Forkable Strings are Rare

Alexander Russell\textsuperscript{1}, Cristopher Moore\textsuperscript{2}, Aggelos Kiayias\textsuperscript{3}, and Saad Quader\textsuperscript{1}

\textsuperscript{1}University of Connecticut
\textsuperscript{2}University of Edinburgh
\textsuperscript{3}Santa Fe Institute

March 20, 2017

A fundamental combinatorial notion related to the dynamics of the \texttt{Ouroboros} proof-of-stake blockchain protocol is that of a forkable string. The original analysis of the protocol \cite{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 $w \in \{0, 1\}^*$,

$$(\lambda(w1), \mu(w1)) = (\lambda(w) + 1, \mu(w) + 1),$$

and

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

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

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 $w_1 \ldots w_n \in \{0, 1\}^n$ defined as above so that $\Pr[w_1 = 1] = (1 - \epsilon)/2$,

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

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

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

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

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

$$\Pr[X_n - X_0 \geq \Lambda] \leq \exp\left(-\frac{\Lambda^2}{2nc^2}\right).$$
**Proof of Bound 1.** Let \( w_1, w_2, \ldots \) 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 \( W_i = (-1)^{1 + w_i} \) and observe that \( \mathbb{E}[W_i] = -\epsilon \).

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

\[
\Phi_i = (\lambda_i, \overline{\mu}_i) \cdot \left( \frac{1}{\alpha} \right) = \lambda_i + \alpha \overline{\mu}_i .
\]

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

Let \( \Delta_i = \Phi_i - \Phi_{i-1} \); we observe that—conditioned on any fixed value \((\lambda, \mu)\) for \((\lambda_i, \mu_i)\)—the random variable \( \Delta_{i+1} \in [-1 + \alpha, 1 + \alpha] \) 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 \( \mu < 0 \), \( \lambda_{i+1} = \lambda + W_{i+1} \) and \( \overline{\mu}_{i+1} = \overline{\mu} + W_{i+1} \), where \( \overline{\mu} = \max(0, \mu) \); then \( \Delta_{i+1} = (1 + \alpha)W_{i+1} \) and \( \mathbb{E}[\Delta_{i+1}] = -(1 + \alpha)\epsilon \leq -\epsilon \). When \( \lambda > 0 \) and \( \mu \geq 0 \), \( \lambda_{i+1} = \lambda + W_{i+1} \) but \( \overline{\mu}_{i+1} = \overline{\mu} \) so that \( \Delta_{i+1} = W_{i+1} \) and \( \mathbb{E}[\Delta_{i+1}] = -\epsilon \). Similarly, when \( \lambda = 0 \) and \( \mu < 0 \), \( \overline{\mu}_{i+1} = \overline{\mu} + W_{i+1} \) while \( \lambda_{i+1} = \lambda + \max(0, W_{i+1}) \); we may compute

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

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

\[
\mathbb{E}[\Delta_{i+1}] = \frac{1 - \epsilon}{2} - \frac{1 + \epsilon}{2}\alpha \leq -\epsilon.
\]

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 \( \Phi_i \) to a related supermartingale by shifting them: specifically, define \( \tilde{\Phi}_i = \Phi_i + \epsilon t \) and \( \tilde{\Delta}_i = \Delta_i + \epsilon \) so that \( \tilde{\Phi}_i = \sum_i^{t_i} \tilde{\Delta}_i \). Then

\[
\mathbb{E}[\tilde{\Phi}_{t+1} | \tilde{\Phi}_1, \ldots, \tilde{\Phi}_t, W_1, \ldots, W_t] \leq \tilde{\Phi}_t, \quad \tilde{\Delta}_i \in [-1 + \alpha, \epsilon, 1 + \alpha + \epsilon],
\]

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

\[
\Pr[w \text{ forkable}] = \Pr[\overline{\mu}_n = 0] \leq \Pr[\Phi_n \geq 0] = \Pr[\tilde{\Phi}_n \geq \epsilon n] \leq \exp\left( -\frac{\epsilon^2 n^2}{2n(1 + \alpha + \epsilon)^2} \right) = \exp\left( -\frac{\epsilon^2}{1 + 3\epsilon + 2\epsilon^2} \cdot \frac{n}{2} \right) \leq \exp\left( -\frac{\epsilon^2}{1 + 3\epsilon} \cdot \frac{n}{2} \right) .
\]

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 \( A(Z) = \sum_{i=0}^{\infty} a_i Z^i \). When \( A(1) = \sum_i a_i = 1 \) 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 functions \( A \) and \( B \) are associated with the random variables \( A \) and \( B \), it is easy to check that \( A \cdot B \) is the 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 \( \sum_{i \leq T} a_i \leq \sum_{i \leq T} b_i \) for all \( T \geq 0 \); we write \( A \preceq B \) 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 \( B_1 \preceq A_1 \) and \( B_2 \preceq A_2 \) then \( B_1 \cdot B_2 \preceq A_1 \cdot A_2 \) and \( \alpha B_1 + \beta B_2 \preceq \alpha A_1 + \beta A_2 \) (for any \( \alpha, \beta \geq 0 \)). Finally, we remark that if \( A(Z) \) is a generating function which
follows that for which \( M \) generating function, \( 0 \), visits \( "\text{gambler’s ruin}\" \) analysis \([1]\), the probability that a negatively-biased random walk starting at 0 ever rises to 1 is exactly \( A \). We likewise consider the generating function \( A(Z) \) for the \( \text{ascent stopping time} \), associated with the first time the walk, starting at 0, visits 1: we have \( A(Z) = pZ + qZA(Z)^2 \) and

\[
A(Z) = \frac{1 - \sqrt{1 - 4pqZ^2}}{2qZ}.
\]

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 \( D' \), the time to descend \( t \) steps. This yields the following generating function \( \hat{M} \) which, as described, stochastically dominates \( M \):

\[
\hat{M}(Z) = pZ \cdot D(Z) + qZ \cdot D(Z) \cdot A(Z \cdot D(Z)).
\]

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 \( \Pr[w_1 \ldots w_n \text{ is forkable}] = O(\exp(-\delta n)) \). A sufficient condition for convergence of \( \hat{L}(z) = \epsilon/(1 - \hat{M}(z)) \) at \( z \) is that 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,

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

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

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

\[\Box\]

References

