## Forkable Strings are Rare

Alexander Russell<sup>1</sup>, Cristopher Moore<sup>2</sup>, Aggelos Kiayias<sup>3</sup>, and Saad Quader<sup>1</sup>

<sup>1</sup>University of Connecticut <sup>2</sup>University of Edinburgh <sup>3</sup>Santa Fe Institute

March 20, 2017

A fundamental combinatorial notion related to the dynamics of the Ouroboros proof-of-stake blockchain protocol is that of a *forkable string*. The original analysis of the protocol [2] established that the probability that a string of length *n* is forkable, when drawn from a binomial distribution with parameter  $(1 - \epsilon)/2$ , is  $\exp(-\Omega(\sqrt{n}))$ . In this note we provide an improved estimate of  $\exp(-\Omega(n))$ .

**Definition** (Generalized margin and forkable strings). Let  $\eta \in \{0, 1\}^*$  denote the empty string. For a string  $w \in \{0, 1\}^*$  we define the generalized margin of w to be the pair  $(\lambda(w), \mu(w)) \in \mathbb{Z} \times \mathbb{Z}$  given by the following recursive rule:  $(\lambda(\eta), \mu(\eta)) = (0, 0)$  and, for all strings  $w \in \{0, 1\}^*$ ,

$$\begin{aligned} (\lambda(w1), \mu(w1)) &= (\lambda(w) + 1, \mu(w) + 1), and \\ (\lambda(w0), \mu(w0)) &= \begin{cases} (\lambda(w) - 1, 0) & \text{if } \lambda(w) > \mu(w) = 0, \\ (0, \mu(w) - 1) & \text{if } \lambda(w) = 0, \\ (\lambda(w) - 1, \mu(w) - 1) & \text{otherwise.} \end{cases} \end{aligned}$$

Observe that  $\lambda(w) \ge 0$  and  $\lambda(w) \ge \mu(w)$  for all strings w. We say that a string w is forkable if  $\mu(w) \ge 0$ .

Our goal is to prove the following theorem.

**Theorem 1.** Let  $w \in \{0, 1\}^n$  be chosen randomly according to the probability law that independently assigns each  $w_i$  to the value 1 with probability  $(1 - \epsilon)/2$  for  $\epsilon > 0$ . Then  $\Pr[w \text{ is forkable}] = \exp(-\Omega(n))$ .

We prove two quantitative versions of this theorem, reflected by the bounds below. The first bound follows from analysis of a simple related martingale. The second bound requires more detailed analysis of the underlying variables, but establishes a stronger estimate.

**Bound 1.** With the random variable  $w_1 \dots w_n \in \{0,1\}^n$  defined as above so that  $\Pr[w_i = 1] = (1-\epsilon)/2$ ,

 $\Pr[w \text{ is forkable}] = \exp(-2\epsilon^4(1 - O(\epsilon))n).$ 

**Bound 2.** With the random variable  $w_1 \dots w_n \in \{0,1\}^n$  defined as above so that  $\Pr[w_i = 1] = (1-\epsilon)/2$ ,

$$\Pr[w \text{ is forkable}] = \exp(-\epsilon^3(1 - O(\epsilon))n/2).$$

We begin with a proof of Bound 1, which requires the following standard large deviation bound for supermartingales.

**Theorem 2** (Azuma; Hoeffding. See [3, 4.16] for discussion). Let  $X_0, \ldots, X_n$  be a sequence of real-valued random variables so that, for all t,  $\mathbb{E}[X_{t+1} | X_0, \ldots, X_t] \le X_t$  and  $|X_{t+1} - X_t| \le c$  for some constant c. Then for every  $\Lambda \ge 0$ 

$$\Pr[X_n - X_0 \ge \Lambda] \le \exp\left(-\frac{\Lambda^2}{2nc^2}\right).$$

*Proof of Bound 1.* Let  $w_1, w_2, ...$  be a sequence of independent random variables so that  $\Pr[w_i = 1] = (1 - \epsilon)/2$  as in the statement of the theorem. For convenience, define the associated  $\{\pm 1\}$ -valued random variables  $W_t = (-1)^{1+w_t}$  and observe that  $\mathbb{E}[W_t] = -\epsilon$ .

Define  $\lambda_t = \lambda(w_1 \dots w_t)$  and  $\mu_t = \mu(w_1 \dots w_t)$  to be the components of the generalized margin for the string  $w_1 \dots w_t$ . The analysis will rely on the ancillary random variables  $\overline{\mu}_t = \min(0, \mu_t)$ . Observe that  $\Pr[w \text{ forkable}] = \Pr[\mu(w) \ge 0] = \Pr[\overline{\mu}_n = 0]$ , so we may focus on the event that  $\overline{\mu}_n = 0$ . As an additional preparatory step, define the constant  $\alpha = (1 + \epsilon)/(2\epsilon) \ge 1$  and define the random variables  $\Phi_t \in \mathbb{R}$  by the inner product

$$\Phi_t = (\lambda_t, \overline{\mu}_t) \cdot \begin{pmatrix} 1 \\ \alpha \end{pmatrix} = \lambda_t + \alpha \overline{\mu}_t$$

The  $\Phi_t$  will act as a "potential function" in the analysis: we will establish that  $\Phi_n < 0$  with high probability and, considering that  $\alpha \overline{\mu}_n \leq \lambda_n + \alpha \overline{\mu}_n = \Phi_n$ , this implies  $\overline{\mu}_n < 0$ , as desired.

Let  $\Delta_t = \Phi_t - \Phi_{t-1}$ ; we observe that—conditioned on any fixed value  $(\lambda, \mu)$  for  $(\lambda_t, \mu_t)$ —the random variable  $\Delta_{t+1} \in [-(1 + \alpha), 1 + \alpha]$  has expectation no more than  $-\epsilon$ . The analysis has four cases, depending on the various regimes of the definition of generalized margin. When  $\lambda > 0$  and  $\mu < 0$ ,  $\lambda_{t+1} = \lambda + W_{t+1}$  and  $\overline{\mu}_{t+1} = \overline{\mu} + W_{t+1}$ , where  $\overline{\mu} = \max(0, \mu)$ ; then  $\Delta_{t+1} = (1 + \alpha)W_{t+1}$  and  $\mathbb{E}[\Delta_{t+1}] = -(1 + \alpha)\epsilon \leq -\epsilon$ . When  $\lambda > 0$  and  $\mu \geq 0$ ,  $\lambda_{t+1} = \lambda + W_{t+1}$  but  $\overline{\mu}_{t+1} = \overline{\mu}$  so that  $\Delta_{t+1} = W_{t+1}$  and  $\mathbb{E}[\Delta_{t+1}] = -\epsilon$ . Similarly, when  $\lambda = 0$  and  $\mu < 0$ ,  $\overline{\mu}_{t+1} = \overline{\mu} + W_{t+1}$  while  $\lambda_{t+1} = \lambda + \max(0, W_{t+1})$ ; we may compute

$$\mathbb{E}[\Delta_{t+1}] = \frac{1-\epsilon}{2}(1+\alpha) - \frac{1+\epsilon}{2}\alpha = \frac{1-\epsilon}{2} - \epsilon\alpha = \frac{1-\epsilon}{2} - \epsilon\left(\frac{1}{\epsilon} \cdot \frac{1+\epsilon}{2}\right) = -\epsilon.$$

Finally, when  $\lambda = \mu = 0$  exactly one of the two random variables  $\lambda_{t+1}$  and  $\overline{\mu}_{t+1}$  differs from zero: if  $W_{t+1} = 1$  then  $(\lambda_{t+1}, \overline{\mu}_{t+1}) = (1, 0)$ ; likewise, if  $W_{t+1} = -1$  then  $(\lambda_{t+1}, \overline{\mu}_{t+1}) = (0, -1)$ . It follows that

$$\mathbb{E}[\Delta_{t+1}] = \frac{1-\epsilon}{2} - \frac{1+\epsilon}{2}\alpha \le -\epsilon.$$

Thus  $\mathbb{E}[\Phi_n] = \mathbb{E}[\sum_i^n \Delta_i] \leq -\epsilon n$  and we wish to apply Azuma's inequality to conclude that  $\Pr[\Phi_n \geq 0]$  is exponentially small. For this purpose, we transform the random variables  $\Phi_t$  to a related supermartingale by shifting them: specifically, define  $\tilde{\Phi}_t = \Phi_t + \epsilon t$  and  $\tilde{\Delta}_t = \Delta_t + \epsilon$  so that  $\tilde{\Phi}_t = \sum_i^t \tilde{\Delta}_t$ . Then

$$\mathbb{E}[\tilde{\Phi}_{t+1} \mid \tilde{\Phi}_1, \dots, \tilde{\Phi}_t] = \mathbb{E}[\tilde{\Phi}_{t+1} \mid W_1, \dots, W_t] \le \tilde{\Phi}_t, \qquad \tilde{\Delta}_t \in [-(1+\alpha) + \epsilon, 1+\alpha+\epsilon],$$

and  $\tilde{\Phi}_n = \Phi_n + \epsilon n$ . It follows from Azuma's inequality that

$$\Pr[w \text{ forkable}] = \Pr[\overline{\mu}_n = 0] \le \Pr[\Phi_n \ge 0] = \Pr[\Phi_n \ge \epsilon n]$$
$$\le \exp\left(-\frac{\epsilon^2 n^2}{2n(1+\alpha+\epsilon)^2}\right) = \exp\left(-\left(\frac{2\epsilon^2}{1+3\epsilon+2\epsilon^2}\right)^2 \cdot \frac{n}{2}\right) \le \exp\left(-\frac{2\epsilon^4}{1+35\epsilon} \cdot n\right).$$

We give a more detailed argument that achieves a bound of the form  $\exp(-\epsilon^3(1+O(\epsilon))n/2)$  (Bound 2 above).

Proof of Bound 2. Anticipating the proof, we make a few remarks about generating functions and stochastic dominance. We reserve the term generating function to refer to an "ordinary" generating function which represents a sequence  $a_0, a_1, \ldots$  of non-negative real numbers by the formal power series  $A(Z) = \sum_{t=0}^{\infty} a_t Z^t$ . When  $A(1) = \sum_t a_t = 1$  we say that the generating function is a probability generating function; in this case, the generating function A can naturally be associated with the integer-valued random variable A for which  $\Pr[A = k] = a_k$ . If the probability generating function associated with the convolution A + B (where A and B are assumed to be independent). In general, we say that the generating function A stochastically dominates B if  $\sum_{t \leq T} a_t \leq \sum_{t \leq T} b_t$  for all  $T \geq 0$ ; we write  $B \leq A$  to denote this state of affairs. Observe that when these are probability generating functions and may be associated with random variables A and B it follows that  $\Pr[A \geq T] \geq \Pr[B \geq T]$  for every T. If  $B_1 \leq A_1$  and  $B_2 \leq A_2$  then  $B_1 \cdot B_2 \leq A_1 \cdot A_2$  and  $\alpha B_1 + \beta B_2 \leq \alpha A_1 + \beta A_2$  (for any  $\alpha, \beta \geq 0$ ). Finally, we remark that if A(Z) is a generating function which

converges as a function of Z for |Z| < R, it follows that  $\lim_{n\to\infty} a_n R^n = 0$  and  $a_n = O(R^{-n})$ ; if A is a probability generating function associated with the random variable A then it follows that  $\Pr[A \ge T] = O(R^{-T})$ .

We define  $p = (1 - \epsilon)/2$  and q = 1 - p and, as above, consider the independent  $\{0, 1\}$ -valued random variables  $w_1, w_2, \ldots$  where  $\Pr[w_t = 1] = p$ . As above we define the associated  $\{\pm 1\}$ -valued random variables  $W_t = (-1)^{1+w_t}$ . Our strategy is to study the probability generating function

$$\mathsf{L}(Z) = \sum_{t=0}^{\infty} \ell_t Z^t$$

where  $\ell_t = \Pr[t \text{ is the last time } \mu_t = 0]$ . Controlling the decay of the coefficients  $\ell_t$  suffices to give a bound on the probability that  $w_1 \dots w_n$  is forkable because

$$\Pr[w_1 \dots w_n \text{ is forkable}] \le 1 - \sum_{t=0}^{n-1} \ell_t = \sum_{t=n}^{\infty} \ell_t.$$

It seems challenging to give a closed-form algebraic expression for the generating function L; our approach is to develop a closed-form expression for a probability generating function  $\hat{L} = \sum_t \hat{\ell}_t Z^t$  which stochastically dominates L and apply the analytic properties of this closed form to bound the partial sums  $\sum_{t \ge n} \hat{\ell}_n$ . Observe that if  $L \le \hat{L}$  then the series  $\hat{L}$  gives rise to an upper bound on the probability that  $w_1 \dots w_n$  is forkable as  $\sum_{t=n}^{\infty} \ell_t \le \sum_{t=n}^{\infty} \hat{\ell}_t$ .

The coupled random variables  $\lambda_t$  and  $\mu_t$  are Markovian in the sense that values  $(\lambda_s, \mu_s)$  for  $s \ge t$  are entirely determined by  $(\lambda_t, \mu_t)$  and the subsequent values  $W_{t+1}, \ldots$  of the underlying variables  $W_i$ . We organize the sequence  $(\lambda_0, \mu_0), (\lambda_1, \mu_1), \ldots$  into "epochs" punctuated by those times t for which  $\lambda_t = \mu_t = 0$ . With this in mind, we define  $M(Z) = \sum m_t Z^t$  to be the generating function for the first completion of such an epoch, corresponding to the least t > 0 for which  $\lambda_t = \mu_t = 0$ . As we discuss below, M(Z) is not a probability generating function, but rather  $M(1) = 1 - \epsilon$ . It follows that

$$L(Z) = \epsilon (1 + M(Z) + M(Z)^{2} + \dots) = \frac{\epsilon}{1 - M(Z)}.$$
 (1)

Below we develop an analytic expression for a generating function  $\hat{M}$  for which  $M \leq \hat{M}$  and define  $\hat{L} = \epsilon/(1 - \hat{M}(Z))$ . We then proceed as outlined above, noting that  $L \leq \hat{L}$  and using the asymptotics of  $\hat{L}$  to upper bound the probability that a string is forkable.

In preparation for defining  $\hat{M}$ , we set down two elementary generating functions for the "descent" and "ascent" stopping times. Treating the random variables  $W_1, \ldots$  as defining a (negatively) biased random walk, define D to be the generating function for the *descent stopping time* of the walk; this is the first time the random walk, starting at 0, visits -1. The natural recursive formulation of the descent time yields a simple algebraic equation for the descent generating function,  $D(Z) = qZ + pZD(Z)^2$ , and from this we may conclude

$$\mathsf{D}(Z) = \frac{1 - \sqrt{1 - 4pqZ^2}}{2pZ}$$

We likewise consider the generating function A(Z) for the *ascent stopping time*, associated with the first time the walk, starting at 0, visits 1: we have  $A(Z) = pZ + qZA(Z)^2$  and

$$\mathsf{A}(Z) = \frac{1 - \sqrt{1 - 4pqZ^2}}{2qZ}$$

Note that while D is a probability generating function, the generating function A is not: according to the classical "gambler's ruin" analysis [1], the probability that a negatively-biased random walk starting at 0 ever rises to 1 is exactly p/q; thus A(1) = p/q.

Returning to the generating function M above, we note that an epoch can have one of two "shapes": in the first case, the epoch is given by a walk for which  $W_1 = 1$  followed by a descent (so that  $\lambda$  returns to zero); in the second case, the epoch is given by a walk for which  $W_1 = -1$ , followed by an ascent (so that  $\mu$  returns to zero), followed by the eventual return of  $\lambda$  to 0. Considering that when  $\lambda_t > 0$  it will return to zero in the future almost surely, it follows that

the probability that such a biased random walk will complete an epoch is  $p + q(p/q) = 2p = 1 - \epsilon$ , as mentioned in the discussion of (1) above. One technical difficulty arising in a complete analysis of M concerns the second case discussed above: while the distribution of the smallest t > 0 for which  $\mu_t = 0$  is proportional to A above, the distribution of the smallest subsequent time t' for which  $\lambda_{t'} = 0$  depends on the value t. More specifically, the distribution of the return time depends on the value of  $\lambda_t$ . Considering that  $\lambda_t \leq t$ , however, this conditional distribution (of the return time of  $\lambda$  to zero conditioned on t) is stochastically dominated by D<sup>t</sup>, the time to descend t steps. This yields the following generating function  $\hat{M}$  which, as described, stochastically dominates M:

$$\widehat{\mathsf{M}}(Z) = pZ \cdot \mathsf{D}(Z) + qZ \cdot \mathsf{D}(Z) \cdot \mathsf{A}(Z \cdot \mathsf{D}(Z)).$$

It remains to establish a bound on the radius of convergence of  $\hat{L}$ . Recall that if the radius of convergence of  $\hat{L}$  is  $\exp(\delta)$  it follows that  $\Pr[w_1 \dots w_n \text{ is forkable}] = O(\exp(-\delta n))$ . A sufficient condition for convergence of  $\hat{L}(z) = \epsilon/(1 - \hat{M}(z))$  at z is that that all generating functions appearing in the definition of  $\hat{M}$  converge at z and that the resulting value  $\hat{M}(z) < 1$ .

The generating function D(z) (and A(z)) converges when the discriminant  $1 - 4pqz^2$  is positive; equivalently  $|z| < 1/\sqrt{1-\epsilon^2}$  or  $|z| < 1 + \epsilon^2/2 + O(\epsilon^4)$ . Considering  $\hat{M}$ , it remains to determine when the second term, qzD(z)A(zD(z)), converges; this is likewise determined by positivity of the discriminant, which is to say that

$$1 - (1 - \epsilon^2) \left( \frac{1 - \sqrt{1 - (1 - \epsilon^2)z^2}}{1 - \epsilon} \right)^2 > 0.$$

Equivalently,

$$|z| < \sqrt{\frac{1}{1+\epsilon} \left(\frac{2}{\sqrt{1-\epsilon^2}} - \frac{1}{1+\epsilon}\right)} = 1 + \epsilon^3/2 + O(\epsilon^4).$$

Note that when the series  $pz \cdot D(z)$  converges, it converges to a value less than 1/2; the same is true of  $qz \cdot A(z)$ . It follows that for  $|z| = 1 + \epsilon^3/2 + O(\epsilon^4)$ ,  $|\hat{M}(z)| < 1$  and  $\hat{L}(z)$  converges, as desired. We conclude that

$$\Pr[w_1 \dots w_n \text{ is forkable}] = \exp(-\epsilon^3 (1 + O(\epsilon))n/2).$$

## References

- [1] Charles M. Grinstead and J. Laurie Snell. Introduction to Probability. American Mathematical Association, 1997.
- [2] Aggelos Kiayias, Alexander Russell, Bernardo David, and Roman Oliynykov. Ouroboros: A provably secure proof-of-stake blockchain protocol. Cryptology ePrint Archive, Report 2016/889, 2016. http://eprint.iacr. org/2016/889.
- [3] Rajeev Motwani and Prabhakar Raghavan. *Randomized Algorithms*. Cambridge University Press, New York, NY, USA, 1995.