ACTA ARITHMETICA
109.3 (2003)

On cluster primes

by

CHRISTIAN ELSHOLTZ (Clausthal-Zellerfeld)

1. Introduction. Blecksmith, Erdős and Selfridge [1] defined a prime $p > 2$ to be a *cluster prime* if every positive even integer $2r \le p-3$ can be written as a difference of two primes, $2r = q - q'$, where $q' \le q \le p$. It is an open question whether there exist infinitely many cluster primes. Guy ([4, Section C1]) attributes this question to Erdős. The attention of the general audience was drawn to this problem by Peterson's article [6] in Science News.

Blecksmith *et al.* [1] proved that the counting function $\pi_C(x)$ of cluster primes can be bounded from above: for all positive s,

$$\pi_C(x) = O_s\left(\frac{x}{(\log x)^s}\right).$$

It is the purpose of this note to prove a better bound, i.e. that cluster primes are rare. This new bound was indeed conjectured by Blecksmith *et al.* [1].

THEOREM. *The number $\pi_C(x)$ of cluster primes below x is bounded by*

$$\pi_C(x) = O\left(\frac{x}{\exp\left(\frac{1}{60}(\log\log x)^2\right)}\right).$$

As Blecksmith, Erdős and Selfridge show, the problem is related to the prime k-tuple conjecture. It is proved that for a cluster prime p the interval $[p-t, p)$ must contain sufficiently many primes, which explains the name cluster prime. This allows us to apply an upper bound sieve. In Blecksmith *et al.* [1], Brun's version of the small sieve is used. The principal problem is that the authors arrive at a constant M whose dependence on the sieve dimension s is not at all clear. This prohibits taking an increasing s.

Filaseta [3] mentioned that an application of Hooley's almost pure sieve proves the result with $s = \varepsilon \log\log\log x$, thus obtaining an upper bound of

$$\pi_C(x) = O\left(\frac{x}{\exp(a\log\log x\log\log\log x)}\right) \quad \text{for some positive constant } a.$$

2000 *Mathematics Subject Classification*: Primary 11N05, 11N36; Secondary 11A41.
Key words and phrases: prime tuples, sieve methods.

In this note we apply the large sieve method, due to Montgomery [5]. In fact, we make use of the following lemma due to Vaughan [7], which is an elaborated version of the large sieve method, perfectly fitting to our application.

LEMMA 1 (Montgomery [5], Vaughan [7]). *Denote by $\mathcal{P}$ the set of primes and let $\omega : \mathcal{P} \to \mathbb{N}$ with $0 \le \omega(p) \le p-1$. Let $\mathcal{A} \subset [1,N]$ denote a set of integers which lies outside $\omega(p)$ residue classes modulo the prime p. Then the number $A(x)$ of elements $n \in \mathcal{A}$ with $n \le x$ satisfies*

$$A(x) \le \frac{2x}{L}, \quad \textit{where} \quad L = \sum_{q \le x^{1/2}} \mu^2(q) \prod_{p \mid q} \frac{\omega(p)}{p-\omega(p)}.$$

Moreover,

$$L \ge \max_{m \in \mathbb{N}} \exp\left(m \log\left(\frac{1}{m} \sum_{p \le x^{1/(2m)}} \frac{\omega(p)}{p} \right)\right).$$

2. Proof of the Theorem. If p is a cluster prime, then the even integers like $p-9$ or $p-15$ are the differences of two primes q, q' with $q, q' \le p$. In particular, there must be a prime in the interval $[p-6, p)$. More generally, an even integer $2r \in [p-t, p-3]$ must be represented by a prime $q \in [p-t, p]$ and a prime $q' \in [1, t]$. By the prime number theorem the number of primes in $[1,t]$ is $(1+o(1))t/\log t$. We see that for any $\varepsilon > 0$ there must be at least $s := (1/2-\varepsilon)\log t$ primes in $[p-t, p)$.

Since the average gap between primes of size x is about $\log x$ we see that this is a useful criterion for $t = O((\log x)^\delta)$ (with $0 < \delta < 1$). On the contrary, for sufficiently large t one expects that an interval of length t has about $t/\log t$ primes so that this criterion becomes useless.

There are (trivially) at most $\binom{t}{s}$ possibilities to place s primes in an interval of length t. For any pattern of s primes in $[p-t, p)$ we will give an upper bound on the number of prime s-tuples below x. This bound will not depend on the particular pattern. So, multiplying this bound by the upper bound for the number of patterns, $\binom{t}{s}$, gives an upper bound on the number of $p \le x$ such that the interval $[p-t, p)$ contains (at least) s primes.

We prove the following lemma:

LEMMA 2. *Let $\delta = 1/(2e)$ and $t = (\log x)^\delta$. Let ε be a sufficiently small positive constant. Let $A(x)$ denote the number of integers $n \le x$ such that the interval $[n-t, n)$ contains at least $s = (1/2-\varepsilon)\log t$ primes. Then*

$$A(x) = O\left(\frac{x}{\exp((1/(4e^2)-\varepsilon)(\log\log x)^2)} \right).$$

We fix a particular pattern $a_1 < \ldots < a_s$. If all $n - a_i$ are prime simultaneously, then the integers n avoid the residue classes $a_i \bmod p$ for $p \le n - t$. If $p > t$, then the number of forbidden classes is $\omega(p) = s$.

We choose $m = \lceil \delta^2 (\log\log x)^2 \rceil$, where x is large. With $y = x^{1/(2m)}$ we have $\log\log y = \log\log x - 2\log\log\log x + O(1)$. So we find that

$$\begin{aligned}\sum_{p \le y} \frac{\omega(p)}{p} &\ge \sum_{t \le p \le y} \frac{\omega(p)}{p} \ge s(\log\log y - \log\log t + o(1))\\ &\ge (1/2 - \varepsilon)\delta(\log\log x)(\log\log x - 3\log\log\log x + O(1))\\ &\ge (1/2 - 2\varepsilon)\delta(\log\log x)^2.\end{aligned}$$

This implies the estimate

$$\begin{aligned}L &\ge \exp\left(\lceil \delta^2(\log\log x)^2\rceil \log\left(\frac{1}{\lceil \delta^2(\log\log x)^2\rceil}\,\delta\left(\frac{1}{2} - 2\varepsilon\right)(\log\log x)^2\right)\right)\\ &\ge \exp\left(\delta^2(\log\log x)^2 \log\left(\frac{1}{\delta}\left(\frac{1}{2} - 3\varepsilon\right)\right)\right)\\ &\ge \exp(\delta^2(\log\log x)^2\log(e - 6e\varepsilon)) \ge \exp(\delta^2(1 - 7\varepsilon)(\log\log x)^2)\\ &\ge \exp\left(\left(\frac{1}{4e^2} - \varepsilon\right)(\log\log x)^2\right).\end{aligned}$$

Therefore, for any fixed pattern $a_1 < \ldots < a_s$ there are at most

$$\frac{2x}{\exp((1/(4e^2) - \varepsilon)(\log\log x)^2)}$$

values $n \le x$ such that all $n - a_i$ are prime. Thus the lemma is proved.

To prove the theorem we only need to recall that

$$\pi_C(x) \le \binom{t}{s}\frac{2x}{\exp(\delta^2(1 - 7\varepsilon)(\log\log x)^2)}.$$

Because of

$$\binom{t}{s} \le t^s \le \exp((1/2 - \varepsilon)\delta^2(\log\log x)^2)$$

we find that

$$\pi_C(x) = O\left(\frac{x}{\exp((1/(8e^2) - \varepsilon)(\log\log x)^2)}\right).$$

3. Further comments. No serious attempt has been made at optimizing the constant $1/60$ or $1/(8e^2) - \varepsilon$ that appears in the Theorem. Some improvement is possible. We only mention the following: Vaughan's argument in Lemma 1 can be refined to

$$L \ge \max_{m \in \mathbb{N}} \exp\left(m \log\left(\frac{e - \varepsilon_m}{m} \sum_{p \le x^{1/(2m)}} \frac{\omega(p)}{p}\right)\right).$$

Here the ε_m are positive constants that tend to 0 as m goes to infinity. This allows using $c_1 \approx \delta/2$ and $\delta \approx 1/2$ and proves the Theorem with $1/8 - \varepsilon$ instead of $1/60$. For details see [2].

The author would like to thank the referees for helpful comments.

References

[1] R. Blecksmith, P. Erdős and J. L. Selfridge, *Cluster primes*, Amer. Math. Monthly 106 (1999), 43–48.
[2] C. Elsholtz, *Combinatorial prime number theory—a study of the gap structure of the set of primes*, Habilitationsschrift, Technische Universität Clausthal, 2002.
[3] M. Filaseta, Review of [1] in the Mathematical Reviews: 2000a:11126.
[4] R. K. Guy, *Unsolved Problems in Number Theory*, 2nd ed., Springer, New York, 1994.
[5] H. L. Montgomery, *The analytic principle of the large sieve*, Bull. Amer. Math. Soc. 84 (1978), 547–567.
[6] I. Peterson, *The scarcity of cluster primes*, Science News 6, February 1999, p. 95.
[7] R. C. Vaughan, *Some applications of Montgomery's sieve*, J. Number Theory 5 (1973), 64–79.

Institut für Mathematik
TU Clausthal
Erzstrasse 1
D-38678 Clausthal-Zellerfeld, Germany
E-mail: elsholtz@math.tu-clausthal.de

Received on 27.5.2002
and in revised form on 26.2.2003 (4298)