J. Math. Crypt. **1** (2007), 329–349

DOI 10.1515 / JMC.2007.016

Statistical distribution and collisions of VSH

Ian F. Blake and Igor E. Shparlinski

Communicated by Phong Q. Nguyen

Abstract. The distributional properties and the collision probability for the hash function VSH recently introduced by Contini, Lenstra and Steinfeld, are studied. The study leads to some interesting number theoretic questions which have apparently not been studied. These and related questions on VSH are considered in this work.

Keywords. Hash function, distribution, collision.

AMS classification. 11N69,11T71,11Y16.

1 Introduction

1.1 Motivation

The *Very Smooth Hash* function, VSH, recently introduced by S. Contini, A. Lenstra and R. Steinfeld [8] works as follows. Let p_i denote the ith prime number and let

$$Q_k = \prod_{i=1}^{k} p_i$$

denote the product of the first k primes and set

$$R_k = Q_{k+1}.$$

Assume that integers k and n satisfy

$$Q_k < n \le R_k. \tag{1.1}$$

Let $\ell < 2^k$, the message length, be a positive integer whose k-bit representation (including all leading zeros) is $\ell = \lambda_1 \dots \lambda_k$ that is

$$\ell = \sum_{i=1}^{k} \lambda_i 2^{i-1}.$$

Then VSH takes an ℓ-bit message $m = \mu_1 \dots \mu_\ell$ and hashes it (in a very efficient way, via a simple iterative procedure) to

$$h_n(m) \equiv \prod_{i=1}^{k} p_i^{e_i} \pmod{n}, \qquad 0 \le h_n(m) < n, \tag{1.2}$$

where $L = \lceil \ell/k \rceil$, $\mu_s = 0$, for $\ell < s \le Lk$, $\mu_{Lk+i} = \lambda_i$, for $1 \le i \le k$, and

$$e_i = \sum_{j=0}^{L} \mu_{jk+i} 2^{L-j}, \qquad i = 1, \ldots, k.$$

It is demonstrated in [8] that h_n is very efficient and also admits a rigorous proof of security against collisions, which is based on some natural number theoretic problems which are presumed to be hard. The above problem is related to factoring, thus it is natural to choose n to be a product of two large primes.

One can roughly separate all known hash functions into two categories, although there are hash functions that are not so clearly categorized (see [15, 19]):

- Hash functions which are based on various Boolean operations and whose design resembles art more than science. Such functions are usually very fast but have no proofs of security behind them.
- Hash functions which are based on various algebraic structures. Such functions are usually much slower but admit at least conditional security proofs.

Thus, the invention of VSH has narrowed the gap between the efficiency of functions from these families. The effort to obtain algebraic hash functions with a security proof related to a mathematical problem, believed to be hard, is felt to be a significant area for further work and has motivated us to consider VSH in greater detail.

Other efforts in this direction include the work of Charles, Goren and Lauter [6] who define a hash function based upon expander graphs. In one of their constructions, the collision resistance depends upon the difficulty of computing isogenies between supersingular elliptic curves. At this point it appears that all such hash functions known to the authors are very slow in comparison to the non-algebraic ones, although VSH is promising.

Several interesting properties and potential applications of VSH h_n have already been outlined in [8].

It has been noticed in [8] that if the factorization of n is known then for any message m one can easily find a second image $\widetilde{m}$ with $h_n(\widetilde{m}) = h_n(m)$. It is also indicated in [8] that indeed finding collisions is equivalent to either factoring n or solving a discrete logarithm problem modulo n. Thus if the factorization of the modulus n is known then collisions can be created by determining the Euler function $\varphi(n)$, since adding a multiple of $\varphi(n)$ to any of the exponents of the small primes, in the hash representation of equation (1.2) results in the same hash value. Other potential weaknesses of VSH are noted in both [8] and [18] and some of these are discussed later.

It is natural that such algebraic hash functions have parameter sets which exhibit potential weaknesses. A similar situation has existed for such cryptographic systems such as RSA and elliptic curves where a significant effort over many years has been expended on precisely the same question of determining suitable parameter sets for security. Our intention here is to continue the examination of VSH with a view to further understanding its properties and to promote its development.

1.2 Our results

To give some rigorous support towards the security and distribution properties of VSH, we show that for almost all n which are products of two primes and any a, the probability that for a random ℓ-bit message m, for sufficiently large ℓ, we have $h(m) \equiv a \pmod{n}$, is negligible. This bounds the collision probability and also the probability of finding a second preimage by brute force. Our results are based on the study of the multiplicative group $\mathcal{H}_n$ defined by:

$$\mathcal{H}_n = \{\langle p_1, \ldots, p_k\rangle \subseteq {\mathbb{Z}_n}^*\}, \tag{1.3}$$

the subgroup of ${\mathbb{Z}_n}^*$ generated by $p_1, \ldots, p_k$. We consider the class $\mathcal{N}_k$ of integers $n = pq$ where p and q are distinct primes, satisfying the inequality (1.1) that is,

$$\mathcal{N}_k = \{n = pq \mid p,\ q \text{ distinct primes},\ \prod_{i=1}^{k} p_i < n \leq \prod_{i=1}^{k+1} p_i\}.$$

Some well-known number theoretic techniques are used to study the probability of collisions. As a by-product, we also show that $\mathcal{H}_n$ is very "massive" for almost all $n \in \mathcal{N}_k$.

Note that VSH enjoys a type of homomorphic property in that if, using the representation of equation (1.2), we define the function

$$\begin{array}{rccccc} \eta: & \mathbb{F}_2^{\ell} & \longrightarrow & \mathbb{Z}_{\geq 0}^{k} & \longrightarrow & \mathbb{Z} \\ & m & \mapsto & (e_1, e_2, \ldots, e_k) & \mapsto & \prod_{i=1}^{k} p_i^{e_i}, \end{array}$$

then this map is an injection and we have

$$h_n(\eta^{-1}(\eta(m_1)\eta(m_2))) \equiv h_n(m_1)h_n(m_2) \pmod{n}. \tag{1.4}$$

We apply this property in Section 4 in an attempt to create collisions. It is obvious that if a message is found that hashes to $1 \pmod{n}$, it allows the creation of collisions for any given message, in an obvious manner. Other aspects of this map are also considered there.

We discuss various approaches to creating collisions which lead to some interesting number theoretic questions. Although our analysis shows that most of them do not seem to present a significant threat, they may possibly be modified for creating trap doors or indicate potentially weak parameters or for other purposes.

In Section 2 we collect some known results about prime numbers (in particular, about the distribution of prime numbers in arithmetic progressions) and also about elements of $\mathcal{N}_k$. We also prove one new result which could be of independent interest. These fundamental results are repeatedly used throughout the paper.

In Section 3, we present some properties of the VSH map (1.2), which can be of cryptographic significance (in particular they help to rule out several straightforward attacks on VSH) and also lead to some interesting number theoretic questions of intrinsic value.

Finally, in Section 4 other approaches to creating collisions in VSH that might be attempted are considered, if only to determine possible weak choices of parameters or to rule the approaches out as infeasible, by obtaining the relevant estimates. In turn, these estimates are based on the results and ideas of Section 3. Since it is already known [8] that finding collisions is equivalent to solving a certain presumably hard problem, direct weaknesses in VSH are not expected.

We also pose some open questions and have some comments on the efficiency of VSH.

It should be noted that although our results are asymptotic in nature, all constants are effective and can be explicitly evaluated. However we doubt the value of such an exercise since our estimates are of theoretic interest and should be considered as just indications that VSH has no hidden weaknesses which could affect its characteristics for large values of parameters. They also indicate that certain naive ways of attacking VSH are bound to fail, but do not imply any lower bounds on the strength of VSH (in the latter case obtaining concrete values for the constant involved would certainly be of great importance).

1.3 Notation

Throughout the paper, the implied constants in symbols 'O' and '$\ll$' may occasionally, where obvious, depend on a positive parameter ε, and are absolute otherwise (we recall that $U \ll V$ and $U = O(V)$ are both equivalent to the inequality $|U| \le cV$ with some constant $c > 0$).

For an integer n, the residue ring modulo n is denoted by $\mathbb{Z}_n$, and the group of invertible elements is denoted by $\mathbb{Z}_n^*$. We always assume that $\mathbb{Z}_n$ is represented by the set $\{0, 1, \ldots, n-1\}$.

As usual, we use $\pi(x)$ to denote the number of primes $p \le x$ and use $\pi(x, d, a)$ to denote the number of primes $p \le x$ in the arithmetic progression $p \equiv a \pmod d$ (although the ordering of the variables in this function is not standard).

We also denote by $\omega(s)$ and $\varphi(s)$ the number of distinct prime divisors and the Euler function, respectively, of an integer $s \ge 2$ (with the usual convention that $\omega(1) = 0$).

Recall that an integer s is called *y-smooth* if all prime divisors of s do not exceed y. As usual, the number of y-smooth positive integers $s \le x$ is denoted by $\psi(x, y)$.

The letters p and q always denote prime numbers.

Finally, for any real number $x > 0$, we define $\log x = \max\{\ln x, 1\}$ (where $\ln x$ is the natural logarithm of x).

2 Analytic number theory background

2.1 Distribution of primes

We recall the *Prime Number Theorem*, PNT, in its simplest form which is sufficient for our purposes, see [20, Section II.4.1],

$$\pi(x) = (1 + o(1))\frac{x}{\log x}, \tag{2.1}$$

as $x \to \infty$.

It is also useful to recall that by the PNT (2.1), we have

$$R_k = \exp\left((1 + o(1))k \log k\right). \tag{2.2}$$

Thus, as has been noticed in [8], we derive from (1.1) and (2.2) that

$$k = (1 + o(1))\frac{\log n}{\log \log n} \tag{2.3}$$

for any $n \in \mathcal{N}_k$.

We also need the *Brun–Titchmarsh bound*, see [20, Section I. 4.6] which asserts a tight upper bound on the number of primes in an arithmetic progression

$$\pi(x, d, a) \ll \frac{x}{\varphi(d) \log(x/d)} \tag{2.4}$$

uniformly over x, d and a. Using this bound and partial summation one can derive that

$$\sum_{\substack{p \le x \\ p \equiv 1 \pmod d}} \frac{1}{p} \ll \frac{\log \log x}{\varphi(d)}, \tag{2.5}$$

(also uniformly over x and d), for example, see the bound (3.1) in [10].

The *Bombieri–Vinogradov theorem*, see [7, Theorem 9.2.1], asserts that for any $A > 0$ there exists some constant B such that for a sufficiently large x

$$\sum_{d \le x^{1/2}(\log x)^{-B}} \max_{\gcd(a,d)=1} \max_{z \le x} \left|\pi(z, d, a) - \frac{\pi(z)}{\varphi(d)}\right| \le x(\log x)^{-A}. \tag{2.6}$$

Also, we need to appeal to the following two famous conjectures.

The *Elliott-Halberstam conjecture*, see [20, Notes, Chapter II.8], suggests that the summation can in fact be extended over all $d \le x^{1-\varepsilon}$ for any $\varepsilon > 0$.

The *prime s-tuplets conjecture*, see [9, Conjecture 1.2.1], asserts that for any number s of linear forms $a_i r + b_i$, $i = 1, 2, \ldots, s$, with integers a_i and b_i such that $a_i > 0$, $\gcd(a_i, b_i) = 1$, such that for each prime $p \le s$ there is an integer m such that p does not divide $\prod_{i=1}^{s}(a_i m + b_i)$, then there are infinitely many integers n such that the $a_i n + b_i$, $i = 1, 2, \ldots, s$ are simultaneously prime. Moreover, it is natural to assume that the number of such values $r \le x$ is about $cx(\log x)^{-s}$ where c depends only on the linear forms.

Although there is little doubt in the correctness of these conjectures, it needs to be noted that neither of them follows from even the *Extended Riemann Hypothesis*.

2.2 Bounds on some arithmetic functions

Since for any integer $s \ge 1$ we have $\omega(s)! \le s$, by the Stirling formula we obtain

$$\omega(s) \ll \frac{\log s}{\log \log s}, \tag{2.7}$$

see [20, Section I.5.3].

Finally, we recall the following well known bound

$$\varphi(s) \gg \frac{s}{\log\log s}, \tag{2.8}$$

see [20, Section I.5.4].

2.3 Distribution of smooth numbers

It is well known that in a very wide range of parameters x and y, namely, for $x \ge y \ge (\log x)^{1+\varepsilon}$, we have

$$\psi(x,y) = xu^{-u+o(u)} \tag{2.9}$$

where

$$u = \frac{\log x}{\log y},$$

see [5]. However here we are mostly interested in smaller values of y which are outside of the range of applicability of (2.9) since in our case y is of order $\log x$. We remark that (8) in [20, Section III.5.1] with $y = \alpha \log x$ reduces to

$$\begin{aligned} Z &= \frac{\log x}{\log(\alpha) + \log\log x}\log(1+\alpha) + \frac{\alpha\log x}{\log(\alpha)+\log\log x}\log(1+1/\alpha) \\ &= (g(\alpha)+o(1))\frac{\log x}{\log\log x} \end{aligned}$$

for any fixed $\alpha > 0$, where

$$g(\alpha) = \log(1+\alpha) + \alpha\log(1+1/\alpha).$$

Accordingly, by [20, Theorem 2, Section III.5.1], we have

$$\psi(x, \alpha\log x) = \exp\left((g(\alpha)+o(1))\frac{\log x}{\log\log x}\right). \tag{2.10}$$

Note that $g(1) = 2\log 2$. Therefore since, by (2.3) and (2.1) we have that

$$p_k = (1+o(1))k\log k = (1+o(1))\log n,$$

the bound (2.10) with $\alpha = 1$ implies that

$$\psi(n, p_k) = \exp\left((2\log 2 + o(1))\frac{\log n}{\log\log n}\right). \tag{2.11}$$

Many more useful bounds on $\psi(x,y)$ can be found in [12, 13, 20].

2.4 Properties of the set $\mathcal{N}_k$

It follows immediately from the classical results about the distribution of integers with a given number of prime factors (see [20, Section II.6.1]) that

$$\#\mathcal{N}_k = (1+o(1))\frac{R_k \log\log R_k}{\log R_k}. \tag{2.12}$$

For an integer $T \geq 1$, let $\mathcal{N}_k(T)$ be the set of $n \in \mathcal{N}_k$ for which

$$\prod_{i=1}^{k} p_i^{e_i} \equiv \prod_{i=1}^{k} p_i^{f_i} \pmod{n} \tag{2.13}$$

for some distinct integer vectors $(e_1, \ldots, e_k)$ and $(f_1, \ldots, f_k)$ with

$$|e_i - f_i| < T, \qquad i = 1, \ldots, k, \tag{2.14}$$

that is,

$$\mathcal{N}_k(T) = \{n = pq \in \mathcal{N}_k \mid \exists (e_1, \ldots, e_k), (f_1, \ldots, f_k) \in \mathbb{Z}^k, \text{ satisfying (2.13) and (2.14)}\}.$$

The following result is similar to several more results of this type which are well-known in the literature (for example, see [14, 16]).

Lemma 2.1. *For an integer $T \geq 1$, we have*

$$\#\mathcal{N}_k(T) \ll \frac{(2T)^{2k+2}(\log k)^2}{(\log T)^2}.$$

Proof. We consider the product

$$W = \prod_{\substack{x_1,\ldots,x_k=0 \\ x_1+\ldots+x_k>0}}^{T-1} \prod_{\mathcal{I},\mathcal{J}} \left(\prod_{i\in\mathcal{I}} p_i^{x_i} - \prod_{j\in\mathcal{J}} p_j^{x_j} \right)$$

where $\mathcal{I}$, $\mathcal{J}$ run over all 2^k disjoint partitions of the set $\{1, \ldots, k\}$; that is, $I \cup J = \{1, \ldots, k\}$ and $I \cap J = \emptyset$.

Clearly, any $n = pq \in \mathcal{N}_k(T)$ is a divisor of W, in particular

$$\#\mathcal{N}_k(T) \leq \omega(W)^2.$$

Thus we see from (2.7) that it only remains to estimate W.

By the PNT (2.1), each term of the above product satisfies

$$\left| \prod_{i\in\mathcal{I}} p_i^{x_i} - \prod_{j\in\mathcal{J}} p_j^{x_j} \right| \leq \left(\prod_{i=1}^{k} p_i \right)^{T} = \exp((1+o(1))kT\log k).$$

Hence

$$W \le \exp\left((1+o(1))k2^k T^{k+1}\log k\right).$$

Putting everything together, we derive

$$\#\mathcal{N}_k(T) \le \omega(W)^2 \ll \left(\frac{\log W}{\log\log W}\right)^2 \ll \frac{(2T)^{2k+2}(\log k)^2}{(\log T)^2}$$

which concludes the proof. □

3 Some properties of the VSH map

3.1 Preimage size

Let $M_n(\ell, a)$ denote the number of ℓ-bit messages m with $h_n(m) = a$.

Let

$$\rho_n(\ell) = \max_{1\le a\le n} M_n(\ell, a)2^{-\ell}$$

be the largest probability that a random ℓ-bit message m has the hash value $h_n(m) = a$. Note that $\rho_n(\ell) \ge 1/n$ for any ℓ.

Theorem 3.1. *For any s such that $s \le \ell - 2k$ and $k \to \infty$, the inequality*

$$\rho_n(\ell) < 2^{-s}$$

holds for all but at most $2^{10k+2s+o(k+s)}$ values of $n \in \mathcal{N}_k$.

Proof. We can assume that k is large enough. Let us fix $n \in \mathcal{N}_k$ and an integer a. Each ℓ-bit message m gives rise to a vector $\mathbf{e} = (e_1, \ldots, e_k)$ in the k-dimensional cube $[0, 2^{L+1} - 1]^k$, where $L = \lceil \ell/k \rceil$. Moreover distinct messages correspond to distinct vectors $\mathbf{e}$.

We put

$$T = \left\lceil 2^{s/k+3} \right\rceil.$$

It is clear that the cube $[0, 2^{L+1} - 1]^k$ can be covered by at most

$$\left(2^{L+1}T^{-1} + 1\right)^k \le \left(2^{L-s/k-2} + 1\right)^k \le 2^{(L-s/k-1)k} < 2^{\ell-s}$$

aligned unit cubes with the side length T.

Assume that $M_n(\ell, a) \ge 2^{\ell-s}$ for some $a \in \mathbb{Z}_n$. In this case, we see that at least one of them contains two vectors $\mathbf{e}$ and $\mathbf{f}$ corresponding to two ℓ-bit messages m_1 and m_2 with $h_n(m_1) = h_n(m_2) = a$, which leads to relations (2.13) and (2.14). So, the inequality $\rho_n(\ell) \le 2^{-s}$ is possible for at most $\#\mathcal{N}_k(T)$ values of $n \in \mathcal{N}_k$. Using Lemma 2.1, we obtain

$$\begin{aligned}\#\mathcal{N}_k(T) &\ll \frac{(2T)^{2k+2}(\log k)^2}{(\log T)^2} \le \frac{2^{(s/k+5)(2k+2)}(\log k)^2}{(\log T)^2}\\ &= 2^{10k+2s+O(s/k+\log\log k)} = 2^{10k+2s+o(k+s)},\end{aligned}$$

which concludes the proof. □

Thus we see from (2.12) that for $s = o(k \log k)$ the preimage size is exponentially smaller than the total number of ℓ-bit messages for the overwhelming majority of $n \in \mathcal{N}_k$.

3.2 Random collision probability

Let

$$\vartheta_n(\ell) = 2^{-2\ell} \sum_{a=0}^{n-1} M_n(\ell, a)^2$$

be the probability of a random collision $h_n(m_1) = h_n(m_2)$ over all ℓ-bit messages m_1 and m_2.

Using Theorem 3.1 and the identity

$$\sum_{a=0}^{n-1} M_n(\ell, a) = 2^{\ell}$$

we derive the following.

Corollary 3.2. *For any s such that $s \le \ell - 2k$ and $k \to \infty$, the inequality*

$$\vartheta_n(\ell) < 2^{-s}$$

holds for all but at most $2^{10k+2s+o(k+s)}$ values of $n \in \mathcal{N}_k$.

Since $\vartheta_n(\ell) \le 1$ for any n we have that the average probability of a random collision of ℓ-bit messages over all $n \in \mathcal{N}_k$ is

$$\frac{1}{\#\mathcal{N}_k} \sum_{n \in \mathcal{N}_k} \vartheta_n(\ell) \le \frac{1}{\#\mathcal{N}_k} \left(2^{2s+o(s)} + 2^{-s} \#\mathcal{N}_k\right)$$

for any s with $s/k \to \infty$. Defining s by the inequalities

$$2^{3s} \le \#\mathcal{N}_k < 2^{3(s+1)}$$

and remarking that $s/k \gg \log k \to \infty$ as $k \to \infty$, we obtain:

Corollary 3.3. *For $\ell \ge k \log k$, we have*

$$\frac{1}{\#\mathcal{N}_k} \sum_{n \in \mathcal{N}_k} \vartheta_n(\ell) \le \#\mathcal{N}_k^{-1/3+o(1)}, \qquad k \to \infty.$$

Taking

$$s = \left\lceil \frac{\log R_k}{4 \log 2} \right\rceil \sim \frac{k \log k}{4 \log 2}$$

in Corollary 3.2, we conclude that even deterministic factoring (in time $n^{1/4+o(1)}$, see [9, Section 5.5]) is a faster way to attack then finding collisions by brute force for the overwhelming majority of $n \in \mathcal{N}_k$.

Corollary 3.4. *We have,*

$$\#\{n \in \mathcal{N}_k \ : \ \vartheta_n(\ell) \ge n^{-1/4}\} \le \#\mathcal{N}_k^{-\kappa+o(1)}, \qquad k \to \infty,$$

where

$$\kappa = \frac{1}{2\log 2} = 0.72134\ldots .$$

In particular, Corollaries 3.3 and 3.4 mean that at least asymptotically, for almost all $n \in \mathcal{N}_k$ finding collisions via brute force search is much slower than factoring the modulus n. However, it is conceivable that there is a way of choosing messages of special structure (for example, short messages) which may increase the probability of collision.

3.3 The cardinality of $\mathcal{H}_n$

Recall the definition of $\mathcal{H}_n$ given by (1.3)

Clearly, for every $n \in \mathcal{N}_k$ we have $\#\mathcal{H}_n \ge \psi(n, p_k)$ and thus from (2.11) we immediately derive

$$\#\mathcal{H}_n \ge \exp\left((2\log 2 + o(1))\frac{\log n}{\log\log n}\right).$$

Here we show that for almost all $n \in \mathcal{N}_k$ the group $\mathcal{H}_k$ is quite "massive".

Theorem 3.5. *The inequality*

$$\#\mathcal{H}_n \ge \frac{n}{(\log n)^2(\log\log n)^5}\exp\left(-4\sqrt{\log 2\log n}\right)$$

holds for all but at most $o(\#\mathcal{N}_k)$ *values of* $n \in \mathcal{N}_k$.

Proof. We say that a prime $p \le R_k$ is *exceptional* if the first k primes generate a subgroup $\mathcal{H}_p$ of size at most

$$\#\mathcal{H}_p \le \frac{p}{\log p(\log\log p)^2}\exp\left(-2\sqrt{\log 2\log p}\right).$$

It follows immediately from a more general Theorem 1.5 of [16] that for any $t \le R_k$ there are at most $o(\pi(t))$ exceptional primes $p \le t$.

Let $L_k = \lceil \log R_k \rceil$. Then we see that the cardinality of the set $\mathcal{E}_k$ of $n \in \mathcal{N}_k$ which

are divisible by an exceptional prime does not exceed

$$\begin{aligned}
\#\mathcal{E}_k &\le \sum_{\substack{p\le R_k\\ p \text{ exceptional}}} \sum_{q\le R_k/p} 1 = \sum_{\substack{p\le R_k\\ p \text{ exceptional}}} \pi(R_k/p)\\
&\ll R_k \sum_{\substack{p\le R_k\\ p \text{ exceptional}}} \frac{1}{p\log(R_k/p)}\\
&= R_k \sum_{i=0}^{L_k} \sum_{\substack{R_k e^{-i-1}<p\le R_k e^{-i}\\ p \text{ exceptional}}} \frac{1}{p\log(R_k/p)}\\
&\ll \sum_{i=1}^{L_k} \frac{e^i}{i} \sum_{\substack{R_k e^{-i-1}<p\le R_k e^{-i}\\ p \text{ exceptional}}} 1 = o(\sigma_k),
\end{aligned}$$

where

$$\begin{aligned}
\sigma_k &\le \sum_{i=1}^{L_k} \frac{e^i}{i}\pi(R_k e^{-i}) \ll R_k \sum_{i=1}^{L_k} \frac{1}{i\log(R_k/e^i)}\\
&\ll R_k \sum_{i=1}^{L_k} \frac{1}{i(L_k-i+1)} \ll R_k \sum_{1\le i\le L_k/2} \frac{1}{i(L_k-i+1)}\\
&\ll \frac{R_k}{L_k} \sum_{1\le i\le L_k/2} \frac{1}{i} \ll \frac{R_k\log L_k}{L_k} \ll \frac{R_k\log\log R_k}{\log R_k} \ll \#\mathcal{N}_k,
\end{aligned}$$

see (2.12). Thus $\#\mathcal{E}_k = o(\sigma_k) = o(\#\mathcal{N}_k)$.

Let $D_k = \lfloor \log\log Q_k \rfloor$. We define $\mathcal{F}_k$ as the set of $n = pq \in \mathcal{N}_k$ with $\gcd(p-1, q-1) \ge D_k$. Then, using the Brun–Titchmarsh bound (2.4), we derive

$$\begin{aligned}
\#\mathcal{F}_k &\le \sum_{d\ge D_k} \sum_{\substack{p\le\sqrt{R_k}\\ p\equiv 1 \pmod d}} \sum_{\substack{p\le q\le R_k/p\\ q\equiv 1 \pmod d}} 1 \le \sum_{d\ge D_k} \sum_{\substack{p\le\sqrt{R_k}\\ p\equiv 1 \pmod d}} \pi(R_k/p, d, 1)\\
&\ll R_k \sum_{d\ge D_k} \sum_{\substack{p\le\sqrt{R_k}\\ p\equiv 1 \pmod d}} \frac{1}{p\varphi(d)\log(R_k/pd)}\\
&\ll \frac{R_k}{\log R_k} \sum_{d\ge D_k} \frac{1}{\varphi(d)} \sum_{\substack{p\le\sqrt{R_k}\\ p\equiv 1 \pmod d}} \frac{1}{p}.
\end{aligned}$$

Now, using (2.5) and then (2.8) we derive

$$\#\mathcal{F}_k \ll \frac{R_k\log\log R_k}{\log R_k} \sum_{d\ge D_k} \frac{1}{\varphi(d)^2} \ll \frac{R_k\log\log R_k(\log\log D_k)^2}{D_k\log R_k}. \tag{3.1}$$

Thus $\#\mathcal{F}_k = o(\#\mathcal{N}_k)$.

Finally, for any $n = pq \in \mathcal{N}_k \setminus (\mathcal{E}_k \cup \mathcal{F}_k)$ we have

$$\#\mathcal{H}_n \ge \frac{\#\mathcal{H}_p \#\mathcal{H}_q}{D_k} \ge \frac{n}{D_k (\log n)^2 (\log\log n)^4} \exp\left(-4\sqrt{\log 2 \log n}\right).$$

Remarking that

$$D_k \le \log\log Q_k \le \log\log n,$$

we conclude the proof. □

Certainly, the power of $\log\log n$ in the bound of Theorem 3.5 can be improved.

We remark that the lower bound of Theorem 3.5 on the size of $\mathcal{H}_n$, combined with well known results about exponential sums over finitely generated subgroups in residue rings (for example, see [14, Theorem 3.4]), implies that the elements of $\mathcal{H}_n$ are very uniformly distributed modulo n for almost all $n \in \mathcal{N}_k$. For instance, for any $\varepsilon > 0$, the statistical distance between binary vectors formed by about a half of the most significant bits of $h \in \mathcal{H}_n$ and random binary vectors of the same dimension is exponentially small.

As before we observe that

$$\#\mathcal{H}_p \ge \psi(p, p_k) \qquad \text{and} \qquad \#\mathcal{H}_q \ge \psi(q, p_k)$$

are distinct modulo p. In the case where $n = pq$ for primes $p = 2ur+1$ and $q = 2vs+1$, r and s distinct primes and

$$2u < \psi(p, p_k) \qquad \text{and} \qquad 2v < \psi(q, p_k)$$

then since $\#\mathcal{H}_p > 2u$ and $\#\mathcal{H}_p \mid p - 1 = 2ur$ then $r \mid \#\mathcal{H}_p \mid \#\mathcal{H}_n$. Similarly, $s \mid \#\mathcal{H}_q \mid \#\mathcal{H}_n$ and hence

$$\mathcal{H}_n \ge rs.$$

It is useful to note that, by using $\alpha = 2$ in (2.11), for $\log p \sim \log q \sim 0.5 \log n$ we have

$$\psi(p, p_k) = \exp\left((0.5 \log(27/4) + o(1)) \frac{\log n}{\log\log n}\right).$$

A similar bound also holds for $\psi(q, p_k)$.

4 Approaches to creating collisions

4.1 General observations

Several problems with VSH in terms of creating collisions have been noted in [8] and [18]. For example, it is noted in [8] that if the message m is short then the hash value

$$h_n(m) = \prod_{i=1}^{k} p_i^{e_i} \pmod n$$

may not "wrap around" modulo n and hence hash inversion (retrieving the message from the hash value) is possible. It is also noted that it is easy to obtain messages m and m' for which $h_n(m) = 2h_n(m') \pmod{n}$. Such problems are easily overcome by, for example, a sufficient number of squarings to ensure wrap around for any message length.

An interesting property of VSH noted in [18] is that if $a,\ b,\ z$ are bit strings of equal length, with z the all-zero string, such that $a \wedge b = z$ then

$$h_n(a \wedge b) \equiv h_n(a)h_n(b) \pmod{n}.$$

This property is used to construct an effective time-memory trade-off attack on VSH. It is also shown there how a partial collision attack, where certain bits of the hash function are constrained to have certain values, can be turned into a full collision attack.

In both [8] and [18] it is noted that VSH produces a relatively long hash and effective methods for shortening it would be needed for certain applications. As well, the fact VSH has a trapdoor (factorization of n) creates certain problems in applications.

Our aim in this section is to consider other potential problems with creating collisions and obtain estimates for their likelihood of occurring.

4.2 False witnesses

In P. Erdős and C. Pomerance [11], an integer $a \in \mathbb{Z}_n$ is called a *false witness* (also called a *Fermat liar* [15]) for an integer $n \in \mathbb{Z}$ if

$$a^{n-1} \equiv 1 \pmod{n} \tag{4.1}$$

(see also [9]). In case (4.1) does not hold, a is called a *witness*, since it immediately implies that n is composite.

Then, much as for adding multiples of $\varphi(n)$ to exponents of a prime in the case the factorization of n is known, in the hash representation of the equation (1.2), adding multiples of $n-1$ to the exponent of any prime p_i, $i = 1, \ldots, k$, that is a false witness modulo n, does not change the hash and hence corresponds to creating a collision. For a given set of small primes, one can think of moduli n which make a given small prime a false witness, as a *weak key*.

Note that if $p-1$ and $q-1$ have only small prime divisors then it may not be too difficult to determine the order of a given small prime, by brute force, and in a similar manner, create collisions.

We recall that n is called a *Fermat pseudoprime* to base a if n is composite and (4.1) holds (see [9]). Pomerance [17] obtained an upper bound on the number of $n \leq x$ which are Fermat pseudoprimes to base $a = 2$. The method can easily be extended to an arbitrary base a and can even give a uniform bound with respect to a. However for $n \in \mathcal{N}_k$ it simplifies and leads to a stronger bound. Let us consider the set

$$\mathcal{W}_k = \{n \in \mathcal{N}_k \mid \text{at least one of } p_1, \ldots, p_k \text{ is a false witness for } n\}.$$

Theorem 4.1. *We have*

$$\#\mathcal{W}_k \ll (\#\mathcal{N}_k)^{3/4+o(1)}.$$

Proof. Let us fix some integer $a \in [2, Q_k^{1/2}]$ and let $t_a(p)$ denote the multiplicative order of a modulo a prime p with $\gcd(a,p) = 1$. We follow the same arguments as in the proof of [17, Theorem 2] which however we adapt to $n \in \mathcal{N}_k$. Clearly, if $n = pq \in \mathcal{N}_k$ and (4.1) holds then

$$n \equiv q \equiv 1 \pmod{t_a(p)}.$$

Thus for every p there are at most $R_k/pt_a(p)$ possibilities for q (even ignoring the fact that q is prime). We can also assume that $p > n^{1/2} \ge Q_k^{1/2}$. Therefore, the number of $n \in \mathcal{N}_k$ which are Fermat pseudoprimes to base a, does not exceed

$$W_{k,a} \le R_k \sum_{R_k \ge p > Q_k^{1/2}} \frac{1}{pt_a(p)} \le \sum_{t \le R_k} \frac{1}{t} \sum_{\substack{R_k \ge p > Q_k^{1/2} \\ t_a(p) = t}} \frac{1}{p}.$$

As in the proof of [17, Theorem 1] we conclude from (2.7) that there are at most

$$J = \omega(a^t - 1) \ll \frac{t \log a}{\log(t \log a)} \ll \frac{t \log a}{\log t}$$

primes p with $t_a(p) = t$, and they all satisfy $p \equiv 1 \pmod t$. Therefore, using (2.5), we obtain

$$\sum_{\substack{R_k \ge p > Q_k^{1/2} \\ t_a(p) = t}} \frac{1}{p} \le \sum_{\substack{p \le p_J \\ p \equiv 1 \pmod t}} \frac{1}{p} \ll \frac{\log\log p_J}{\varphi(t)} \ll \frac{\log\log(t \log a)}{\varphi(t)}.$$

Finally, the bound (2.8) implies that

$$\sum_{\substack{R_k \ge p > Q_k^{1/2} \\ t_a(p) = t}} \frac{1}{p} \ll \frac{(\log\log t + \log\log\log a)\log\log t}{t}. \tag{4.2}$$

We also have the following trivial bound:

$$\sum_{\substack{R_k \ge p > Q_k^{1/2} \\ t_a(p) = t}} \frac{1}{p} \le \frac{\omega(a^t - 1)}{Q_k^{1/2}} \ll \frac{t \log a}{Q_k^{1/2} \log t}. \tag{4.3}$$

Let us fix some sufficiently large T. Using (4.3) for $t \le T$ and (4.2) for $t \ge T$, we obtain

$$\begin{aligned} W_{k,a} &\ll R_k \left(\frac{\log a}{Q_k^{1/2}} \sum_{t \le T} \frac{1}{\log t} + \sum_{t > T} \frac{(\log\log t + \log\log\log a)\log\log t}{t^2} \right) \\ &\ll R_k \left(\frac{T \log a}{Q_k^{1/2} \log T} + \frac{(\log\log T + \log\log\log a)\log\log T}{T} \right). \end{aligned}$$

We now choose

$$T = Q_k^{1/4}(\log Q_k)^{1/2}(\log a)^{-1/2}\log\log Q_k$$

getting

$$W_{k,a} \ll R_k Q_k^{-1/4}(\log Q_k)^{-1/2}(\log a)^{1/2}\left(\log\log Q_k + \log\log\log a\right).$$

Therefore

$$\#\mathcal{W}_k \le \sum_{j=1}^{k} W_{k,p_j} \ll kR_kQ_k^{-1/4}(\log Q_k)^{-1/2}(\log k)^{1/2}\log\log Q_k$$

and using (2.12), (2.2) and (2.3), we conclude the proof. □

One easily verifies that if for an integer r both $p = 6r+1$ and $q = 12r+1$ are prime then, since 3 is a quadratic residue modulo q, we have

$$3^{6r} \equiv 1 \pmod{pq}.$$

Hence 3 is a false witness modulo $n = pq = 18r(4r+1)+1$. Therefore any reasonably quantitative form of the prime s-tuplets conjecture implies that $\#\mathcal{W}_k \gg (\#\mathcal{N}_k)^{1/2+o(1)}$ which in fact can be the right order of growth of $\#\mathcal{W}_k$.

While, as Theorem 4.1 shows, it is easy to avoid the possibility a prime p_i is a false witness modulo n, it does point out that some care is required in choosing the primes. Note that it is simple to test for a prime being a false witness for a given $n \in \mathbb{Z}$.

Rather than consider a single small prime being a false witness, one may try to consider products

$$a = \prod_{i=1}^{k} p_i^{d_i} \in \mathcal{H}_n \tag{4.4}$$

for some randomly chosen integers $d_1, \dots, d_k$ in a hope that at least one of these products is a false witness for $n \in \mathcal{N}_k$ which leads to a question of estimating the probability of this event.

For every positive integer n, the total number of false witnesses is given by

$$\gamma(n) = \prod_{p \mid n} \gcd(n-1, p-1) \tag{4.5}$$

and has been studied in the literature (see [11] and the references therein). For example, the average value, the normal order, and some other properties of $\gamma(n)$ are studied in [11].

If $n = pq \in \mathcal{N}_k$ then by (4.5) we see that

$$\begin{aligned}\gamma(n) &= \gcd(pq-1, p-1)\gcd(pq-1, q-1)\\ &= \gcd(p-1, q-1)^2.\end{aligned} \tag{4.6}$$

Thus, as in the proof of Theorem 3.5 (see the bound (3.1) in particular) we see that the number $E_k(\Gamma)$ of $n \in \mathcal{N}_k$ with $\gamma(n) \ge \Gamma$ is at most

$$E_k(\Gamma) \ll \#\mathcal{N}_k \Gamma^{-1/2} \log\log \Gamma. \tag{4.7}$$

We now show that although for almost all n this quantity is rather small, on average over $n \in \mathcal{N}_k$ it is quite large (but probably not large enough to become a real weakness).

Theorem 4.2. *We have*

$$R_k^{1/2+o(1)} \gg \frac{1}{\#\mathcal{N}_k} \sum_{n \in \mathcal{N}_k} \gamma(n) \gg R_k^{1/4+o(1)}.$$

Proof. To prove the upper bound we use (4.7) and also note that

$$\sum_{n \in \mathcal{N}_k} \gamma(n) = \sum_{\Gamma=1}^{R_k} \left(E_k(\Gamma) - E_k(\Gamma+1)\right)\Gamma \le \sum_{\Gamma=1}^{R_k} E_k(\Gamma).$$

To prove the upper bound we fix an arbitrary $\varepsilon > 0$ and define $y = R_k^{1/4-\varepsilon}$. Using the Bombieri–Vinogradov theorem (2.6) with $x = \sqrt{R_k}$ and $A = 3$ we see that

$$\pi(\sqrt{R_k}, r, 1) \ge \frac{\pi(\sqrt{R_k})}{2\varphi(r)} \tag{4.8}$$

holds for all but at most $O(y(\log x)^{-2})$ primes $r \in [y, 2y]$. Let $\mathcal{R}$ be set of prime numbers in the interval $[y, 2y]$ for which we have (4.8). Then we have

$$\#\mathcal{R} = \pi(2y) - \pi(y) + O(y(\log x)^{-2}) = (1 + o(1))y(\log y)^{-1}. \tag{4.9}$$

For $r \in \mathcal{R}$, we now consider the set $\mathcal{P}_r$ of primes $p \in [\sqrt{Q_k}, \sqrt{R_k}]$ such that $p \equiv 1 \pmod r$ and $p \not\equiv 1 \pmod s$ for any other $s \in \mathcal{R}$. We have

$$\#\mathcal{P}_r = \pi(\sqrt{R_k}, r, 1) - \pi(\sqrt{Q_k}, r, 1) + O\left(\sum_{s \in \mathcal{R}} \pi(\sqrt{R_k}, rs, 1)\right).$$

Using the Brun–Titchmarsh theorem (2.4), we obtain

$$\begin{aligned}\sum_{s \in \mathcal{R}} \pi(\sqrt{R_k}, rs, 1) &\ll \sum_{s \in \mathcal{R}} \frac{\sqrt{R_k}}{rs \log(\sqrt{R_k}/rs)} \le \frac{\pi(2y)\sqrt{R_k}}{ry \log(\sqrt{R_k}/y^2)} \\ &\ll \frac{\sqrt{R_k}}{r \log R_k \log y},\end{aligned}$$

and also

$$\pi(\sqrt{Q_k}, r, 1) \ll \frac{\sqrt{Q_k}}{r \log Q_k}.$$

Therefore, recalling (4.8), we deduce

$$\#\mathcal{P}_r = \pi(\sqrt{R_k}, r, 1) + O\left(\frac{\sqrt{R_k}}{r \log R_k \log y} + \frac{\sqrt{Q_k}}{r \log Q_k}\right).$$

Hence

$$\#\mathcal{P}_r \geq \frac{\pi(\sqrt{R_k})}{3\varphi(r)}, \tag{4.10}$$

provided that k is large enough.

Clearly the sets $\mathcal{P}_r$, $r \in \mathcal{R}$, are disjoint. Also for two distinct primes $p, q \in \mathcal{P}_r$ we have $n = pq \in \mathcal{N}_k$ and also we see from (4.6) that $\gamma(pq) \geq r^2$. Hence,

$$\begin{aligned}\frac{1}{\#\mathcal{N}_k} \sum_{n \in \mathcal{N}_k} \gamma(n) &\geq \frac{1}{\#\mathcal{N}_k} \sum_{r \in \mathcal{R}} r^2 \frac{\#\mathcal{P}_r(\#\mathcal{P}_r - 1)}{2} \\ &\gg \frac{\pi(\sqrt{R_k})^2 \#\mathcal{R}}{\#\mathcal{N}_k} \gg \frac{R_k \#\mathcal{R}}{\#\mathcal{N}_k \log R_k}.\end{aligned}$$

Now using the bounds (2.12) and (4.9) together with the fact that ε is arbitrary we conclude the proof. □

We note that P. Erdős and C. Pomerance [11] have shown that the average value of $\gamma(n)$ over all composite $n \leq x$ is much higher:

$$x \exp\left(-(1 + o(1)) \frac{\log x \log\log\log x}{\log\log x}\right) \geq \frac{1}{x} \sum_{\substack{n \leq x \\ n \text{ composite}}} \gamma(n) \gg x^{15/23}$$

and they conjecture that the upper bound is in fact tight. Moreover, using modern bounds [2] on shifted primes without large prime divisors the exponent $15/23 = 0.6521\ldots$ can be replaced with 0.7039.

We also believe that the upper bound of Theorem 4.2 gives the correct order of magnitude of the average value of $\gamma(n)$ over $n \in \mathcal{N}_k$. Furthermore, under some standard number theoretic conjectures (such as the Elliott-Halberstam conjecture or the prime s-tuplets conjecture) this indeed can be proven.

In any case the bound (4.7) and Theorem 4.2 show that the set of false witnesses is very small for most of $n \in \mathcal{N}_k$. Together with Theorem 3.5 we can now conclude that for a "typical" $n \in \mathcal{N}_k$ the proportion of false witnesses inside of $\mathcal{H}_n$ is negligible. So random products (4.4) are very unlikely to generate a false witness.

Several very interesting results about the distribution of witnesses and false witnesses can be found in [1, 4, 3].

Slightly more general than the form of (4.1) and (4.4) to find collisions, one can ask about a possibility of more general relations involving all primes $p_1, \ldots, p_k$ simultaneously:

Open Question 4.3. Given k polynomials $f_1(X), \ldots, f_k(X) \in \mathbb{Z}[X]$, determine the size of the set

$$\{n \in \mathcal{N}_k \mid \prod_{i=1}^{k} p_i^{f_i(n)} \equiv 1 \pmod{n}\}.$$

4.3 Creating preimages for a given hash value

It is noted in [8] and [18] that by multiplying the hash value of a message by the inverse modulo n of the power of a small prime or product of a collection of small primes, it may be possible to determine the hash of a modified (hopefully meaningful) message, without knowing the message. This might pose a problem in some circumstances.

A simple extension of this observation is the following. Choose a (fairly large) set of t binary sequences and their corresponding hash values:

$$m_j \in \mathbb{F}_2^{\ell} \sim (e_1^{(j)}, e_2^{(j)}, \ldots, e_k^{(j)}) \mapsto a_j \in \mathbb{Z}_n, \qquad j = 1, 2, \ldots, t.$$

Then it is easy to find the preimage of any product of powers of the hash values:

$$\begin{aligned} \prod_{j=1}^{t} a_j^{f_j} \pmod{n} &\mapsto \sum_{j=1}^{t} f_j(e_1^{(j)}, \ldots, e_k^{(j)}) \\ &\mapsto \prod_{i=1}^{k} p_i^{\sum_{j=1}^{t} f_j e_i^{(j)}} \mapsto m \in \mathbb{F}_2^{\ell_0} \end{aligned}$$

for some overall message length ℓ_0. This is a simple application of the hash homomorphism (1.4). Thus preimages of a great many hash values can be created. To find a collision for a given hash value with this approach however, would seem to require the ability to express the given hash value multiplicatively in terms of the given set of hash values, modulo n, whose preimages are known. This appears to be a difficult problem, implying it is unlikely a collision could be found with this approach. However, the approach remains one of interest for further consideration.

4.4 A smooth numbers approach to creating collisions

As noted previously, the homomorphic property of VSH in (1.4), implies that any preimage of $1 \in \mathbb{Z}$ leads to the easy creation of collisions for any (message,hash) pair. Suppose $a \in \mathbb{Z}_n$ is a hash obtained from a message corresponding to exponents e_i, $i = 1, 2, \ldots, k$ in the representation (1.2)

$$a \equiv \prod_{i=1}^{k} p_i^{e_i} \pmod{n}. \tag{4.11}$$

If a is viewed as an element of $\mathbb{Z}$ (rather than in $\mathbb{Z}_n$), and if it is smooth in $\mathbb{Z}$ with respect to p_k, so that in addition to equation (4.11) we also have

$$a = \prod_{i=1}^{k} p_i^{d_i} \in \mathbb{Z}$$

then (as noted previously) the right hand side of this last equation can be multiplied by inverses of the $p_i^{e_i} \pmod{n}$ to determine a message with a hash of unity modulo n which would be disastrous. It is thus of interest to determine the likelihood that the hash of a randomly chosen message (of any length) is smooth with respect to p_k. Fortunately this probability is easily shown to be negligible. We note that by (2.11) the number of p_k-smooth positive integers $b < n$ is a negligible quantity compared to the total number n of such integers.

It is also natural to assume that the special shape (4.11) does not change this probability in a substantial way, however a rigorous proof of this is not immediate. This leads us to the more general:

Open Question 4.4. Given a subgroup $\mathcal{H} \in {\mathbb{Z}_n}^*$ and a real $y > 0$, estimate how many elements of $\mathcal{H}$ are y-smooth.

5 Comments

Certain aspects of VSH have been shown to involve interesting number theoretic questions and some of these have been pursued in this work to obtain estimates relevant to the security of VSH. Many other interesting questions remain and we note a few of these here.

The VSSR (and VSDL) assumptions of [8] are new and interesting ones and central to the security arguments for VSH. Further work to determine greater insight and a more precise understanding of their complexity and their relation to other standard computational problems, such as factoring and modular square roots, would be of great interest.

The efficiency of VSH is a critical factor in determining its adoption. Any techniques that improve on the iterative techniques given in [8] would be of value in promoting VSH.

To conclude, some natural questions arising in considering the security of VSH have led to interesting number theoretic and implementation questions. It is hoped the estimates obtained in this work related to these questions support the security arguments for VSH and motivate further interest on this system.

Acknowledgments. This paper was initiated during a very enjoyable visit of I. S. at the Department of Electrical and Computer Engineering at the University of Toronto whose hospitality and support are gratefully appreciated. Research of I. B. was supported in part by NSERC grant 7382 and that of I. S. by ARC grant DP0556431. The authors are also very grateful to two reviewers of the paper who provided useful and illuminating comments that greatly improved the presentation.

References

[1] W. R. Alford, A. Granville, and C. Pomerance, *On the difficulty of finding reliable witnesses*. Algorithmic Number Theory, First International Symposium, Lect. Notes in Comp. Sci. 877, pp. 1–16. Springer-Verlag, Berlin, 1994.

[2] R. C. Baker and G. Harman, *Shifted primes without large prime factors*, Acta Arith. 83 (1998), pp. 331–361.

[3] R. J. Burthe, *The average witness is 2*, Acta Arith. 80 (1997), pp. 327–341.

[4] ———, *Upper bounds for the least witnesses and generating sets*, Acta Arith. 80 (1997), pp. 311–326.

[5] E. R. Canfield, P. Erdős, and C. Pomerance, *On a problem of Oppenheim concerning "Factorisatio Numerorum"*, J. Number Theory 17 (1983), pp. 1–28.

[6] D. Charles, E. Goren, and K. Lauter, *Cryptographic hash functions from expander graphs*, J. Cryptology, to appear.

[7] A. Cojocaru and M. R. Murty, *An Introduction to Sieve Methods and their Applications*, LMS Student Texts 66. Cambridge Univ. Press, Cambridge, 2006.

[8] S. Contini, A. K. Lenstra, and R. Steinfeld, *VSH, an efficient and provable collision-resistant hash function*. Advances in Cryptology – Eurocrypt 2006, Lect. Notes in Comp. Sci. 4004, pp. 165–182. Springer-Verlag, Berlin, 2006.

[9] R. Crandall and C. Pomerance, *Prime Numbers: A Computational Perspective*. Springer, New York, 2005.

[10] P. Erdős, A. Granville, C. Pomerance, and C. Spiro, *On the normal behaviour of the iterates of some arithmetic functions*, Analytic Number Theory, Birkhäuser, Boston, 1990, pp. 165–204.

[11] P. Erdős and C. Pomerance, *On the number of false witnesses for a composite number*, Math. Comp. 46 (1986), pp. 259–279.

[12] A. Granville, *Smooth numbers: Computational number theory and beyond*. Proc. MSRI Conf. Algorithmic Number Theory: Lattices, Number Fields, Curves, and Cryptography, Berkeley 2000. Cambridge Univ. Press, (to appear).

[13] A. Hildebrand and G. Tenenbaum, *Integers without large prime factors*, J. de Théorie des Nombres de Bordeaux 5 (1993), pp. 411–484.

[14] S. V. Konyagin and I. E. Shparlinski, *Character Sums with Exponential Functions and their Applications*. Cambridge Univ. Press, Cambridge, 1999.

[15] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone, *Handbook of Applied Cryptography*. CRC Press, Boca Raton, FL, 1996.

[16] F. Pappalardi, *On the order of finitely generated subgroups of* $\mathbb{Q}^*$ $(\mathrm{mod}\ p)$ *and divisors of* $p-1$, J. Number Theory 57 (1996), pp. 207–222.

[17] C. Pomerance, *On the distribution of pseudoprimes*, Math. Comp. 37 (1981), pp. 587–593.

[18] M.-J. O. Saarinen, *Security of VSH in the real world*. Progress in Cryptology – INDOCRYPT 2006, Lect. Notes in Comp. Sci. 4329, pp. 95–103. Springer-Verlag, Berlin, 2006.

[19] D. R. Stinson, *Cryptography: Theory and Practice*. CRC Press, Boca Raton, FL, 2006.

[20] G. Tenenbaum, *Introduction to Analytic and Probabilistic Number Theory*. Cambridge Univ. Press, 1995.

Received 6 July, 2006; revised 31 March, 2007

Author information

Ian F. Blake, Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada.
Email: ifblake@comm.toronto.edu

Igor E. Shparlinski, Department of Computing, Macquarie University, Sydney, NSW 2109, Australia.
Email: igor@comp.mq.edu.au