Counting the Palstars

L. Bruce Richmond
Combinatorics and Optimization
University of Waterloo
Waterloo, ON N2L 3G1
Canada
lbrichmo@uwaterloo.ca

Jeffrey Shallit
School of Computer Science
University of Waterloo
Waterloo, ON N2L 3G1
Canada
shallit@cs.uwaterloo.ca

Submitted: Jun 12, 2014; Accepted: Jul 28, 2014; Published: Aug 13, 2014
Mathematics Subject Classifications: 05A05, 05A15, 05A16, 68R15

Abstract

A palstar (after Knuth, Morris, and Pratt) is a concatenation of even-length palindromes. We show that, asymptotically, there are $D_k \alpha_k^n$ palstars of length $2n$ over a k-letter alphabet, where D_k and α_k are positive constants with $2k-1 < \alpha_k < 2k - \frac{1}{2}$. In particular, $\alpha_2 \doteq 3.33513193$.

Keywords: palindrome, palstar, prime palstar unique factorization, generating function, enumeration.

1 Introduction

We are concerned with finite strings over a finite alphabet Σ_k having $k \geqslant 2$ letters. A *palindrome* is a string x equal to its reversal x^R, like the English word `radar`. If T, U are sets of strings over Σ_k then (as usual) $TU = \{tu \ : \ t \in T, u \in U\}$. Also $T^i = \overbrace{TT \cdots T}^{i}$ and $T^* = \bigcup_{i \geqslant 0} T^i$ and $T^+ = \bigcup_{i \geqslant 1} T^i$.

We define
$$P = \{x\, x^R \ : \ x \in \Sigma_k^+\},$$
the language of nonempty even-length palindromes. Following Knuth, Morris, and Pratt [3], we call a string x a *palstar* if it belongs to P^*, that is, if it can be written as the concatenation of elements of P. Clearly every palstar is of even length.

We call x a *prime palstar* if it is a nonempty palstar, but not the concatenation of two or more palstars; alternatively, if $x \in P^+ - P^2 P^*$ where $-$ is set difference. Thus, for example, the the English word `noon` is a prime palstar, but the English word `appall` and the French word `assailli` are palstars that are not prime. Knuth, Morris, and Pratt [3] proved that no prime palstar is a proper prefix of another prime palstar, and, consequently, every palstar has a unique factorization as a concatenation of prime palstars.

A nonempty string x is a *border* of a string y if x is both a prefix and a suffix of y and $x \neq y$. We say a string y is *bordered* if it has a border. Thus, for example, the English word `ionization` is bordered with border `ion`. Otherwise a word is *unbordered*. Rampersad et al. [6] recently gave a bijection between the unbordered strings of length n and the prime palstars of length $2n$. As a consequence they obtained a formula for the number of prime palstars.

Despite some interest in the palstars themselves [4, 1], it seems no one has enumerated them. Here we observe that the bijection mentioned previously, together with the unique factorization of palstars, provides an asymptotic enumeration for the number of palstars.

2 Generating function for the palstars

Again, let $k \geqslant 2$ denote the size of the alphabet. Let $p_k(n)$ denote the number of palstars of length $2n$, and let $u_k(n)$ denote the number of unbordered strings of length n.

Lemma 1. *For $n \geqslant 1$ and $k \geqslant 2$ we have*

$$p_k(n) = \sum_{1 \leqslant i \leqslant n} u_k(i) p_k(n-i).$$

Proof. Consider a palstar of length $2n > 0$. Either it is a prime palstar, and by [6] there are $u_k(n) = u_k(n)p_k(0)$ of them, or it is the concatenation of two or more prime palstars. In the latter case, consider the length of this first factor; it can potentially be $2i$ for $1 \leqslant i \leqslant n$. Removing this first factor, what is left is also a palstar. This gives $u_k(i)p_k(n-i)$ distinct palstars for each i. Since factorization into prime palstars is unique, the result follows. □

Now we define generating functions as follows:

$$\begin{aligned} P_k(X) &= \sum_{n \geqslant 0} p_k(n) X^n \\ U_k(X) &= \sum_{n \geqslant 0} u_k(n) X^n. \end{aligned}$$

The first few terms are as follows:

$$\begin{aligned} P_k(X) &= 1 + kX + (2k^2 - k)X^2 + (4k^3 - 3k^2)X^3 + (8k^4 - 8k^3 + k)X^4 + \cdots \\ U_k(X) &= 1 + kX + (k^2 - k)X^2 + (k^3 - k^2)X^3 + (k^4 - k^3 - k^2 + k)X^4 + \cdots . \end{aligned}$$

Theorem 2.

$$P_k(X) = \frac{1}{2 - U_k(X)}.$$

Proof. From Lemma 1, we have

$$\begin{aligned}
U_k(X)P_k(X) &= \left(\sum_{n\geqslant 0} u_k(n)X^n\right)\left(\sum_{n\geqslant 0} p_k(n)X^n\right) \\
&= 1 + \sum_{n\geqslant 1}\left(\sum_{0\leqslant i\leqslant n} u_k(i)p_k(n-i)\right)X^n \\
&= 1 + \left(\sum_{n\geqslant 1}\sum_{1\leqslant i\leqslant n} u_k(i)p_k(n-i)X^n\right) + \sum_{n\geqslant 1} p_k(n)X^n \\
&= 1 + \left(\sum_{n\geqslant 1} p_k(n)X^n\right) + \sum_{n\geqslant 1} p_k(n)X^n \\
&= 2P_k(X) - 1,
\end{aligned}$$

from which the result follows immediately. □

3 The main result

Theorem 3. *For all $k \geqslant 2$ there are positive constants D_k and α_k, with $2k-1 < \alpha_k < 2k - \frac{1}{2}$, such that the number of palstars of length $2n$ is asymptotically $D_k\alpha_k^n$.*

Proof. From Theorem 2 and the "First Principle of Coefficient Asymptotics" [2, p. 260], it follows that the asymptotic behavior of $[X^n]P_k(X)$, the coefficient of X^n in $P_k(X)$, is controlled by the behavior of the roots of $U_k(X) = 2$. Since $u_k(0) = 1$ and $U_k(X) \to \infty$ as $X \to \infty$, the equation $U_k(X) = 2$ has a single positive real root, which is $\rho = \rho_k = \alpha_k^{-1}$. We first show that $2k-1 < \alpha_k < 2k - \frac{1}{2}$.

Recalling that $u_k(n)$ is the number of unbordered strings of length n over a k-letter alphabet, we see that $u_k(n) \leqslant k^n - k^{n-1}$ for $n \geqslant 2$, since k^n counts the total number of strings of length n, and k^{n-1} counts the number of strings with a border of length 1. Similarly

$$u_k(n) \geqslant \begin{cases} k^n - k^{n-1} - \cdots - k^{n/2}, & \text{if } n \geqslant 2 \text{ is even;} \\ k^n - k^{n-1} - \cdots - k^{(n+1)/2}, & \text{if } n \geqslant 2 \text{ is odd,} \end{cases}$$

since this quantity represents removing strings with borders of lengths $1, 2, \ldots, n/2$ (resp., $1, 2, \ldots, (n-1)/2$) if n is even (resp., odd) from the total number. Here we use the classical fact that if a word of length n has a border, it has one of length $\leqslant n/2$.

It follows that for real $X > 0$ we have

$$\begin{aligned}
U_k(X) = \sum_{n\geqslant 0} u_k(n)X^n = 1 + kX + \sum_{n\geqslant 2} u_k(n)X^n &\leqslant 1 + kX + \sum_{n\geqslant 2}(k^n - k^{n-1})X^n \\
&= \frac{kX^2 - 1}{kX - 1}.
\end{aligned}$$

Similarly for real $X > 0$ we have

$$\begin{aligned}
U_k(X) &= \sum_{n\geqslant 0} u_k(n)X^n \\
&= 1 + kX + \sum_{l\geqslant 1} u_k(2l)X^{2l} + \sum_{m\geqslant 1} u_k(2m+1)X^{2m+1} \\
&\geqslant 1 + kX + \sum_{l\geqslant 1}(k^{2l} - k^{2l-1} - \cdots - k^l)X^{2l} + \sum_{m\geqslant 1}(k^{2m+1} - \cdots - k^{m+1})X^{2m+1} \\
&= \frac{1-2kX^2}{(kX-1)(kX^2-1)}.
\end{aligned}$$

This gives, for $k \geqslant 2$, that

$$2 < \frac{(2k-1)(4k^2-6k+1)}{(k-1)^2(4k-1)} \leqslant U_k\left(\frac{1}{2k-1}\right)$$

and

$$U_k\left(\frac{1}{2k-\frac{1}{2}}\right) \leqslant \frac{16k^2-12k+1}{(4k-1)(2k-1)} < 2.$$

It follows that $\frac{1}{2k-\frac{1}{2}} < \rho_k < \frac{1}{2k-1}$ and hence $2k-1 < \alpha_k < 2k-\frac{1}{2}$.

To understand the asymptotic behavior of $[X^n]P_k(X)$, we need to rule out other (complex) roots with the same magnitude as ρ. Here we follow the argument of the anonymous referee, which replaces our earlier and more complicated argument [7].

Suppose $X = -\rho$ is a solution. Then, since $u_k(n) > 0$ for all n, we have

$$2 = \sum_{n\geqslant 0} u_k(n)(-\rho)^n < \sum_{n\geqslant 0} u_k(n)\rho^n = 2.$$

This is a contradiction, so if there exists another solution it must be of the form

$$X = \rho e^{i\psi} = \rho(\cos n\psi + i\sin n\psi)$$

with $0 < \psi < 2\pi$. Since $2 - U_k(X) = 0$, the imaginary part of $U_k(\rho e^{i\psi})$ must equal 0. Therefore

$$\begin{aligned}
0 &= 2 - \sum_{n\geqslant 0} u_k(n)\rho^n \cos n\psi \\
&= 2 - \sum_{n\geqslant 0} u_k(n)\rho^n + \sum_{n\geqslant 0} u_k(n)\rho^n(1-\cos n\psi) \\
&= \sum_{n\geqslant 0} u_k(n)\rho^n(1-\cos n\psi).
\end{aligned}$$

Since $u_k(n) > 0$ and $\rho^n > 0$ for all n, we must have $1 - \cos n\psi = 0$ for all n. So for all n there exists an integer ℓ_n such that $n\psi = 2\pi\ell_n$. Hence

$$\psi = (n+1)\psi - n\psi = 2\pi\ell_{n+1} - 2\pi\ell_n = 2\pi(\ell_{n+1} - \ell_n).$$

Since ℓ_{n+1} and ℓ_n are both integers, this contradicts the assumption $0 < \psi < 2\pi$. Therefore $P_k(X)$ has only one singularity with $|X| = \rho$.

It remains to determine the order of the zero ρ. From above $U_k(X) = 2$ has a solution α_k^{-1} which satisfies $2k - 1 < \alpha_k < 2k - \frac{1}{2}$. Nielsen [5] showed that $u_k(n) = \Theta(k^n)$, and so $U_k(X)$ has radius of convergence $1/k$. Therefore $1/\alpha_k$ lies in the region where U_k is analytic. Hence $2 - U_k(X)$ is analytic at $1/\alpha_k$ and has a zero at $1/\alpha_k$ of multiplicity m. If $m \geqslant 2$, then the derivative of $2 - U_k(X)$ equals 0 at $X = 1/\alpha_k$. However $C_k := U_k'(\alpha_k^{-1}) > 0$ since $u_k(n) > 0$ for some n. Thus $2 - U_k(X)$ has a simple zero at $X = 1/\alpha_k$, and so $P_k(X)$ has a simple pole at $X = 1/\alpha_k$. Near α_k^{-1} the generating function $U_k(X)$ has the expansion $2 + C_k(X - \alpha_k^{-1}) + C_k'(X - \alpha_k^{-1})^2 + \cdots$. Furthermore

$$P_k(X) = \frac{1}{2 - U_k(X)} = \frac{1}{-C_k(X - 1/\alpha_k) - C_k'(X - 1/\alpha_k)^2 + \cdots}.$$

As we have seen,

$$P_k(X) - \frac{\alpha_k}{C_k}\left(\frac{1}{1 - \alpha_k X}\right)$$

has no singularity on the circle $|X| = 1/\alpha_k$, and so it has radius of convergence $> 1/\alpha_k$. Now

$$P_k'(X) = \frac{U_k'(X)}{(2 - U_k(X))^2},$$

so from standard results (e.g., [2, Thm. IV.7, p. 244]) there exists a positive δ such that

$$\begin{aligned}
[X^n]P_k(X) &= [X^n]\frac{1}{C_k(1/\alpha_k - X)} + \cdots \\
&= [X^n]\frac{\alpha_k}{C_k}\left(\frac{1}{1 - \alpha_k X}\right) + \cdots \\
&= \frac{\alpha_k^{n+1}}{C_k} + O\left((\alpha_k - \delta)^n\right).
\end{aligned}$$

Now, setting $D_k = \alpha_k / C_k$ completes the proof. □

4 Numerical results

Here is a table giving the first few values of $P_k(n)$.

$n =$	0	1	2	3	4	5	6	7	8	9	10
$k = 2$	1	2	6	20	66	220	732	2440	8134	27124	90452
$k = 3$	1	3	15	81	435	2349	12681	68499	370023	1998945	10798821
$k = 4$	1	4	28	208	1540	11440	84976	631360	4690972	34854352	258971536

By truncating the power series $U_k(X)$ and solving the equation $U_k(X) = 2$ we get better and better approximations to α_k^{-1}. For example, for $k = 2$ we have

$$\begin{aligned}
\alpha_2^{-1} &\doteq 0.2998382135935269050615511181457960391930318236478173036633919933306520 2\\
\alpha_2 &\doteq 3.335131930033579367667896261037624484236327063440561157710444730851186 0\\
C_2 &\doteq 6.278652437421018217684895562492005276088368718322063642652328654828673\\
D_2 &\doteq 0.531185945276419575719915203572875899822069417373160261548729841 7
\end{aligned}$$

To determine an asymptotic expansion for α_k as $k \to \infty$, we compute the Taylor series expansion for $P_k(n)/P_k(n+1)$, treating k as an indeterminate, for n large enough to cover the error term desired. For example, for $O(k^{-10})$ it suffices to take $k = 16$, which gives

$$\alpha_k^{-1} = \frac{1}{2k} + \frac{1}{8k^2} + \frac{3}{32k^3} + \frac{1}{16k^4} + \frac{27}{512k^5} + \frac{93}{2048k^6} + \frac{83}{2048k^7} + \frac{155}{4096k^8} + \frac{4735}{131072k^9} + O(k^{-10})$$

and hence

$$\alpha_k = 2k - \frac{1}{2} - \frac{1}{4k} - \frac{3}{32k^2} - \frac{5}{64k^3} - \frac{31}{512k^4} - \frac{25}{512k^5} - \frac{23}{512k^6} - \frac{683}{16384k^7} + O(k^{-8}).$$

Acknowledgment

We thank the anonymous referee for his/her very helpful report.

References

[1] Z. Galil and J. Seiferas. A linear-time on-line recognition algorithm for “palstar”. *J. ACM* **25** (1978), 102–111.

[2] P. Flajolet and R. Sedgewick. *Analytic Combinatorics.* Cambridge University Press, 2009.

[3] D. E. Knuth, J. Morris, and V. Pratt. Fast pattern matching in strings. *SIAM J. Comput.* **6** (1977), 323–350.

[4] G. Manacher. A new linear-time “on-line” algorithm for finding the smallest initial palindrome of a string. *J. ACM* **22** (1975), 346–351.

[5] P. T. Nielsen. A note on bifix-free sequences. *IEEE Trans. Info. Theory* **IT-19** (1973), 704–706.

[6] N. Rampersad, J. Shallit, and M.-w. Wang. Inverse star, borders, and palstars. *Info. Proc. Letters* **111** (2011), 420–422.

[7] L. B. Richmond and J. Shallit. Counting the palstars. Preprint, June 11 2014, available at `arXiv:1311.2318`.