involve
a journal of mathematics

Statistics for fixed points of the self-power map

Matthew Friedrichsen and Joshua Holden

msp

2019 vol. 12, no. 1

INVOLVE 12:1 (2019)
dx.doi.org/10.2140/involve.2019.12.63

Statistics for fixed points of the self-power map

Matthew Friedrichsen and Joshua Holden

(Communicated by Anant Godbole)

The map $x \mapsto x^x$ modulo p is related to a variation of the ElGamal digital signature scheme in a similar way as the discrete exponentiation map, but it has received much less study. We explore the number of fixed points of this map by a statistical analysis of experimental data. In particular, the number of fixed points can in many cases be modeled by a binomial distribution. We discuss the many cases where this has been successful, and also the cases where a good model may not yet have been found.

1. Introduction and motivation

The security of the ElGamal digital signature scheme against selective forgery relies on the difficulty of solving the congruence $g^{H(m)} \equiv y^r r^s \pmod{p}$ for r and s, given m, g, y, and p but not knowing the discrete logarithm of y modulo p to the base g. (We assume for the moment the security of the hash function $H(m)$.) Similarly, the security of a certain variation of this scheme given in, e.g., [Menezes et al. 1997, Note 11.71] relies on the difficulty of solving

$$g^{H(m)} \equiv y^s r^r \pmod{p}. \tag{1}$$

It is generally expected that the best way to solve either of these congruences is to calculate the discrete logarithm of y, but this is not known to be true. In particular, another possible option would be to choose s arbitrarily and solve the relevant equation for r. In the case of (1), this boils down to solving equations of the form $x^x \equiv c \pmod{p}$. We will refer to these equations as "self-power equations", and we will call the map $x \mapsto x^x$ modulo p the "self-power map". This map has been studied in various forms in [Anghel 2013; 2016; Balog et al. 2011; Cilleruelo and Garaev 2016a; 2016b; Crocker 1966; 1969; Somer 1981; Holden 2002a; 2002b; Holden and Moree 2006; Friedrichsen et al. 2010; Holden and Robinson 2012;

MSC2010: primary 11Y99; secondary 11-04, 11T71, 94A60, 11A07, 11D99.
Keywords: self-power map, exponential equation, ElGamal digital signatures, fixed point, random map, number theory.

Kurlberg et al. 2015]. In this work we will investigate experimentally the number of fixed points of the map, i.e., solutions to

$$x^x \equiv x \pmod{p} \tag{2}$$

between 1 and $p-1$. In particular, we would like to know whether the distribution across various primes behaves as we would expect if the self-power map were a "random map". We do this by creating a model in which values of a map are assumed to occur uniformly randomly except as forced by the structure of the self-power map. We can then predict the distribution of the number of fixed points of this random map and compare it statistically to the actual self-power map. If there is "nonrandom" structure in the self-power map, it may be possible to exploit that structure to break the signature scheme mentioned above or others like it.

In this paper, we will give a general heuristic (based on Heuristic 1 below) for the number of fixed points of the self-power map and show that for most cases it appears to accurately predict the behavior of the map. The outlying cases mostly appear to involve elements with order d that are relatively small or large compared to p. We will first show that the number of fixed points for elements with orders 1, 2, $p-1$, and $(p-1)/2$ can be predicted exactly. For other small orders which largely don't follow the general heuristic, we specifically look at the orders 3, 4, and 6 and give a separate model for them. For large orders, we make predictions for the orders $(p-1)/3$ and $(p-1)/4$.

Some theoretical work has also been done on bounding the possible number of fixed points of the self-power map. If we denote the number of solutions to (2) which fall between 1 and $p-1$ by $F(p)$, then we have:

Theorem 1.1 [Cilleruelo and Garaev 2016b, Corollary 2]. *For some absolute constant $c > 0$,*

$$F(p) \leq p^{1/3-c+o(1)}$$

as $p \to \infty$.

Remark 1. The corollary in [Cilleruelo and Garaev 2016b] is more general and puts a bound on the number of solutions for $x^{f(x)} \equiv 1 \pmod{p}$ for any nonconstant polynomial in $\mathbb{Z}[x]$ without multiple roots in $\mathbb{C}$.

Remark 2. In the related case of solutions to $x^x \equiv 1 \pmod{p}$, [Cilleruelo and Garaev 2016a] shows that the exponent can be taken to be $\frac{27}{82} + o(1)$ and that is likely also the case here.

As far as a lower bound, every p has at least $x = 1$ as a solution to (2), and at least some primes have only this solution. However, while [Kurlberg et al. 2015; Felix and Kurlberg 2017] give good reason to believe that there are infinitely many such primes, they also prove that these primes are fairly rare:

Theorem 1.2 [Felix and Kurlberg 2017, Corollary 1.2]. *Let $\pi(N)$ be the number of primes less than or equal to N as usual. Let $\mathcal{A}(N)$ denote the set of primes less than or equal to N such that $F(p) = 1$. Then*

$$\#\mathcal{A}(N) \le \frac{\pi(N)}{(\ln\ln\ln N)^{1-1/e+o(1)}}$$

as $N \to \infty$.

2. Models and experimental results

Heuristics and normality. Theorem 1.1 gives us a range in which the number of fixed points $F(p)$ can lie, but does not say anything about the distribution of the values within that range. As described above, our goal is to create a random model for the self-power map much like was done for the discrete exponential map in [Holden 2002a; 2002b; Holden and Moree 2006]. Our first attempt assumed that $F(p)$ was normally distributed around the predicted value $\sum_{d\,|\,(p-1)} \phi(d)/d$. (The normality assumption had been successfully used for the discrete exponential map in, e.g., [Cloutier and Holden 2010]; see also [Holden and Lindle 2008]. Furthermore, it appeared to be justified by the central limit theorem, given the number of primes we were intending to test.)

In order to calculate the variance of $F(p)$, we use the following heuristic, which is related to those in [Holden and Moree 2006, Section 6], and can also be derived from the assumptions in [Kurlberg et al. 2015, Section 4.1].

Heuristic 1. The map $x \mapsto x^x \mod p$ is a random map in the sense that for all p, if x, y are chosen uniformly at random from $\{1, \dots, p-1\}$ with $\operatorname{ord}_p x = d$, then

$$\Pr[x^x \equiv y \pmod p] \approx \begin{cases} 1/d & \text{if } \operatorname{ord}_p y \mid d, \\ 0 & \text{otherwise.} \end{cases}$$

As some justification, one can use the methods of [Holden and Robinson 2012, Corollary 6.2] to prove the following lemma. This shows that the heuristic holds exactly over the range $1 \le x \le (p-1)p$ rather than $1 \le x \le p-1$:

Lemma 2.1. *For all p, given fixed $d \mid (p-1)$ and fixed $y \in \{1, \dots, (p-1)p\}$, $p \nmid y$, such that $\operatorname{ord}_p y \mid d$, we have*

$$\#\left\{x \in \{1, \dots, (p-1)p\} : p \nmid x,\ x^x \equiv y \pmod p,\ \operatorname{ord}_p x = d\right\} = (p-1)\frac{\phi(d)}{d}.$$

Similar methods are used in [Holden et al. 2016] to prove the following theorem:

Theorem 2.2 [Holden et al. 2016, Corollary 4]. *Let $G(p)$ be the number of solutions to (2) with $1 \le x \le (p-1)p$ and $p \nmid x$. Then*

$$G(p) = (p-1) \sum_{n\,|\,(p-1)} \frac{\phi(n)}{n}.$$

For more on the self-power map over the range $1 \le x \le (p-1)p$, see [Somer 1981, Theorem 1; Holden and Robinson 2012, Sections 6 and 7; Holden et al. 2016].

As far as using Heuristic 1, note that it implies that the "experiment" of testing whether x is a fixed point behaves as a Bernoulli trial. Let $F_d(p)$ be the number of solutions to (2) with $1 \le x \le p-1$ and $\operatorname{ord}_p x = d$. Assuming independence of the Bernoulli trials (which is not completely accurate, as we shall see), $F_d(p)$ is distributed as a binomial random variable with $\phi(d)$ trials and success probability $1/d$. (We denote by $\phi(d)$ the Euler ϕ function and it occurs here because it gives the number of elements with order d when $d \mid (p-1)$.)

This distribution has mean $\phi(d)/d$, as expected, and variance $\phi(d)(d-1)/d^2$. Summing over $d \mid (p-1)$ gives the predicted mean and variance of $F(p)$.

We tested the hypothesis that $F(p)$ was normal with this mean and variance by collecting data for 16,405 primes from 100,003 to 299,993 and 10,314 primes from 1,000,003 to 1,142,971. The number of fixed points for each prime was determined using C code originally written by Cloutier [Cloutier and Holden 2010] and modified by Lindle [2008], Hoffman [2009], and Friedrichsen, Larson, and McDowell in [Friedrichsen et al. 2010]. Postprocessing was done using a Python script written by the first author. This data set combined a preliminary set of data from code run on servers maintained by the Rose-Hulman Computer Science & Software Engineering and Mathematics Departments and data from code run on the Tufts High Performance Computing Cluster. The code took a few hours of computational time, with about a day postprocessing work to fully put together the data sets. The postprocessing was the limiting factor in the number of primes we could feasibly work with.

Once the values of $F(p)$ were collected, they were normalized to a z-statistic by subtracting the predicted mean and dividing by the predicted standard deviation (square root of the variance). The z-statistics were grouped separately for the six-digit and seven-digit primes and tested to see if they conformed to the expected standard normal distribution. As you can see in Figures 1 and 2, the distributions appear to be roughly normal to the naked eye, and the standard deviations are close to 1 as expected. The means are a little higher than the expected 0, and there are a few bars which seem significantly off, but these features could be attributed to certain known properties which appear below in Theorem 2.3.

More troubling is the lack of normality revealed by probability plots in Figures 3 and 4. Perfectly normal distributions would lie along the diagonal lines in these figures, and Ryan–Joiner tests confirm that it is very unlikely that $F(p)$ is obeying a normal distribution for these primes. In fact there appear to be more primes in the "tails" than expected, that is, a larger than expected number of primes with significantly more or fewer fixed points than expected. Felix and Kurlberg [2017, Section 1.2] studied the same phenomena with two data sets comprised of seven-digit and ten-digit primes, respectively. They also broke up each data set into

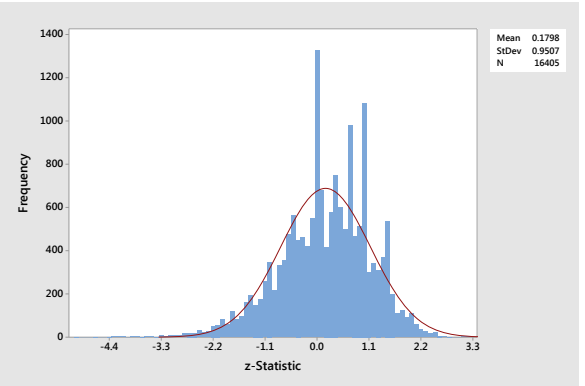

Figure 1. Histogram of z-statistics for six-digit primes.

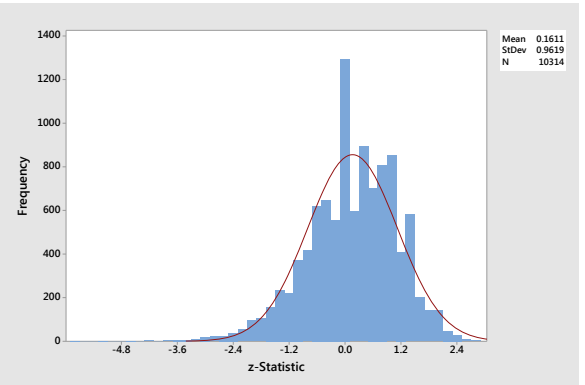

Figure 2. Histogram of z-statistics for seven-digit primes.

different subgroups based on the number of unique prime divisors of $p-1$. Their analysis matches ours, including a deviation from the binomial model at the tails.

Binomial distribution and goodness of fit. Some modification of the code by the first author allowed us to collect the values of $F_d(p)$ for the same primes as above, in order to see if particular orders were behaving less "randomly" than others. We excluded certain orders where $F_d(p)$ is known to behave predictably:

Theorem 2.3. (1) $F_1(p) = 1$ *for all* p.

(2) $F_2(p) = 0$ *for all* p.

(3) $F_{p-1}(p) = 0$ *for all* p.

(4) $F_{(p-1)/2}(p) = \begin{cases} 0 & \text{if } p \equiv 3 \text{ or } 5 \pmod 8, \\ & \quad \text{or if } p \equiv 1 \text{ or } 7 \pmod 8 \text{ and } \operatorname{ord}_p 2 \neq (p-1)/2; \\ 1 & \text{if } p \equiv 1 \text{ or } 7 \pmod 8 \text{ and } \operatorname{ord}_p 2 = (p-1)/2. \end{cases}$

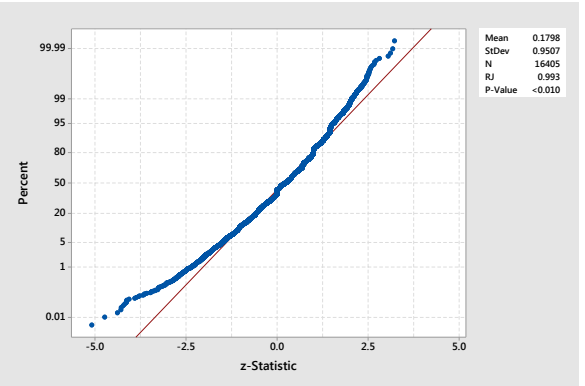

Figure 3. Probability plot of z-statistics for six-digit primes.

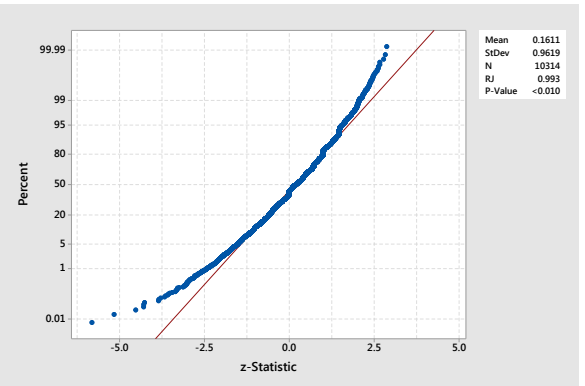

Figure 4. Probability plot of z-statistics for seven-digit primes.

To prove this we use the following lemmas:

Lemma 2.4 [Friedrichsen et al. 2010, Proposition 7]. *Let p be prime. The number x is a solution to (2) if and only if $x \equiv 1 \pmod{\operatorname{ord}_p x}$.*

Corollary 2.5. *Let $d \mid (p-1)$. The solutions to (2) of order d are exactly the elements of $\mathcal{P} = \{1, d+1, 2d+1, \ldots, p-d\}$ which have order d.*

Proof of Theorem 2.3. Parts (1) and (2) are clear from the definition. Part (3) is Proposition 6 of [Friedrichsen et al. 2010]. If x is a fixed point such that $\operatorname{ord}_p x = (p-1)/2$, then Corollary 2.5 implies $x = (p+1)/2$. Then Proposition 2 of [Friedrichsen et al. 2010] tells us x is a fixed point if and only if 2 is a quadratic residue modulo p, which is if and only if $p \equiv 1$ or $7 \pmod 8$. Combining this with the fact that $\operatorname{ord}_p (p+1)/2 = \operatorname{ord}_p 2$ gives part (4). □

Remark 3. Note that the behavior of fixed points in safe primes, that is, primes where $(p-1)/2$ is also prime, is completely explained by Theorem 2.3. Safe primes

are important for discrete logarithm-based algorithms because the group $(\mathbb{Z}/p\mathbb{Z})^\times$ will have a subgroup with large prime order. Specifically, it will have a subgroup with order $(p-1)/2$.

We collected values of $F_d(p)$ for each prime and each value of $d \mid (p-1)$ other than $d = 1,\ 2,\ p-1$, and $(p-1)/2$. We then attempted to normalize this data, but the resulting z-statistics turned out to be too highly clustered and did not resemble normal data. We therefore decided to do a chi-squared goodness-of-fit test on the data. We used the formula for the mass function of a binomial distribution to predict:

Prediction 1. $$\Pr[F_d(p) = k] = \binom{\phi(d)}{k}\left(\frac{1}{d}\right)^k\left(\frac{d-1}{d}\right)^{\phi(d)-k}.$$

We chose to use the categories $k = 0,\ k = 1,\ k = 2$, and $k > 2$ for our test in order to make sure the categories with large k did not get too small. We summed the predictions over p and d for each of the categories and compared them with the observed numbers of p and d which fell into each category. An initial test using only the primes between 100,003 and 102,677 gave a chi-squared statistic of 4.66 and a statistical p-value of 0.198.[1] Using the common cutoff of 0.05 for statistical significance of p-values, we do not see statistical evidence that our predictions are incorrect. However, using the full set of primes between 100,003 and 299,993 gave a much larger chi-squared statistic of 491.14 and a p-value of less than 10^{-100}.

We hypothesized that not all values of p and d fit the predictions equally well. We tested this by sorting in various ways the values of $F_d(p)$ collected for p between 100,003 and 102,667, and $d \mid (p-1)$ other than $d = 1,\ 2,\ p-1$, and $(p-1)/2$. After each sort, we calculated the chi-squared statistics and p-values for a sliding window of 100 values, with predictions and observations calculated as above. (The size of the window was chosen in order to make sure there were enough data points in the window for the chi-squared test to be valid.)

The strongest evidence of a pattern was seen when the data was sorted by value of d. This was confirmed for the full range of primes between 100,003 and 299,993, as can be seen in Figure 5. For data randomly generated according to the relevant binomial distributions, p-values should be evenly distributed between 0 and 1. When p-values are biased towards 0 it indicates statistically significant divergence from the predicted distributions. In other words, dots on the same (approximate) horizontal line should be evenly distributed between the left- and right-hand sides of the graph. (Note that the value of d used to place the dot on the plot is the largest value of d in the window of 100 pairs, so some dots would more accurately "belong" to more than one line.) Horizontal lines where the dots are clustered towards the left-hand side indicate statistically significant divergence.

[1] We will use the term "p-value" in this paper when referring to the statistical concept in order to distinguish it from use of p to indicate a prime.

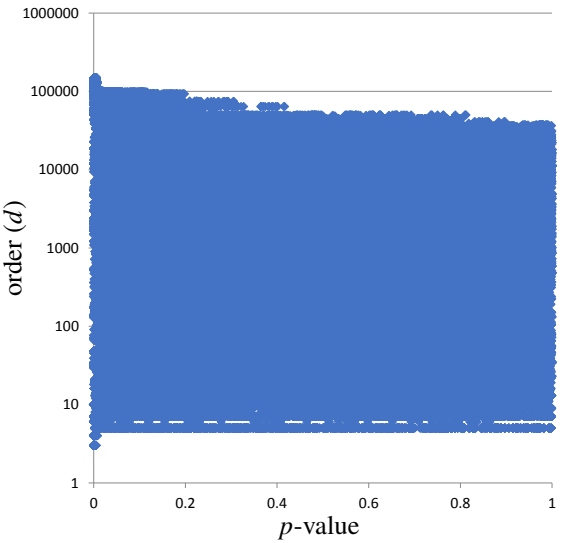

Figure 5. Logarithmic plot showing p-values of the sliding-window goodness-of-fit test, data sorted by order, for six-digit primes.

As you can see, the strongest divergence from the predictions occurs with particularly small and particularly large values of d. (Since the value of d used to place the dot on the plot is the largest value in the window, the effect for small d is even larger than it appears in the plot.) We therefore looked for theoretical explanations of these effects. We observed two significant properties that affected whether or not a given order d followed the formula in Prediction 1. The first is the size of $\phi(d)$ and the second is the size of the set $\mathcal{P} = \{1, d+1, 2d+1, \ldots, p-d\}$. On the smaller end of the spectrum, the size of $\phi(d)$ is the most influential. On the larger end, the size of the set $\mathcal{P}$ is the most influential. In the next section, we will discuss specific examples of both small and large orders.

3. Small and large orders

Small orders. For $d = 3$ we observed that while $F_3(p) = 2$ should occur roughly one-ninth of the time according to Prediction 1, it never occurred at all in our data. A similar but less striking effect was observed for $d = 4$, while for $d = 6$ it was $F_6(p) = 1$ which was never observed, despite Prediction 1 saying it should happen over one-quarter of the time. It turns out that there is a significant lack of independence in the fixed points for these orders, as we were able to show.

Theorem 3.1. (1) $F_3(p) = 0$ *or* $F_3(p) = 1$ *for all* p *such that* $3 \mid (p-1)$.

(2) $F_4(p) = 0$ *or* $F_4(p) = 1$ *for all* p *such that* $4 \mid (p-1)$.

(3) $F_6(p) = 0$ *or* $F_6(p) = 2$ *for all* p *such that* $6 \mid (p-1)$.

Proof. If $3 \mid (p-1)$, then by Lemma 2.4 the fixed points of order 3 are exactly the elements congruent to 1 modulo 3. In this case there are two elements of order 3, and a direct computation shows that if x is one of them, then $p-1-x$ is the other. Thus the elements of order 3 add up to $p-1 \equiv 0 \pmod 3$. So at most one of the elements of order 3 can be a fixed point, proving part (1). Part (2) is similar except that the elements of order 4 add up to $p \equiv 1 \pmod 4$. In part (3) the elements of order 6 add up to $p+1 \equiv 2 \pmod 6$ so if one is a fixed point then the other must be also. □

The following lemma says that the elements of a given order f are approximately uniformly distributed across the residue classes modulo any given r.

Lemma 3.2. *Let a, r, and f be positive integers such that $0 \le a < r \le p-1$ and $f \mid (p-1)$. Let*

$$\mathcal{Q} = \left\{ a, r+a, 2r+a, \ldots, \left\lfloor \frac{p-1-a}{r} \right\rfloor r + a \right\}.$$

Let $\mathcal{Q}' = \{x \in \mathcal{Q} : \operatorname{ord}_p(x) = f\}$. Then

$$\left| \#\mathcal{Q}' - \frac{\phi(f)}{r} \right| \le 1 + \tau(f)\sqrt{p}(1 + \ln p),$$

where $\tau(f)$ is the number of divisors of f.

Proof. The proof is the same as the proof of equation (7) from [Cobeli and Zaharescu 1999] with the order equal to f instead of $p-1$. □

In particular, we would expect the elements of order d to be equally likely to be of any residue class modulo d. Since Theorem 3.1 shows that the fixed points of orders $d = 3$ and $d = 4$ are entirely determined by their residue classes modulo d, this leads us to predict:

Prediction 2. (1) $\Pr[F_3(p) = 0] = \frac{1}{3}$ *and* $\Pr[F_3(p) = 1] = \frac{2}{3}$.
(2) $\Pr[F_4(p) = 0] = \frac{1}{2}$ *and* $\Pr[F_4(p) = 1] = \frac{1}{2}$.
(3) $\Pr[F_6(p) = 0] = \frac{5}{6}$ *and* $\Pr[F_6(p) = 2] = \frac{1}{6}$.

This is in fact what we observe in the data, as shown in Figure 6. This figure shows the number of primes such that $d \mid (p-1)$ for $d = 3$, 4, and 6, the number of primes for each d with $F_d(p) = 0$, 1, and 2, and the p-value given by a chi-squared test against the distribution predicted above. We do not see statistical evidence that our predictions are incorrect.

Remark 4. Not all small orders seem to exhibit this lack of independence in a statistically significant way. For example, $d = 5$ fits the distribution of the original model with $p = 0.90$ and $d = 7$ fits with $p = 0.48$. However, $d = 8$, $d = 12$, and $d = 18$ do not appear to fit the original model. For $d = 8$ and $d = 12$ the four elements of order d come in pairs which each have a dependence similar to

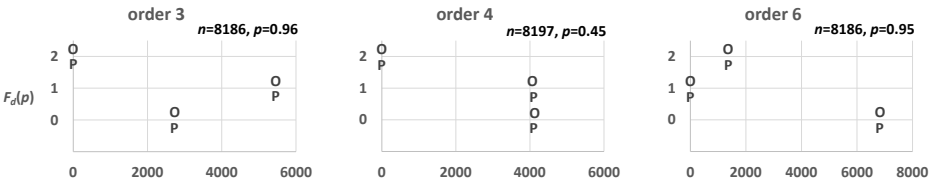

Figure 6. Predicted (P) and observed (O) numbers of primes for fixed points of orders 3, 4, and 6 in six-digit primes.

that for order 4, but we have not worked out the exact model. Other values of d which are multiples of 4 also have dependent pairs but the effect is apparently not large enough to be detected in our data. For $d = 9$ and $d = 18$ the six elements of order d come in two sets of three which each add up to 0 modulo p, producing a dependence pattern related to the ones for orders 3 and 6. We have not worked out the exact model, and it is not clear why the results are statistically significant for $d = 18$ but not $d = 9$. It may be due to chance.

Large orders. We also observed significant deviation from our predictions in the case of large orders. Recall that part (4) of Theorem 2.3 used Proposition 2 of [Friedrichsen et al. 2010] to prove that there was at most one fixed point of order $(p-1)/2$. In fact, that proposition also showed that the fixed point exists if and only if 2 is a quadratic residue modulo p. Similarly, if $3 \mid (p-1)$ then Corollary 2.5 shows that there are at most two fixed points of order $(p-1)/3$, namely $(p+2)/3$ and $(2p+1)/3$. Using methods similar to the above we can show that these residue classes will be fixed points when they are cubic residues modulo p.

Proposition 3.3. *Let p be a prime number equivalent to 1 modulo 3. The residue class $(p+2)/3$ is a fixed point if and only if it is a cubic residue modulo p, and similarly for $(2p+1)/3$.*

Proof. Note that since $1 \le x \le p-1$, (2) is equivalent to

$$x^{x-1} \equiv 1 \pmod{p}. \tag{3}$$

Then $(p+2)/3$ is a fixed point if and only

$$\left(\frac{p+2}{3}\right)^{(p-1)/3} \equiv 1 \pmod{p},$$

which by Euler's criterion is equivalent to $(p+2)/3$ being a cubic residue.

Similarly, if $(2p+1)/3$ is a fixed point then

$$\left(\frac{2p+1}{3}\right)^{(2p-2)/3} \equiv 1 \pmod{p}.$$

But then

$$\left(\frac{2p+1}{3}\right)^{(p-1)/3} \equiv \left(\frac{2p+1}{3}\right)^{(4p-4)/3} \equiv 1 \pmod{p}$$

also, where the first equivalence is just Fermat's little theorem. So Euler's criterion is satisfied again. Conversely, if

$$\left(\frac{2p+1}{3}\right)^{(p-1)/3} \equiv 1 \pmod{p}$$

then certainly

$$\left(\frac{2p+1}{3}\right)^{(2p-2)/3} \equiv 1 \pmod{p}$$

so $(2p+1)/3$ is a fixed point. □

More simplifications show that $(2p+1)/3 \equiv 3^{-1} \pmod{p}$ and $(p+2)/3 \equiv 2(3^{-1}) \pmod{p}$ so $(2p+1)/3$ will be a cubic residue whenever 3 is a cubic residue, and both $(p+2)/3$ and $(2p+1)/3$ will be cubic residues when both 2 and 3 are cubic residues. These same methods can be used to show that all numbers of the form $(m(p-1)/k)+1$ where $1 \le m < k$ will be fixed points in the self-power map when the number is a k-th residue.

This is not quite enough to investigate $F_{(p-1)/3}(p)$ since not all cubic residues have order equal to $(p-1)/3$. We thus estimate the probability that a given element of $\{(p+2)/3, (2p+1)/3\}$ has order equal to exactly $(p-1)/3$. Lemma 3.2 suggests that elements of order d occur in $\mathcal{P}$ in approximately the same proportion that they occur in the whole range $1 \le x \le p-1$, namely $\phi(d)/(p-1)$. (A more precise statement on the frequency of p such that $kd+1$ has order d would appear to require some variation on Artin's primitive root conjecture.)

We again use a binomial distribution to predict:

Prediction 3. (1) $\Pr[F_{(p-1)/3}(p) = 0] = \left(1 - \dfrac{\phi((p-1)/3)}{p-1}\right)^2$.

(2) $\Pr[F_{(p-1)/3}(p) = 1] = 2\left(\dfrac{\phi((p-1)/3)}{p-1}\right)\left(1 - \dfrac{\phi((p-1)/3)}{p-1}\right)$.

(3) $\Pr[F_{(p-1)/3}(p) = 2] = \left(\dfrac{\phi((p-1)/3)}{p-1}\right)^2$.

If $4 \mid (p-1)$, Corollary 2.5 shows that there are at most three fixed points of order $(p-1)/4$, namely $(p+3)/4$, $(p+1)/2$, and $(3p+1)/4$. However, it turns out that they cannot all be fixed points at the same time.

Theorem 3.4. *Let p be a prime number equivalent to* 1 *modulo* 4:

(1) *If* $p \equiv 1 \pmod 8$ *and* $p \equiv 1 \pmod 3$, *then* $F_{(p-1)/4}(p) \le 2$.

(2) *If* $p \equiv 1 \pmod 8$ *and* $p \equiv 2 \pmod 3$, *then* $F_{(p-1)/4}(p) \leq 1$.

(3) *If* $p \equiv 5 \pmod 8$ *and* $p \equiv 1 \pmod 3$, *then* $F_{(p-1)/4}(p) \leq 1$.

(4) *If* $p \equiv 5 \pmod 8$ *and* $p \equiv 2 \pmod 3$, *then* $F_{(p-1)/4}(p) = 0$.

Proof. Suppose $p \equiv 1 \pmod 8$. Since $(p+1)/2 \equiv 2^{-1} \pmod p$ and $(3p+1)/4 \equiv 4^{-1} \pmod p$, these two can only be both fixed points of order $(p-1)/4$ if $\mathrm{ord}_p 2 = \mathrm{ord}_p 4 = (p-1)/4$. But we know $8 \mid (p-1)$, so if $\mathrm{ord}_p 2 = (p-1)/4$ then $\mathrm{ord}_p 4 = (p-1)/8$. On the other hand, if $p \equiv 5 \pmod 8$, then we know $\mathrm{ord}_p 2 \nmid (p-1)/2$ so neither $\mathrm{ord}_p 2$ nor $\mathrm{ord}_p 4$ can be $(p-1)/4$. Now, suppose $p \equiv 2 \pmod 3$. Then $(p+3)/4$ can only be a fixed point if it is a quartic residue. We know $(p+3)/4 = 3(4^{-1})$ and 4^{-1} is a quadratic residue, but 3 is not a quadratic residue. So, $(p+3)/4$ cannot be quartic since it is not quadratic. □

To make predictions on the probabilities of each number of fixed points, we again use a binomial distribution. If $p \equiv 1 \pmod 8$, we keep in mind that the orders of $(p+1)/2$ and $(3p+1)/4$ are dependent so we can treat them together. If $p \equiv 1 \pmod 3$ also, we know that $(p+3)/4$ might be a fixed point, which is independent of the behavior of $(p+1)/2$ and $(3p+1)/4$:

Prediction 4. *Assume* $p \equiv 1 \pmod 8$ *and* $p \equiv 1 \pmod 3$; *i.e.*, $p \equiv 1 \pmod{24}$:

(1) $\Pr[F_{(p-1)/4}(p) = 0] = \left(1 - \dfrac{2\phi((p-1)/4)}{p-1}\right)\left(1 - \dfrac{3\phi((p-1)/4)}{(p-1)/2}\right)$.

(2) $\Pr[F_{(p-1)/4}(p) = 1] =$
$\left(\dfrac{2\phi((p-1)/4)}{p-1}\right)\left(1-\dfrac{3\phi((p-1)/4)}{(p-1)/2}\right)+\left(1-\dfrac{2\phi((p-1)/4)}{p-1}\right)\left(\dfrac{3\phi((p-1)/4)}{(p-1)/2}\right)$.

(3) $\Pr[F_{(p-1)/4}(p) = 2] = \left(\dfrac{2\phi((p-1)/4)}{p-1}\right)\left(\dfrac{3\phi((p-1)/4)}{(p-1)/2}\right)$.

Assume $p \equiv 1 \pmod 8$ *and* $p \equiv 2 \pmod 3$; *i.e.*, $p \equiv 17 \pmod{24}$:

(1) $\Pr[F_{(p-1)/4}(p) = 0] = \left(1 - \dfrac{3\phi((p-1)/4)}{(p-1)/2}\right)$.

(2) $\Pr[F_{(p-1)/4}(p) = 1] = \left(\dfrac{3\phi((p-1)/4)}{(p-1)/2}\right)$.

If $p \equiv 5 \pmod 8$, then we simply have:

Prediction 5. *Assume* $p \equiv 5 \pmod 8$ *and* $p \equiv 1 \pmod 3$; *i.e*, $p \equiv 13 \pmod{24}$:

(1) $\Pr[F_{(p-1)/4}(p) = 0] = \left(1 - \dfrac{2\phi((p-1)/4)}{p-1}\right)$.

(2) $\Pr[F_{(p-1)/4}(p) = 1] = \left(\dfrac{2\phi((p-1)/4)}{p-1}\right)$.

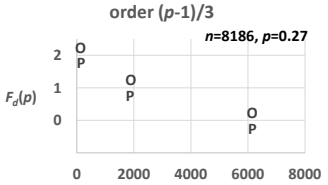

Figure 7. Predicted (P) and observed (O) numbers of primes for fixed points of order $(p-1)/3$ in six-digit primes.

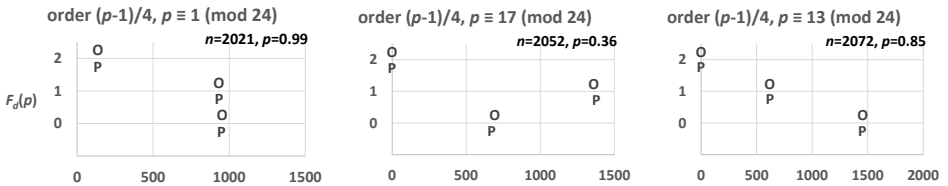

Figure 8. Predicted (P) and observed (O) numbers of primes for fixed points of order $(p-1)/4$ in six-digit primes.

Assume $p \equiv 5 \pmod 8$ *and* $p \equiv 2 \pmod 3$; *i.e.*, $p \equiv 5 \pmod{24}$:

(1) $\Pr[F_{(p-1)/4}(p) = 0] = 1$.

Chi-squared tests on the observed data from six-digit primes against the distributions predicted for orders $(p-1)/3$ and $(p-1)/4$ do not show significant deviation, as shown in Figures 7 and 8.

4. Conclusion and future work

In practice, it would certainly be possible for a user of the variant ElGamal digital signature scheme to simply make sure p is a safe prime, or alternatively arrange for r to always be a primitive root. In this way one could avoid the issue of fixed points altogether. However, we feel that it is very likely that a better understanding of the self-power map will help us better understand the security of this and other similar schemes.

We have given some bounds on the number of fixed points of the self-power map and attempted to predict the distribution of the fixed points using a binomial model whose mean is related to these proven bounds. When the order of x is moderate, this binomial model is a good predictor according to the data we collected. When the order of x is small, in particular when it is 3, 4, or 6, the independence assumption of the binomial model is violated in a significant way. However, we were able to find another model which appears to successfully predict the distribution.

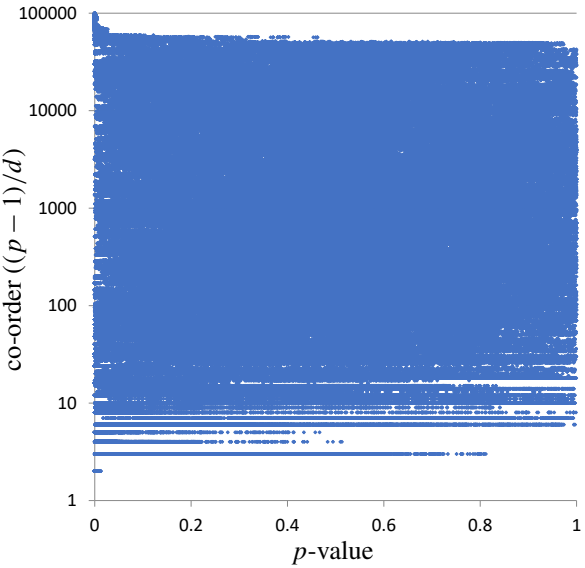

Figure 9. Logarithmic plot showing p-values of the sliding-window goodness-of-fit test, data sorted by co-order, for six-digit primes.

When the order of x is $(p-1)/3$ or $(p-1)/4$, we once again have a significant deviation from our first binomial model. However, a closer look at the set of possible fixed points in each case leads to another binomial model which appears to be successful. Some orders in the range $(p-1)/5$ to $(p-1)/9$ also appear to be showing significant deviation from the original model, as can be seen more clearly in Figure 9. In addition, the sliding-window chi-squared test shows evidence of likely divergence from the predictions in the neighborhood of $(p-1)/16$ and possibly other orders between $(p-1)/20$ and $(p-1)/50$. It is not clear yet whether all of these are true problems with the model, or just "random" consequences of the particular primes that we picked. Further investigation of these orders would appear to be the first item to be considered in future work.

Another very important item of future work would be to consider two-cycles, namely solutions to the equations

$$h^h \equiv a \pmod{p} \quad \text{and} \quad a^a \equiv h \pmod{p}, \tag{4}$$

or more generally k-cycles. Some data has been collected for these larger cycles but the binomial distribution has not yet been calculated or checked. The paper [Friedrichsen et al. 2010] also examined other graph-theoretic statistics of the functional graphs created by the self-power map, especially the number of components. This was also found to obey a nonnormal distribution and one could explore how that distribution is related to the one found here for fixed points.

Acknowledgements

Many thanks to Richard Layton for the design of Figures 6, 7, and 8. We also thank the Rose-Hulman Computer Science & Software Engineering and Mathematics Departments and Tufts University Technology Services for the use of their computers.

References

[Anghel 2013] C. V. Anghel, *The self-power map and its image modulo a prime*, Ph.D. thesis, University of Toronto, 2013, https://search.proquest.com/docview/1501462750. MR

[Anghel 2016] C. V. Anghel, "The self-power map and collecting all residue classes", *Math. Comp.* **85**:297 (2016), 379–399. MR Zbl

[Balog et al. 2011] A. Balog, K. A. Broughan, and I. E. Shparlinski, "On the number of solutions of exponential congruences", *Acta Arith.* **148**:1 (2011), 93–103. MR Zbl

[Cilleruelo and Garaev 2016a] J. Cilleruelo and M. Z. Garaev, "The congruence $x^x \equiv \lambda \pmod{p}$", *Proc. Amer. Math. Soc.* **144**:6 (2016), 2411–2418. MR Zbl

[Cilleruelo and Garaev 2016b] J. Cilleruelo and M. Z. Garaev, "Congruences involving product of intervals and sets with small multiplicative doubling modulo a prime and applications", *Math. Proc. Cambridge Philos. Soc.* **160**:3 (2016), 477–494. MR Zbl

[Cloutier and Holden 2010] D. Cloutier and J. Holden, "Mapping the discrete logarithm", *Involve* **3**:2 (2010), 197–213. MR Zbl

[Cobeli and Zaharescu 1999] C. Cobeli and A. Zaharescu, "An exponential congruence with solutions in primitive roots", *Rev. Roumaine Math. Pures Appl.* **44**:1 (1999), 15–22. MR Zbl

[Crocker 1966] R. Crocker, "On a new problem in number theory", *Amer. Math. Monthly* **73** (1966), 355–357. MR Zbl

[Crocker 1969] R. Crocker, "On residues of n^n", *Amer. Math. Monthly* **76** (1969), 1028–1029. MR Zbl

[Felix and Kurlberg 2017] A. T. Felix and P. Kurlberg, "On the fixed points of the map $x \mapsto x^x$ modulo a prime, II", *Finite Fields Appl.* **48** (2017), 141–159. MR Zbl

[Friedrichsen et al. 2010] M. Friedrichsen, B. Larson, and E. McDowell, "Structure and statistics of the self-power map", *Rose-Hulman Undergrad. Math J.* **11**:2 (2010), art. id. 6.

[Hoffman 2009] A. Hoffman, "Statistical investigation of structure in the discrete logarithm", *Rose-Hulman Undergrad. Math. J.* **10**:2 (2009), art. id. 7. Zbl

[Holden 2002a] J. Holden, "Addenda/corrigenda: fixed points and two-cycles of the discrete logarithm", preprint, 2002. arXiv

[Holden 2002b] J. Holden, "Fixed points and two-cycles of the discrete logarithm", pp. 405–415 in *Algorithmic number theory* (Sydney, 2002), edited by C. Fieker and D. R. Kohel, Lecture Notes in Comput. Sci. **2369**, Springer, 2002. MR Zbl

[Holden and Lindle 2008] J. Holden and N. Lindle, "A statistical look at maps of the discrete logarithm (abstract only)", *ACM Commun. Comput. Algebra* **42**:1-2 (2008), 57–59.

[Holden and Moree 2006] J. Holden and P. Moree, "Some heuristics and results for small cycles of the discrete logarithm", *Math. Comp.* **75**:253 (2006), 419–449. MR Zbl

[Holden and Robinson 2012] J. Holden and M. M. Robinson, "Counting fixed points, two-cycles, and collisions of the discrete exponential function using p-adic methods", *J. Aust. Math. Soc.* **92**:2 (2012), 163–178. MR Zbl

[Holden et al. 2016] J. Holden, P. A. Richardson, and M. M. Robinson, "Counting fixed points and two-cycles of the singular map $x \mapsto x^{x^n}$ modulo powers of a prime", preprint, 2016. arXiv

[Kurlberg et al. 2015] P. Kurlberg, F. Luca, and I. E. Shparlinski, "On the fixed points of the map $x \mapsto x^x$ modulo a prime", *Math. Res. Lett.* **22**:1 (2015), 141–168. MR Zbl

[Lindle 2008] N. W. Lindle, *A statistical look at maps of the discrete logarithm*, senior thesis, Rose-Hulman Institute of Technology, 2008, https://scholar.rose-hulman.edu/math_mstr/35.

[Menezes et al. 1997] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone, *Handbook of applied cryptography*, CRC Press, Boca Raton, FL, 1997. MR Zbl

[Somer 1981] L. Somer, "The residues of n^n modulo p", *Fibonacci Quart.* **19**:2 (1981), 110–117. MR Zbl

Received: 2017-04-22 Revised: 2018-01-31 Accepted: 2018-02-14

matthew.friedrichsen@tufts.edu *Department of Mathematics, Tufts University, Medford, MA, United States*

holden@rose-hulman.edu *Department of Mathematics, Rose-Hulman Institute of Technology, Terre Haute, IN, United States*

involve

msp.org/involve

INVOLVE YOUR STUDENTS IN RESEARCH

Involve showcases and encourages high-quality mathematical research involving students from all academic levels. The editorial board consists of mathematical scientists committed to nurturing student participation in research. Bridging the gap between the extremes of purely undergraduate research journals and mainstream research journals, *Involve* provides a venue to mathematicians wishing to encourage the creative involvement of students.

See inside back cover or msp.org/involve for submission instructions. The subscription price for 2019 is US $/year for the electronic version, and $/year (+$, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Involve (ISSN 1944-4184 electronic, 1944-4176 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

Involve peer review and production are managed by EditFlow® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

involve

2019 vol. 12 no. 1