a kódok felbonthatóságának automataelméleti alapokon való megközelítésével kapcsolatosak. Bemutattam a felbonthatóságot eldöntő legelső algoritmust, mely általában Sardinas–Patterson algoritmusként ismert. A továbbiakban három automatát ismertettem, melyek mindegyike megoldja a problémát:

- A virág automata esetében a [BP85]-ben közölt példát végig számoltam és megmutattam, hogy a kód felbontható.
- A König által definiált automaták összefűzött szorzata esetén megadtam egy példát és annak megoldását, mivel az eredeti cikkben nem szerepelt az algoritmus alkalmazása.
- A Tsuji által megadott példában korrigáltam a téves mozgási szabályokat, valamint egy saját példán szemléltettem egy olyan esetet, amikor a kód nem lesz felbontható.

A dolgozatban megmutattam egy új λ -véges automatán alapuló általam alkotott algoritmust a felbonthatóság eldöntésére, majd ismertettem az algoritmus kapcsolatát a Sardinas–Patterson algoritmussal, mindezt példákkal alátámasztva.

Definiáltam egy új kódolást, azaz megadtam a kvázikódok fogalmát. Bevezettem a felbonthatóság egy lehetséges kiterjesztését a kvázikódokra. Megadtam egy algoritmust, melynek segítségével megkonstruálható egy adott kvázikód esetén a kvázikód által definiált nyelvet elfogadó véges automata. Megmutattam néhány összefüggést a kvázikódhoz tartozó automata és a kvázikód között. Definiáltam a kvázikódokra a λ -prefixmentesség fogalmát, majd vizsgáltam ennek a relációnak a tulajdonságait. Megmutattam, hogy ha egy λ -a $\{\lambda\}$ halmazt nem tartalmazó λ -suffixiál halmazokból álló kvázikód λ -prefixmentes, akkor felbontható is. Megadtam a nemfelbontható kvázikód néhány tulajdonságát. Bebizonyítottam, hogy a Sardinas–Patterson algoritmus kvázikódokra alkalmazott változata nem dönti el a felbonthatóságot. Bevezettem kvázikódokra a dekompozíciós halmazok fogalmát és megmutattam, hogy bármely véges halmaz esetén eldönthető az, hogy adott kvázikódra nézve felbontható-e. Végül megadtam a kvázikódok felbonthatóságának egy szükséges feltételét, mely a hagyományos felbonthatóságot vizsgáló algoritmusomon alapszik, azaz itt is az automatákat alkalmaztam, mint megoldó eszközöket. Megmutattam, hogy a feltétel nem elégséges. Végül megadtam a kvázikód egy lehetséges gyakorlati alkalmazását.

További kutatási feladat tárgyát képezik a következők:

- Szükséges és elegendő feltételek kidolgozása a kvázikódok felbonthatóságára
- További általánosítások
- Végtelen kvázikódokra való alkalmazhatóságok vizsgálata.

- Egyéb gyakorlati alkalmazási lehetőségek felkutatása

5. fejezet

Summary and further research tasks

We reviewed the main results concerning the decipherability of codes, to narrow the results which are based on automata theory. We showed the foremost algorithm for decipherability, which is known as Sardinas–Patterson algorithm. We presented three automata. All of them solve the problem of decipherability.

- In the case of the flower automaton we completed the example in [BP85], and we showed, that the code is decipherable.
- In the case of the latticed products of automata defined by König, we gave an example and its solution, since in that paper the application of the algorithm was not presented.
- In the example given by Tsuji, we corrected the mistaken transaction rules. By an example we presented the case of non-decipherable code.

In the dissertation we showed our new algorithm for deciding the uniquely decipherability of codes. This algorithm is based on theory of

finite automata. We showed the connection of our algorithm and the Sardinas–Patterson algorithm and we gave some examples for these states.

We defined a new coding method, that is we gave the definition of quasi-codes. We extended the decipherability property to the quasi-codes. We gave an algorithm to construct a finite automaton for a given quasi-code. The automaton and the quasi-code are equivalent, that is the language defined by automaton and the language over the quasi-code are the same. We showed some connections between the automaton of quasi-code and the quasi-code. We defined the notion of λ -prefix-free relation for quasi-codes, and we investigated the properties of this relation. We proved, that every non-empty λ -prefix-free quasi-code consist of suffixial sets -which does not contain λ set- is decipherable. We gave some properties of the non-decipherable quasi-code. We proved, that the Sardinas–Patterson algorithm modified for quasi-codes can not decide the question of decipherability. We introduced the definition of decomposition sets for quasi-codes. We introduced the notion of decomposition sets for quasi-codes. We showed, that the problem of decipherability is decidable for finite sets. Finally, we gave a necessary condition of decipherability of quasi-codes, which is based on our algorithm for codes. We use finite automaton, too. We showed, that the condition is only necessary, but is not sufficient. Finally, we gave a possible application of quasi-code. Objects of further research:

- Development of necessary and sufficient conditions for the decipherability of quasi-codes.
- Further generalizations.
- Investigation of applicability for the infinite quasi-codes.
- Further possible applications

Irodalomjegyzék

- [AF05] S.V. Avgustinovich and A.E. Frid. A unique decomposition theorem for factorial languages. *International Journal of Algebra and Computation*, 15(1):149–160, 2005.
- [AG84] A. Apostolico and R. Giancarlo. Pattern matching machine implementation of a fast test for unique decipherability. *Inform. Process. Lett.*, 18(3):155–158, 1984.
- [AHU75] Alfred V. Aho, John E. Hopcroft, and Jeffrey D. Ullman. *The design and analysis of computer algorithms*. Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, 1975. Second printing, Addison-Wesley Series in Computer Science and Information Processing.
- [AL99] X. Augros and I. Litovsky. Algorithms to test rational ω code. In *Mathematical Foundations of Informatics'99 Conference*, Hanoi, 1999.
- [Ban63] G. Bandyopadhyay. A simple proof of the decipherability criterion of Sardinas and Patterson. *Information and Control*, 6:331–336, 1963.

- [BP85] J. Berstel and D. Perrin. *Theory of codes*, volume 117 of *Pure and Applied Mathematics*. Academic Press Inc., Orlando, FL, 1985.
- [BPR09] J. Berstel, D. Perrin, and C. Reutenauer. *Codes and Automata*. Cambridge University Press, 2009. <http://www-igm.univ-mlv.fr/~berstel/LivreCodes/CodesAutomata.pdf>.
- [BR05] F. Burderi and A. Restivo. Varieties of codes and kraft inequality. *Lecture Notes in Comput. Sci.*, 3404:545–556, 2005.
- [Brz67] J. A. Brzozowski. Roots of star events. *J. Assoc. Comput. Mach.*, 14:466–477, 1967.
- [DDP99] J. Demetrovics, J. Denev, and J. Pavlov. *A számítástudomány matematikai alapjai*. Nemzeti Tankönyvkiadó, Bp., 1999.
- [dL76] Aldo de Luca. A note on variable length codes. *Information and Control*, 32(3):263–271, 1976.
- [Eil74] Samuel Eilenberg. *Automata, languages, and machines. Vol. A*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York, 1974. Pure and Applied Mathematics, Vol. 58.
- [Fal04] J. Falucskai. DFA of non unique decodable code. In *6th International Conference on Applied Informatics*, pages 303–309, Eger, 2004.
- [Fal06] J. Falucskai. On a test for codes. *Acta Mathematica Academiae Paedagogicae Nyíregyháziensis*, 22:121–125, 2006.

- [Fal07] J. Falucskai. Some algorithms concerning uniquely decipherable codes. In *7th International Conference on Applied Informatics*, volume 2, pages 229–236, Eger, 2007.
- [Fal08a] J. Falucskai. On equivalence of two tests for codes. *Acta Mathematica Academiae Paedagogicae Nyíregyháziensis*, 24:249–256, 2008.
- [Fal08b] J. Falucskai. Változó hosszúságú kódok. In *II. Nyíregyházi Doktorandusz (PhD/DLA) Konferencia*, pages 219–222, Nyíregyháza, 2008.
- [Fal11a] J. Falucskai. A new interpretation of the decipherability. Közlésre benyújtva, *Acta Informatica*, 2011.
- [Fal11b] János Falucskai. A novel test for unique decipherability of codes. *Publ. Math. Debrecen*, 78(3-4):535–541, 2011.
- [Gal85] Zvi Galil. Open problems in stringology. In *Combinatorial algorithms on words (Maratea, 1984)*, volume 12 of *NATO Adv. Sci. Inst. Ser. F Comput. Systems Sci.*, pages 1–8. Springer, Berlin, 1985.
- [GGV02] L. Györfi, S. Györi, and I. Vajda. *Információ- és Kódelmélet*. TypoT_EX kiadó, 2002.
- [GM59] E. N. Gilbert and E. F. Moore. Variable-length binary encodings. *Bell System Tech. J.*, 38:933–967, 1959.
- [Hof84] Christoph M. Hoffmann. A note on unique decipherability. In *Mathematical foundations of computer science, 1984 (Prague, 1984)*, volume 176 of *Lecture Notes in Comput. Sci.*, pages 50–63. Springer, Berlin, 1984.

- [HU79] John E. Hopcroft and Jeffrey D. Ullman. *Introduction to automata theory, languages, and computation*. Addison-Wesley Publishing Co., Reading, Mass., 1979. Addison-Wesley Series in Computer Science.
- [KPS95] Charlie Kaufman, Radia Perlman, and Mike Speciner. *Network Secutity: Private Communication in a Public World*. Englewood Cliffs, Prentice/Hall, 1995.
- [Kön94] R. König. Lectures on codes. Internal Reports of the IMMD I, Friedrich Alexander University of Erlangen-Nürnberg, 1994.
- [Lev64] V. I. Levenšteĭn. Some properties of coding and self-adjusting automata for decoding messages. *Problemy Kibernet. No.*, 11:63–121, 1964.
- [Mar62] Al. A. Markov. On alphabet coding. *Soviet. Phys. Dokl.*, 6:553–554, 1962.
- [McC96] Robert McCloskey. An $O(n^2)$ time algorithm for deciding whether a regular language is a code. In *Proceedings of the 8th International Conference of Computing and Information, ICCI'96 (Waterloo, ON)*, volume 2, pages 79–89 (electronic), 1996.
- [McM56] B. McMillan. Two inequalities implied by unique decipherability. *IRE Transactions Information Theory*, 2:115–116, 1956.
- [MS77] F. J. MacWilliams and N. J. Sloane. *The theory of Error Correcting Codes*. North-Holland, 1977.

- [MSY98] Alexandru Mateescu, Arto Salomaa, and Sheng Yu. On the decomposition of finite languages. Technical Report 222, TUCS, December 1998.
- [Niv66] Maurice Nivat. Éléments de la théorie générale des codes. In *Automata Theory*, pages 278–294. Academic Press, New York, 1966.
- [NV91] T. Nemetz and I. Vajda. *Bevezetés az algoritmikus adatvédelembe*. Akadémiai Kiadó, 1991.
- [Ril67] John A. Riley. The Sardinas-Patterson and Levenshtein theorems. *Information and Control*, 10:120–136, 1967.
- [Rod82] M. Rodeh. A fast test for unique decipherability based on suffix trees. *IEEE Trans. Inform. Theory*, IT-28:648–651, 1982.
- [RS97] G. Rosenberg and A. Salomaa, editors. *Handbook of formal languages. Vol. 1,2,3*. Springer-Verlag, Berlin, 1997. Beyond words.
- [Sal73] Arto Salomaa. *Formal languages*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1973. ACM Monograph Series.
- [Sch55] M. Schützenberger. Une théorie algébrique du codage. *Séminaire Dubreil-Pisot 1955-56*, page Exposé 15, 1955.
- [Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423, 623–656, 1948.
- [Shy01] H. J. Shyr. *Free Monoids and Languages*. Hon Min Book Company, Taichung, Taiwan, 2001.

- [SP53] A. A. Sardinas and C. W. Patterson. A necessary and sufficient condition for the unique decomposition of coded messages. *IRE Internat. Conv. Rec.*, 8:104, 1953.
- [Spe76] J. C. Spehner. *Quelques problèmes d'extension, de conjugaison et de présentation des sous-monoïdes d'un monoïde libre*. PhD thesis, Université Paris 7, 1976.
- [SW49] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana, Ill., 1949.
- [SW86] C. E. Shannon and W. Weaver. *A kommunikáció matematikai elmélete. Az információ elmélet születése és távlatai*. OMIK, Bp., 1986.
- [Tsu01] Kayoko Tsuji. An automaton for deciding whether a given set of words is a code. *RIMS Kokyuroku*, 1222:123–127, 2001.

Publikációs lista

- [1] J. Falucskai. A novel test for unique decipherability of codes. *Publicationes Mathematicae Debrecen*, volume 78, Fasc.: 3–4, 535–541, 2011.
DOI: 10.5486/PMD.2011.4716
- [2] J. Falucskai. On the k-reversibility of finite automata. *Annales Mathematicae et Informaticae*, 36:71–75, 2009.
ZENTRALBLATT Zbl 1212.68088
- [3] J. Falucskai. On equivalence of two tests for codes. *Acta Mathematica Academiae Paedagogicae Nyíregyháziensis*, 24:249–256, 2008.
ZENTRALBLATT Zbl 1164.94303
- [4] J. Falucskai. Some algorithms concerning uniquely decipherable codes. In *7th International Conference on Applied Informatics*, volume 2, pages 229–236, Eger, 2007.
ZENTRALBLATT Zbl 1179.94075
- [5] J. Falucskai. On a test for codes. *Acta Mathematica Academiae Paedagogicae Nyíregyháziensis*, 22:121–125, 2006.
ZENTRALBLATT Zbl 1120.94013

- [6] J. Falucskai. DFA of non unique decodable code. In *6th International Conference on Applied Informatics*, volume 1, pages 303–309, Eger, 2004.
ZENTRALBLATT Zbl 1074.68507
- [7] J. Falucskai. An application of regular expressions. In *microCAD 2003 International Scientific Conference*, volume Section N: Applied Information Engineering, pages 59–64, Miskolc, 2003.
- [8] J. Falucskai. Unique decodability of codes. In *microCAD 2002 International Scientific Conference*, volume Section H: Applied Information Engineering, pages 77–82, Miskolc, 2002.

Előadások

- [1] J. Falucskai. A New Interpretation of the Decipherability. *Brno University of Technology*, 2011.07.21
- [2] J. Falucskai. Some Algorithms Concerning Uniquely Decipherable Codes. *Brno University of Technology*, 2011.03.29
- [3] J. Falucskai. Változó hosszúságú kódok. In *II. Nyíregyházi Doktorandusz (PhD/DLA) Konferencia*, pages 219–222, Nyíregyháza, 2008.
- [4] J. Falucskai. Kódok felbonthatóságát vizsgáló program. *Magyar Tudomány Napja*, Nyíregyháza, 2004.
- [5] J. Falucskai. Reguláris kifejezések alkalmazása. *Az MTA Sz-Sz-B Megyei Tudományos Testület tudományos ülése*, Nyíregyháza, 2003.
- [6] J. Falucskai. Reguláris kódok alkalmazása. *Magyar Tudomány Napja*, Nyíregyháza, 2003.
- [7] J. Falucskai. Nemprefix kódok automatái. *Az MTA Sz-Sz-B Megyei Tudományos Testület tudományos ülése*, Nyíregyháza, 2002.

- [8] J. Falucskai. Reguláris kifejezések. *Magyar Tudomány Napja*, Nyíregyháza, 2002.
- [9] J. Falucskai. Reguláris halmaz egyenletrendszerek és nem felbontható kódok. *Magyar Tudomány Napja*, Nyíregyháza, 2001.
- [10] J. Falucskai. Kódok és automaták összefüggéseinek néhány vizsgálata. *Magyar Tudomány Napja*, Nyíregyháza, 2000.
- [11] J. Falucskai. Véges mondattal rendelkező nyelvek és a kódelmélet néhány összefüggése. *Az MTA Sz-Sz-B Megyei Tudományos Testület tudományos ülése*, Nyíregyháza, 2000.

Egyéb publikációk

- [1] J. Falucskai. A new interpretation of the decipherability. Submitted for publication, Acta Informatica, 2011.
- [2] J. Falucskai. A számítástudomány néhány analóg rendszere. *Természettudományi közlemények, Nyíregyházi Főiskola*, 4:89–99, Nyíregyháza, 2004.
- [3] Á. Kuki, J. Falucskai, and K. Tarnay. *Bevezetés a formális nyelvek és automaták alkalmazásába*, volume 7 of *Az MTA Sz-Sz-B Megyei Tudományos Testületének Közleményei*. Nyíregyháza, 1993.