Journal of Discrete Algorithms 28 (2014) 31–40

Contents lists available at ScienceDirect

Journal of Discrete Algorithms

www.elsevier.com/locate/jda

The lexicographically smallest universal cycle for binary strings with minimum specified weight

Joe Sawada [1], Aaron Williams [2], Dennis Wong *

School of Computer Science, University of Guelph, Canada

ARTICLE INFO

Article history:
Available online 2 July 2014

Keywords:
De Bruijn cycle
Universal cycle
Necklace
FKM algorithm
Minimum weight

ABSTRACT

H. Fredricksen, I.J. Kessler and J. Maiorana discovered a simple but elegant construction of a universal cycle for binary strings of length n: Concatenate the aperiodic prefixes of length n binary necklaces in lexicographic order. We generalize their construction to binary strings of length n whose weights are in the range $c, c+1, \ldots, n$ by simply omitting the necklaces with weight less than c. We also provide an efficient algorithm that generates the universal cycles in constant amortized time per bit using $O(n)$ space. Our universal cycles have the property of being the lexicographically smallest universal cycle for the set of binary strings of length n with weight $\geq c$.

1. Introduction

Let $\mathbf{B}(n)$ denote the set of all binary strings of length n. A *universal cycle* for a set $\mathbf{S}$ is a cyclic sequence $u_1u_2 \ldots u_{|\mathbf{S}|}$ where each substring of length n corresponds to a unique object in $\mathbf{S}$. When $\mathbf{S} = \mathbf{B}(n)$, these sequences are commonly known as *de Bruijn sequences* since they were proven to exist and counted by de Bruijn [5] (also see [6]). These sequences were also independently discovered by Good [10] in the same year. As an example, the cyclic sequence 0000100110101111 is a universal cycle (de Bruijn sequence) for $\mathbf{B}(4)$; the 16 unique substrings of length 4 when considered cyclicly are:

$$0000, 0001, 0010, 0100, 1001, 0011, 0110, 1101, 1010, 0101, 1011, 0111, 1111, 1110, 1100, 1000.$$

When considering universal cycles for a specific set $\mathbf{S}$, there are several important questions: Does a universal cycle exist for $\mathbf{S}$? What is the number of universal cycles for $\mathbf{S}$? How can a specific universal cycle for $\mathbf{S}$ be constructed? Is there an efficient algorithm that constructs a universal cycle for $\mathbf{S}$? The last two questions can also be put for the lexicographically smallest universal cycle for $\mathbf{S}$. By *lexicographically smallest*, we mean that the linear representation is the smallest possible in lexicographic order. For instance, the universal cycle from our example is the lexicographically smallest for $\mathbf{B}(4)$. (The term *minimal* is also used in the literature [19,20] for the same concept.)

The lexicographically smallest universal cycle for $\mathbf{B}(n)$ was first constructed by Martin in the 1930s [18]. The author showed that the lexicographically smallest universal cycle for $\mathbf{B}(n)$ can be constructed by a greedy algorithm that uses exponential space. Later, Fredricksen, Kessler and Maiorana provided a more direct method in [8] for constructing this

* Corresponding author.
E-mail addresses: jsawada@uoguelph.ca (J. Sawada), haron@uvic.ca (A. Williams), cwong@uoguelph.ca (D. Wong).
1 Research supported by NSERC.
2 Research supported in part by the NSERC Accelerator and Discovery Programmes, and a basic research grant from ONR.

http://dx.doi.org/10.1016/j.jda.2014.06.003
1570-8667/

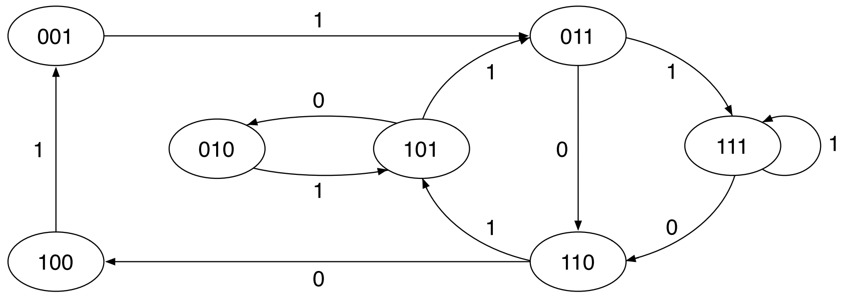

Fig. 1. The de Bruijn graph $G(\mathbf{B}_2^4(4))$.

universal cycle, and this method is now referred to as the FKM construction. Ruskey, Savage, and Wang [21] provided an algorithm for generating the FKM construction and analyzed its efficiency. Due to its importance and interesting history, Knuth refers to the lexicographically smallest universal cycle for $\mathbf{B}(n)$ as the *grand-daddy* of de Bruijn sequences [16].

Universal cycles have been studied for a variety of combinatorial objects including permutations, partitions, subsets, multisets, labeled graphs, various functions, and passwords [1,2,4,12–17,24,27]. Fredricksen, Kessler and Maiorana generalize their results to construct the lexicographically smallest universal cycle for k-ary strings of length n [9]. Many papers have focused on finding constructions and efficient algorithms to generate universal cycles for interesting subsets of k-ary strings of length n [7,11,17,23,25,26,28].

Let $\mathbf{B}_c^d(n)$ denote the set of length n binary strings whose weights (number of 1s) are in the range $c, c+1, \ldots, d$. A *universal cycle for binary strings with a minimum specified weight* is a cyclic sequence of length $\binom{n}{c} + \binom{n}{c+1} + \cdots + \binom{n}{n}$ that contains each string in $\mathbf{B}_c^n(n)$ exactly once as a substring. We refer to these universal cycles as *minimum-weight universal cycles* for simplicity. For example, the circular sequence 00110101111 is a minimum-weight universal cycle for $\mathbf{B}_2^4(4)$ since its 11 substrings of length 4 include each element in

$$\mathbf{B}_2^4(4) = \{0011, 0101, 0110, 1001, 1010, 1100, 0111, 1011, 1101, 1110, 1111\}$$

exactly once. Similarly, a *universal cycle for binary strings with a maximum specified weight*, or simply a *maximum-weight universal cycle*, is a cyclic sequence of length $\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d}$ that contains each string in $\mathbf{B}_0^d(n)$ exactly once as a substring. A maximum-weight universal cycle for $\mathbf{B}_0^d(n)$ can be obtained by complementing each bit of a minimum-weight universal cycle for $\mathbf{B}_{n-d}^n(n)$ [25].

In this paper, a universal cycle has an *efficient algorithm* if each successive symbol of the sequence can be generated in constant amortized time (CAT) while using a polynomial amount of space with respect to n. A universal cycle for $\mathbf{B}_{d-1}^d(n)$ is known as a *dual-weight universal cycle*, and more generally a universal cycle for $\mathbf{B}_c^d(n)$ is known as a *weight-range universal cycle*. Algorithms to generate universal cycles with various weight-ranges have previously been studied in the sequence of the following articles:

- an efficient algorithm for dual-weight universal cycles is given in [23],
- an efficient algorithm for minimum-weight and maximum-weight universal cycles is given in [25],
- an efficient algorithm for weight-range universal cycles is given in [26].

Although efficient algorithms for generating minimum-weight and maximum-weight universal cycles are given in [25] (and generalized in [26]), there are several advantages to our new results. Firstly, our new universal cycles are the lexicographically smallest, whereas the constructions in [23,25,26] are not. Secondly, the constructions in [25,26] are based on cutting and pasting dual-weight universal cycles from [23], whereas our new construction is much simpler. Thirdly, our new constructions are based on lexicographic order, whereas the constructions in [25,26] are complicated by their use of 'cool-lex' order. (The construction in [25] was simplified by a generalized version of cool-lex order found in [28], although that article did not include an efficient algorithm.)

The *de Bruijn graph* $G(\mathbf{S})$ for a set of length n strings $\mathbf{S}$ is a directed edge-labeled graph whose vertex set consists of the length $n-1$ strings that are a prefix or a suffix of the strings in $\mathbf{S}$. For each string $b_1 b_2 \ldots b_n \in \mathbf{S}$ there is an edge labeled b_n that is directed from the vertex $b_1 b_2 \ldots b_{n-1}$ to the vertex $b_2 b_3 \ldots b_n$. Thus, the graph has $|\mathbf{S}|$ edges. As an example, the de Bruijn graph $G(\mathbf{B}_2^4(4))$ is illustrated in Fig. 1. It is well known that $\mathbf{S}$ admits a universal cycle if and only if $G(\mathbf{S})$ is directed Eulerian. The de Bruijn graph $G(\mathbf{B}_c^d(n))$ is directed Eulerian for all $0 \le c < d \le n$ [25,26].

The problem of finding a directed Euler cycle of lexicographically minimal labels of an edge-labeled directed graph has been applied to find the optimal encoding in a DRAM address bus [19]. The problem is proven to be NP-complete with respect to the number of edges for general directed graphs [19]. For the de Bruijn graph $G(\mathbf{B}(n))$, the Euler cycle of lexicographically minimal labels can be constructed in $O(E)$ time where E denotes the number of edges in $G(\mathbf{B}(n))$ [21]. Before this paper, it was not known if the lexicographically minimal Euler cycle can be constructed similarly in $O(E)$ time for $G(\mathbf{B}_c^n(n))$.